2001 A Transportation ODYSSEY

2001 Une odyssée en TRANSPORT



CANADIAN TRANSPORTATION RESEARCH FORUM

CTRF/GRTC

LE GROUPE DE RECHERCHES SUR LES TRANSPORTS AU CANADA

Vancouver, British Columbia • May 6-9 mai • 2001

Proceedings compiled by
W.G. Waters II, the University of British Columbia
Vice President Program & Publications, CTRF

PROCEEDINGS

of the 36th Annual Conference
de la 36è Conférence annuelle

RECEIVED
MAY 23 2001
TRANSPORTATION
DIVISION

CANADIAN TRANSPORTATION RESEARCH FORUM
LE GROUPE DE RECHERCHES SUR LES TRANSPORTS AU CANADA

PROCEEDINGS
of the 36th Annual Conference

ACTES
de la 36è Conférence annuelle

2001 - A Transportation Odyssey
2001 – Une odyssée en transport

Vancouver, British Columbia
May 6-9 mai
2001

Proceedings compiled by
W.G. Waters II, the University of British Columbia
Vice President Program & Publications, CTRF

Proceedings Printed by

University of Saskatchewan
Printing Services

The Canadian Transportation Research Forum assumes no responsibility for the content of the papers published in these Proceedings, or for any errors or omissions in the printing of this volume. These papers are published by the CTRF as a means of disseminating transportation research and to encourage discussion on a broad range of issues during the CTRF Annual Conference.

President's Message

Welcome to the 36th Annual Meeting of the Canadian Transportation Research Forum. This event promises to be *A Transportation Odyssey*. The conference organizers are to be commended for their efforts which culminate in this educational and enjoyable 3 to 4 days in Vancouver. Congratulations to Bill Waters, Paul Ouimet and their team.

It has been a great honour and pleasure for me to serve as CTRF President for the year 2000-2001. I hereby claim to have been president during the first year of the new millennium, whichever year you believe that to be.

I take this opportunity to recognize and acknowledge all those who have contributed to another successful year for CTRF. Thank you to all board members, sponsors, conference organizers, attenders, panelists, authors, presenters and volunteers.

These conference proceedings contain the results of research conducted by numerous transportation professionals over the past very interesting and exciting year for transportation in Canada.

I know you will find the 2001 Transportation Odyssey to be a worthwhile, informative, and memorable journey; by attending in Vancouver or by reading these proceedings at a later date, or both.

Sincerely,



Gordon E. Tufts, P. Eng.
el presidente

Message du Président

Bienvenue à la 36è Conférence annuelle du Groupe de recherches sur les transports au Canada. Cet événement promet d'être *Une odyssée et transport.* Les organisateurs de la conférence méritent tout un hommage pour leurs efforts qui ont abouti à ces 3 ou 4 jours aussi instructifs qu'agréables à Vancouver. Félicitations à Bill Waters, Paul Ouimet et leur équipe.

Cela a été un grand honneur et un plaisir pour moi de servir comme président pour l'année 2000-2001. Par la présente, je revendique le fait d'avoir été président durant la première année du nouveau millénaire, quelle que soit l'année qui vous convient.

Je profite de cette occasion pour souligner la contribution de tous ceux qui ont assuré une autre année de succès pour le GRTC. Merci à tous les membres du conseil d'administration, aux commanditaires, aux organisateurs de la conférence, aux congressistes, aux panélistes, aux auteurs, aux conférenciers et aux volontaires.

Les présents Actes contiennent les résultats de la recherche entreprise par de nombreux professionnels en transport durant l'année écoulée, laquelle a été des plus intéressantes et des plus passionnantes en matière de transport au Canada.

Je sais que vous trouverez que l'Odyssée en transport 2001 est un voyage utile, instructif et mémorable, que ce soit par votre participation à Vancouver même, ou en lisant ces Actes ou bien les deux.

Sincèrement,

Gordon E. Tufts, ing.
El presidente



CONFERENCE SPONSORS
COMMANDITAIRES DE LA CONFÉRENCE

Province of British Columbia
Transport Canada / Transports Canada
Industry Canada / Industrie Canada
Canadian Transportation Agency / Office des transports du Canada
Statistics Canada / Statistique Canada
Canadian Pacific Railway / Canadien Pacifique
Canadian National Railways / Canadien National
GE Capital
National Steel Car Ltd.
Trimac
Trinity Industries, Inc.
Reefer Sales & Service
CGTX Inc.
Thermo King of Eastern Canada
Mi-Jack Products
Top Lift Enterprises
The Greenbrier Companies
KPMG Consulting
CANAC Inc.
Bison Transport
VEMAX Management Inc.
IBI Group
Vancouver Airport Authority
University of British Columbia
Intervistas Consulting Inc.
Canadian Freightways Limited

This list was updated April 3, 2001. Additional sponsors will be acknowledged in the Forumation, on the CTRF Website, and at the conference.

DONORS TO THE CTRF SCHOLARSHIP PROGRAM
PARRAINS AU PROGRAMME DE BOURSES DU GRTC
2001-2002

Canadian Transportation Education Foundation (CTEF)
Canada Maritime Agencies Limited
Canadian National Railways
Canadian Pacific Charitable Foundation
Government of Quebec
Marine Atlantic
plus many individual donors

2000 STUDENT PAPER COMPETITION DONORS
PARRAINS DU CONCOURS DES ÉTUDIANTS EN 2000

Air Canada
Albert Stevens Fund
Jim Davey Trust Fund

CANADIAN TRANSPORTATION RESEARCH FORUM
LE GROUPE DE RECHERCHES SUR LES TRANSPORTS AU CANADA

EXECUTIVE - EXECUTIF
2000/2001

President
Gordon Tufts
Manitoba Highways and
Government Services

Executive Vice President
Eric Hildebrand
University of New Brunswick

Secretary
Gail R. Sparks
VEMAX Management Inc.

Vice President Finance & Treasurer
Jan B. Bowland
KPMG Management Consulting

Vice President External
Gerry Kolaitis
VIA Rail Canada

Vice President Meetings
Paul Ouimet
Intervistas Consulting Inc

Vice President Programs
Bill Waters
Transportation Studies, UBC

Vice President Awards
Clarence Woudsma
University of Calgary

Immediate Past President
Lloyd Ash
Trimac Consulting Services Ltd

Vice President Organization and Development
Bill Raney
Ministry of Transportation Ontario

Vice President Sponsorship
Marcel Leduc
Canadian National Railway

Councillors:

Harold Hefferton
Transport Canada

Paul Larouche
Bombardier Transportation

Tony Nardi
Neptune Bulk Terminals Ltd.

James Nolan
University of Saskatchewan

Barry Prentice
Transport Institute, University of Manitoba

Rachid Raffa
Quebec Ministry of Transport

Jeff Rosenblum
Canadian Pacific Railway

Roger Roy
Transport Canada

Doug Shea
NF & LB Dept. of Works, Services & Transportation

Tricia Trépanier
Statistics Canada

CTRF is ...
a broadly based network of transportation professionals with an interest in research and commentary on public policy issues. Membership is open to anyone interested in any aspect of transportation.

CTRF's objectives are ...
- to encourage transportation research in those disciplines which address themselves to the economic, managerial, technical, political, and social aspects of transportation
- to provide a common meeting ground to exchange ideas and research results
- to provide a forum for presenting and discussion transportation issues
- to bring together persons involved with Canada's transportation systems.

CTRF's program consists of ...
- a national conference with research papers, panel discussions, and invited speakers held annually in a different Canadian location
- published proceedings of the national conference
- a one-day semi-annual conference on a topic of national interest
- regional seminars promoted and co-sponsored with other organizations
- an annual CTRF student paper competition
- an annual CTRF scholarship competition
- a newsletter published five times a year.

Le GRTC est ....
un groupe de professionnels(les) en transport intéressés(ées) à la recherche et aux questions de politiques publiques. Etre membre du GRTC convient à quiconque s'intéresse aux transports.

Les objectifs de GRTC sont ...
- d'encourager la recherche dans les domaines suivants liés aux transports: économie, gestion, technologie, politique et aspect social
- de fournir un lieu de rencontre pour échanger idées et résultats de recherche
- de créer un forum pour présenter et discuter de questions de transport
- de réunir les personnes intéressées au système canadien de transport.

Au programme du GRTC ....
- une conférence nationale annuelle avec site variable au Canada, qui consiste en une présentation de textes, tables rondes, conférenciers invités
- la publication des textes présentés à la conférence nationale
- une conférence semi-annuelle d'une journée sur un juget d'intérêt national
- la promotion et le parrainage de colloques régionauz avec d'autres organismes
- un coucours annuel de textes d'étudiants(es)
- un fonds permettant de décerner des bourses
- un bulletin de nouvelles publié cinq fois l'an

..... ANNOUNCING .....

2002 CTRF ANNUAL CONFERENCE TO BE HELD IN ST. JOHN'S, NEWFOUNDLAND

WATCH FOR CALL FOR PAPERS COMING IN SEPTEMBER 2001

For further information, contact:

Gail Sparks
(306)242-6199
e-mail: gail.sparks@ctrf.ca

CTRF website: www.ctrf.ca

CANADIAN TRANSPORTATION RESEARCH FORUM
209 - 15 Innovation Blvd.
Saskatoon, Saskatchewan
S7N 2X8
www.ctrf.ca

Telephone: (306) 242-6199
Fax: (306) 242-6199

INVOICE FOR 2001 CTRF MEMBERSHIP

CTRF Membership for Calendar Year 2001 ($60.00)	$ _____
GST at 7% ($4.20)	$ _____
Student Membership for Calendar Year 2001 ($18.00)	$ _____
GST @ 7% ($1.26)	$ _____
TOTAL	$ _____

GST Registration No. R126274562 (If GST exempt, please indicate)

Please make cheque payable to Canadian Transportation Research Forum and mail to the address noted above. **PLEASE RETURN A COPY OF THIS INVOICE WITH PAYMENT IN ORDER THAT OUR DATABASE WILL CONTAIN ACCURATE INFORMATION RELATING TO YOUR MEMBERSHIP.**
VISA and MasterCard are also acceptable.

- ☐ **VISA**
- ☐ **MasterCard**

Card Number _____
Expiry Date _____
Name on Card _____

Please complete the following to ensure accurate membership data. The information below will appear in the next published CTRF/TRF Directory.

Name: _____ Title: _____

Company: _____

Address: _____

_____

_____

Telephone: _____ Fax: _____

e.mail: _____

Please indicate your language preference: ☐ English ☐ French

Le Groupe de Recherches sur les Transports au Canada
209 - 15 Innovation Blvd.
Saskatoon, Saskatchewan
S7N 2X8
www.ctrf.ca

Téléphone: (306) 242-6199
Télécopieur: (306) 242-6199

FACTURE 2001

Membre du GRTC pour l'année 2001 (60.00 $)	$ _____
TPS (4.20$)	$ _____
Membre étudiant du GRTC pour l'année 2001 (18.00$)	$ _____
TPS (1.26$)	$ _____
TOTAL	$ _____

Numéro d'enregistrement pour la TPS R126274562. (Si vous êtes exempté de la TPS, veuillez l'indiquer.)

Veuillez libeller votre chèque au nom du Groupe de Recherches sur les Transports au Canada et poster le tout a l'adresse ci-haut. **SVP RETOURNEZ UNE COPIE DE CETTE FACTURE AVEC VOTRE PAIEMENT AFIN QUE NOTRE BASE DE DONNÉES CONTIENNE UNE INFORMATION EXACTE SUR VOUS COMME MEMBRE DE L'ASSOCIATION.**

VISA et MasterCard sont accéptes.

- ❒ **VISA**
- ❒ **MasterCard**

Numéro de Carte _____
Date d'Expiration _____
Nom Figurant sur la Carte _____

fournir les renseignements demandée ci-dessous afin de nous permettre d'avoir une information complète sur nos membres. Cette information va apparaitre dans le prochain répertoire du GRTC.

Nom: _____ Titre: _____

Compagnie:
_____

Adresse:
_____
_____
_____

Téléphone: _____ Télécopieur: _____

Courrier électronique _____

Veuillez indiquer la langue de correspondance de votre choix.

❒ anglais ❒ français

Volume I

Session 1 **Plenary Session**
"Consolidation and Continentalization in North American Air and Rail Transportation"

Speakers to be announced

Session 2.1 Airport Policy

- One Size Does Not Fit All: A Critique of Canada's National Airport Policy. *John Brander and B.A . Cook.* 1
- Airport Privatization in Brazil : Questions and Answers. *Respicio A. Espírito Santo Jr., Fabio Cardoso Correia and Guilherme L. Palhares* 17

Session 2.2 Transportation and Supply Chain Efficiency

- The Internet-Enabled Supply Chain and Transportation. *Garland Chow and Michael Haughton* 32
- GIS Application in Transportation. *David Spearin.* 56
- The Japanese Takuhaibin System – Prototype for a Universal Solution of the Distribution Challenges in the Age of E-Business. *Claus Eberhard* 69

Session 2.3 Transportation Policy – Canada and Communities

- Canada Transportation Act : A Snapshot of Key Stakeholder Opinion. *A.J. Phillips and D.B. Duncan* 85
- Regulatory Reforms in Canadian Transportation since 1987. *Joseph Monteiro* 101
- Planning a Transportation and Technology Museum: Concept and Market Analysis. *Linda Christian* 121

Session 3.1 Evolution of Airline Industry and Performance

- The Evolution of Airlines in Canada with Emphasis on the Role of the Competition Bureau's Role in the Restructuring of the Airline Industry. *Joseph Monteiro and Richard Annan and Andrew Eckert* 136
- Adverse Effects of Mergers on Airline Performance – The Case of Canadian Airlines International, Ltd. *William Jordan* 156
- Assessment of Recent Performance of Canadian Carriers in Comparison with U.S. Carriers. *Tae H. Oum and Chunyan Yu* 173

Session 3.2 Trade and Transportation

- Mexico's Rail Privatization: Implications for Modal Competition in North American Grain Trade. *Mark Ojah and Barry E. Prentice* 174
- Trade and Transportation: The Impact of the 1995 Transborder Air Services Accord. *Gordon Baldwin and Michael W. Pustay* 189
- Economics of Congestion at a Border Gateway: The Case of Laredo. *Barry E. Prentice and Wanda Heron* 205

Session 3.3 Technology and Rail Transportation

- The Fight Between Technology and Survival: Heavy Cars versus Short Line Railroads. *Kenneth L. Casavant, Jerry Lenzi, and Denver Tolliver.* 221
- A Preliminary Feasibility Study about High-Speed Rail in Canada. *Masatoshi Hatoko* 235
- Oil Prices and Railway Electrification. *F.H. Howard* 251

Session 4.1 Air Cargo

- Innovative Air Cargo Transfer Flexibility at Ted Stevens Anchorage International Airport. *Darren Prokop* 262
- The Blurred Lines of Responsibility in Air Cargo Customer Service. *Meyer D. Lapkin and Michael A. Crockatt* 278

Session 4.2 Intermodal Transportation and International Linkages

- Vertical Integration in Liner Shipping: Strategy and Status. *Trevor D. Heaver* 292
- The Future for Canada-U.S. Container Port Rivalries. *Doug O'Keefe* 309
- An Assessment of North American Cabotage Laws: The Carriage of Goods in Containers. *Erica Vido and Jake Kosior* 325

Session 4.3 Transportation Safety

- Heads Up! Enhancing the Safety Audit. *Lennart Henriksson* 337
- Designing Programs for Reducing Aggressive Driving, Beyond Law Enforcement. *Richard A. Raub and Richard I. Wark* 347
- Allocating Emergency Response Facilities to Mitigate Risks of Dangerous Goods Transportation. *Ghada Hamouda, F. Saccomanno, and L. Fu.* 363

Session 5.1 Airline Alliances and Competition

- Global Alliance Airline Network: Traffic Routing Patterns and Impacts on Airports. *Tae H. Oum and Chunyan Yu* 377
- Simulating Airline Alliances in an International Context. *James Nolan, Pamela Ritchie and John Rowcroft,* 378
- Des alliances stratégiques à la desserte des Îles- de-la-Madeleine : les enjeux de la déréglementation du secteur aérien au Canada (From Global Alliances to Servicing Magdalene Islands: Issues of the Canadian Air Transportation Deregulation Policy). *Yann Alix* 380

Session 5.2 Motor Carriers

- Truck Safety and Safety Regulations. *Fred Nix* 396
- Evolution des conditions économiques des conducteurs – propriétaires, 1991-1998 (The Evolution of the Economic Conditions of Owner Operators 1991-1998). *Gilles Paré* 412
- Economic Impact of For-Hire and Private Trucking in Manitoba, 1998 Results and Methodology. *Darren R.Gorman, Barry E. Prentice, Scott J. Shurvell* 414

Session 6.1 Airport Management and Performance

- Private Airports and Public Needs: There is a Connecting Flight. *Michael Crockatt and Darren Prokop* 430
- Transport Canada's Delegation to New Commercialised Entities: the Cases of Montréal Airports and Port. *Michel Boucher* 446
- A Comparison Study of Medium Airport Management in Different Countries. *Guilherme Lohmann Palhares and Respicio A. Espírito Santo Jr.* 448

Session 6.2 Efficient Investment and Maintenance of Road Infrastructure

- Tender Evaluation – Quality-Price Trade-Off Method. *Gordon A. Sparks, Paul Christensen, Curtis F. Berthelot* 465
- The Impact of Transportation on Local Economies: Two Case Studies. *Tom Kearns.* 480
- Addressing Infrastructure Maintenance Issues Through Management Innovations. *Mohammad Tayyaran and Ata M. Khan.* 496

Session 6.3 Energy and Efficiency in Transportation

- Rail Freight Competition and Efficiency Gains in Australia. *Philip Laird* 512
- The Economic and Environmental Benefits from Vehicle Speed Reduction. *F.R. Wilson, A.M. Stevens and A. Ambo* 530
- Transportation Energy Demand Model: The Application of a Complete Decomposition Method. *Jason Randall and Mohamed Nouhi.* 549

Volume II

Session 7.1 Urban Transportation

- Exploring the Effectiveness of Land-Use Policy on Air Quality: A Simulation. *Terence Lam and Debbie Niemeier* 565
- A Model of Curb Rights in Private Urban Transit Markets. *André De Palma and Robin Lindsey* 581
- Trucking in Canadian Cities: Planning for The Future. *Clarence G. Woudsma.* 597

Session 7.2 Competition and Regulation

- Mediation and Regulatory Efficiency. *Arun Thangaraj, C. McClellan and G. Nera* 614
- Importance des modifications · la loi sur la concurrence portant sur les comportements d'eviction et les ordonnances provisoires relativement à un services aérien. (The Importance of Amendments to the Competition Act on Temporary Orders in Relation to a Domestic Airline Service in Canada). *Jacques Picard* 630
- The Use of Economic Analyses in Competition Cases (Part II) Analysis of a Conspiracy and a Merger in a Transportation Cases. *Joseph Monteiro and Andrew Eckert* 646

Session 7.3 Environment and Transportation: What Next?

- The Green Mile: Approaches and Attitudes toward Environment Sustainability among Canadian Transport Faculty. *Anthony Perl and Michael Bates.* 666
- Modern Roundabout, Global Warming, and Emissions Reductions: Status of Research, and Opportunities for North America. *Tony Redington.* 682

Session 8.1 Traffic Congestion and Transportation Planning

- A Time Series Analysis of Vessel Traffic Service Utilization. *Louis Le Blanc and Nicholas A. Beadles.* 698
- Economic Modelling of Congestion in Transportation Network. *Yuri Yevdokimov* 705
- The Foundation of a New Transportation Planning Model for the Manitoba Capital Region. *Kean Hoe Lew, Alan Clayton and Doug Hurl.* 721

Session 8.2 Rail Competition and Access

- Clearing the Track: Appraising Railway Perspectives on the Issues of Regulation and Competition in the North American Rail Industry. *William Drew and James Nolan* 732
- Railway Competition and Continental Consolidation. *Kells Boland.* 748
- Railway Reform in Australia: Access Regimes, Changes in Ownership, and Structural Change. *Derek Scrafton,* 761
- Rail Access Pricing for Freight : International Practice. *Malcolm Cairns.* 777

Session 9.1 Planning and Paying for Roads and Their Operation

- Modeling the Rationalization of Rural Road Networks: The Case of Saskatchewan. *Paul Christensen, James Nolan and Gordon Sparks.* 797
- Planning, Operations and Maintenance of Toll Road Systems. *Ata M. Khan and Mohammad R. Tayyaran,* 812
- User Charges and Cost Recovery: The State of Idaho/American Trucking Association Case. *James R. Jones and Selahattin Bekmez* 828

Session 9.2 Integrated Transportation and Economy Modeling (ITEM) : A Status Report

- Integrated Transportation and Economy Models: Version 0 *Marc Gaudry, et al.* 844
- A computable General Equilibrium Model for Transport in Canada. *Marc Gaudry, et al.* 845
- A Modal Accounting System for Transport with First National and Regional Results. *Marc Gaudry, et al.* 846
- Analysis: Some ITEM System Tools. *Marc Gaudry,et al.* 847

Session 10.1 Changing Policies and Practice in Grain Transport

- The Development of Western Grain Rates. *Jim Riegle* 848
- Regional Railway and Grain System Project. *Roy Atkinson, Javier Caceres, Gary Housch and Darrell Richards.* 862
- Grain Car Canada: Electronic Railcar Market Allocation System. *Barry E. Prentice and Tamara Thomson* 878

Session 10.2 Transportation Safety

- Identifying High-Hazard Locations for Pedestrian and Bicycle Crashes. *Richard A. Raub.* 895
- The Role of Inclement Weather in Heavy Truck Accident Causation: Implications for Driver Training, LCV's and Safety Programs. *J.M.Kosior and Shannon Summerfield.* 896
- Simulation of Heavy Trucks in Inclement Weather. *Shannon Summerfield and J.M. Kosior* 913

Student Prize-Winning Papers from the 1999-2000 CTRF Student Paper Competition

- The A to DC of Order Fulfillment in E-Tailing. *Ignatius V. Hsu, University of British Columbia, Mark Bunting Award for 1st Prize Undergraduate Category* 923
- The Civil Aviation Industry: A Discussion of Air Pollution Problems. *Samantha Sawyer, University of British Columbia, 2nd Prize Undergraduate Category* 939
- A Theoretical Analysis of Low-Cost Entry in the Air Transport Industry. *Bonnie-Lynn Robertson, Albert Stevens Award for 1st Prize Masters Category. Also won the Air Canada Award for the best Air Transport Paper* 951

Exploring the Effectiveness of Land-use Policy on Air Quality: A Simulation

Terence Lam and D. A. Niemeier*
University of California, Davis

Abstract

The relationship between land use and transportation, particularly how to optimize it, has been studied at great length. Less defined is how the relationship changes when there is a concomitant interest in optimizing air quality. This paper develops a simple theoretical model to simulate the impact of land use policies on the transportation activities and changes in air quality. We assume all residents in two neighboring cities commute to work in one employment center. Given planning policies to reduce air pollution by controlling vehicle miles traveled (VMT) or vehicle hours traveled (VHT), firms are assumed to relocate for optimal employment at no cost. Decreases in vehicle emissions due to reduced VMT, however, may drive up housing price in both cities. While the increase in housing prices is a transfer of benefit from travel time savings, the rigidity of housing supply may result in larger-than-optimal housing patterns in both cities. If one of the cities has zoning laws that restricts re-development, residents may be forced to move to the adjacent city; this will tend to dampen the original policy effect. This paper measures the change in emissions level before and after changes in land use policies.

Keywords: Land use, transportation, air quality, simulation

---

* Corresponding author: We would like to acknowledge financial support from NSF, and to thank Dan Sperling and Ken Small for their helpful comments. All the errors and opinions expressed herein are ours.

1.0 Introduction

For more than thirty years, technological innovations have substantially reduced new per-mile vehicle emissions in compliance with the Clean Air Acts (1970, 1977, and 990). In recent years, efforts at improving air quality have shifted more towards changing travel behavior (Howitt and Altshuler 1999). For example, Environmental Protection Agency and Department of Transportation have commissioned several studies to explore the use of market mechanisms and land use patterns to reduce vehicle miles traveled (VMT) and vehicle emissions. (EPA (1997), DOT(1994), Apogee(1994)); one significant focus has been the possibility of reducing vehicle emissions through land use policies.

Until recently, prior studies usually discuss the impacts of land use on VMT and vehicle emissions on policy level, but rarely have data for empirical analysis. (Bae (1993), Frank (1998)) Recent empirical studies, however, show little consensus about the strength of the link between land use and transportation. For example, Ewing and Cervero (2001) survey 50 recent empirical studies and find that there is significant but minor effect of land use patterns on VMT, whereas Greenwald and Boarnet (2001) can only find significant land use effect on walking behaviors but little evidence on other non-walking travel behaviors. Pickrell (1999) also finds little compelling evidence to support a strong link between land use patterns and travel behaviors. This is probably because of the practical difficulty in eliminating the effects of other confounding variables (e.g., population, income, more women in workforce.). Regardless of the strength of the relationship, it is fairly clear that the omission of land use effects in travel demand modeling results in large errors in estimates of VMT and the concomitant emissions. (Rodier and et al. (2001))

In this study, we modify the simulation model by Liu and McDonald (1998) to examine the interactions of land use and transportation and the impacts of policy changes on mobile emissions. Vehicle emissions are assumed to be linearly proportional to VMT, and incorporated in an individual's transportation cost. Residential choice includes decisions about housing costs among 6

zones of equal size in two cities, 3 in each city. The residential demands in each city are negatively affected by price of the city, but positively by the cross-town price. One city is assumed to be more residentially desirable than the other, and thus that city has a higher demand for residential housing. Residents in each city commute the same distance to the central business district before implementation of land use policies. Three scenarios are constructed to explore the mobility of residents when land use policy is enacted in the more popular city to designate the zone of city center for industrial use. Each scenario is compared to a base scenario where the price of living in each zone is equal to their respective marginal costs.

The study results indicate that interactions between mobility and housing markets can significantly impact the effectiveness of land use policies in achieving cleaner air quality. This paper, however, does not consider the relocation cost of firms and people, nor specifies transportation cost as a function of density.

2. Theoretical Model

Figure 1 shows two cities A and B, which are identical in size with suburban center s kilometers apart. Let us assume that people in both cities commute to work in the central business district (CDB), u kilometers away from either city center; for simplicity, the city center is the only access to CBD. Each city is assumed to comprise three equal zones, namely, a_1, a_2, a_3 in city A, and b_1, b_2 and b_3 in city B. The radius of the innermost zone is d, and the radii of each subsequent zones increase such that their areas are the same as those of a_1 and b_1, i.e., πd^2 kilometers squared. The distance of each zone to the CBD (x_{ij}) is thus

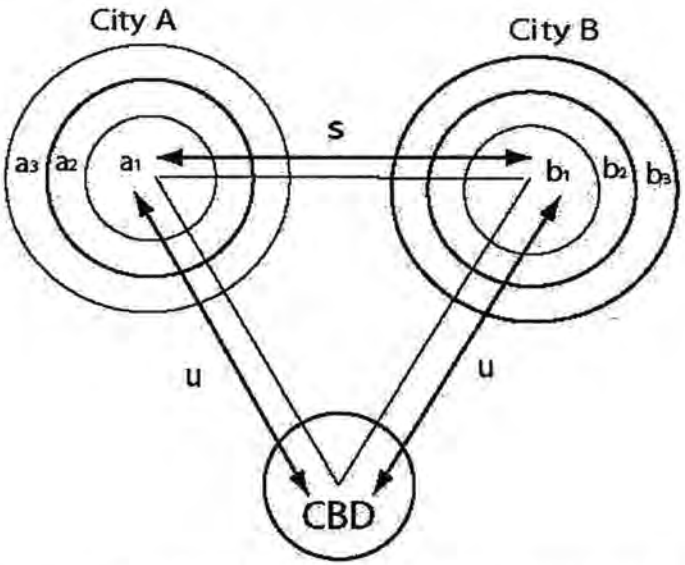
$$x_{i1} = u + d$$

$$x_{i2} = u + \sqrt{2}d$$

$$x_{i3} = u + \sqrt{3}d$$

where i, j are the indices for city and zone respectively, that is, i= a, b; j=1, 2, 3. This implies that $u > \sqrt{3}d$ and $s > 2\sqrt{3}d$.

Fig 1. Residential Location



The residential demand functions for each city are given by $N_a = f_a(P_a, P_b)$ and $N_b = f_b(P_a, P_b)$; the demands are assumed to be negatively related to their own prices, $(\partial N_i/\partial P_i) < 0$, but positively related to prices in the other city $(\partial N_i/\partial P_r) > 0$ for $i \neq r$; and income effects are assumed to be negligible. The inverse demand functions for each city can be derived:

$$P_a = P_a(N_a, N_b)$$
$$P_b = P_b(N_a, N_b)$$

Consequently, the gross social benefits of the two cities, denoted by B, is the sum of the areas of the two inverse demand functions:

$$B = \int_{0,0}^{(N_a,N_b)} P_a(z_a,z_b)dz_a + P_b(z_a,z_b)dz_b \quad (3)$$

which is uniquely defined and independent of the path chosen under the integrability condition $\partial P_a/\partial N_b = \partial P_b/\partial N_a$.

The average cost borne by each individual in each zone consists of transportation and housing costs. The housing cost is assumed to be a function of population in each zone:

$$c_{ij}(N_{ij}) \quad (4)$$

The total cost is thus defined as the sum of all the costs in each zone:

$$C = \sum_{i=a}^{b} \sum_{j=1}^{3} N_{ij} c_{ij} \quad (5)$$

Given equations 3 and 5, the maximum welfare can be formulated as a problem of constrained optimization,

$$\text{Max } W = B - C$$

The technical nuances of the problem are discussed in the appendix.

3.0 Simulation Study

We can simulate the effects of changing land use policy on transportation, and consequently air quality through changes in mobile emissions. To conduct the simulation, we must specify the cost and demand functions and develop the baseline scenario.

3.1 Specification of Cost and Demand Functions

Similar to the monocentric models of Mills (1967) and Muth (1969), the average residential cost is assumed to consist of both transportation and housing costs. The transportation cost is specified as a simple linear function of commute length whereas the housing cost is usually an increasing function of density:

$$C(N_{ij}) = t(x)x_{ij} + C_1\left(\frac{N_{ij}}{L}\right)^k \quad (7)$$

where C_1 and k (k>1) are parameters; k is the reciprocal of the product of price elasticity of housing demand and labor elasticity of housing supply in Muth (1969). $t(x)$ is the per-mile cost of travel, $L_{ij}=\pi i^2$. The marginal cost of living in each city zone is therefore,

$$MC = \frac{\partial N_{ij} C(N_{ij})}{\partial N_{ij}} = t(x)x_{ij} + \frac{(k+1)C_1}{L^k} N_{ij}^k \quad (8)$$

The linear demand functions in both cities are defined as follows:

$$N_a = Q_a - \beta_{aa}P_a + \beta_{ab}P_b \quad (9)$$

$$N_b = Q_b + \beta_{ab}P_a - \beta_{bb}P_b \quad (10)$$

where β_{aa}, β_{bb} and β_{ab} are all positive to ensure negative own-price effects and positive cross-price effects. The derived inverse demand functions are also linear:

$$P_a(N_a,N_b) = \theta_a - \alpha_{aa}N_a - \alpha_{ab}N_b \quad (11)$$

$$P_a(N_a,N_b) = \theta_b - \alpha_{ab}N_a - \alpha_{bb}N_b \quad (12)$$

where

$$\theta_a = \frac{\beta_{bb}Q_a + \beta_{ab}Q_b}{h}$$

$$\theta_b = \frac{\beta_{ab}Q_a + \beta_{aa}Q_b}{h}$$

$$\alpha_{aa} = \frac{\beta_{bb}}{h}$$

$$\alpha_{ab} = \frac{\beta_{ab}}{h}$$

$$\alpha_{bb} = \frac{\beta_{bb}}{h}$$

$$h = \beta_{aa}\beta_{bb} - 2\beta_{ab}$$

We make two more assumptions in the simulation. The first assumption is that city A is more attractive to city B so that the potential demand is higher in city A, $Q_a > Q_b > 0$. The second assumption is that own-price effects outweigh the cross-price effects, $\beta_{aa}\beta_{bb} - 2\beta_{ab} > 0$.

3.2 Base Scenario(BP)

In the base scenario of perfect competition, equilibrium conditions exist when the price of demand in each city equals the marginal cost in each zone of the respective city. That is,

$$P_i^*(N_a, P_b) = MC(N_{ij}) + tx_{ij} \qquad (13)$$

The total number of people living in both cities is denoted by $\overline{N} = N_a + N_b$.

3.3 Land-use policy

In this section we construct three scenarios after policies are enacted to buy back land from residents in zone a_1 in city A, and leased with subsidy to firms whose employees live in city A. The aim of the policies is to reduce vehicle miles traveled (VMT) and thus, mobile emissions. The three scenarios are: a no constraint scenario in which the total number of people is not held equal to $\overline{N}$. The last two scenarios project short-term and long-term changes in populations and prices in each city holding total population equal to $\overline{N}$.

No Constraint (NC)

The optimization is exactly the same as that in base scenario, except that an additional constraint applies,

$$N_{a1} = 0 \qquad (14)$$

The decrease in land supply will induce price increases in both cities. The difference of the resulting total population and $\overline{N}$ can be regarded as individuals who have moved out of the system.

Short Term (SR)

Unlike the preceding scenario (NC), in this scenario we constrain the total number of people in both cities equal that of the base scenario, $\overline{N}$:

$$\sum_{j=1}^{3} N_{aj} + \sum_{j=1}^{3} N_{bj} = \overline{N} \qquad \text{where} \quad N_{a1}=0 \qquad (15)$$

This allows us to study changes in VMT and emissions when people move to city B after the land-use policy. As shown in appendix, maximizing the welfare function of equation (6) in both cities requires $MC_{ij} \geq P_i$; this implies that both cities have to subsidize their residents in each zone to accommodate everyone.

Long Term (LR)

Finally, we study the long-term effects of the land use policy assuming that city A does not increase land supply whereas city B expands to 4 zones. The demand function for city B is also assumed to shift until all the equilibrium conditions as well as the constraint in equation (15) are satisfied, that is,

$$MC_a = P_a(N_a, N_b) = \theta^*_a - \alpha_{aa} N_a - \alpha_{ab} N_b$$
$$MC_b = P_a(N_a, N_b) = \theta^*_b - \alpha_{ab} N_a - \alpha_{bb} N_b$$

where

$$\theta_a^* = \frac{\beta_{bb} Q_a + \beta_{ab}(Q_b + \Delta Q_b)}{h}$$

$$\theta_b^* = \frac{\beta_{ab} Q_a + \beta_{aa}(Q_b + \Delta Q_b)}{h}$$

3.4 Relocation Cost

It occurs that the computation of welfare will be overstated if cost of relocation is ignored, but specification of any cost also seems arbitrary. As a solution to the conundrum, we simply assume the cost as the least amount of money government of city A must pay the incumbent residents for the service of the land in zone a_1 in the base scenario.

$$C_r = N_{a1} P_a \qquad (16)$$

4 Simulation Results

Ideally, values for the parameters can be obtained from prior studies so that the results can be compared to the real world. This proves difficult as specifications of cost and demand functions are not commonly available in literature related to this problem. As a result, we choose parameter values so that the results in the base scenario closely replicate the demographics of two neighboring cities in northern California—Davis and Woodland. These two cities have similar sizes, 22 km^2 for Davis and 23.8 km^2 according to data from US Census Bureau 1990, and are both within 20 kilometers from the employment center in downtown Sacramento metropolis.

The second column of Table 1 lists the simulation results for the base scenario (BP): As discussed earlier, the population in city A and B, N_a and N_b are very close to the true populations in Davis and Woodland (47,209 and 39,802). Since city A is more popular than B, one has to pay roughly 52% more to live in city A than in B ($14,867 to $9,480) annually, or about one-third of the median household income ($29,044[1] in Davis and $31,671 in Woodland.) according to US Census Bureau 1990.

The next three scenarios reflect demographic changes after policy is enacted designating zone 1 in city A for industrial use. Column 3 of the scenario (NP) shows that decrease in land

[1] The median income is lower than that of Woodland because a large fraction of population in Davis city are students of University of California, Davis.

availability increases prices in city A by almost 40% while prices in city B are affected very little. Since the total population is not constant (by assumption), a net out-flow of 5,179 from the two cities or 6.4% of the total population, results. In reality, it is very likely that the policy in a popular city will impact residents in its poorer neighboring city as its "refugees" flee to push up housing prices in the adjacent cities. For example, it's been reported that the residents in Silicon Valley Northern California displace the incumbent poor residents in southern parts of San Francisco city.

The fourth column of Table 1 depicts another hypothetical scenario in which the total population in the two cities are held equal to that in the base scenario (BP). This forces both cities to subsidize their residents in order to accommodate all of them in the short run. Compared to scenario (NC) of column 3, prices in both cities drop. Almost 8.2% (3,839) of the former residents of city A move to city B. If we assume that the distance between A and B is the same as they are to the CBD, i.e. $s = u = 20$, or in simple terms, no change in commute length for those moved to city B, the projected decrease of VMT (residents in city A no longer have to commute to CBD) will still have to be discounted at least by 8.2%. This has not yet taken into account of the welfare loss and the relocation cost of firms from CBD to city A.

In the long run, we expect that, if land supply in city A is not replenished, prices in city A will rise to meet the demand (from \$14,867 in (BP) to \$28,591 in (LR), and supply of land increase in city B to meet the increase in demand. Not surprisingly, the population in city A shrinks 26.2% to 34,786. This, assuming $s = u = 20$ again, will imply downward adjustment of 26.2% in the projected VMT in the long run.

Lastly, it is worth pointing out that the welfare estimates do not account for such costs as relocation of firms and individuals, development of land etc, neither does the paper specify transportation costs as function of density. Regardless of these simplified assumptions, the paper shows that, in a simple set up of two cities, the mobility of people and its subsequent impacts on housing markets can significantly affect effectiveness of land use policies in reducing

vehicle emissions. It is also very likely that the people of lower income might suffer the most—with rising housing prices excluding them from the market.

In conclusion, because of the interactions of the mobility of people and the housing markets, policies, which alter the land-use patterns, may not be a tool as effective or economically efficient as believed in achieving cleaner air quality.

Table 1: Values of Parameters Used in the Simulation

Scenario[a]	(BP)	(NC)	(SR)	(LR)
Demand Functions				
β_{aa}	0.9			
β_{bb}	1.25			
β_{ab}	0.02			
Q_a	2500*3L=60,319			
Q_b	1900*3L=45,842			62,028
Cost Function				
U	20km			
D	1.6km			
T	\$0.35/km			
C_l	\$0.001/km^4			
L	$(1.6)^2\pi$=8.043km^2			
K	2			

[a] **BP**: Base Scenario, **NC** Land-Use Policy for $\Sigma_{i=a,b} N_i \neq \overline{N}$, **SR** and **LR**: Short-Run and Long-Run Land-Use Policy Scenarios.

Table 1 (continued): Simulation Results

Scenario	(BP)	(NC)	(SR)	(LR)
$P_a(\$)$	14,867	20,713	19,066[b]	28,591
$P_b(\$)$	9,480	9,509	6,476[b]	8,070
N_a	47,128	41,867	43,289	34,786
N_b	34,289	34,370	38,128	46,637
ΔN_a	–	-5,261	-3,839	12,342
B(million)	2,759.3	2,666.5	2,724.8	3163.9
C[c](million)	527.55	487.19	555.94	435.38
Cr(million)	–		234.51	
W=B-C	2,231.7	2,179.3	2,168.8	2,728.5
% of W1	100	97.65	97.18	122.3

Appendix: Land-Use Scenario

Since the problem of constrained optimization is similar in each scenario, only the solutions for short run land-use scenario (SR) are derived below. Maximize Welfare *W*

$$B = \int_0^{(N_a,N_b)} P(N_a,N_b)dN$$

$$W = \int_0^{(N_a,N_b)} P(N_a,N_b)dN - C(N_a,N_b)$$

[b] The marginal costs are \$22,116 and \$10,912 for city A and B respectively.

[c] In order to streamline housing and transportation costs for comparison, we consider C as the total cost to live in either city in one year, therefore, the transportation cost is computed as: *t** (Number of working days)* (round trip length). For example, for people in zone 1 of city B, C=\$0.35x220x2x21.6.

subject to the following constraints:

$$N_{ij} > 0$$

$$Ni = \sum_{j=1}^{3} N_{ij}$$

$$P_i \geq 0$$

$$N_{a1} = 0$$

$$\sum_{j=2}^{3} N_{aj} + \sum_{j=1}^{3} N_{bj} = \overline{N}$$

The Lagrangian can be written as:

$$L(N_{ij}, N_a, N_b) = B(N_a, N_b) - C(N_{ij}) - \mu_1[N_a - N_{a1} - N_{a2} - N_{a3}] - \mu_2[N_b - N_{b1} - N_{b2} - N_{b3}] - \lambda_1[N_{a1} - 0] - \lambda_2[\overline{N} - N_a - N_b] \quad (17)$$

The first order conditions yield:

$$\frac{\partial L}{\partial N_{a1}} = -MC_{a1} + \mu_1 - \lambda_1 = 0$$

$$\frac{\partial L}{\partial N_{a2}} = -MC_{a2} + \mu_1 = 0$$

$$\frac{\partial L}{\partial N_{a3}} = -MC_{a3} + \mu_1 = 0$$

$$\frac{\partial L}{\partial N_{b1}} = -MC_{b1} + \mu_2 = 0$$

$$\frac{\partial L}{\partial N_{b2}} = -MC_{b2} + \mu_2 = 0$$

$$\frac{\partial L}{\partial N_{b3}} = -MC_{b3} + \mu_2 = 0$$

$$\frac{\partial L}{\partial N_a} = P_a - \mu_1 + \lambda_2 = 0$$

$$\frac{\partial L}{\partial N_b} = P_b - \mu_2 + \lambda_2 = 0$$

$$\frac{\partial L}{\partial \mu_1} = N_a - N_{a1} - N_{a2} - N_{a3} = 0$$

$$\frac{\partial L}{\partial \mu_2} = N_b - N_{b1} - N_{b2} - N_{b3} = 0$$

$$\frac{\partial L}{\partial \lambda} = -\overline{N} - N_a - N_b = 0$$

Simplification of the above equations gives rise to the following reduced equations:

$$P_a = MC_{aj} - \lambda_2 \qquad (18)$$

$$P_b = MC_{bj} - \lambda_2 \qquad (19)$$

$$N_{a1} = 0 \qquad (20)$$

$$N_a = \sum_{j=1}^{2} N_{aj} \qquad (21)$$

$$N_b = \sum_{j=1}^{3} N_{bj} \qquad (22)$$

$$\overline{N} = N_a + N_b \qquad (23)$$

Equations (18) and (19) imply that both cities have to subsidize their residents so that the same number of people as in the base scenario can be accommodated.

References

1. Apogee Research, Costs and Effectiveness of Transportation Control Measures (TCMs): A Review and Analysis of the Literature," Bethesda, MD., 75, 1994.

2. Bae, Christine Chang-Hee, "Air Quality and Travel Behavior: Untying the Knot," *Journal of the American Planning Association*, Vol.59, No.1, Winter 1993, pp.65-74

3. Ewing, Reid and Robert B. Cervero, "Travel and the Built Environment: Synthesis," Paper Proceedings (01-3515) of the 80th Transportation Research Board, Washington, D.C, 2001

4. Frank, Lawrence D., " Improving Air Quality through Growth Management and Travel Reduction Strategies," *Journal of Urban Planning and Development*, Vol.124, No.1, (March) 1998, pp. 11-32

5. Greenwald, Michael J. and Marlon G. Boarnet, "Built Environments Determinant of Walking Behavior: Analyzing Non-work Pedestrian Travel in Portland, Oregon," Paper Proceeding (01-2792) of 80th Transportation Research Board, Washington, D.C, 2001.

6. Howitt, Arnold M. and Alan Altshuler, "The Politics of Controlling Auto Air Pollution," in: Jose Gomez-Ibanez and William B.Tye, and Clifford Winston, eds., Essays in

Transportation Economics and Policy, Chapter 7, Brookings Institution Press, 1999.

7. Liu, Louie Nan and John F. McDonald, "Efficient Congestion Tolls in the Presence of Unpriced Congestion: A Peak and Off-Peak Simulation Model," *Journal of Urban Economics* No.44, 352-366 (1998)

8. Mills, Edwin S. "An Aggregate Model of Resource Allocation in a Metropolitan Area." *American Economic Review*, No.57,1967, pp.197-210.

9. Muth, Richard., *Cities and Housing*, Chicago: University of Chicago Press, 1969

10. Pickrell, Don, "Transportation and Land Use," in: Jose Gomez-Ibanez and William B.Tye, and Clifford Winston, eds., Essays in Transportation Economics and Policy, Chapter 12, Brookings Institution Press, 1999.

11. Rodier, Carolin J., John E. Abraham, Robert A. Johnston and John D Hunt, "Anatomy of Induced Travel Using an Integrated Land Use and Transportation Model in the Sacramento Region," Paper Proceedings (01-2582) of the 80th Transportation Research Board, Washington, D.C, 2001

12. US Department of Transportation Federal Highway Administration, *The Effects of Land Use and Travel Demand Management Strategies on Commuting Behavior*, February 1994

13. US Environmental Protection Agency., *Guidance on The Use of Market Mechanisms to Reduce Transportation Emissions,* Office of Policy, Planning and Evaluation Environment and Transportation Sectors Division, Washington D.C. 1997

A Model Of Curb Rights In Private Urban Transit Markets

André de Palma and Robin Lindsey

1. Introduction

Urban transit service has been privatized in a number of countries over the last two decades. Local bus service was deregulated in Britain in 1986, and the impacts have been widely studied (e.g. Nash, 1993). Partial transit deregulation has been implemented in the Nordic countries (Alexandersson *et al*, 1998), and privatization is widespread in Latin America (Rebelo and Benvenuto, 1995; Zubieta, 1998; Lee, 1999) and developing countries elsewhere in the world (Reja, 2000). Arguments have been made for deregulation in the United States (Lave, 1985; Viton, 1989; Winston and Shirley, 1998), and limited experiments with deregulation have been conducted in a few areas (Reinke, 1986; Teal, 1986).

Deregulation has had a number of positive impacts in transit operation, including reductions in operating costs, innovations such as the use of minibuses, and expansion of service measured in route-miles. But several problems have also emerged. Firms headrun or jockey for riders by attempting to run their buses just in front of rivals' buses. Jitneys and other informal service providers predate on scheduled service by picking up riders at bus stops just before the scheduled bus arrives. And in attempts to forestall entry, incumbent firms have added buses to their routes and allegedly engaged in other preemptive or predatory behaviour.

Schedule jockeying, frequent schedule changes, and difficulties in obtaining information have left travellers uncertain about bus arrival times. This increases waiting time, and may discourage individuals from traveling by bus. And predation by jitneys reduces the profitability of scheduled bus service and may drive it into insolvency. The resulting loss of service frequency and reliability

provided by buses may reduce demand so much that jitneys also become unprofitable, leaving travellers without any transit service.

To avoid these potential problems with transit service deregulation, Klein, Moore and Reja (1997) (KMR) have proposed a system of property rights that they call curb rights. Each operator offering scheduled service on a route is given the exclusive right to pick up and discharge passengers at specified locations and/or times of day. Exclusive zones for scheduled bus companies can alternate with commons at which jitneys and other free-wheeling operators can stop. The commons facilitates entry, and enhances competition without dissolving incentives for companies to establish scheduled service.

The purpose of this paper is to lay some groundwork for a location-theoretic analysis of curb rights. Consistent with KMR's vision, the ultimate objective is to assess the merits of a system of curb rights in transit markets that are otherwise unregulated. Thus, while fare regulation has often continued in transit markets where limited deregulation and privatization have taken place (Britain being a notable exception), it is assumed that firms compete in fares. It is also assumed that free-entry conditions prevail, despite some evidence that transit markets are not perfectly contestable.

To meet this agenda it is necessary to have a temporal model of fare competition between transit service providers. There are relatively few theoretical studies of price competition in transit services. Evans (1987) develops a Hotelling-type model of horizontal product market competition in which bus companies provide service at different times of day and compete in fares. Transit demand is price-elastic and travellers know the bus schedules/timetables. Ireland (1991) adds a mode choice dimension by assuming that travellers choose between bus and car. Ellis and Silva (1998) focus on headrunning in bus markets. They consider both the case where travellers know the schedule and the case where they do not. In their model all travellers have the same reservation price to use transit. Market demand is therefore price inelastic, and cannot capture the potential downward

spiral in demand that infrequent and/or unpredictable service (and high fares) may induce.

This paper builds on the work of Evans, Ireland, and Ellis and Silva. A single transit route is considered with one stop at which transit riders can board. Each traveller has a preferred time at which to travel, and incurs a schedule delay cost from starting either earlier or later. Transit demand per unit time is a decreasing function of the expected cost of a trip — which includes the fare, expected schedule delay cost, and expected waiting time at the stop.

Three competitive markets are analyzed in this setting. In Market 1, bus companies provide service according to a schedule that travellers know. Market 2 is the same as Market 1 except that travellers do not know the schedule. Finally, in Market 3 jitneys provide transit service with random arrival times. To enable a quantitative comparison of the three markets, a numerical example is used and equilibrium headways, fares, and welfare (social surplus) are computed.

While the analytical focus of the paper is on fare competition, three potential roles of a system of curb rights are examined. It is well known that competitive free-entry markets can produce either too much or too little product variety. In a transit market, variety means service frequency (the inverse of headways). Evans, Ireland, and Ellis and Silva all show that equilibrium service frequencies in their respective models are not optimal. This is also true of the bus markets considered here. A system of time-based curb rights could be used to control headways by stipulating when buses may arrive at the stop. Where free-entry competition results in excessive crowding, a welfare improvement can be gained by forcing an increase in headways, although with competitive fare setting the second-best (subject to breakeven) optimum cannot be reached. The welfare gain from time-based curb rights in Markets 1 and 2 is evaluated using the numerical example.

A second potential role of curb rights is to support a competitive equilibrium with a stable schedule. As explained below, such an

equilibrium is problematic in Market 2. Finally, curb rights may be used to restrict transit service to efficient operators. This is investigated by comparing social surplus with jitney-only service in Market 3 to social surplus with bus-only service in either Market 1 or Market 2. Such a comparison provides some indication as to whether curb rights should be restricted to buses, or whether jitneys should be allowed too. But since the comparison is limited to all-or-nothing service of either type, it does not address the potential merits of mixed service.

The paper is organized as follows. In the next section the basic features of the model that are common to all three markets are described. Sections 3-5 analyze competition in Markets 1, 2 and 3 respectively. Section 6 provides a summary. Due to space constraints, derivations and details are omitted. These can be found in De Palma and Lindsey (2000).

2. The Basic Model

The basic model is adopted from Evans (1987). There is a transit route on which travellers can board at a single location. Each traveller has a preferred time to board. Preferred times are distributed uniformly in the population over an infinitely long day. A traveler who boards earlier than desired will be said to "board early"; "boarding late" is defined analogously. Travellers incur a schedule delay cost of β for each minute that they board early or late. Demand to travel per unit time has the exponential form $L\exp(-g/v)$, where L is a measure of demand intensity, v is a parameter, and g is the generalized cost of travel which includes the fare, f, schedule delay costs and waiting time costs. Travellers are assumed to be indifferent between riding a bus and riding a jitney, and also indifferent between firms that provide transit service of either type.

The cost of operating a bus is $F + mq$, where F is fixed cost, q is the number of passengers, and m is marginal cost per passenger. The cost

for a jitney is $F_J + mq$, where F_J may be lower than F. Capacity constraints do not bind in equilibrium for either buses or jitneys. Free entry prevails. Given an infinitely long day and no costs to rescheduling buses, transit companies earn zero profits in equilibrium.

3. Bus Service with Uniform Headways and a Known Schedule

Market 1 is analyzed in Evans (1987) and the description here is accordingly brief. Evans assumes that firms each operate one bus. But the equilibrium he derives also applies if firms operate more than one bus as long as no firm serves the stop with consecutive buses. Firms play a two-stage game: choosing schedules for their buses in Stage 1, and setting fares in Stage 2. In equilibrium headways between buses are uniform. Firms charge a common fare that is an increasing function of the headway. Travellers incur an (unweighted) average schedule delay cost of $\beta h/4$. They do not incur waiting time costs because they know when buses arrive.

Following Evans, market efficiency is measured by social surplus per unit time and per unit of potential demand. Evans shows that free-entry equilibrium does not result in socially optimal headways. Except in thin markets characterized by low demand intensity (low L), time headways are too short and corresponding service frequencies too high. This bias towards excessive entry could be rectified by a system of time-based curb rights whereby access to the bus stop is rotated between firms at a prescribed rate; *e.g.* every 20 minutes. The optimal rotation period is determined by maximizing social surplus with respect to h, given competitive fare-setting and a nonnegative profit constraint. If the optimal headway is less than the free-entry level, the profit constraint is binding, the free-entry equilibrium prevails, and no increase in surplus is achieved. But if the optimal headway exceeds the free-entry level, the profit constraint is not binding and surplus can be increased. The excessive profits that accrue to firms could be extracted through a system of competitive bidding for the curb rights.

As a benchmark with which to compare the free-entry equilibrium and system of time-based curb rights, the second-best optimum subject to a breakeven constraint is also computed. To convey an idea of magnitudes, as well as to enable a quantitative comparison with the equilibria of Markets 2 and 3, a numerical example is developed. Parameter values in US dollars are set at v =\$3.00/trip, β =\$9.00/hr, m=\$0.25/passenger and F=\$40.00 per bus trip.[1] Demand intensity, L, is treated as a variable in order to explore the effects of market thickness.

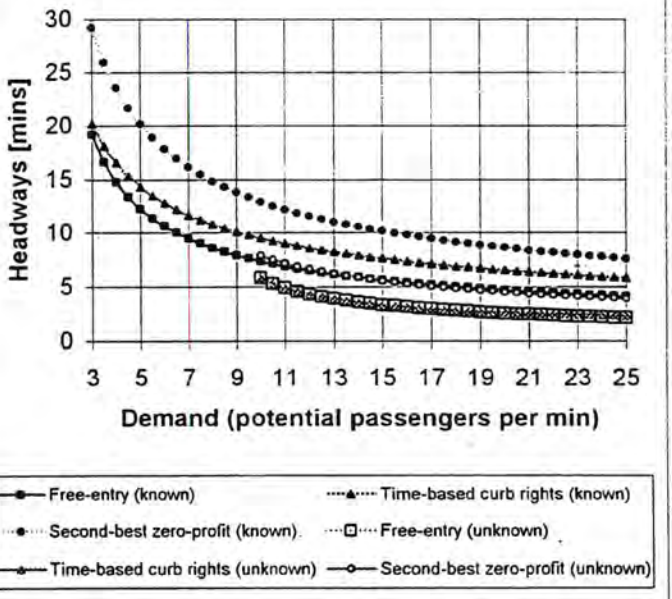
At low demand intensities, long time headways between buses and correspondingly high fares are required for firms to cover costs. As demand intensity rises (the market gets thicker), economies of traffic density are gained because of the fall in fixed operating costs per passenger. As a result, in all three administration regimes equilibrium headways and fares fall, and social surplus rises. Headways and social surplus are shown in Figures 1 and 2 respectively. (To economize on space, these figures include the results for Market 2 with unknown schedules, which will be considered later. The three curves for Market 1 are labeled "(known)".)

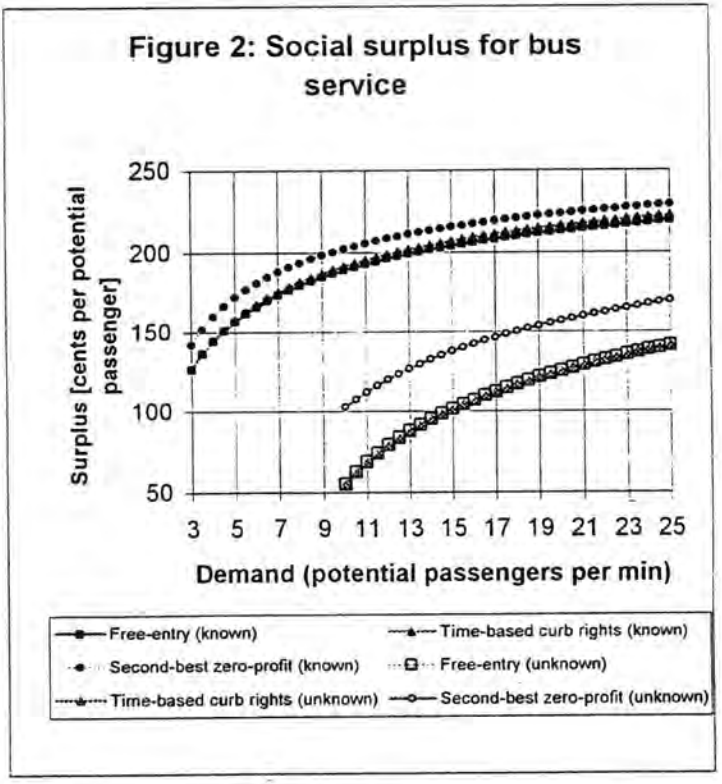
Within the range $L \in [3,25]$ free-entry equilibrium drives headways well below (and fares well above) their second-best optimal levels. This is consistent with the fact that competition in British bus markets has been manifest more in frequent service than in low fares. The optimal time-based curb rights regime entails a headway longer than the equilibrium headway. But because fares continue to be set competitively, fares end up even higher than with free entry. As a result, social surplus is only slightly higher than with free entry. The gain ranges from zero for $L \leq 3$ to 36% of the difference between

[1] Evans adopts a numerical example with parameter values v=£0.60/trip, β =£1.80/hr., m=£0.10/passenger and F=£8.00. These numerical values are scaled up here by a factor of 5 to adjust (approximately) for exchange rate and inflation. The exception is parameter m, which is increased by a factor of 2.5. This is line with relative values for fixed and per-passenger bus costs in Sweden cited by Jansson (1997).

surplus in the second-best optimum and free-entry equilibrium at $L = 25$.

Figure 1: Headways for bus service





4. Bus Service with Uniform Headways and an Unknown Schedule

Travellers do not always bother with transit timetables because of the inconvenience of memorizing them, or keeping a current timetable at hand and learning how to use it. This is particularly so when transit service is frequent and waiting times are correspondingly short.

When travellers do not know the schedule, they minimize their expected travel cost by arriving at the stop before their preferred boarding times. Given a uniform distribution of desired boarding times, this means that they arrive at the stop in a continuous flow, rather than *en masse* as in Market 1. It is then most profitable for a firm to serve the stop just before a rival arrives when the largest numbers of passengers are waiting.

Ellis and Silva (1998) show that under these conditions there does not exist a pure-strategies equilibrium to Stage 1 of the game in which firms choose schedules. What happens in such a situation is not obvious, but in practice headrunning and irregular service can prevail, which is inconvenient for travellers. In what follows it is assumed that a system of time-based curb rights is established to support a stable schedule, with uniform headways in order to minimize travellers' expected waiting time and schedule delay costs.

The determination of equilibrium fares is complicated by the fact that travelers prefer to board a bus that has just arrived rather than wait for the next, or switch to an alternative transport mode. Consequently, a bus operator can raise its fare slightly above the going rate without losing business. Indeed, if there are many operators, one can raise its fare without perceptibly reducing market demand. As a result, there may not exist a price equilibrium.[2]

To circumvent this problem it is assumed that just two firms provide alternating service on the route.[3] If one cuts its fare deeply enough, it can then induce travellers to wait for one of its buses if one of the rival's buses arrives at the stop first. The solution concept adopted is Morgan and Shy's (1998) undercut-proof equilibrium. According to this concept, a firm sets a price that maximizes its profits subject to

---

[2] This problem has long been recognized in the context of cruising taxi markets. A recent synthesis and review of the taxi literature is found in Bergantino and Longobardi (2000).

[3] This is consistent with experience in British bus markets after deregulation in which no more than a few companies ever compete on the same route.

the constraint that a rival does not find it profitable to undercut it. In the undercut-proof equilibrium, average waiting time is one half the headway between buses. Average waiting time cost is therefore $(\alpha/2)h$, where α is the unit cost of waiting time. Travellers also incur an average schedule delay cost of $(\beta/4)h$. Average trip cost is thus $f+(\alpha/2+\beta/4)h$. By comparison, the (unweighted) average cost of a trip in Market 1 is just $f+(\beta/4)h$.

To implement the numerical example, a value for waiting time of α=\$0.30/min. is chosen. This is twice the unit cost of schedule delay, broadly in line with the limited empirical evidence (Nuzzolo and Russo, 1996).

Equilibrium headways and social surplus are plotted in Figures 1 and 2 respectively. In contrast to Market 1, where free-entry equilibrium is viable at demand intensities above $L \equiv 1$, Market 2 is viable only for $L \geq 9.5$. This difference is attributable to two factors. First, as just noted travellers incur not only an average schedule delay cost of $(\beta/4)h$, but also an average waiting time cost of $(\alpha/2)h$. Second and relatedly, travellers' uncertainty about bus arrival times weakens fare competition so that fares are considerably higher. As a result, the market is viable only at demand intensities high enough to support frequent service and short waiting times.

As demand intensity rises, headways fall and surplus rises in all three administration regimes. At any given demand intensity, headways and surplus are considerably lower than in Market 1. The curb rights regime again features longer headways than in free-entry equilibrium. But the gain in social surplus is very small, so that in Market 2 use of curb rights to optimize service frequency may not be worthwhile.

5. Jitney Service with Random Time Headways

Unlike buses, jitneys typically do not adhere to a schedule but rather provide informal freewheeling service. To capture this it is assumed that the timing of jitney arrivals at the stop is governed by a Poisson

process with parameter $1/h$, so that time headways between jitneys and passenger waiting times are both distributed exponentially with a mean value of h. As in Markets 1 and 2, h is determined by the free-entry zero-profit condition.

In Market 2 the problem of fare determination was resolved by assuming a duopoly in which firms have an appreciable influence on the average fare, and can attract travellers if they cut their fare enough. In the absence of route associations or large fleets of jitneys, however, this is not a reasonable assumption about a jitney market.

An alternative approach adopted here is motivated by the fact that two or more jitneys can arrive at a stop nearly simultaneously. If the jitneys become visible from the stop before they arrive, travellers can safely bypass the first jitney and take the next without incurring much delay. A jitney that raises its fare even slightly above the going rate then risks the occasional loss of business to a rival following just behind it.[4]

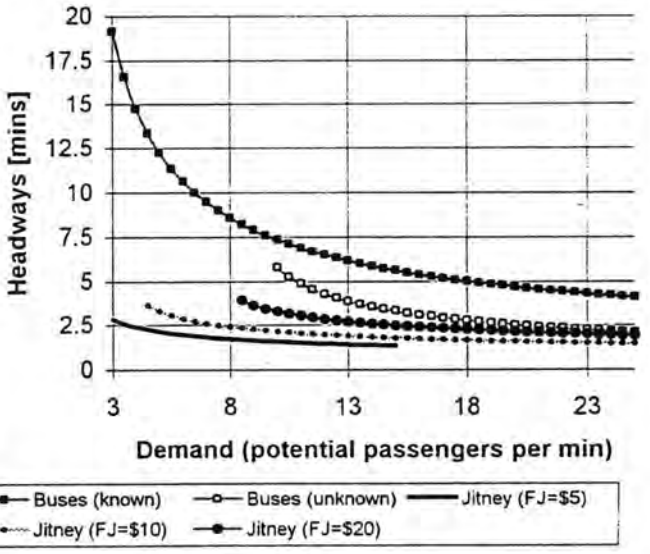
For a wide range of parameter values a Nash equilibrium fare exists that is an increasing function of h, and a decreasing function of the length of time that jitneys can be seen prior to reaching the stop. In the free-entry equilibrium expected trip cost is $f + (\alpha + \beta \ln 2)h$. By comparison, expected trip cost in Market 2 is $f + (\alpha/2 + \beta/4)h$. For the same headway, the expected cost of a jitney trip is more than twice the corresponding cost of a bus trip. The inefficiency of jitney service derives from the random headways.

To compute free-entry equilibrium it is assumed that jitneys become visible two minutes before they arrive. A value must also be chosen for F_J: the fixed cost per jitney. Jitneys are typically more

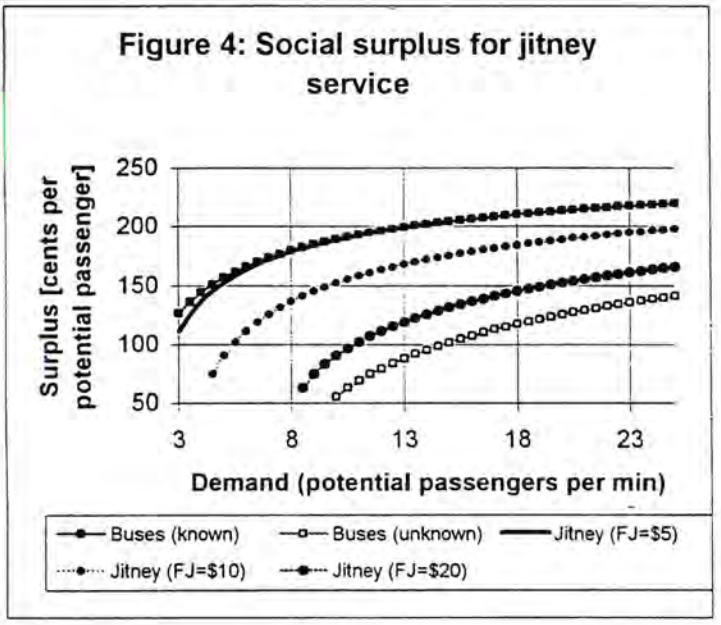
[4] The inspiration for this approach comes from Schreiber's (1975, p.271) observation that if taxi fares rise enough, empty taxi cabs become so numerous that more than one will sometimes be within hailing distance of a customer, and fare competition may ensue.

economical than buses because of their smaller size and cheaper labour, and operating costs can be quite low for used vehicles. Three values of F_J are considered: \$20, \$10 and \$5. Jitney service is viable for demand intensities above $L \cong 8.2$, $L \cong 4.1$ and $L \cong 2.1$ respectively (the minimum intensity is proportional to fixed cost).

Figure 3: Headways for jitney service



Equilibrium headways and social surplus are plotted in Figures 3 and 4, with the corresponding free-entry equilibrium values for Markets 1 and 2 superimposed to facilitate comparison. Headways with jitney service are lower than for bus service, and are driven down to about



80 seconds with F_J = \$5 and L=15.[5] Social surplus is quite sensitive to F_J. With F_J = \$20, surplus is closer to surplus with an unknown bus schedule, whereas with F_J = \$5, surplus is indistinguishable in thicker markets from surplus with a known bus schedule. This demonstrates that, despite the inefficiency of random arrival times, jitney service can be competitive or superior to bus service provided fixed costs are low enough to support short headways and low fares.

[5] A Nash fare equilibrium exists with the base-case parameter values only for headways greater than about $1.5\hat{h}$. This condition is violated with F_J =\$5 for demand intensities more than 15. For this reason, results for F_J =\$5 are reported only for $L \in [3,15]$.

6 Conclusions

To avoid some problems with deregulated transit markets, Klein, Moore and Reja (1997) have proposed a system of curb rights whereby operators are given exclusive rights to load and unload passengers at specified locations and times. The purpose of this paper is to lay some foundations for a model of curb rights by studying fare and schedule competition under free-entry conditions. Three market settings are considered: bus service with uniform time headways when travellers know the schedule (Market 1), bus service with uniform headways when travellers do not know the schedule (Market 2), and jitney service with random time headways (Market 3).

Market 2 operates less efficiently than Market 1 because travelers have to wait for buses, and also because traveler uncertainty about bus arrival times weakens fare competition. For given vehicle costs Market 3 is in turn less efficient than Market 2 because travelers suffer greater schedule delays and waiting times. Nevertheless, this disadvantage of jitney service may be outweighed by lower vehicle costs, which translate into shorter average headways.

In the course of the analysis three potential roles of a system of curb rights were identified. First, in all but thin markets free entry in Markets 1 and 2 results in excessive service frequency. This inefficiency can be alleviated by using time-based curb rights to rotate access to the bus stop at a period equal to the optimal headway. Second, Market 2 is prone to headrunning and changeable timetables, both of which impair service performance. Time-based curb rights could be implemented to enforce stable service with uniform time headways. Finally, curb rights can be used to limit access to firms that provide the most effective service.

The analysis in this paper could be fruitfully extended in several directions. Most pressing is to examine the possibility of combined bus and jitney service. Combined service has two opposing effects on welfare. The benefit is that it can better match the needs of a heterogeneous traveling public. Travellers with a low cost of waiting

time may favour a cheap-priced jitney over a bus. Alternatively, those with a high cost of in-vehicle travel time may favour high-priced express service in uncrowded vehicles that make few stops.[6] The main disadvantage of two (or more) modes of service is that, because the market is split, economies of traffic density are sacrificed.

References

Alexandersson, G., Hulten, S. and Folster, S. (1998) The effects of competition in Swedish local bus services. **Journal of Transport Economics and Policy** 32(2) 203-219.

Bergantino, A.S. and Longobardi, E. (2000) Rethinking deregulation. The case of the taxi industry. 6th RSAI World Congress. Lugano.

De Palma, A. and Lindsey, R. (2000) Fare competition and curb rights in urban transit markets. 6th RSAI World Congress. Lugano.

Ellis, C.J. and Silva, E.C.D. (1998) British bus deregulation: Competition and demand coordination. **Journal of Urban Economics** 43(3), 336-361.

Evans, A. (1987) A theoretical comparison of competition and other economic regimes for bus services. **Journal of Transport Economics and Policy** 21, 7-36.

Gronau, R. (2000), Optimum diversity in the public transport market, **Journal of Transport Economics and Policy** 34(1), 21-42.

Ireland, N.J. (1991) A product differentiation model of bus deregulation, **Journal of Transport Economics and Policy** 25(2), 153-162.

Jansson, J.O. (1997) Theory and practice of transport infrastructure and public transport pricing, in G. de Rus and C. Nash (eds.), **Recent Developments in Transport Economics**, Ashgate: Aldershot, 74-134.

Klein, D.B., Moore, A. and Reja, B. (1997) **Curb Rights: A Foundation for Free Enterprise in Urban Transit**. Brookings Institution Press, Washington, D.C.

---

[6] Gronau (2000) provides an insightful analysis of this dimension of heterogeneity for the case of public-sector transit provision.

Lave, C.A. (ed.) (1985) **Urban Transit: The Private Challenge to Public Transportation**. Ballinger: Cambridge, Mass.

Lee, D.A. (1999) Models of government regulation of privatized transit systems in five Latin American cities. **Transportation Quarterly** 53(1), 21-39.

Morgan, P. and Shy, O. (1998) Under-cut equilibria, working paper.

Nash, C.A. (1993) British bus deregulation. **Economic Journal** 103, 1042-1049.

Nuzzolo, A. and Russo, F. (1996) Stochastic assignment models for transit low frequency services: Some theoretical and operative aspects, in L. Bianco and P. Toth (eds), **Advanced Methods in Transportation Analysis**. Springer-Verlag, Berlin. 321-339.

Rebelo, J.M and Benvenuto, P.P. (1995) Concessions of busways to the private sector: The Sao Paulo metropolitan region experience. World Bank working paper.

Reinke, D. (1986) Update on taxicab and jitney regulation in San Diego, **Transportation Research Record** 1103, 9-11.

Reja, B. (2000) The industrial organization of public transport in developing countries: Bus franchising in the presence of a competitive informal transport sector, 79th Annual Meeting of the Transportation Research Board. Washington, D.C.

Schreiber, C. (1975) The economic reasons for price and entry regulation of taxi-cabs. **Journal of Transport Economics and Policy** 9(3), 268-293.

Teal, R.F. (1986) Impacts of comprehensive urban transportation deregulation in Arizona. **Transportation Research Record** 1103, 11-14.

Viton, P. (1989) The desirability of private-sector urban transportation, **International Journal of Transport Economics** 16(2), 131-163.

Winston, C. and Shirley, C. (1998) **Alternate Route: Towards Efficient Urban Transportation**. Brookings, Washington, D.C.

Zubieta, L. (1998) A network equilibrium model for oligopolistic competition in city bus services. **Transportation Research** 32B(6), 413-422.

Trucking in Canadian Cities: Planning For The Future

Clarence G. Woudsma
Centre for Transportation, Van Horne Institute, Calgary
Department of Geography, University of Calgary, Calgary, Alberta, Canada, T2N 1N4, cwoudsma@ucalgary.ca

Abstract

The movement of goods or freight is an essential component of any urban area and the majority of that movement is truck related. However, it can be argued that our understanding of this essential component has not received the level of attention devoted to people movement, nor, more importantly, kept pace with the massive changes impacting upon it. This paper addresses to key aspects of the goods movement question: 1) what is our current level of understanding related to urban goods movement and how does it vary 2) what are the implications of this understanding with respect to the issue of global climate change.

The first component involves a comparative analysis of goods movement planning studies drawn from a number of cities. The studies provide a basic level of understanding on issues such as time of day activity, trip characteristics, and vehicle characteristics. However, they fall short in areas related to understanding the key relationships that underlie goods movement activity including those related to land use/transport interaction. It is argued that basic research into an integrated modeling effort that is freight related is required to address pressing issues associated with congestion, air quality, and related emissions questions (climate change). On this front, this paper also addresses the Canadian response to the Kyoto protocol and the nature of this response with respect to the goods movement sector.

Context [1]

In Canada, and elsewhere in the developed world, it is safe to say that our research efforts directed at understanding urban goods movement (the bulk of which travels by truck) have not kept pace with our efforts directed at people movement. The following quote was issued over 20 years ago but through the course of this paper, its relevance for today will be established.

"There is no reliable, comprehensive understanding of the cost of the goods movement system in urban areas. There is no systematic comprehension at the Metropolitan level of the complex interactions between land use, type of goods moved, nature of transfer operations, relationships between goods type, truck type and operating patterns etc. . Hence, it is extremely difficult to develop analytical techniques (whether by hand or by computer) to analyze the effects of possible changes to the present system". (Transport Canada, 1979, p.32)

The reasons associated with this shortcoming are varied. First, congestion issues emphasize cars since they make up the bulk of traffic flow. Depending on the time of day, trucks only make up from 7 to 18% of traffic (DOT, 1995). Subsequently, the priority is to understand automobile movements. Second, the elements of freight movement and the actors involved are much more complex, which makes it a difficult subject to research. Third, there is a lack of reliable data on urban freight movements (Byrne and Mulhall, 1995). Government agencies have attempted to provide data at the intercity or regional scale, but intracity data is much harder to obtain. The data that are available are often viewed as unreliable because transportation firms that are surveyed are skeptical about the eventual use of the data. Finally, the phrase "important yet mundane matters" encapsulates the political view toward urban freight issues (O'Sullivan, p.1662, 1992). Firms that operate in the city are reluctant to speak of their concerns for fear of further government intervention. Local citizens are not concerned about inefficiencies in the local transport network and the eventual indirect price they pay.

Politically, issues related to urban freight transportation do not carry much weight, unless safety is involved.

This discussion does not imply that there is no research effort. The goal of this paper is to establish what we do know about trucking in Canadian cities. A secondary goal is to relate this understanding to our current climate change strategies.

Planning and Urban Goods Movement (UGM)

Table 1: Planning Objectives

Objectives	Examples of Specific Elements
Economic	• Develop and improve the freight system towards improving the local, regional, and national economy • Focus on trade exposed sectors with urban economy
Efficiency	• Minimization or reduction in transportation operation costs related to en-route travel • Focus on congestion; role of freight and costs to freight • Road network deficiencies including road design and geometry, maintenance, signage, local area traffic management and arterial capacity • End point costs associated with loading and unloading, parking, terminal activities, hours of operation and site access and egress • Energy costs associated with vehicle speed and character and shipment type
Road Safety	• Minimization of property damage, injury and fatality related accidents • Focus on policy related to traffic management, road design, vehicle design, driver training and land use
Environmental	• Focus on mitigation of noise, air, and vibration pollution • Perceived threat of large vehicles and intrusive activity in residential areas
Infrastructure and Management	• Explore government influence through regulations, pricing controls, taxation and investment • Road construction and maintenance and its relationship with freight sector
Urban Structure	• Focus on interaction between freight facilities and urban structure including interaction between freight and urban structure, city size and its effects on freight costs, and freight as a user of urban land
Derived from Ogden, K.W. (1992) Urban Goods Movement: A Guide to Policy and Planning, Ashgate Publishing Limited, London, pgs 89-132	

Table 1 summarizes the elements of the major policy and planning objectives associated with UGM (Ogden, 1992). Odgen's categories span issues involving the broader economy through specific business management issues (cost minimization = efficiency) to external impacts of safety and pollution. These objectives represent an ideal set of objectives that are typically not fully addressed in most urban areas. They may be recognized, but the challenge is how do we make headway on these issues?. How clear is our understanding of the role of trucks in congestion and their contribution to urban air quality issues? Which of these objectives take priority? To explore these questions, a review of Canadian truck planning studies is now presented.

Comparing Truck Planning Efforts

This section compares a number of planning studies that have been carried out for major Canadian cities (N.D. Lea and Associates Ltd., 1974: Metropolitan Toronto, 1987, 1988: Vancouver City Engineering Department, 1990 : Delcan Corporation, 1991a, 1991b: City of Edmonton, 1996). These studies represent the most recent completed for major cities (above 500,000 population) in Canada. It is not an exhaustive listing but is meant to provide a representative selection of efforts.

One of the first observations is that these planning studies tend to be sporadic, rather than regularly scheduled. For example, Calgary's first was in 1974, followed by 1986 and the City is currently conducting a major trucking study. Secondly, the objectives in these studies range from Vancouver's narrow focus on describing movements (where, when, and what), to a broader look at system deficiencies (Ottawa, Edmonton) and linkages with local economic conditions (Toronto). None cover all the topics in Table 1, yet each deals with some of the topics.

In terms of methodology, the focus in all cases is on the trip itself rather than on the good being moved. Traditionally, this has been the most common metric studied although there is some debate as to the

merits of using commodity vs. shipment. The planning studies reviewed tend to vary in the level of spatial aggregation employed but can be generalized as using large zones. The common approach is to define the sample population as being all licensed commercial vehicles within the urban area and to mail out a trip diary survey to a selection of those vehicle owners. Part of the problem in this approach is that vehicles active in an urban area may not necessarily be licensed within that area (Lau, 1995). Some studies supplemented the basic survey with firm level interviews of management staff of transportation providers or firms that utilized transportation services. Due to the derived nature of the demand for services, the decision making involved goes beyond the transportation firm.

The heterogeneity associated with trucking is evident in the range categorizations employed. Some studies adopt a body style categorization while others use a weight class (e.g. Light /Medium/Heavy). The variation in defining vehicle classes is indicative of the lack of consistency across planning approaches. "Trip generation" is one of the 4 phases of the classic urban transportation modeling system that includes trip generation, trip distribution, mode choice, and network assignment (Ozutar and Williams, 1994). Most of the truck planning efforts utilized basic modeling of trip generation. The trip distribution component would be calibrated using origin-destination data from the surveys completed. This highlights another point of contention in UGM research: these models were developed for the study of automobiles, not freight movement. Less elegant approaches convert truck movements to automobile movements via axle equivalents in order to operationalize the models. Ozutar and Williams (1994) provide a more detailed discussion of model adaptation for freight.. Anecdotal evidence suggests that planning agencies are aware of other modeling approaches but lack the resources and the necessary data to tackle them.

The relationship between urban structure and the movement of freight is recognized in a number of the studies. For example, the role of land use in establishing trip generation rates and the development of

clusters of land use activity related to freight movement. However, none of the studies explicitly deal with land use planning in the context of being utilized to address freight movement issues. That is, an integrated effort to model land use and freight transportation. To illustrate this, the influence of land use on trip generation may be somewhat understood --- x number of square meters of x land use would result in x number of trips per day --- however, the same cannot be said for the influence of freight trip activity or system performance on land use evolution. Discussion of the global competitiveness of urban areas has emphasized transportation factors as playing an important role in city competitiveness. These factors include the overall quality of accessibility in an area, intermodal opportunities, developable land, connectivity to other regions, and supporting logistics business activity. Yet, there is little evidence to link these factors back to land use evolution or location choice within these cities.

Planning Studies: Conclusions

The analysis and conclusions of the Canadian planning studies outlined in the previous section are difficult to address. Given the range of approaches used and the differences in sampling methodology, comparisons are difficult. There is it little commonality between these reports in terms of measures reported. In the following section, their analytical results will be explored.

Table 2 displays a few common statistical results from these studies. They are sorted on the basis of city /metropolitan region size because the size of study area (measured by population) is hypothesized to relate to characteristics of UGM. The Average Daily Trips per Truck is a measure of the daily activity of the goods movement fleet. With the exception of Toronto, there appears to be a trend of increasing trip rates with decreasing city size. The highest trip rate is associated with the smallest city – 1974 Calgary. The date of that particular study is also a factor. At that time, the freight movement industry was highly regulated and unionized and the high trip rates may reflect inefficiencies associated with this situation. Alternatively, the

smaller city size may account for shorter and more frequent trips. This does not indicate more effective utilization, however, since information on load factors is not provided.

Table 2: Findings of Canadian Truck Planning Studies

Urban Area	Pop. Millions	Study Year	Average Daily Trips Per Truck	Avg. Km Per Truck Trip	Truck % of traffic flow	Truck Trips Per Person	Truck Trips Per Job
Toronto (Metropolitan Toronto, 1987)	3.4	1987	9.7	24.3	10-20%		
Vancouver (Vancouver City Engineering Department, 1990)	1.6	1990	7.7	12.5	3%		
Ottawa/Hull (Delcan Corporation, 1991a, 1991b)	0.9	1989	8.7	19.8	11%	0.16	0.34
Edmonton (City of Edmonton, 1996)	0.86	1996	8.4	13.8	7.3 - 10.3%	0.17	0.36
Calgary (City of Calgary 1984)	0.67	1984	9.1				
Calgary (N.D. Lea and Associates Ltd., 1974)	0.4	1974	12.6	5.8			

The argument for shorter and more frequent trips in smaller cities is weakened upon review of the Average Kilometre per Truck Trip variable. The inverse relationship between trip length with city size is

not evident as that between the number of trips and city size. Exceptions such as Ottawa/Hull at 19.8 km and Vancouver at 12.5 km may be due to the boundaries of the study area or the urban form. Ottawa/Hull is a regional study that includes activity beyond the defined borders of the urban area and Vancouver is a city that can be described as being relatively more compact (higher density) by North American standards.

The final three variables in Table 2 illustrate a wider range of findings and lack of commonality. The truck percentage of traffic flow figures range from 3% to 20% and vary with local conditions and time of day considerations. The truck percentage also depends on the classification of a goods movement vehicle. For example, light trucks, or pick-up trucks often make up a large portion of the "fleet" and if they are not included, the truck percentage of traffic flow is decreased. The Truck Trips per Person and Job provide an aggregate measure of activity that may be useful in comparing different cities.

Another common result found in these studies concerns truck traffic variations with the time of day. Figure 1 depicts the hourly percentages of truck activity for selected Canadian studies and exemplifies a common mid-morning, mid-afternoon peak with a considerable lack of activity outside of normal business hours. The overall trend in Figure 1 suggests that trucks start their activity earlier in the day and generally try and avoid the peak commuter periods. As indicated previously, depending on the time of day, trucks only make up from 7 to 18% of traffic (DOT, 1995). Together, these facts suggest that trucks are not a big part of the congestion issue. However, the move to just-in-time delivery and the demands of integrated supply chain management has meant that the flexibility to travel at off-peak times is eliminated (Barnes, 2000). Subsequently, trucks are an increasing component of the traffic stream during peak periods. They also move much differently than automobiles and have a more severe impact on traffic flow. There have been calls to establish separate truck lanes in order to deal with this issue.

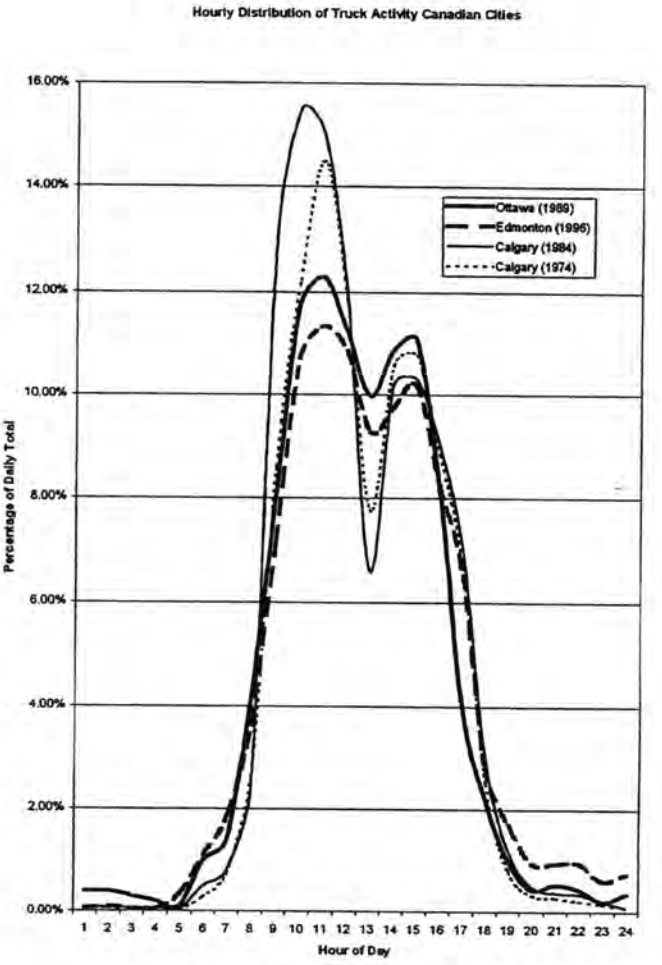


Figure 1: Comparing Time of Day Activity Patterns

In summary, these few quantitative results discussed in this section suggest that city size and form obviously play a role in influencing the characteristics of UGM, as do the definition of the spatial scope and key elements such as defining what constitutes a commercial vehicle. It can be argued that there are more similarities inherent in the goods movement system but the inconsistencies in how the studies are carried out and reported on, and the obvious difficulty with time frames precludes recognition or understanding of them.

In order to get a better understanding of the scope of these studies, it is useful to consider a broader range of their conclusions and compare them to the issues and objectives of Table 1. A selection of other key results from the Canadian studies is therefore presented in Table 3.

The age of fleet has important implications for air quality issues in urban areas and the Edmonton study suggests that a considerable portion of the "fleet" is quite old – 42% older than 11 years. This point was not addressed in the other studies, which is unfortunate. The Edmonton study also examined accident activity, reporting a significantly lower accident rate for commercial vehicles; a fact that has widespread recognition. In terms of vehicle use, the Vancouver result of heavier vehicles making less frequent and longer trips was echoed in the other studies. They also tend to be more active outside of regular business hours.

With respect to land use, the studies vary in what are considered the prime land use activities associated with traffic generation. The Ottawa/Hull study suggests that industrial land uses are high generators while their CBD and residential areas are low. The Calgary study in contrast suggests that retail and residential generate more traffic. The variation in the levels of zonal aggregation in these studies may contribute their differing results, especially in areas of mixed land use activity. A number of the studies employed traffic generation models that are not based on physical land use area but on employment levels in zones. A related result from the Toronto study provides a contrast between trip generation rates in CDB versus suburban locations. The order of magnitude difference is not entirely

Table 3: Major Findings: Canadian Studies

Study	Major Findings
Ottawa (1989, Goods Movement Study)	• Industrial zones are high generators of goods movement while CBD office and residential are low • Truck traffic is growing at a rate slightly higher than automobile traffic • 10% of transport costs related to congestion (50 million annually)
Vancouver (1990, Truck Study)	• heavy trucks make less trips with larger trip lengths – light trucks make more trips with shorter trip lengths • dominant trip destination a wholesaling or warehouse facility (20%)
Edmonton (1996, Truck Route Study)	• 26% of fleet older than 16 years, and 42% is older than 11 years • 70% of light trucks carry half loads • congestion costs estimated at 22 million annually, (1.3% of total costs) • trucks comprise 5% of all vehicles yet were found to be involved in under 2% of accidents • accident rate for trucks is 7 times lower than for automobiles (1.5 accidents per million kilometres of travel
Toronto (1987 Goods Movement Study)	• Trip generation is 1750 trips/sq.km in CBD, 100 trips/sq.km. outlying regions • 20% of firms operated 24 hours a day • truck operating costs do not rise in proportion to vehicle size • congestion mentioned as biggest issue • transportation is a key relocation issue for core located firms • growth in movements much higher in suburbs than CDB • expect 80% increase in goods movement traffic over the next 10 years • price elasticity of truck travel is .9 with respect to time (10% increase in travel time leads to a 9% drop in demand)

Toronto (1987 Goods Movement Study)	• congestion costs estimated at 30% of total cost of moving goods (projected cost to society, 15 Billion over 10 years) • 30-50% of small package movements in CBD delivered by couriers (foot, bike, transit) or taxis • 3% of trips involve Dangerous Goods
Calgary (1974)	• 80% of trips involve outlying areas, 11% the CBD • retail land use is the largest trip generator (25%) followed by residential (19%) • light trucks make most trips but have lowest average load factors

surprising but it raises questions about activity in suburban nodes versus suburban areas in general. The inclusion of large tracts of residential land in a zone can contribute to much lower rates of traffic generation on an area basis. A subsequent study in this city has chosen as a focus "activity nodes" and "freight centres" rather than the traditional zonal approach (Gorys et al., 1999).

This question of land use and traffic generation is most relevant to discussions of congestion. As suggested by the survey results from the Toronto study, congestion is the most common issue in urban goods movement. This ties back to the earlier discussion in this paper about the planning objectives for UGM and the focus on efficiency. The estimation of the costs of congestion in relation to the overall cost for goods movement is presented for the Toronto, Ottawa and Edmonton studies at 30, 10 and 1.3 % respectively. This result supports the argument that the costs of congestion for freight movement increase with increasing city size. This result has important implications in terms of the evolution of urban areas and the competitiveness of large metropolitan areas at a global level. It is also of prime concern given the apparent continued growth of our reliance on road transport and by no coincidence, congestion in Europe and North America (Pucher and Lefevre, 1996)

Discussion and Conclusions

What is known about urban goods transportation is the result of studies to address the planning needs of particular urban areas. These studies are not carried out on a regular basis and do not use standard definitions and approaches. Traditional issues have included efficiency and routing and a minimization of the negative aspects of freight movement. The studies reported here all include very specific recommendations on ways to deal with these issues but tend to be more reactive than proactive, responding to an obvious problem. Key issues related to 1) land use/transportation interactions and 2) the tremendous changes impacting goods movement are dealt with in a very superficial manner or not at all.

The urban goods movement literature fails to adequately address the link between land use and goods movement (Miller et al., 1999; Transportation Research Board (TRB), 1991). A handbook on the "Characteristics of Urban Freight Systems" exemplifies the state of knowledge of urban goods movement (DOT, 1995) in the U.S. and is consistent with the comparative results presented in this paper. General information on the percentage of freight vehicles on urban roads at various points in time throughout the day and crude estimates of the number of freight trips associated with various land use types is available. However, by their own admission, they rely on outdated studies conducted in the 1970's (DOT, 1995). This fact ignores the massive changes that have occurred with respect to goods movement between the late 1970's and the present.

This lack of understanding has serious consequences given that from a "Global Warming" standpoint, freight transportation is a major contributor (27% of Canada's total). In the extensive consultation process associated with the landmark report, "Transportation and Climate Change, Options for Action", it became clear that the lack of attention to land use issues was again a major source of concern for participants. Similarly, the research efforts on trucking were directed at measures with a more technological focus (more efficient vehicles,

scrapage programs). A consistent lament in this research and elsewhere concerns the lack of reliable data on trucking at both the regional and local scales. If we are going to be able to increase our understanding of trucking in urban environments, the land use / transport relationship must be explored in a meaningful way and more reliable and consistent data gathering on the industry must take place.

Endnote 1: This research is sponsored by the Social Sciences and Humanities Research Council of Canada. This paper is an edited and altered version of a more detailed paper forthcoming in Urban Studies.

References:

Anderson, W.P., Kanaroglou, P.S. and E.J. Miller (1996) Urban Form, Energy and the Environment: A Review of Issues, Evidence and Policy, Urban Studies, 33, pp.7-35.

ASCE, American Society of Civil Engineers, Committee on Urban Goods Movement (1989) Issues and Problems of Moving Goods in Urban Areas, Journal Of Transportation Engineering, 115(1), pp. 4-19.

Barnes, D. (2000) Shipping Trends Have Trucks Facing More Congestion, Transport Topics, Jan. 24, 2000

Byrne, G.E., and S.M. Mulhall (1995) Congestion Management Data Requirements and Comparisons, Transportation Research Record, no. 1499, pp. 28-36.

Chatterjee, A, Fisher, P.J. and Stanley R.A. (1989) Goods Transportation in Urban Areas, Proceedings of the Fifth Conference sponsored by the Engineering Foundation. New York: American Society of Civil Engineers.

City of Calgary (1984) 1981 Truck Travel Survey. Calgary: City of Calgary

City of Edmonton (1996) Truck Route Study. Edmonton: City of Edmonton Transportation Department.

Delcan Corporation (1991a) National Capital Region Goods Movement Study: Summary Report. Ottawa: Trans: A Joint Technical Committee on Transportation Systems Planning.

Delcan Corporation (1991b) National Capital Region Goods Movement Study: Technical Appendices. Ottawa: Trans: A Joint Technical Committee on Transportation Systems Planning.

DOT, U.S. Department of Transportation (1995) Characteristics of Urban Freight Systems, Office of Environment and Planning, report DOT-T-96-22, Washington D.C.: Federal Highway Administration.

FHA, Federal Highway Administration (1998) U.S. Freight: Economy in Motion:1998, U.S. Department of Transportation, Washington D.C.: Federal Highway Administration.

FHA, Federal Highway Administration (1999) Heavy-Duty Truck Activity Data. Washington D.C.: Office of Highway Information Management

ECMT, European Conference of Ministers of Transport (1995) Transforming the Structure of the Freight Transport Sector, Report of the Ninety-ninth Round Table on Transport Economics, Paris: OECD Publications Service.

Gorys, J., Poulos, N., Gray, R. and Ruggero A. (1999) Defining a Strategic Freight Network for the Greater Toronto Area, in Proceedings of the Transportation Research Board 78th Annual Meetings, Washington D.C.: Transportation Research Board .

Hunt, J.D., Kriger, D.S. and Miller, E.J. (1999) Current Operational Urban Land-Use Transport Modelling Frameworks, in Proceedings of the Transportation Research Board 78th Annual Meetings, Washington D.C.: Transportation Research Board .

Lau, S.W. (1995) Truck Travel Surveys: A Review of the Literature and State of the Art, Planning Section, Oakland : Metropolitan Transportation Commission.

Metropolitan Toronto (1987) Metropolitan Toronto Goods Movement Study: Technical Report. Toronto: Metropolitan Toronto, Roads and Traffic Department.

Metropolitan Toronto (1988) Metropolitan Toronto Goods Movement Study: Executive Summary. Toronto: Metropolitan Toronto, Roads and Traffic Department.

Miller, E.J., Kriger, D.S. and Hunt, J.D. (1999) A Research and Development program for Integrated Urban Models, in Proceedings of the Transportation Research Board 78th Annual Meetings, Washington D.C.: TRB

NCHRP , National Cooperative Highway Research Program, (1995) Characteristics and Changes in Freight Transportation Demand, Project 8-30, Washington D.C.: National Cooperative Highway Research Program.

N.D. Lea and Associates Ltd. (1974) A Profile of Urban Goods Flow in Calgary, Urban Transportation Research Branch of the Canadian Surface Transportation Administration, Transport Canada, Ottawa Canada

Ogden, K.W. (1992) Urban Goods Movement: A Guide to Policy and Planning. London: Ashgate Publishing Limited.

Ortuzar, J., and L.G. Willumsen (1994) Modelling Transport. Toronto: John Wiley.

O`Sullivan, P. (1992) A review of "Urban Goods Movement: A Guide to Policy and Planning, by K.W. Ogden,", Environment and Planning A, 24(11), pp. 1662-1663.

Pucher J. and Lefevre C. (1996) The Urban Transport Crisis in Europe and North America. London: Macmillan Press Ltd..

Transport Canada (1979) Urban Goods Movement Research: A Framework and Results, Urban Transportation Research Branch of Canadian Surface Transportation Administration Report TP 1834, Ottawa: Transport Canada.

TRB, Transportation Research Board (1991) Transportation, Urban Form, and the Environment, TRB Special Report 231, Proceedings of a Conference, Beckman Centre, Irvine California, Washington D.C. TRB National Research Council.

Vancouver City Engineering Department (1990) Truck Study. Vancouver: Vancouver City Engineering Department, Greater Vancouver Regional District.

Mediation and Regulatory Efficiency[i]

The resolution of conflict within a society and the systems established to resolve that conflict, are crucial for effective governance. The resolution of conflict is the underlying rationale behind systems of administrative justice. To this end, governments create judicial systems and specialized administrative tribunals, such as the Canadian Transportation Agency (Agency) to resolve conflict. The Agency's role as a conflict resolver is heightened by its additional responsibility to provide regulatory oversight of the transportation sector.

Much of the regulation in the transportation sector focusses on providing a dispute resolution mechanism to parties. The Agency handles numerous regulatory disputes each year. The cost of these disputes, from the perspective of all of the parties involved, is significant - not only in monetary terms, but also in terms of the opportunity cost of resources. Disputes cause parties to direct scarce resources towards an adjudicative dispute resolution process and away from the transportation system and the organizations' strategic objectives.

It is not only disputes that appear before administrative bodies that prevent organizations from using their scarce resources optimally. Private negotiations of regulatory and quasi-regulatory disputes are often ineffective due to poor historical working relationships between the parties, which cause negotiations to become hostile and unproductive. Even if negotiations are successful, unresolved conflict may undermine

[i] The opinions expressed are those solely of the authors and do not reflect the opinion of the Canadian Transportation Agency

Arun Thangaraj
Christa McClellan
Gerry Nera

the preexisting and future working relationships. Parties often continue to conflict with each other even after a negotiated agreement has been reached. This conflict may extend to one or both parties attempting to subvert the agreement altogether. This has a high cost and has the effect of diverting scarce resources away from where they are needed most.

The significant costs associated with adversarial negotiation and adjudication has led to an increased focus on collaborative processes. Collaborative dispute resolution processes are believed to reduce the costs and preserve valuable relationships between parties. In this paper, we will explore our understanding of negotiation and adjudication. We will then examine whether mediation can be used to improve the efficiency of the dispute resolution process, specifically as it relates to the rail transportation sector.

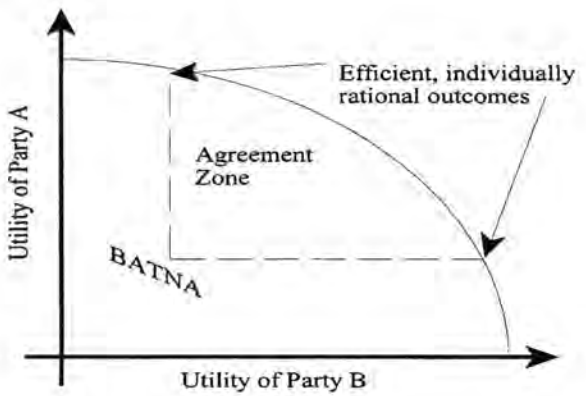
Negotiation

The first step in dispute resolution almost always takes the form of a negotiation. Parties in a dispute will often attempt to negotiate a solution. Noise disputes are a good example of this, where railway companies and affected parties get together to attempt to find a solution to their concern. The current regulatory regime is based on the desire for parties to settle disputes without regulatory intervention. This is premised on the belief that negotiated agreements and the process through which they are arrived at are efficient. We assert that negotiated agreements and the process used to attain them may not always be efficient.

Game theory suggests that two parties in a non-cooperative game will engage in strategic behaviour in order to maximize their outcome based upon their understanding of what they believe the other party will do. The actions and strategies of one party depends on the actions and

strategies of the other party. This model assumes that the negotiators will behave in a rational fashion and focuses on the prediction of whether or not an agreement will be reached and the nature of that agreement. When there is complete information and an agreement zone between the parties exist, economic models predict that an agreement will occur. The agreement will also be Pareto optimal, that is no party can be made better off without the other party being made worse off in terms of utility.[1]

Bargaining Set for a Two-party Negotiation[2]



The above diagram illustrates this point. Each party's utility derived from a given outcome is plotted on each axis. There also exists an

agreement zone, which is the area where a negotiated settlement can be struck. The boundaries of that agreement zone, plotted by the dotted lines, represents the parties BATNA - their best alternative to a negotiated agreement. It is reasonable to assume that no party would voluntarily accept an agreement that leaves themselves worse off than their BATNA. Each point on the curve of the potential outcomes of the negotiation within the agreement zone (between the arrows) are said to be efficient.[3]

Nearly all economic and game-theory models of negotiation assume that an efficient settlement is reached. This assumption is based on the fundamental assumptions of efficiency and the belief that inefficient outcomes are incompatible with rational behaviour by well informed negotiators.[4] The current regulatory framework is based on this assumption - that if left to their own devices, parties can negotiate an economically efficient outcome.

However, there is much evidence to suggest that actual negotiator behaviour does not comply with the assumption of rationality. This evidence suggests that even small deviations from rational behaviour may prevent negotiated agreements from being reached even when there is an agreement zone. Irrational negotiator behaviour may also cause an agreement to be Pareto inefficient.[5] Often behavioural irrationality can be considered rational negotiating behaviour. Brams states, "...posturing, emotional outbursts, and even outright deception are often the best prescription for doing well in negotiations."[6] These types of negotiation tactics are often counterproductive as meaningful gains are often forgone.[7] Young argues that, "[b]ecause there is a natural incentive for the negotiators to misrepresent and exaggerate their bottom lines, and they all know this, they may be unable to overcome their well founded suspicions and reach an agreement that really is there - if only they could discover it."[8]

A secondary problem for the model is that while the outcome of a negotiation may be Pareto optimal, it may not be fair. Pareto optimality does not mean that the outcome is 'good' in any absolute sense. Many economists and commercial negotiators believe that fairness is an irrelevant criteria. When consideration for the fairness of outcome is ignored, the power relationship between the parties becomes determinative of the outcome. This leads parties to use irrational behaviour and negotiating tactics to gain advantage over the other, further undermining the ability to achieve an efficient outcome. While fairness is not an economic criteria for efficiency, notions of fairness are deeply ingrained in all individuals. If a party's notion of fairness with respect to a particular agreement is violated, it is very likely that the agreement itself will eventually be violated. This implies that the initial agreement could not have been efficient to begin with.

Adjudication

Adjudication is the most commonly used dispute resolution mechanism when negotiation fails. Parties turn to courts and administrative tribunals to provide the wisdom of Solomon for disputes they cannot resolve themselves. The process of adjudication requires that parties advance rights-based claims against each other, with the party who advances the best position being the winner.

The major shortcoming with adjudication as a means of dispute resolution is evident in basic economic theory. Adam Smith's theory of the invisible hand essentially argues that an equilibrium produced by competitive markets will exhaust all possible gains from exchange. This theory lends itself quite nicely to the examination of disputes. A dispute is essentially a conflict over a trade-off. A dispute over the value at which a railway line is transferred to governments centres around the inability to tradeoff an amount of money for a quantity of assets.

Although negotiations are not always 'efficient', in an ideal negotiation, parties are able to agree on the trade-off based on their individual indifference curves and achieve a Pareto optimal outcome. This solution is seen to be 'economically efficient' as parties are able to determine their trade-offs based on their own preferences.

When parties are unable to agree on the trade-offs, they usually request an adjudicated solution. Adjudicative solutions do not consider the efficiency of the solution for both parties; instead, they allocate the trade-offs for the parties based on the judgement of an outside arbiter. The difficulty with this practice is that institutions rarely have a complete understanding of the shapes and locations of individuals' indifference curves which determine their individual preferences. As the dispute before an adjudicative body is rights-based, the decision by that body is a decision that is often only rights-based. While individuals may consider their legal rights or power important, these dimensions may only have a small impact on their overall indifference curve. Therefore, it is unlikely that a decision by an adjudicative body, which may exclude factors that impact on the underlying sources of the dispute (such as an issue of trust between the parties), will be an optimal solution.

A second shortcoming with an adjudicative solution relates to uncertainty. In the above diagram mapping the bargaining set of a two-party negotiation, all possible outcomes in a negotiation are represented by the space under the curve. A negotiated process narrows down the potential outcomes by providing an agreement zone. By agreeing to negotiate, parties eliminate the probability of arriving at a solution outside of the agreement zone. In an adjudicative setting, what the parties would or could agree to is irrelevant. The major considerations in an adjudicative process are precedents and legal principles that are applicable in determining a solution. Therefore, the number of potential outcomes increase. No longer are possible outcomes limited to those in

the agreement zone. Now, all points under the curve represent a potential outcome. Bodie *et al.* present a simple utility equation as follows: $U=E(o) - \frac{1}{2}A\sigma^2$, where E(o) is the expected outcome with a variance of σ^2 and A being an individual's risk aversion.[9] This equation illustrates that utility is increased by high expected outcomes and diminished by high risk. Even if an individual expects an outcome that lies on the utility boundary, given the uncertainty of an adjudicative decision, the value of σ^2 will be high. Therefore, an individual who is risk neutral or risk averse (i.e $0 \leq A \leq \infty$) would find their utility reduced due to the uncertainty of the adjudicative process. Given the fundamental assumption of economics that individuals will act to maximize their utility, a rational economic actor would not choose to enter into an adjudicative process unless they were almost certain of the outcome and could assign all other potential outcomes small probability values.

The economic inefficiency of adjudication does not mean it is an inappropriate dispute resolution mechanism. The focus of an adjudicative process is to determine a legal solution to a dispute. While this may not be optimal, in some cases it is. There is often need for precedent or social learning that can only come from an adjudicated decision. We will discuss this further in our conclusion to the paper.

Impact of Inefficiency

The preceding sections have argued that traditional negotiation practices and adjudicative decisions may not be as efficient as economic models predict. It is difficult to quantify these expenditures with any degree of precision given the number of disputes and parties involved. Railway companies do not track their expenditures that are specific to dispute resolution. In the case of organizations where individuals volunteer their time, it becomes more difficult to attach a dollar amount to the time and

effort expended in resolving disputes that are formally before the Agency for an adjudicative decision. That said, railway companies we interviewed have conservatively estimated that they spend up to 1% of total rail expenditures on resolving disputes. This could potentially translate to several million dollars per year for each company. This figure excludes the cost of the time of very senior officers, lost business opportunities, or the resource and financial costs from negotiations leading into these disputes.

Our research also found that certain types of disputes are reoccurring. It was conveyed to us that the ongoing relationships in this sector have not been addressed within the current dispute resolution procedures or the regulatory framework. The adjudicative process is believed to be a process where, in the end, there is a winner and a loser. Our interview subjects indicated that such a process is not conducive to fostering future relationships and negotiations.

Mediation

The above analysis indicates that there are efficiencies to be gained in the way disputes are handled in the rail transportation sector. The question is, how do we address the inefficiency? We propose that the answer lies in the creation of a dispute resolution system that assists parties to achieve economically efficient outcomes.

In the previous sections on negotiation and adjudication we have identified two significant problems. Negotiation may not lead to a solution due to negotiator behaviour, and even if a solution is reached, it may not be optimal or fair. If negotiation fails, parties may elect - or are forced - to have their dispute adjudicated. The ensuing problem is that parties lose control over the agreement and, therefore, face the

potential for a utility reducing decision. We believe that both problems can be addressed through the use of mediation.

Mediation has been defined as, "...the intervention in a negotiation or a conflict of an acceptable third party who has limited or no authoritative decision-making power but who assists the involved parties in voluntarily reaching a mutually acceptable settlement of issues in dispute."[10] Assisting parties to voluntarily reach agreements is crucial to addressing the shortcomings of negotiation and adjudication. The role of a mediator is not to impose, decide or evaluate. This would cause the same problems as identified in the adjudicative model. For an efficient solution to be reached, parties must have control over the solution, and they must be rational economic actors. We explore both of these criteria below.

The role of the mediator is to give the parties control over their dispute and to encourage them not to take actions that are counterproductive to an agreement. Control is essential for two reasons: parties know what their own interests are, and when parties have control, agreements are more stable. Negotiation is based on the premise that when parties act rationally, they will achieve an agreement that fulfills the best interest criterion. However, when parties do not act rationally, negotiations become unsuccessful, and the control over the dispute is transferred into an adjudicative process, where the interests of the party are weighted less than rights-based concerns. Mediation is a process that allows parties to retain control over their dispute by forcing them to consider their complete interests.

Economic models of negotiation focus almost exclusively on tradeable goods or benefits. However, disputes are seldom only about tradeable assets or considerations. In almost all disputes, there is an underlying current of mistrust, mis-communication, unspoken concerns and

unarticulated desires. Fisher and Ury have called these *interests* or "the silent movers" in a dispute.[11] If the positions of parties are the sole concern in a negotiation or adjudication, the underlying sources of the conflict - or the non-tradeable goods - will not be addressed. Failure to consider these in a negotiation leads to irrational behaviour which inhibits parties from reaching an efficient outcome.[12] The focus of disputants is their conflicting positions. One of the disadvantages with this approach is that positional bargaining nearly always entails one party losing more than they win. This is clearly seen in adjudicated decisions or negotiations between parties of unequal strength. The goal of the mediation process and the mediator is to encourage the disputing parties to consider the underlying interests of each side. In order to achieve this outcome, parties must alter their perception of the conflict. When successful, the mediation process results in agreements that create additional value and are more stable than those imposed by an adjudicative body or negotiated to settlement, as these approaches do not address underlying issues.

Economic theory indicates that a free exchange between two rational parties will exhaust all potential gains from trade. In essence, the argument is that no additional value exists. However, an interest-based dispute resolution process, such as mediation, leads to value creation. Attempting to determine optimal trade-offs in a negotiation or adjudication may result in a Pareto optimal solution if all the gains from trading are exhausted *and* if all potential trading opportunities are known. Mediation and interest-based processes change the way disputes and conflict are viewed, and as a result, they have the potential to create additional value and additional opportunities for trade-offs not previously envisaged. The creation of value is difficult, if not impossible, through traditional dispute resolution processes, as courts exist primarily to enforce existing legal rights which are based on the status quo.[13]

Bounded rationality constrains the way individuals and organizations view conflict and the solutions to that conflict. For example, when solutions to noise disputes between railway companies and neighbourhoods are conceptualized, the focus is almost always on noise abatement measures, financial compensation or establishing different operational procedures in the railway yard. In settling the dispute through a negotiated or adjudicative process, trade-offs are made that attempt to balance the desires of the community and the operational realities of the railway company. However, noise disputes may have as much to do with communication and the relationship between the parties as they do with the noise produced. The failure to resolve these underlying issues indicates that all possible trade-offs were not considered by the negotiators or the adjudicative body. Therefore, all gains from exchange would not be exhausted, and the utility of both parties will be sub-optimal.

If negotiation or adjudication resulted in efficient outcomes, the agreements and decisions would be stable. By stable, we mean that because parties have exhausted all gains from trade and maximized their utility, they will voluntarily adhere to the agreement or decision. However, we know this is not always the case. In positional bargaining, which focusses on a narrow definition of issues and where fairness concerns are ignored, a party who has unresolved interests or believes the outcome to be unfair would feel that the conflict was not resolved, even though an economically efficient agreement has been reached or a judicially imposed decision has been rendered. There is a tendency to dismiss this residual displeasure to be no more than a party not accepting the outcome of a negotiation. However, it is often the case that this residual displeasure will result in poor future working relationships. Residual displeasure potentially leads to the re-occurrences of conflict, as parties breach or push the limits of the agreement. This is clearly inefficient as resources are again diverted away from where they could

be used best to re-resolve the initial dispute. Organizations which use mediation as a dispute resolution tool have indicated to us that the agreements reached through a mediated process have high degrees of voluntary compliance. Accordingly, we believe that mediated agreements will have higher degrees of compliance than negotiated agreements that are viewed as unfair by one or both parties. Although compliance with judicial decisions is high, due to the sanctions for non-compliance, we believe that parties to mediated agreements will comply on their own volition. This is because parties have more control over the outcomes and tend to be more satisfied with a process that completely resolves the dispute by addressing their underlying interests.

Mediation places disputing parties in the role of joint problem solvers, thus, giving them control over the resolution of their dispute. Mediation also establishes a process whereby all aspects of a dispute - the positional and underlying - are to be resolved. It leads to outcomes that are more stable than those achieved through positional bargaining or adjudication. These benefits of mediation allow parties to mutually pursue a solution that not only lies within the agreement zone, but lies on the boundary of the efficient rational payoff curve. In the section on negotiation, we argued that when notions of fairness are ignored, sub-optimal outcomes are usually the consequence. The mediation process is more likely to result in a fair outcome, not because the mediator determines or imposes it, rather, because when parties better understand their own interests and the interests of each other, they are more likely to achieve a fair outcome and one which is more likely to be stable and serve as the basis for ongoing future relationships.

Conclusion

It is clear that mediation can result in more efficient outcomes, more stable agreements and improved working relationships. However, the

question of the impact that mediation can have on the efficiency of the transportation sector or regulatory framework has not yet been answered. The efficiency gains from mediation come from improved dispute resolution, which leads to conflict avoidance. The potential gains that can be achieved by improving the efficiency of the dispute resolution process are enormous.[14] As stated previously, railway companies in Canada spend up to 1% of their total rail expenditures on resolving conflict. However, this is only one side of the dispute equation. Short-line railway companies, shippers, municipalities and other parties who engage in disputes with railway companies also face significant costs. These costs prevent these organizations from using funds more productively. If we were to add the opportunity costs of all resources expended on disputes for all stakeholders in the sector, the dollar amount would be much higher and the impact of the inefficient use of those resources would be that much more significant. Mediation provides a way, not to eliminate, but to reduce the cost of resolving conflict. This is accomplished by addressing all sources of conflict - the positions *and* underlying interests of the parties.

It is clear that mediation is beneficial, but what about the regulatory bodies who are, as stated in the introduction, essential to effective governance? If they impose inefficient outcomes, why bother having them decide at all? The answer is that the alternative is not any better. We have demonstrated that negotiation often leads to sub-optimal agreements and an exacerbation and escalation of conflict. That said, mediation is not a panacea. It is not applicable in all circumstances, only appropriate ones. There are matters for which adjudication is clearly the most appropriate course of action. Matters establishing precedent or having broad public interest are not appropriate matters for mediation. Neither are complex technical or complex multi-party/multi-interest disputes. The reality is that even a mediation process does not always produce agreements. Therefore, the need for adjudicative bodies is

clear. The primary purpose of administrative and regulatory bodies is to resolve conflict within societies and mitigate against market abuses by dominant industries. As long as industries have market dominance and there are systemic conflicts within a society, administrative tribunals and regulatory bodies will continue to exist.

Therefore, we believe that the existence of systemic conflict and the resulting need for adjudicative and regulatory bodies necessitates that, for sectors needing regulation, mediation occur within a regulatory context that is backed by an adjudicative process. This allows a clear BATNA for parties in a mediation to be established. Parties in a mediation will know that if they continue to employ bargaining strategies that are counterproductive to settlement, the uncertainty of an adjudicative decision awaits them. Parties are, therefore, more mindful of having the power to exercise choice taken away from them if they enter into a mediation process with less than good faith or act in ways to prevent a win-win solution from being achieved.

The challenge, therefore, is to make their application of regulation and adjudication more efficient. Opening the adjudicative and regulatory process to allow parties the opportunity for mediation is crucial for this to occur. Parties who have a dispute have the opportunity to collaboratively determine a solution that meets their positional concerns and underlying interests before it becomes imposed. The obligation of an adjudicative body is to resolve conflict. We believe that this obligation extends beyond the resolution of disputes that are properly before it. This obligation extends to the ongoing relationship between the disputing parties. A decision which extinguishes conflict for only a short period of time, only to have it reignite at a point in the future, indicates a failure to fully resolve the dispute. The adjudicative body also has an obligation to ensure that disputes within its area of expertise,

even those that are not before it, are resolved before relationships between parties become embedded with systemic conflict.

The policy statement contained in Section 5 of the *Canada Transportation Act* states that an efficient transportation system is vital to maintain the economic well being of Canada and its regions. The existence of disputes indicate that resources are being diverted away from transportation and towards litigation and confrontation. Regulatory bodies have begun to recognize that this diversion of resources is inefficient and have introduced mediation as a dispute resolution tool in the rail and marine transportation sector.

Effective governance requires an administrative body, which has an obligation to reduce conflict and promote efficiency, to be responsible for the mechanism that will reduce conflict and improve efficiency. We believe, therefore, that regulatory and adjudicative bodies such as the Canadian Transportation Agency are in ideal positions to provide mediation services, as they have expertise in the technical elements of disputes *and* have developed skills in enabling parties to identify and resolve underlying sources of conflict. When implemented effectively, a dispute resolution system, which has mediation as a fundamental component, will increase the efficiency of how regulation is applied and, in the case of the Agency, of the transportation sector as a whole.

Endnotes

1. Max H. Brazerman and Margaret A. Neale, "Negotiator Rationality: The Interactive Roles of Prescriptive and Descriptive Research" in Peyton Young ed., Negotiation Analysis, (Ann Arbor: The University of Michigan Press, 1991), pg. 109

2. H. Peyton Young, "Negotiation Analysis", in Peyton Young ed., Negotiation Analysis, pg. 3.

3. Ibid.

4. V.P Crawford, "A Theory of Disagreement in Bargaining", in Ken Binmore and Partha Dasgupta eds., The Economics of Bargaining, (Oxford: Basil Blackwell Ltd., 1987), pg. 122.

5. Max H. Brazerman and Margaret A. Neale, "Negotiator Rationality: The Interactive Roles of Prescriptive and Descriptive Research" in Peyton Young ed., Negotiation Analysis, pg. 110.

6. Stephen J. Brams, Negotiation Games: Applying Game Theory to Bargaining and Arbitration, (New York: Routledge, 1990), pg. xiii.

7. H. Peyton Young, "Negotiation Analysis" in Peyton Young ed., Negotiation Analysis, pg. 12.

8. Ibid, pg. 14.

9. Zvi Bodie et al, Investments, (Toronto: McGraw-Hill Ryerson, 2000), pg. 151-152.

10. Christopher W. Moore, The Mediation Process: Practical Strategies for Resolving Conflict, (San Francisco: Jossey-Bass Publishers, 1996) pg. 15.

11. Roger Fisher and William Ury, Getting to Yes: Negotiating an Agreement Without Giving In, (New York: Penguin Books USA, 1991), pg. 41.

12. These behaviours are outlined in Part I of Max Bazerman and Margaret A. Neale, Negotiating Rationally, (New York: The Free Press, 1992) and in Max Bazerman and Margaret A. Neale, "Negotiator Rationality and Negotiator Cognition: The Interactive Roles of Prescriptive and Descriptive Research", in H. Peyton Young ed., Negotiation Analysis.

13. Jeffrey W. Stemple, "Beyond Formalism and False Dichotomies: The Need for Institutionalizing a Flexible Context of the Mediator's Role", in Florida State Law Review, (Vol. 24:949, 1997), pg. 976

14. V.P. Crawford, "A Theory of Disagreement in Bargaining", in Ken Binmore and Partha Dasgupta eds., The Economics of Bargaining, pg. 122.

IMPORTANCE DES MODIFICATIONS À LA LOI SUR LA CONCURRENCE PORTANT SUR LES COMPORTEMENTS D'ÉVICTION ET LES ORDONNANCES PROVISOIRES RELATIVEMENT À UN SERVICE AÉRIEN

Jacques Picard
Département Stratégie des affaires
Université du Québec à Montréal

I. Introduction

Suite aux préoccupations sérieuses causées par la situation financière précaire dans laquelle se trouvait l'entreprise Canadian Airlines Corporation (Lignes aériennes Canadien International Limitée) le gouvernement du Canada adoptait le 13 août 1999 en vertu de l'article 47 de la Loi sur les transports du Canada[1], un décret pour une durée de 90 jours suspendant l'application de certaines dispositions de la Loi sur la concurrence[2]. Dans la même foulée, le Ministre des transports David M. Collenette entreprenait un processus ayant pour objectif l'établissement d'une politique de restructuration de l'industrie du transport aérien. Dès le 30 août il demandait son point de vue au Commissaire de la Concurrence, qui dans sa réponse datée du 22 octobre 1999, souligne que "la disparition de Canadien ou un fusionnement avec Air Canada soulèverait de très sérieuses préoccupations en matière de concurrence sur le marché des transports aériens". À cet effet il suggère un certain nombre de propositions sur des modalités possibles de la restructuration, des changements aux politiques gouvernementales, et des modifications

[1] L.C. 1996 ch. 10.
[2] L.R.C. 1985 ch. C-34.

en matière législative et réglementaire, ceci afin de maintenir un niveau de concurrence adéquat sur le marché aérien intérieur. Après avoir également consulté et rencontré des représentants de l'industrie, des employés et des utilisateurs de transport aérien, le gouvernement établit le 26 octobre 1999 une politique-cadre traitant entre autres de la nécessité de contrôle canadien de l'industrie, des services aux petites collectivités, mais surtout de favoriser la concurrence. On y souligne en particulier que "le gouvernement veillera à ce que des mesures efficaces soient mises en place pour contrer l'agissement abusif au sens de l'industrie du transport aérien". Le ministre des transports demande au Comité permanent des transports de la Chambre des Communes et au Comité Sénatorial Permanent des Transports et des Communications de se pencher sur les principales questions en soulignant qu'elles pourraient éventuellement "nécessiter des engagements de la part d'un transporteur dominant et l'imposition de conditions, l'adoption de mesures législatives ou prise de règlements par le gouvernement". Ces derniers publieront leurs rapports en décembre 1999.

Le 5 novembre 1999, son projet de restructuration ayant été déclaré illégal par la Cour Supérieure du Québec, Onex (Airco) retire son offre d'achat d'actions d'Air Canada, et seule demeure la proposition d'acquisition par cette dernière de Canadien International. Pour qu'elle se réalise, il faut l'accord du Bureau de la Concurrence. Confronté à deux alternatives dont l'une serait la faillite de Canadien, le Commissaire préfère approuver la transaction proposée après avoir obtenu des engagement relatifs à la concurrence de la part d'Air Canada. Le ministre des transports, suite aux engagements d'Air Canada donne également son aval, ce qui fait de cette dernière le transporteur aérien très largement dominant sur le marché du transport intérieur aérien. Il annonce cependant que des changements législatifs vont être proposés afin d'assurer entre autres un renforcement des mesures de prévention de pratiques d'éviction par le transporteur aérien dominant.

Le 17 février 2000, le gouvernement fédéral dépose un projet de loi (Bill C-46) pour la restructuration de l'industrie du transport aérien.

Entre autres, celui-ci propose la modification de la Loi sur la Concurrence afin que les dispositions sur l'abus de position dominante offrent des recours plus efficaces contre des comportements d'éviction dans le secteur du transport aérien.

Le projet est adopté, et la Loi, sanctionnée le 29 juin 2000[3], entre en vigueur le 5 juillet. Le règlement d'application sera enregistré le 23 août et publié le 30 août de la même année.

II Nature des modifications législatives et réglementaires sur l'abus de position dominante

Les principaux changement à la Loi sur la Concurrence, relativement à l'abus de position dominantes sont de deux ordres, l'un substantif et l'autre procédural.

a) Changements substantifs

Au niveau substantif l'article 78 de la Loi sur la Concurrence définissant les agissement anti-concurrentiels pouvant donner lieu à une ordonnance d'interdiction dans le cas d'abus de position dominante (article 79) est modifié,[4] pour y inclure des agissements que le gouvernement précisera par règlements, ce qu'il fait d'ailleurs dès le mois d'août 2000[5].

On peut se demander quelle était la nécessité d'effectuer cette modification dans la mesure où la liste originelle des agissements anti-concurrentiels prévue par l'article 78 n'est pas limitative, ceci tel que l'a d'ailleurs affirmé le Tribunal de la concurrence dans l'affaire

---

[3] "Loi modifiant la Loi sur les transports au Canada, la Loi sur la concurrence, la Loi sur le Tribunal de la Concurrence et la Loi sur la participation publique au Capital d'Air Canada et modifiant une autre Loi en conséquence". L.C. 2000 ch. 15.

[4] Id. art. 13.

[5] "Règlement sur les agissements anti-concurrentiels des exploitants de service intérieur" DORS/2000-32h, 23 août 2000.

Nutra Sweet[6] "Il est évident que cette liste ne se veut pas exhaustive et la défenderesse admet que d'autres conduites qui ne sont pas expressément à l'article 78 peuvent constituer des agissements anti-concurrentiels. Certains de ces agissements ont des éléments communs mais comme le reconnaissent le directeur et la défenderesse, un seul est commun à tous : l'agissement anti-concurrentiel doit avoir un but particulier dont il est nécessaire de faire la preuve. Le but commun à tous ces agissements (sauf celui de l'alinéa 78 f) est l'effet négatif intentionnel sur un concurrent et cet effet doit être abusif, viser une exclusion ou une mise au pas."[7]

C'est ainsi qu'il ajoute que l'expression "agissement anti-concurrentiel" à l'article 78 est suffisamment vaste pour englober d'autres pratiques d'établissement prédatoire "que celui prévu par l'article 78". De la même manière le Tribunal de la Concurrence a donné une interprétation large de la notion d'agissement anti-concurrentiel dans l'affaire Nielsen[8].

Cependant l'article 78 (aujourd'hui 78 (1)) de la Loi sur la Concurrence souffre d'une lacune, c'est celle de ne pas incorporer de définition. Il aurait été alors hasardeux compte tenu du manque d'abondance de jurisprudence relative à cet article, de se baser uniquement sur lui, tel qu'ainsi formulé, pour éviter des comportements d'éviction dans l'industrie du transport aérien interne.

De la même façon la disposition criminelle contre les prix déraisonnablement bas (article 50 (1) c de la Loi) risquait aussi de rester lettre morte dans la mesure où d'une part, le fardeau de la preuve (au-delà de tout doute raisonnable) incombant à la Couronne est considérable et que d'autre part, lorsque l'entreprise poursuivie peut prouver que son prix est supérieur à son coût variable (moyen),

---

[6] Canada (directeur des enquêtes et recherches) c. The Nutra Sweet Company (4 octobre 1990) C.T.-89/2 Motifs et ordonnances (Tribunal de la concurrence) p. 65; (1990) 32 CPR (3d).

[7] Id. p. 84.

[8] Canada (directeur des enquêtes et recherche) c. D&B Companies of Canada. (1995) 64 C.P.R. (3d) 216.

la preuve de prix déraisonnablement bas est relativement difficile à faire[9]. En effet, il existe des motifs souvent valables (d'un point de vue économique) de vendre un produit à un prix inférieur à son coût total moyen tant et aussi longtemps qu'il dépasse le coût marginal (couramment estimé par le coût variable). Or, la notion de coût variable dans le transport aérien n'est pas absolument limpide. Ainsi serait-il possible de prétendre qu'il s'agit du coût subi par le transporteur aérien lorsqu'il fournit l'accès à un passager additionnel sur ses vols, ce qui constitue évidemment une dépense très minime. Finalement, la jurisprudence semble indiquer de manière tout à fait claire que le prix d'un produit ne peut pas être considéré comme déraisonnablement bas, même s'il est inférieur à son coût, lorsqu'il correspond au prix déjà demandé par un concurrent[10]. Ceci signifie que selon l'article 50 (1) c, un transporteur dominant qui serait défié sur un trajet particulier par une entreprise plus modeste offrant un tarif inférieur, pourrait toujours abaisser son prix sur ce trajet, sans égard à son coût d'opération, au niveau de ceux du concurrent en question, l'éliminant de la sorte.

Il semble cependant curieux que l'on ne réfère pas dans les différents mémoires présentés par le gouvernement, le Parlement, le Sénat et le Commissariat de la concurrence, à l'utilisation de l'article 50 (1) b sur la Discrimination régionale par les prix (appelée communément "subvention croisée"), pour combattre des politiques de prix d'éviction qui pourraient être conçues par le transporteur dominant contre des entreprises plus petites qui tenteraient de le concurrencer sur une base régionale[11]. Cet article pourrait être utilisé soit directement dans une poursuite criminelle, soit indirectement comme agissement anticoncurrentiel donnant lieu à un abus de position dominante. En effet, il ne fait aucun doute qu'un agissement qui peut

---

[9] Canada, Directeur des enquêtes et recherches (Loi sur la Concurrence), Prix d'éviction lignes directrices pour l'application de la Loi, Ottawa, Ministère des Approvisionnements et services. 1992.

[10] Boehringer Ingelheim (Canada) Inc. v. Bristol-Myers Squibb Canada Inc. 83 C.P.R. (3d) 51, p. 61.

[11] C'est ce que soutient d'ailleurs le transporteur CAN JET, (voir plus bas).

donner lieu à des poursuites criminelles parce qu'anti-concurrentiel, est inclus dans l'article 78 (aujourd'hui 78(1)).

L'amendement au niveau de la définition d'agissements anti-concurrentiels et l'adoption de règlement qui le précise vont-il constituer une barrière étanche à des comportements d'éviction? Ce n'est pas certain. En fait si l'on ne fait que considérer les aliénas a, b, et c, de l'article 1 du Règlement on s'aperçoit que l'on utilise la notion de "coûts évitables de prestation du service en cause" pour souligner la limite au-dessous de laquelle les prix de la prestation pourraient être considérés suffisamment bas pour constituer un acte anti-concurrentiel. Or ces notions sont imprécises. Dans l'ébauche de ce règlement (13 avril 2000), il avait été précisé que le coût évitable de prestation de service était constitué par tous les coûts qui pourraient être évités si le service n'était pas offert. Il est entendu que cette notion convient mieux que celle de coût variable pour l'industrie du transport aérien dans la mesure où c'est le coût marginal du service dans son ensemble dont on tient compte et non pas le coût marginal par passager. Néanmoins un transporteur dominant qui à long terme ne pourrait maintenir ses prix sur une route à un niveau où ils ne font que couvrir les coûts évitables parce que ne contribuant pas suffisamment au remboursement des coûts communs à toutes les routes (publicité, dépenses reliées au siège social etc...), pourrait à court terme réduire ses prix à ce niveau afin de discipliner un transporteur local. Ce dernier devant couvrir, avec les seuls revenus qu'il tire de l'exploitation de son unique itinéraire, tous les coûts de l'entreprise, pourrait très bien être exclu de la course, même à un niveau de productivité excédant celui du transporteur dominant.

b) Changements procéduraux

La Loi de modification adoptée en juin 2000[12] incorpore un changement procédural majeur. En effet le nouvel article 104.1 de la Loi sur la Concurrence donne au Commissaire un pouvoir d'une nature radicalement nouvelle. Si auparavant le Tribunal de la

[12] Supra. note 3, art. 15.

Concurrence pouvait après demande du Commissaire, rendre des ordonnances provisoires, en vertu des articles 104 (et 100 pour le cas des fusions), l'article 104.1 innove dans la mesure où le Commissaire a le pouvoir dans certaines circonstances de rendre une ordonnance provisoire sans en référer immédiatement au Tribunal de la Concurrence. Cette ordonnance en vigueur pendant vingt jours (art. 104.1 (4)) peut être prolongée pour deux délais additionnels de 30 jours, soit un total de 80 jours (Art. 104.1 (5)). S'il est vrai que le transporteur objet d'une telle mesure peut en demander la modification ou l'annulation au Tribunal de la Concurrence, il n'en reste pas moins que nous sommes en présence d'une disposition juridique exceptionnelle dans notre droit dans la mesure où la partie qui effectue l'enquête possède également un pouvoir de jugement. Cette conséquence a d'ailleurs provoqué des critiques sérieuses[13], d'autant plus que l'ordonnance peut être rendue pour interdire un acte ou une activité qui pourrait constituer un "agissement concurrentiel" (Art. 78.1) et non pas nécessairement un "abus de position dominante" qui requiert additionnellement une preuve de pratique qui "a, a eu ou aura vraisemblablement pour effet d'empêcher ou de diminuer sensiblement la concurrence" (Art. 79 (c)).

Il ne faut cependant pas perdre de vue que l'adoption du nouvel article 104.1 s'est faite dans le cadre de la restructuration de l'industrie du transport aérien au Canada. Ce n'est pas de gaieté de cœur comme nous l'avons mentionné, mais plutôt par manque d'alternatives, que le gouvernement canadien acceptait l'acquisition de Canadien par Air Canada, aboutissant ainsi à la naissance d'un transporteur aérien dont la position dominante sur le marché intérieur devenait écrasante (90% des revenus de passager de l'ensemble de l'industrie). Ce fusionnement suscitait beaucoup d'inquiétude quant à ses conséquences potentielles au niveau des prix et de la concurrence en général. Il était donc naturel qu'en plus des engagements souscrits pas Air Canada, des mesures législatives spéciales soient prises pour

[13] Voir par exemple : M. SANDERSON and M.J. TREBILCOCK" Bad Policy, Bad Law : Bill C-26 Amendments to the Competition Act on Airline Predation" Canadian Competition Record Spring/Summer 2000. Vol. 20 No 1, pp. 32-37.

gérer cette situation exceptionnelle, afin d'offrir des moyens efficaces contre un comportement d'éviction, qui réduirait à néant l'objectif de favoriser la concurrence dans ce marché; cet objectif demeurant (plutôt qu'une réglementation de l'industrie) la meilleure garantie, selon le gouvernement, de tarifs aériens raisonnables.

Ceci est d'autant plus vrai qu'il est facile pour une compagnie aérienne de déployer rapidement ses avions de manière à discipliner ou exclure un concurrent par une pratique de prix d'éviction. Le Commissaire de la Concurrence soulignait d'ailleurs à ce sujet que "contrairement aux entreprises dans d'autres industries, un transporteur aérien important peut discriminer par les prix dans une mesure beaucoup plus large, ajuster les prix beaucoup plus rapidement, et déplacer des ressources entre les marchés beaucoup plus promptement... Dans le contexte particulier de l'industrie du transport aérien, les disposition actuelles de la Loi sont inadéquates pour obtenir un redressement rapide et efficace. Pour obtenir un redressement qui arrête le prédateur avant qu'il ne tue sa proie, il est essentiel que des mesures de redressement provisoires puissent être obtenues rapidement"[14].

III Développements jurisprudentiels

Quinze jours à peine après l'entrée en vigueur du nouveau règlement concernant les transporteurs aériens, une première plainte était produite le 7 septembre 2000 auprès du Bureau de la Concurrence. Celle-ci provenait de la compagnie Can Jet Airlines créée en avril 2000 dans le but de fournir des services aériens à bas prix dans l'Est du Canada. Dès le 31 juillet elle annonce qu'elle offrira un service à tarif réduit sur certains trajets à partir du 5 septembre. Air Canada en réaction à cette initiative propose également des tarifs réduits spéciaux (LI4 EASTS) sur ces mêmes trajets. Le Commissaire commence, dès réception de la plainte le 7 septembre, une enquête qu'il poursuit formellement dès le 28 septembre sous le régime de l'article 10 (1)c de la Loi sur la Concurrence, afin d'établir si Air

[14] Lettre du 22 octobre 1999 au ministre des transports. pp. 7 et 8.

Canada aurait contrevenu aux articles de la Loi relatifs aux prix d'éviction (articles 50 et 79).

Le 12 octobre il émet une ordonnance temporaire en vertu du pouvoirs que lui confère le nouvel article 104.1(1) de la Loi, exigeant qu'Air Canada arrête d'offrir ses tarifs spéciaux sur ces trajets. Celle-ci sera renouvelée pour une période de 30 jours, le 31 octobre. Après avoir tenté sans succès de faire suspendre l'ordonnance temporaire par la Cour Supérieure du Québec en attendant l'issue de la contestation constitutionnelle, Air Canada demande le 1er novembre 2000 au Tribunal de la Concurrence l'annulation de l'ordonnance en vertu de l'article 104.1 (7). Celui-ci maintient l'ordonnance temporaire jusqu'au 31 décembre[15]. Quoique ce délai ait expiré, Air Canada au moment de l'écriture de ces lignes n'avait pas repris la tarification spéciale. L'enquête cependant se poursuit.

Un certain nombre d'éléments du jugement par le Tribunal de la Concurrence semblent renforcer la portée de l'article 104.1. Nous en soulignerons trois. Tout d'abord le Tribunal interprète de façon stricte l'alinéa 104.1 (7) rejetant toute suggestion qu'à ce stade il puisse examiner la nature ou le bien fondé de l'enquête du commissaire ou du bien fondé de son avis sur la nature des actes commis.

Ensuite l'utilisation du conditionnel ("pourrait constituer des agissements anti-concurrentiel") dans le paragraphe 104.1 (1) indique clairement que le Commissaire n'a pas à être convaincu que l'acte correspond à un agissement anti-concurrentiel. Dans notre affaire le Commissaire n'avait donc pas besoin d'acquérir la certitude que l'offre des tarifs spéciaux constituait une "exploitation de la capacité sur une ou plusieurs routes à des prix qui ne couvrent pas les coûts évitables de prestation de service en cause"[16]. Il suffisait qu'il s'agisse d'une hypothèse raisonnable.

---

[15] Air Canada c. Canada (Commissaire de la Concurrence). [2000] D.T.C.C. no. 24.

[16] Alinéa 78 (1) j) de la Loi sur la Concurrence et alinéa 1a) du règlement.

Finalement, contrairement aux prétentions d'Air Canada, le Tribunal souligne qu'au stade de l'ordonnance provisoire, le Commissaire n'a pas à se faire une opinion sur le caractère abusif (au sens de l'article 79 de la Loi) des actes qu'il considère comme pouvant constituer des agissements anti-concurrentiels. En d'autres termes, même si en fin de parcours le commissaire devra pour obtenir une ordonnance permanente faire la preuve d'abus de position dominante c'est-à-dire de tous les éléments du paragraphe 79(1) de la Loi. À ce stade il n'a qu'à démontrer le caractère raisonnable de l'hypothèse selon laquelle les actes ou activités faisant l'objet de son ordonnance provisoire pourraient constituer des agissements anti-concurrentiel selon l'article 78(1) (anciennement 78) de la Loi.

Conclusion

Nous avons tenté dans cet article de présenter après en avoir décrit le processus historique, les changements apportés à la Loi sur la Concurrence en matière de comportement prédateur pouvant donner lieu à un abus de position dominante dans l'industrie du transport aérien interne au Canada. L'importante modification procédurale permettant dans le cas particulier de cette industrie au Commissaire de la Concurrence d'émettre dans certaines circonstances, une ordonnance provisoire de son propre chef, sans en référer à un tribunal du transport aérien a également été souligné.

Ces modifications législatives quoique ne garantissant aucune étanchéité, devraient, en liaison avec les engagements précis prix par Air Canada, néanmoins doter le public d'un arsenal solide qui limitera les effets pervers que pourrait avoir la concentration d'environ 30% du marché aérien interne dans les mains d'une entreprise unique.

Les ordonnances d'interdiction émises par le Commissaire de la Concurrence lors de l'automne 2000, et l'appui qu'il a reçu du Tribunal de la Concurrence à ce sujet semblent indiquer qu'effectivement ces mesures législatives ne resteront pas lettre morte et seront au contraire déployées avec succès contre quiconque

tenterait d'éliminer ou de réduire indûment les velléités de concurrence saine dans l'industrie du transport aérien interne au Canada.

ANNEXE I

Loi modifiant la Loi sur les transports au Canada la Loi sur la concurrence, la Loi sur le Tribunal de la concurrence et la Loi sur la participation publique au capital d'Air Canada et modifiant une autre Loi en conséquence

13. (1) L'article 78 de la même loi devient le paragraphe 78(1).

(2) Le paragraphe 78(1) de la même loi est modifié par adjonction, après l'alinéa *i*), de ce qui suit :

j) à l'égard des exploitants d'un service intérieur, au sens du paragraphe 55(1) de la *Loi sur les transports au Canada,* les agissements précisés à l'alinéa *(2)a);*

k) le fait pour l'exploitant d'un service intérieur, au sens du paragraphe 55(1) de la *Loi sur les transports au Canada,* de ne pas donner accès, à des conditions raisonnables dans l'industrie, à des installations ou services essentiels à l'exploitation dans un marché d'un service aérien, au sens de ce paragraphe, ou de refuser de fournir ces installations ou services à de telles conditions.

(3) L'article 78 de la même loi est modifié par adjonction, après le paragraphe (1), de ce qui suit :

(2) Le gouverneur en conseil peut, par règlement pris sur recommandation du ministre et du ministre des Transports :

a) préciser des agissements pour l'application de l'alinéa *(1)j);*

b) préciser des installations ou services essentiels pour l'application de l'alinéa *(1)k).*

15. La même loi est modifiée par adjonction, après l'article 104, de ce qui suit :

104.1 (1) Le commissaire peut rendre une ordonnance provisoire interdisant à une personne exploitant un service intérieur, au sens du

paragraphe 55(1) de la *Loi sur les transports au Canada,* d'accomplir tout acte, ou de mener toute activité, qui, selon lui, pourrait constituer des agissements anti-concurrentiels ou lui enjoignant de prendre les mesures qu'il estime nécessaires pour ne pas nuire à la concurrence ou pour éviter de causer des dommages à une autre personne lorsque, à la fois :

a) il a commencé une enquête en vertu du paragraphe 10(1) en vue de déterminer si les agissements de la personne ont donné lieu à une situation visée à l'article 79;

b) il estime qu'en cas de non-prononcé de l'ordonnance :

(i) soit la concurrence subira vraisemblablement un préjudice auquel le Tribunal ne pourra adéquatement remédier,

(ii) soit un compétiteur sera vraisemblablement éliminé ou une personne subira vraisemblablement une réduction importante de sa part de marché, une perte importante de revenu ou des dommages auxquels le Tribunal ne pourra adéquate- ment remédier.

(2) Le commissaire peut rendre l' ordonnance sans préavis et sans donner au préalable à qui que ce soit la possibilité de présenter des observations.

(3) Le commissaire envoie un avis écrit de l'ordonnance et des motifs de celle-ci, dans les meilleurs délais après son prononcé, aux personnes qui en font l'objet et aux autres personnes directement touchées.

(4) Sous réserve des paragraphes (5) et (6), l'ordonnance demeure en vigueur pendant vingt jours.

(5) Le commissaire peut, à deux reprises, proroger l'ordonnance d'une période supplémentaire de trente jours et peut, en tout temps, annuler l'ordonnance. Dans les meilleurs délais, il avise par écrit de la prorogation ou de l'annulation les personnes qui ont été avisées au titre du paragraphe (3). ...

ANNEXE II
RÈGLEMENT SUR LES AGISSEMENTS ANTI-CONCURRENTIELS DES EXPLOITANTS DE SERVICE INTERIEUR

AGISSEMENTS ANTI-CONCURRENTIELS

1. Pour l'application de l'alinéa 78(1)*j*) de la *Loi sur la concurrence,* constituent des agissements anti-concurrentiels les agissements ci-après de l'exploitant d'un service intérieur, au sens du paragraphe 55 (I) de la *Loi sur les transports au Canada* :

a) l'exploitation de la capacité sur une ou plusieurs routes à des prix. qui ne couvrent pas les coûts évitables de prestation du service en cause;

b) l'augmentation de la capacité sur une ou plusieurs routes à des prix qui ne couvrent pas les coûts évitables de prestation du service en cause;

c) l'utilisation d'un transporteur secondaire à moindre coût d'une façon visée aux alinéas *a)* ou *b);*

d) la préemption d'installations ou de services aéroportuaires nécessaires à un autre transporteur aérien pour l'exploitation de son entreprise, dans le but de retenir ces installations ou ces services hors d'un marché;

e) dans la mesure où elle n'est pas régie par un règlement - pris en vertu d'une autre loi - concernant les créneaux de décollage ou d'atterrissage, la préemption de créneaux de décollage ou d'atterrissage nécessaires à un autre transporteur aérien pour l'exploitation de son entreprise, dans le but de retenir ces créneaux hors d'un marché;

f) l'utilisation, pour la vente ou l'achat de ses vols, de commissions, de primes ou d'autres incitatifs dans le but de discipliner ou d'éliminer un concurrent, ou d'empêcher l'entrée ou la participation accrue d'un concurrent dans un marché ou d'y faire obstacle; ...

ANNEXE III
Concurrence, Loi sur la

50. (1) Commet un acte criminel et encourt un emprisonnement maximal de deux ans toute personne qui, exploitant une entreprise, selon le cas :

b) se livre à une politique de vente de produits, dans quelque région du Canada, à des prix inférieurs à ceux qu'elle exige ailleurs au Canada, cette politique ayant pour effet ou tendance de réduire sensiblement la concurrence ou d'éliminer dans une large mesure un concurrent dans cette partie du Canada ou étant destinée à avoir un semblable effet;

c) se livre à une politique de vente de produits à des prix déraisonnablement bas, cette politique ayant pour effet ou tendance de sensiblement réduire la concurrence ou éliminer un concurrent, ou étant destinée à avoir un semblable effet.

78. Pour l'application de l'article 79, «agissement anti-concurrentiel» s'entend notamment des agissements suivants :

a) la compression, par un fournisseur intégré verticalement, de la marge bénéficiaire accessible à un client non intégré qui est en concurrence avec ce fournisseur, dans les cas où cette compression a pour but d'empêcher l'entrée ou la participation accrue du client dans un marché ou encore de faire obstacle à cette entrée ou à cette participation accrue;

b) l'acquisition par un fournisseur d'un client qui serait par ailleurs accessible à un concurrent du fournisseur, ou l'acquisition par un client d'un fournisseur qui serait par ailleurs accessible à un concurrent du client, dans le but d'empêcher ce concurrent d'entrer dans un marché, dans le but de faire obstacle à cette entrée ou encore dans le but de l'éliminer d'un marché; ...

h) le fait d'inciter un fournisseur à ne vendre uniquement ou principalement qu'à certains clients, ou à ne pas vendre à un concurrent ou encore le fait d'exiger l'une ou l'autre de ces attitudes de la part de ce fournisseur, afin d'empêcher l'entrée ou la participation accrue d'un concurrent dans un marché;

(*i*) le fait de vendre des articles à un prix inférieur au coût d'acquisition de ces articles dans le but de discipliner ou d'éliminer un concurrent.

79. (1) Lorsque, à la suite d'une demande du commissaire, il conclut à l'existence de la situation suivante :

a) une ou plusieurs personnes contrôlent sensiblement ou complètement une catégorie ou espèce d'entreprises à la grandeur du Canada ou d'une de ses régions;

b) cette personne ou ces personnes se livrent ou se sont livrées à une pratique d'agissements anti-concurrentiels;

c) la pratique a, a eu ou aura vraisemblablement pour effet d'empêcher ou de diminuer sensiblement la concurrence dans un marché,

le Tribunal peut rendre une ordonnance interdisant à ces personnes ou à l'une ou l'autre d'entre elles de se livrer à une telle pratique. ...

The Use of Economic Analyses in Competition Cases (Part II) - Analysis of a Conspiracy and a Merger in Transportation Cases

Joseph Monteiro, Competition Bureau, Industry Canada*
Andrew Eckert, University of Alberta

1. Introduction

The underpinnings of Competition law in Canada can be traced in part to economic thought before the turn of the nineteenth century. At that time, the teachings of Adam Smith and his support for free competition were well known and were gaining ground over the view that saw merit in combinations and aggregation of wealth. As a result, combinations were not only against the common law in Canada but also in England and the U.S. This is also suggested in the discussions that occurred between Members of the House and Senate during the enactment of the first competition law in Canada in 1889. The underpinnings of competition law also became evident when this law was extended to include other forms of economic behaviour that restrained trade in the decades that followed. Even today, a great deal of discussion occurs among economists when revising the *Competition Act* in light of developments in economic theory. There is thus a close and fundamental relationship between economics and competition law.

Despite this close association, the explicit use of economic reasoning and economists in Courts began on a rather uncertain foundation. This can be seen in the jurisprudence. For example, in *Crown Milling Company v. The King*, Lord Finlay stated: "It is not the judge's task to adjudicate between conflicting theories of political economy." This opinion was reiterated in cases such as *R. v. Container Materials Ltd.* and *R. v. Howard Smith Paper Mills Ltd.*; and in one 1976 case - *Allied Chemical et al v. The Queen* - evidence of the Crown economist was not admitted.

After this initial setback, the attitude towards the treatment of expert economic evidence by the Courts began to gradually change and though economic evidence was accepted, the weight to be attributed to it was a matter to be determined by the Court. A more dramatic change in attitude occurred with the inclusion of civil provisions into the competition law in 1976 and a sharp increase in the use of economic experts. These developments were re-enforced with the establishment of the Competition Tribunal in 1986, (a quasi-judicial body, comprising of judges, experts from the business, academic and

civil service communities). In fact, provisions for the use of expert evidence are now specifically provided for under sections 47 and 48 of the *Competition Tribunal Rules*. Thus the acceptance of economics and economists by Canadian courts have now become firmly established. In this paper we shall illustrate how economic analyses is used in a conspiracy case and a merger case.

2. Economic Analysis in a Conspiracy Case

The case chosen for economic analysis is about a complaint to the Competition Bureau involving allegations of an alleged agreement between several freight forwarders engaged in the business of assembling and consolidating less-than-carload shipments which were transported by rail in boxcars thereby violating the conspiracy provision of the *Competition Act* (i.e., Act). Before describing this case, the elements of this provision will be provided and finally an economic analysis of this case. However, first some background information on this provision will be briefly provided.

The conspiracy provision was first included in the Act of 1889 entitled *An Act for the Prevention and Suppression of Combinations in Restraint of Trade* which later became part of the *Criminal Code* in 1892.[1] In this provision, the usage of the words 'unlawfully', 'unduly' and 'unreasonably' led to difficulties in its interpretation. As a result, the words 'unduly' and 'unreasonably' were deleted in the changes to the *Criminal Code Amendments Act* of 1899. This was reversed a year later, as these words were reintroduced in the *Criminal Code Amendments Act* of 1900 but this time the word 'unlawfully' was omitted. A few years later in 1910, the first *Combines Investigation Act* came into effect on May 4 and the Act now contained a definition of combine in section 2(c).[2] This Act was replaced by the *Board of*

---

1 "Every person who conspires, combines, agrees or arranges with any other person, or with any railway, steamship, steamboat or transportation company, unlawfully, (a) to unduly limit the facilities for transporting, producing, manufacturing, supplying, storing or dealing in any article or commodity which may be a subject of trade or commerce; or - (b) to restrain or injure trade or commerce in relation to any such article or commodity; or - (c) to unduly prevent, limit or lessen the manufacture or production of any such article or commodity, or to unreasonably enhance the price thereof; or - (d) to unduly prevent or lessen competition in the production, manufacture, purchase, barter, sale, transportation or supply of any such article or commodity, or in the price of insurance upon person or property. Is guilty of a misdemeanour and liable, on conviction, to a penalty not exceeding four thousand dollars and not less than two hundred dollars, or to imprisonment for any term not exceeding two years; and if a corporation, is liable on conviction to a penalty not exceeding ten thousand dollars and not less than one thousand dollars.

2 'Any contract, agreement, arrangement or combination which has or is designed to have, the effect of increasing or fixing the price or rental of any article or commerce or the cost of the storage or transportation thereof, or of restricting the competition in or of controlling the production, manufacture, transportation, storage, sale or supply thereof to the detriment of

Commerce Act and the Combines and Fair Prices Act of 1919. In it was included a comprehensive definition of 'combine' - as one which in the opinion of the Board of Commerce of Canada operated or was likely to operate to the detriment or against the interest of the public and was deemed to include several situations.[3] A few years later, the second *Combines Investigation Act* was passed in 1923 basically retaining the definition of 'combine' used in the 1919 legislation. In 1935, after an inquiry into price spreads, the *Dominion Trade and Industry Commission Act* was passed which resulted in a section permitting price and production agreements. As a result, any agreement approved by Cabinet that may be the subject of investigation under section 498 or section 498A of the Criminal code required approval of the Commission. Shortly thereafter, this section conferring authority on the Commission was ruled unconstitutional leading to further amendments to the *Combines Investigation Act* in 1937. Apart from this brief period of uncertainty, the legislation of 1923 relating to conspiracy continued to remain in force till 1960. Before that year, important changes were introduced in the Act in 1952 following the MacQuarrie Committee Report, however the conspiracy provision was left unaltered other than its definition. Further amendments were introduced in 1960, the most important related to the introduction of six limited defences (pertaining to exchange of statistics, product standards, etc.) together with an exemption relating to export agreements subject to certain conditions. Several years later in 1976, the conspiracy provision was further clarified by the addition of a section whereby it was not necessary to establish that competition was virtually eliminated by

---

consumers or producers of such article of trade or commerce and includes the acquisition, leasing or otherwise taking over the business, or any portion of the business, of any person, and also includes what is known as a trust, monopoly or merger'.

3 "The expression "combine" is used in this Act with intended relation to articles of commerce, and it shall be deemed to have reference only to such combines, immediately hereinafter defined, as, with relation as aforesaid, have, in the opinion of The Board of Commerce of Canada (or of a single member thereof acting under authority of and for the purposes of section eight of this Act) operated, or are likely to operate, to the detriment of or against the interest of the public, consumers, producers or others, and, limited as aforesaid, the said expression as used in this Act shall be deemed to include,- (a) mergers, trust and monopolies, so called, and (b) the relation resulting from the purchase, lease or other acquisition by any person of any control over or interest in the whole or part of the business of any other person, and, (c) any actual or tacit contract, agreement, arrangement or combination which has or is designed to have the effect of (1) limiting facilities for transporting, producing, manufacturing, supplying, storing or dealing; or (2) preventing, limiting or lessening manufacture or production; of (3) fixing a common price, or a resale price, or a common rental, or a common cost of storage or transportation, or enhancing the price, rental or cost of article, rental storage or transportation; or (4) preventing or lessening competition in, or substantially controlling, within any particular district, or generally, production, manufacture, purchase, barter, sale, transportation, insurance or supply; or (5) otherwise restraining or injuring commerce."

the agreement if carried into effect or that was the object of the agreement. Finally in 1986, the provision was further clarified and strengthened. It was no longer necessary to establish that the firms intended to lessen competition unduly; further inference of an agreement was permitted without evidence of communications.

a) The Conspiracy provision of the *Competition Act*

The conspiracy provision is contained in sections 45(1) to 45(8) of the *Competition Act* (formerly sections 32(1) to 32(7) of the *Combines Investigation Act*. Section 45(1) states:

"Every one who conspires, combines, agrees or arranges with another person

> (a) to limit unduly the facilities for transporting, producing, manufacturing, supplying, storing or dealing in any product,
> (b) to prevent, limit or lessen, unduly, the manufacture or production of a product or to enhance unreasonably the price thereof,
> (c) to prevent or lessen, unduly, competition in the production, manufacture, purchase, barter, sale, storage, rental, transportation or supply of a product, or in the price of insurance on persons or property, or
> (d) to otherwise restrain or injure competition unduly,

is guilty of an indictable offence and liable to imprisonment for a term not exceeding five years or to a fine not exceeding ten million dollars or to both.

For an offence to be committed under this provision, two basic elements must be satisfied. First, the Crown must prove beyond a reasonable doubt that the accused intended to enter into the combination, conspiracy, agreement or arrangement (i.e., CCAA). Secondly, the Crown must show that the CCAA if carried into effect would prevent or lessen competition unduly. Jurisprudence on these two elements has been provided by the Supreme Court of Canada in its decision in *R. v. Nova Scotia Pharmaceutical Society* (1992). Regarding the first element, a subjective and an objective component must be satisfied. To satisfy the subjective component, the Crown must prove that the accused had the intention to enter into the CCAA and had knowledge of the terms of that CCAA. To satisfy the objective component, the Crown has to prove that on an objective view of the evidence, the accused was aware or ought to have been aware that the effect of the agreement would be to prevent or lessen competition unduly. Regarding the second element, the Supreme Court maintained that it is a combination of market power and behaviour injurious to competition that makes a lessening of competition **undue**. "As a two step process, one begins by determining if the parties to the CCAA

possess market power, or are they likely to attain market power by virtue of the CCAA. Consistent with economic theory the Court made it clear that market share alone is not sufficient to demonstrate market power. They noted that possessing only a moderate amount of market power may be sufficient to support a finding of undueness." Then one inquires if the CCAA has or was likely to injure competition (eg., limiting supplies, enhancing price, preventing or lessening competition, etc).

In addition, section 45(2.1) provides for the inference of the existence of a conspiracy, combination, agreement or arrangement from circumstantial evidence with or without direct evidence of communication between the alleged parties. The issue of inferential conspiracy also raises the question of the status of conscious parallelism. This type of behavior, in the absence of other information, is not considered sufficient to satisfy the existence of a CCAA. However, the status of conscious parallelism with 'plus' factors has not been conclusively resolved.

This provision also contains several defences. First, according to section 45(3), certain types of CCAA are exempt from conviction if they relate only to exchange of statistics, definition of product standards, exchange of credit information, definition of trade terminology, cooperation in research and development, restrictions on limitations or advertising, packaging, system of weights or measures to protect the environment. These defences do not apply if the CCAA has lessened or is likely to lessen competition unduly with respect to prices, quantity or quality of production, markets or customers, or channels or methods of distribution or if it is likely to restrict any person from entering into or expanding a business in a trade, industry or profession. Second, CCAAs that relate solely to 'exports of products from Canada' shall be exempt from conviction so long as it does not result in a reduction or limitation of the real value of exports, or does not restrict any person from entering into or expanding the business of exporting or does not prevent or lessen competition unduly in the supply of services facilitating that export or is likely to have that effect. Third, the conspiracy, combination, agreement or arrangement is exempt to the extent that it relates only to a 'service' and to 'standards of competence and integrity' that are reasonably necessary for the 'protection of the public' in the practice of a trade or profession or in the collection and dissemination of information.

Besides the above defences, there are two additional situations to which the conspiracy provision does not apply: those that relate to CCAAs between federal financial institutions with respect to Banks described in section 49(1); and those that relate to agreements or arrangements between affiliates. Further, no prosecution may be commenced under this provision where an order is sought under the abuse of dominance or merger provisions on the same or substantially the same facts.

If the above elements of the conspiracy provision are satisfied and the defences do not apply, a person who conspires, combines, agrees or arranges

with another person is guilty of an indictable offence. The penalty for this offence is imprisonment not exceeding five years or a fine not exceeding ten m. dollars or to both. In addition, the court can impose a prohibition order.

b) The complaint and its evaluation under the key elements of section 45(1)(c)

It was alleged by the complainant that several companies and individuals conspired to lessen competition in the supply of pool car freight forwarding services contrary to s. 45(1)(c). Later, charges were laid against five pool car freight forwarding corporations - Clarke Transport Canada Inc., Consolidated Fastfrate Transport Inc., Northern Pool Express Ltd., TNT Canada Inc, and Cottrell Transport Inc. These pool car corporations were basically engaged in the consolidation and forwarding of freight for transportation by railway boxcars. The consolidation occurs at the terminals in boxcars after the shipments are picked up by the pool car operators or after shipments are delivered at the terminal by the customer. The boxcars are then taken from the terminal and moved by CN railway and CP rail to major Canadian cities. The reverse process occurs at the destination terminals (i.e., unloading and delivery to final destination by truck). For the use of the railway service, the pool car operators pay the railways according to their tariffs. This service is offered for long-haul routes primarily over 500 miles. The other services competitive with pool-car services are long-haul trucking services and intermodal railway transportation services. The long-haul trucking services can be divided into truckload and less-than-truckload, the latter is very similar to the pool car service. The intermodal railway services are basically provided through trailer on flat car (TOFC) owned by the trucker, railway or shipper.

In attempting to establish the first element, it was shown that the freight forwarders employed a customer protection scheme (agreement) designed to shield an incumbent carrier's rates from price competition by the other members of the alleged conspiracy. This agreement was implemented and maintained through the exchange of Special Rate Notices (SRNs) and through access to information from an individual freight forwarder regarding their existing customers, rates being charged, etc. If this led to cheating, retaliatory action could be employed by the members of the conspiracy. Further, the Crown also pointed to other forms of anti-competitive behavior designed to further the success of the alleged conspiracy (for example, establishment of a common tariff, co-ordination of general rate increase, imposition of a cube rule, the imposition of standard ancillary charges, attempts to discourage new entrants into the railway tariff, collective action to confront the competitive threat posed by new entrants and other pressure tactics designed to bear on new entrants to join or at least respect the agreement). Based on this evidence, the Court found that the accused entered into an agreement (between February 1976 and May 1987) the effect of which would be to lessen competition in the supply of pool car freight forwarding

services. Further, the Court was fully satisfied that the accused intentionally entered into the agreement and that if implemented would lessen competition among themselves and their confederates. In the words of the Judge, "the evidence in support of these findings is overwhelming."

To establish the second element, undueness, the Crown argued that it could be established either through: (i) evidence that undueness was the object of the agreement; (ii) direct evidence that undueness was the effect of the agreement without the need to define the relevant market; or (iii) evidence that undueness was the effect of the agreement, using the PANS framework, without resort to the market-power behaviour analysis. Regarding the first, the Crown pointed out that there was substantial documentary evidence to support its findings that the object of the agreement was to maintain high rates by virtually eliminating rate competition in the pool car industry without inquiring into its effects. The Court indicated "With respect, I do not accept this [that the *actus reus* component of undueness can be satisfied upon proof of 'an agreement having as its object the undue prevention or lessening of competition'] as an accurate statement of the law." Regarding the second, the Crown argued that evidence of actual detrimental effects of an agreement that was fully realized in the past, obviates the need for inquiry into market definition to prove market power. Further, the Crown pointed out that the test was met as the freight forwarders implemented the agreement and maintained it for several years. Regarding the third, the Crown argued that the accused had 100 percent of the pool car market and there were barriers to entry. The Crown's argument that the pool car market was the relevant market was based on the treatment given by the railways to the pool car industry, minimum substitutability between the pool cars and trucks or intermodal railway service and expert economic evidence. Unfortunately for the Crown, the Court declined to accept the Crown's definition of the relevant market as pool car operators. It adopted a broader market concept which not only included pool car operators but also trucking and intermodal railways and within this broader market the Court was of the opinion that the accused did not have the requisite degree of market power to establish undueness needed to satisfy the requirements under section 45(1)(c).[4]

The defences raised in the other sections were not applicable to this conspiracy and were therefore not considered in its evaluation here. Further, section 45(2.1) was not part of the law at that time and therefore did not apply, however, even if it did apply the matter was not relevant since there was ample

---

4 "...I have concluded that the Crown has failed to establish, as a fact, that the relevant product market is the product for pool car services as particularized in the indictment. Based upon the evidence, it is obvious to me that during the requisite time frame, pool car operators, truckers and intermodal operators were equally capable of moving and indeed did move the identical kinds of freight on behalf of shipper/customers from Ontario to points west." *Id.* p. 315.

evidence of an agreement and there was no need for any inference.

The case is important as it established jurisprudence in a number of areas: 1) It is not necessary to establish relevant market beyond a reasonable doubt; 2) It set out important guidelines in defining the relevant market; and 3) the undueness requirement of the law cannot be satisfied by establishing that the object of the agreement was to unduly lessen competition.

c) The economic analysis of the complaint

The rationale for collusion is that firms have an incentive to coordinate their production and pricing activities to increase their collective and individual profits by restricting market output and raising market price. This is because a firm's profit goes up when it forms a cartel even though competitive firms may be "maximizing their profits".[5] As each firm in the competitive situation ignores the increase in profits to other firms from a reduction of its own output, which it believes to be insignificant since it cannot affect price. In contrast, a cartel is able to capture the benefits of a reduction of output by its members. To visualize the nature of this collective gain, consider two polar cases. Consider a industry made up of many identical competitive firms having identical cost curves with each firm acting as a price taker. In contrast, consider another similar situation where the same firms form a cartel and act as a monopoly. In the first case, the equilibrium output of the competitive industry (Qc) is determined by the intersection of the industry demand and supply curves (i.e. the average revenue curve and aggregate marginal cost curve). Note that the industry's marginal cost curve, which is the aggregation of their individual marginal cost curves, is the industry's supply curve.

In the second case, the equilibrium output (Qm) of the cartel is determined by the intersection of the industry marginal revenue curve and the supply curve. This equilibrium cartel output is less than the equilibrium output of competitive firms and it earns a profit.[6] The reason for the reduction in output of the cartel is because the cartel's marginal cost is greater than its marginal revenue at the competitive output. On the other hand, firms in the competitive industry on its

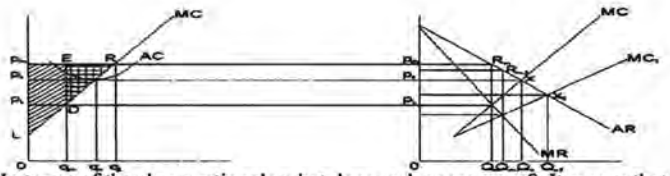
---

5 Assuming a competitive case.

6 The gain to the cartel from a reduction in output comes about because it faces a downward sloping demand curve. On the other hand, each competitive firm is considered to face a horizontal demand curve. More precisely, a nearly horizontal demand curve (i.e. with a small downward slope). While this small downward slope can be ignored for a single competitive firm, it cannot be ignored when taking all firms collectively or for the cartel. The competitive firms place no value on the gain that each other derives from a reduction in output of one unit by each firm because its effect is considered to be negligible. However, the gains to a cartel from a reduction of one unit is internalized, namely the cartel realizes the gain from its reduction.

own do not reduce its output because the competitive firm sets its output where marginal revenue equals marginal cost. A lowering of its output will cause it to lose more than it gains, as the marginal revenue on the last unit exceeds its cost. In view of this, there is an incentive for cartel members to cheat once in a cartel. Since at a higher price, each cartel member would like to sell more and would want to cheat by producing more than the cartel permits it to do.[7]

The above rationale can be seen in the following diagram:



In terms of the above rationale what does undueness mean? It means that the agreement affects the entire industry or a significant part of it. As a result, the members of the cartel will be able to capture all or a fair amount of the benefits that could arise from a reduction in output. The smaller the cartel, the less would be its market power and its impact would not be undue. In this latter scenario, prices would be lower than when a cartel encompasses the entire industry (i.e., less than Pm) and the output would be higher (between Qc and Qm). Further, since non-cartel members would like to sell more at the higher price this might defeat the purpose of the cartel members if it affects only a small part of the industry.

Similarly, what would happen if the market defined above now included other substitutable products? The supply curve[8] would now become much more elastic (MC2) as a result the competitive equilibrium output (Qc2) would be larger and the competitive equilibrium price would be lower. Similar results would apply if the entire industry was cartelized (i.e., Qm2 - output and

---

7 Carlton, Dennis W., and Perloff, Jeffrey M., Modern Industrial Organization, 1990, p. 233. In terms of the first diagram, The firms equilibrium output is qm at price pm, but it would like to sell ql at price pm. The cross-hatched area indicates the gain from chiselling. At price pm the individual firm in the cartel earns a profit LOEpm, the competitive firms sell an output ql at price pm and earns a profit LRpm. The latter profit is greater than the former.

8 Factoring demand into consideration results in a more elastic demand curve if the products are substitutes, thus the market power of the cartel would be lower, re-enforcing the results obtained in considering supply.

Pm2 - price). However, since the cartel does not effect the entire industry including the substitutable product, the output would be somewhat larger and prices lower than those indicated. Thus, the effectiveness of the cartel in raising price depends on the effectiveness of the substitutable product. In terms of this case, the initial situation described would result if a narrower market definition is adopted as the pool car operators had a 100 percent of it. When the market is defined to include trucking and intermodal rail, the final situation described would result. Further, since the pool car cartel accounted for an insignificant part of the market its market power would be negligible.

Thus in terms of our analysis, for the detrimental effect of a reduction in output on total welfare to occur, the firms to the agreement must have a moderate degree of market power. It should be noted however, that the approach to conspiracy law is not based on a maximization of total welfare. Since, the law does not provide for the consideration of efficiencies or any balancing of a reduction in output with any potential increase in efficiencies. It is this opinion that have led to suggestions to revise the conspiracy laws.

The above provides the standard rationale in most conspiracy cases and is generally assumed. Besides this, one issue that was the subject of considerable economic analysis in this case was the relevant market. The relevant market according to the Crown was the pool car market. The relevant market according to the defence also included trucking and intermodal rail services. The Crown's claim was based on direct evidence that was anecdotal consisting of buyer price sensitivity that there was no close substitute, together with the evidence of the expert witness and the Clarke Rate Quote Analysis to confirm the reliability of shipper/customer testimony. The defence's claim was based on indirect evidence of functional interchangeability and inter-industry competition. The Judge ruled in favour of the defence as he found the evidence of the Crown was not very convincing. There were a few fatal flaws with the evidence of the Crown. First, the differences in rates and services were not substantial to convince the Judge that they were sufficient to exclude trucking and intermodal services from the definition of the relevant market.[9] Second, more fatal to the Crown's case was the testimony of seventeen shipper/customers to establish the above. Further, no convincing reasoning was provided by the Crown how these seventeen shipper/customers were chosen and why they constituted a representative sample. In addition, their testimony was not of much assistance to the Crown's case: two of the witnesses provided no evidence and the other fifteen provided evidence that also supported the case of the defence. The Crown should have requested an independent agency, like Statistics Canada, to conduct a survey to establish this

---

9 The Judge indicated that the differences in rates and services of air cargo and expedited trucking vs pool car were substantial to exclude these alternatives from the definition of the relevant market.

matter based on a sound sampling methodology. It appears that the Crown was over confident that its market definition would be acceptable. This appears to be substantiated by the Judge's claim that he had no way of knowing what the share of the pool car operators was in a broader market context.[10] Third, the Crown made no attempt to show that product substitutability was illusory because the pool car operators through their collective power priced up to the elastic portion of the demand curve. While this is a potential argument, the Crown by failing to do so protected itself from further embarrassment as the evidence did not support it.[11]

3. Economic Analysis in a Merger Case

The case chosen for economic analysis is the acquisition of Cast North America Inc. and related companies by Canadian Pacific Limited, its affiliates and subsidiaries. When the matter was brought to the attention of the Competition Bureau, the Acting Director of Investigation filed an application with the Competition Tribunal under section 92 of the *Competition Act*. Before describing this case, the elements of this provision will be provided and finally its economic analysis. However, first some historical background on this provision will be briefly given.

A separate merger provision appeared for the first time in the 1960 amendments to the *Combines Investigation Act*. Following these amendments, mergers and monopolies were made offences only when they were likely to be, or to operate 'to the detriment of the public'. Before this period, the approach to mergers was through the conspiracy provision. The first time the word 'acquisition' or 'merger' was used in competition legislation was in the first *Combines Investigation Act* of 1910 in the definition of combine under subsection 2(c). Later, in the Act of 1919 it appeared as a stand alone subsection in the definition as subsections 2 (a) and (b). This definition with minor changes was also used in the definition in the Act of 1923. It was not until the MacQuarrie Committee Report in 1952, that other changes were made to the merger definition which now appeared under subsection 2(a)(iv). Apart from this earlier approach to mergers, the merger provision of 1960 remained virtually unchanged till 1986 except for the inclusion of services in the 1976 amendments. Notwithstanding these minor modifications, the criminal merger

---

10 The Judge's claim here was not absolutely accurate as the Crown's submission did directly or indirectly mention that the share of the pool car operator's market in the broadly defined market was 18%. See "Argument of the Attorney General of Canada," pp. 161-164.

11 The evidence discloses that the pool car operators were constantly concerned about increases in the railway line-haul tariff and how it would enable them to remain competitive with truckers together with the pool car operators financial difficulties and the bankruptcy of several operators.

provision proved to be unenforceable as it was extremely difficult to establish 'detriment to the public' beyond a reasonable doubt, required by criminal standards. As a result, between 1960 and 1986 the Competition Bureau did not win a single conviction under the criminal merger provision. It is therefore not surprising that the approach to merger enforcement took a dramatic turn, both as a result of the above noted difficulty and changes in economic philosophy towards mergers in 1986. These new provisions introduced in the 1986 amendments will be described hereafter.

a) The Merger provisions of the *Competition Act*

The merger provisions are contained in sections 91 to 103 of the *Competition Act* (formerly the merger provision was contained in section 33 of the *Combines Investigation Act)*.

For the Competition Tribunal to make an Order, the following elements must be satisfied. First, the relevant market or markets must be determined. Second, it must be determined that the merger or proposed merger prevents or lessens, or is likely to prevent or lessen 'competition substantially'. In determining the above, the Tribunal in accordance with section 93 may have regard to the following factors: the extent of foreign competition, the likelihood of the business failing, the availability of acceptable substitutes, barriers to entry (tariff/non-tariff, interprovincial barriers and regulatory entry control), the extent of effective remaining competition, the likelihood of removal of a vigorous and effective competitor, the extent of change and innovation in a relevant market and any other factor that is relevant to competition in a market. Further, the finding that the merger or proposed merger operates or lessens competition substantially shall not be based solely on the basis of evidence of concentration or market share. In *Hillsdown*, the Competition Tribunal stated "What will constitute a likely 'substantial' lessening will depend on the circumstances of each case. It is difficult to articulate criteria which might be applicable apart from the obvious ones of degree and duration. The degree of lessening can in some circumstances be assessed by reference to factors such as number of competitors left in the market, the amount of harm which can be done before the market is likely to again become competitive, for example, as a result of new entry. Various tests have been proposed: a likely price rise sustainable for one year; a 5% price rise sustainable over two years; a small but significant and non-transitory price rise. The Tribunal does not find it useful to apply rigid numerical criteria although these may be useful for enforcement purposes."

The third element that must be satisfied is whether the merger or proposed merger is likely to bring about gains in efficiency that will be greater than and will offset the effects of any prevention or lessening of competition that will result or likely result form the above transaction and that the gain in efficiency would not likely be attained if the Tribunal made the order. In considering whether the transaction will bring about the gain in efficiency, the

Tribunal shall consider whether the gains result in a significant increase in real exports, or a significant substitution of domestic products for imported products. Gains in efficiency shall not be considered brought about as a result of a redistribution in income. In *Hillsdown* the Competition Tribunal provided its views on the interpretation of the efficiency defence. The Tribunal would not exclude wealth or distribution effects from its assessment of the impact of the merger and could not assume that such effects are neutral, questioning whether the redistribution from Canadian consumers to foreigners was non neutral. Further, since the efficiency exception requires efficiency to be balanced against 'effects of any lessening of competition that will result or is likely to result from the merger or proposed merger', any reading down of 'effects of substantial lessening of' competition could result in exclusion of transfers from consumers or producers.[12]

The fourth element that must be satisfied is that the combination formed or proposed is not the result of a joint venture. However, this element is conditional and applies only if: the project or program would not otherwise have taken place; change of control of any party to the venture does not result; all parties to the joint venture enter into an agreement in writing which imposes an obligation to contribute assets; the agreement governs a continuing relationship; the joint venture provides for a termination date on completion; and the agreement restricts the range of activities that may be carried out.

The final element that must be satisfied is whether the transaction was made under the *Bank Act*, the *Trust and Loan Companies Act* or the *Insurance Companies Act* and certified by the Minister of Finance to the Commissioner that it is in the best interest of Canada or whether it was completed before the coming into force of this section (i.e., June 19, 1986) which is presumed to be met given the fact that the section came into force nearly fifteen years ago.

In addition to the above elements, no application for an order shall be made by the Commissioner if the merger has occurred more than three years after it has been completed and where an order is sought under the conspiracy or abuse of dominance provisions on the same or substantially the same facts.

The Orders that the Competition Tribunal can make range from dissolving the merger to divestiture. In designing the Order, the Tribunal has to fashion an Order which in its judgment serves the purpose of the Act or that

---

12

See *Hillsdown* cited above. The Tribunal's reasoning may be taken to imply one of three standards for the efficiencies defence: 1) Efficiencies must exceed the loss in consumer surplus resulting from the merger; 2) All efficiency gains must be passed on to consumers; and 3) The merger must not result in any reduction in consumer surplus (i.e., the U.S. approach or price standard). The efficiencies required under the US standard are larger than the efficiencies required under the approach suggested by the Bureau Merger Guidelines, i.e., the aggregate economic welfare approach to trade-off analysis.

the Order will be effective to accomplish the objectives of the Act. This does not imply that the Tribunal can impose any remedy, only those can be imposed which seek to restore competition to before the substantial lessening of competition occurred. In addition, the Tribunal also has the jurisdiction to grant relief from an earlier Order and to grant a new Order in accordance with the terms of section 92. The Commissioner can also seek a conditional Order or an interim consent Order or consent Order.

b) The complaint and its evaluation under the key elements of section 92

In 1995, Canadian Pacific Limited, its affiliates and subsidiaries acquired Cast North America Inc. and related companies which led to the filing of an application to the Competition Tribunal on Dec. 20, 1996, by the Acting Director of Investigation (on behalf of the Director). These two entities operate fully integrated intermodal container shipping service companies known as Cast and Canada Maritime Limited (i.e., CML). The application further alleged a substantial lessening of competition in the North Atlantic intermodal containerized shipping industry involving carriers operating between the Port of Montreal and ports in Northern Europe and the U.K.

First, in determining the relevant transportation service market for antitrust purposes, the alternatives or substitutes to containerized services were considered that is bulk or break-bulk ocean transportation services and air cargo services. The former services were considered lower in quality, slower and more labour intensive, whereas goods using containerized services are higher valued and can be more easily damageable. The latter services (air cargo) are more costly and the goods transported are very high valued and time sensitive. As a result, a substantial number of shippers would not switch to these alternative forms of transportation services in response to a significant, nontransitory increase in the price of containerized ocean shipping services. This led the Director to conclude that the product market was point-to-point shipping services for non-refrigerated containerized shipping services for each origin/destination market (i.e., Canada through the Port of Montreal and Northern Continental Europe or the United Kingdom) where the merged parties compete. In considering the relevant geographic market, the alternative routings serving the same origins and destinations were considered to be the services through the Port of Halifax and through the United States East Coast ports. The former port suffers from cost disadvantage (i.e., the costs of rail are 10 to 30 percent more) and time disadvantage (three days by rail from Halifax to Toronto or Montreal). The latter ports suffer an even greater cost disadvantage (i.e., the costs of rail or truck are 10 to 50 percent more) together with higher container handling fees. Based on these facts, it was concluded that these alternatives could not provide a supply response to effectively constrain a significant, nontransitory price increase by the merged parties and were therefore determined not to be part of the relevant geographic market. Second, whether the merger or proposed merger prevents or lessens, or is likely

to prevent or lessen 'competition substantially' had to be determined. Before the merger, the shipping lines offering services in the market were Cast, CML, Orient Overseas Container Line (OOCL), Hapag-Lloyd A.G. (HG) and BOLT. The market shares of these shipping lines were 31.0%, 31.7%, 14.7%, 7.8% and 14.8%, respectively. These three lines, CML, OOCL and HL, were members of the St. Lawrence Coordinated Service (i.e., SLCS had a market share of 54.2%) and were members of four conferences. With the proposed merger, the SLCS market share would rise to 85.2% or the market share of the merged parties would rise to 63%. Further, Cast was CSL's main rival and chief competitor and was an important factor in the tariff setting process within the conferences. In addition, the Herfindahl-Hirshman Index (HHI), a measure of concentration in a market, of SLCS increased from 4,118 to 8,220 and the HHI of the parties rose from 2,462 to 4,427.

In assessing, the factors to be considered under section 93, it was determined: that foreign competition from US East Coast ports was not likely to constrain the alleged parties from imposing a significant, nontransitory price increase; that Cast was not a failing firm at the time of the acquisition and alternatives may have been available even if it was a failing firm; that there were no acceptable alternatives as indicated in the above paragraphs; that there were a number of barriers to entry to the market due to complex logistics and significant capital investments, sunk costs due to the narrow ships required to enter Montreal, the length of service contracts which precludes a new entrant from securing business immediately, strategic behaviour of established competitors, failed past entry, and failure to attract shippers due to loss of cargo by failure of entrants; that the remaining competitors (HL, OOCL and Bolt) were not considered as effective remaining competitors, as HL does not operate any vessels in the market operating on a slot-charter arrangement, OOCL operates only two ships which are insufficient to operate a weekly service and BOLT operates only three ships which are insufficient to operate a weekly service in addition to its financial difficulties in the past; and that the merger would result in the removal of Cast as a vigorous and effective competitor; and that there was no change or innovation in the industry which is likely to increase competition in this market. In light of all these factors, the Director indicated that the merger will prevent or lessen, or is likely to prevent or lessen competition substantially in the market. Third, in determining the issue of balancing the gains in efficiency with any prevention or lessening of competition, no special issue arose as the merged parties did not claim that the merger would result in any efficiency. Consequently, this element was not an issue in evaluating this merger. Finally, the fourth and fifth elements needed to be examined were also met as the merger was not the result of a joint venture nor was it made under various financial acts and certified by the Minister of Finance. Further, the time duration for disqualifying the Director's application had not been met as the merger did not occur three years after it had been completed.

On September 9, 1997, the Competition Bureau announced that it was seeking a stay of proceedings regarding its application. The application was filed in response to the planned entry of a competing service in the market by a consortium consisting of Maersk Line, Sea-Land Service Inc., and P&O Nedlloyd Container Ltd. As the Bureau did not apply to the Tribunal to lift the stay by March 31, 1998, the complaint was dismissed permanently and the Bureau dropped the matter.

This case had several interesting aspects. First, the stay of proceedings sought by the Director indicates that an application is not written in stone, if market conditions change or if the situation no longer warrants, the Director can seek a stay in proceedings. Second, an application can be made by the Director even if another federal agency approves the application. In this case, the Canadian Transportation Agency had approved the acquisition. This indicates that the regulated conduct defence may not necessarily hold in civil matters. Third, the market was defined with regard to origin-destination pairs. In this case, the Port of Montreal and the Ports of Northern Europe and UK.

c) The economic analysis of the complaint

The acquisition of Cast North America Inc. (henceforth Cast) by Canadian Pacific is an example of a horizontal merger, involving firms that sell potentially competing products. This is in contrast with a vertical merger, in which, eg., a firm that produces a product and sells it at the wholesale level merges with a firm that purchases the product at wholesale and resells it at the retail level. In the case of a horizontal merger, the primary concern is whether the combination of two firms into one will reduce the overall level of competition in a market, by increasing the prices paid by the customers of the firms in the market and reducing societal welfare (according to some standard). In this context, the concern is whether the acquisition of Cast by CP would result in a reduction in the intensity of competition between cross-Atlantic shippers.

As an example of what the effect of such a merger might be, and the possible cause for concern, suppose that there are a number of potential shipping companies that transport goods from point A to B. These companies offer differing products from the point of view of the potential customer. Some firms ship from A to B using a combination of land transport and air transport, which will appeal to customers for whom rapid transport is important. Alternatively, other shippers will employ land and ocean transport, which is slower. Within this group of shippers, there is further differentiation, as some shippers offer containerized, bulk, and break-bulk services, which differ for example according to the protection offered to the cargo. In addition, to the extent that these shippers have different schedules and operate out of different locations, their products will be viewed as further differentiated.

Suppose that two of these shipping firms merge, so that a single firm operates the two different shipping lines (that is, suppose that the merged firm

does not choose to cancel all of the services provided by one of the two firms). When setting prices, the merged firm will take into account the effect the price of each service has on the quantities demanded, and hence profits, earned by the other service, something which the firms did not consider when they were separate entities. That is, the merged firm will recognize that lowering the price of shipping on one line may decrease to some degree the number of customers wishing to ship products on the other line. This recognition may result in higher prices charged by both shipping lines. In fact, in the end result, it could be that shipping firms other than the merged firms will also have higher prices, since merging two of the services has a potentially negative effect on the degree of competition as a whole, and does not just effect the merged firms. Therefore, to evaluate the full effect of the merger, we would need to evaluate the likely effect on the behavior of all firms, both the merging parties and those not party to the transaction.

The degree to which the prices of the merged services will increase, and the degree to which the prices of other shippers will increase, will depend in part on the degree of differentiation between the services offered by the merging parties. This differentiation could be both regarding the characteristics of the service, or the distance between the locations from which each party offers its service. For example, suppose the merger involved two firms with services that are viewed by consumers to be vastly different (examples in this circumstance might be an air shipper and bulk ocean transport, or shipping offered through ports that are far apart). If consumers do not view these two services to be close substitutes for each other, then the amount by which the price of one influences the sales and profits of the other is small. As a result, the merger would have a smaller impact than if the services were considered to be close substitutes. In this case, the price of one has a large effect on the volume of sales of the other, so that prices would be expected to rise significantly post merger. In the Director of Investigation and Research's application to the Competition Tribunal regarding the acquisition of Cast, it was alleged that since the merging parties both offered non-refrigerated containerized transportation services between the same origin and destination points, they could be viewed as offering products with only a small degree of differentiation.

As well, the amount by which prices will be expected to increase post merger will depend upon the similarity between the merging services and services offered by other firms, and the number of other firms offering similar services. If for example the other firms besides the merging parties offer a very similar product to that offered by the merging firms, then the anticipated price increase by the merged shipping lines would be smaller than if the other firms offered a very different product. Similarly, the merger would have a smaller expected effect on price if the number of close competitors was large. In the Director's Application, it was argued that there were only three other major firms offering similar products from ports sufficiently close to those used by the

merging parties. In addition it was argued that since two of these were members of the same shipping consortium as Canadian Pacific (i.e., CP), the degree of competition offered by these firms would be small. The parties naturally disagreed with these arguments. In their reply to the Director's application, CP argued that the Director had too narrowly defined the set of shippers offering close substitutes, both by ignoring substitutable forms of shipping and by ignoring the competitive pressure on the merging firms exerted by shippers operating out of ports outside of Ontario and Quebec.

In addition to existing competitors, the effect of the merger on prices would need to consider the potential for new competitors, that is new firms who would offer similar products within a short geographic distance. One can easily see why a merger might attract entry, since a lessening of competition would result in higher profits for all firms in the market. Again, the parties and the Director disagreed on the barriers to entry in this case, and in fact it was the planned entry of competing service that lead the Director to seek a stay of proceedings and to ultimately discontinue with the matter.

Finally, our discussion so far has said nothing about the possible role of costs. Consider again our example, but suppose that the merger will result in cost savings to the merging parties. Cost savings might be for example due to cost synergies from combining certain production and transportation processes. In the analysis of a merger, cost savings could have more than one effect. The first is that cost savings might put downward pressure on the prices charged by the merging parties (and indirectly the parties not involved in the merger). This would be the case if the merger were to result in lowering the parties' marginal costs, that is the costs of producing an additional unit of the product being sold. This decrease in marginal cost would have a downward effect on the prices being charged post-merger, possibly resulting in lower prices than pre-merger. In the Cast example, it appears that neither Canadian Pacific nor the Competition Bureau made any reference in their submissions to reductions in marginal costs.

Secondly, even though the effect on costs does not result in lower final prices, it may still lead to a merger having effects that benefit society as a whole. Suppose as an example that the merger resulted in savings in overhead costs that would have no direct effect on the prices being charged by the merged firm. It is possible that these savings may still outweigh the detrimental effect on society of the higher prices. In their submission, Canadian Pacific alleged that the transaction would lead to cost savings of approximately $20 million, although they do not document the sources of these savings (with the exception of a mention of reduced inventory costs). Unfortunately, the question of how these two effects should be measured and the method of comparison that should be used is, still a question of debate among economic and legal communities.

4. Conclusion

Today economics provides a substantive basis for competition law, and economic analyses generally plays an important role in these cases. Economics provides a screening device for identifying practices that undermine competitive behaviour. Judges deciding competition cases not only accept expert economic testimony but also seek expert guidance that will point to the correct result as often the economic reasoning is quite complex.

In conspiracy cases, there are two basic steps used in the economic analysis once it has been established that the accused entered into an agreement to lessen competition. First, one begins by determining if the parties to the agreement possess market power or are they likely to obtain market power by virtue of the agreement. In other words, do the accused have the power to enable them to behave relatively independent of the market, in a passive way. It is not necessary for the accused to have a monopoly, possession of a moderate degree of market power may be sufficient to support a finding of undueness. In determining this matter, market share alone is not sufficient to demonstrate market power. It is also important to determine if barriers to entry exist as it enables one to identify whether the accused have market power. Second, the agreement must be or must have been likely to injure competition. Some agreements that merely involve exchange of product standards, or define trade terminology, etc. may not injure competition. Turning back to market power, the question is market power in what and where? It is here that defining the relevant product and geographic market becomes important. The determination of the relevant product and geographic market is an art and science in which practitioners in economics have developed considerable experience over nearly a hundred years of antitrust work. To do so requires considerable skill and training in economics and generally expert economic testimony is brought in by contestants in antitrust cases. This was evident in the case examined - *R. v. Clarke Transport Canada Inc.* - where the determination of the final outcome rested on a central issue - was the pool car operator services the relevant market as explained by the Crown? The Judge was of the view that it was not. As a result, in the broader market definition which included trucking and intermodal rail services, the Judge was not convinced that the accused had market power to satisfy the undueness test.

In merger cases, several analytical steps are used in determining the outcome of a merger. First, the relevant market or markets must be determined together with any changes in concentration that will occur. Second, the competitive concerns must be assessed to determine if the merger prevents or lessens competition substantially or is likely to do so. In determining the latter, a number of factors should be considered such as the extent of foreign competition, the likelihood that the merging firm would fail, acceptable substitutes, barriers to entry, the extent to which effective competition remains, the extent to which the merger would remove a vigourous competitor, the extent of innovation in the market and any other factor relevant to competition

in the market. Third, the efficiencies resulting from the transaction and the balancing of these efficiencies with the effect of any lessening of competition have to be addressed to determine if it will offset the latter and if the efficiencies would not occur had the merger not been permitted. All these analytical steps call for an increasing use of practioners in economics with antitrust experience.

The above cases, clearly demonstrate how the use of economic analyses and expert economic witnesses is not only accepted in the Courts but is also quite normal in the examination of other competition law complaints. This abodes well for future development of competition law as the underlying reasons behind various provisions of the Act are becoming increasingly complex. Further, in the second or new millennium, the increased use of econometrics as a tool in antitrust analysis will be increasingly witnessed in the courts.

* *The views expressed are those of the authors and not necessarily those of the Competition Bureau.*

The Green Mile: Approaches and Attitudes Toward Environmental Sustainability Among Canadian Transport Faculty

Anthony Perl
&
Michael Bates

Research Unit for Public Policy Studies
University of Calgary

The authors would like to acknowledge the support of the Government of Canada's Climate Change Action Fund, the Centre for Sustainable Transportation and the research assistance of John Roslinski.

INTRODUCTION

The following analysis was undertaken to assess current attitudes and approaches of faculty at Canadian post secondary institutions whose teaching includes at least some transportation content. The survey is part of the Centre for Sustainable Transportation's "Enabling Tomorrow's Transportation Professionals to Address Climate Change" project. A previous phase of research arrived at two primary findings: first, environmental education and transportation education exist as two solitudes in Canadian post secondary education, and second; no 'canon' of education materials exists to support educators in equipping students to attain environmentally sustainable transportation.

The project's objective is to improve Canadian university capacity to train professionals for the challenge of mitigating the transportation sector's contribution to climate change. A more detailed analysis of this survey is available online at http://www.cstctd.org/CSTadobefiles/Research_Report.PDF

The faculty survey reported on here was intended to provide an in-depth and nuanced analysis of what is actually taking place in Canadian post-secondary transportation teaching. The survey was originally drafted in April 1999 by RUPPS project staff and then reviewed at a focus group of transportation faculty attending the Canadian Transportation Research Forum's May 1999 meeting. Significant changes were made to the survey following the focus group.

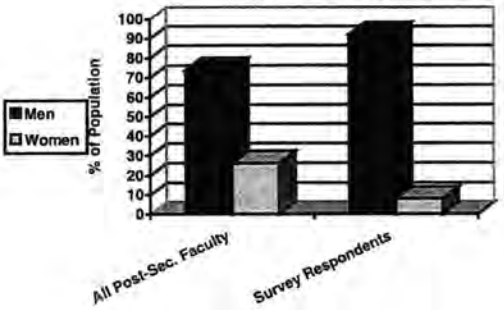
262 instructors were invited to participate in the survey. Additionally, 26 department heads were contacted to publicize the survey among their staff. 113 other contacts were made with administrators and support staff in search of respondents and to promote the survey. In total, 133 persons visited the survey website with 119 providing at least some response, representing a response rate of 45.4%. Of the 119 respondents, 110 visited every page and question with the remaining nine visiting only some of the questions. Respondents did not necessarily answer every question and therefore N values will vary from question to question.

In total, the 119 respondents reported on 268 courses for an average of 2.25 courses per respondent. 44 respondents reported teaching a single course related to transport, 35 listed two courses, 19 taught three, 8 taught four courses and 13 respondents reported on five transport related courses, the maximum allowed by the survey.

Demographics

Respondents were asked 19 questions about themselves, ranging from their age and gender to employment status. Ninety-seven men participated versus only eight women, although 14 respondents did not complete the question on gender. This reveals both an absolute gender imbalance among our population of transportation faculty and a relative gap with the gender distribution of Canada's post-secondary faculty as a whole. Figure 1 compares the gender split of faculty as a whole (74% male versus 26% female) to those teaching transportation

Figure 1
Gender of Canadian
Post-Secondary Faculty



(92% versus 8%).[1] Some implications of this gender gap are considered below.

Eighty percent of respondents were over the age of 40 and 51% were over the age of 50, revealing a 'graying' tendency among transport instructors. There is relative symmetry between the age distribution of faculty teaching transportation in our survey and the entire population of university educators, which emphasizes the gender gap shown in Figure 1. Given that a slightly higher share of transportation faculty in our survey can be found in the young (31 – 40) and very young (30 and under) age cohorts compared to faculty as a whole, the gender gap in transportation teaching cannot be attributed to a lack of "new blood." Despite the fact that gender is a recognized criterion in employment equity policies at universities, younger transportation faculty do not appear to be making much, if any, difference to the gender gap.

Not surprisingly, given the age of respondents, most were tenured (57/106) and more were Full Professors (42/109), rather than

[1] Data on faculty gender and age distribution obtained from Association of Universities and Colleges of Canada, http://www.aucc.ca/en/acuindex.html.

Associate or Assistants (18 and 22 respectively). Eighty-five were full time employees of their university versus 23 who were not.

When asked for general comments on the environmental aspects of their transportation teaching, respondents provided a full range of responses. Several respondents offered comments that illustrated the 'two solitudes' problem revealed in our earlier analysis of environmental content in transportation courses:

"To go very far into environmental issues requires a significant amount of knowledge I feel I do not have...I need to develop expertise in the areas I am strongest. Other specialists can deal with these other issues or, if academic institutions believe the topics you address to be important, provide resources for the development of this expertise and course material."

"An environmental economics course is offered in our department. In this light, minimal attention is paid to the environmental impacts, although environmental concerns are often discussed... in associated areas in class.

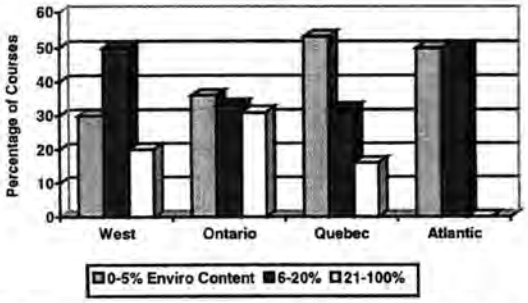
ANALYSIS

Below is an exploration of where opportunities for change in Canada's university transport education system exist, who might lead that change, and how it might be launched.

Where Should Curriculum Development Begin?

Our survey reached into nine of the ten provinces, with only Prince Edward Island not represented. Of the 119 respondents, 41 originated in the West, 41 in Ontario, 20 in Quebec, and 14 in the Atlantic region. Within the regions, provincial representation was relatively balanced in relation to population, suggesting an encompassing national perspective was obtained on those teaching transport and related content at Canadian universities.

Figure 2
% of Environmental Content in Transport Related Courses by Region
n=112

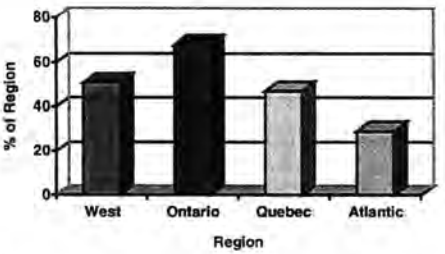


Very little variation of modal focus in courses existed by region.[2] The West exhibited a relatively high level of aviation content and, not surprisingly, the Atlantic region had a large number of maritime transport courses. But overall, these differences were slight. When looking at discipline, the number of Engineering and Geography courses was relatively equal in the West and Quebec, whereas the transportation and related courses in Ontario and the Atlantic region were dominated by Engineering. Thus, more interdisciplinary opportunities to bridge the two solitudes of transport and environmental education (generally found in Engineering and Geography) would seem to exist in the West and Quebec.

Figure 2 shows that transport and related courses taught in Quebec and the Atlantic region exhibit a relatively low amount of environmental content. Fifty two percent of courses taught in Quebec, and 50% in the Atlantic region, featured only 0-5% environmental content. Only 16% of Quebec courses, and none of Atlantic courses contained more than 20% environmental content, as opposed to 20% in the West and 30% in Ontario. In addition, Ontario respondents were most likely to strongly agree with the

[2] All references to courses refer to those listed first by respondents, who could provide information about up to five courses in the survey, except where otherwise noted.

Figure 3
Region Distribution of Respondents Strongly Agreeing that
"Environmental issues are important in transportation education"
n=58



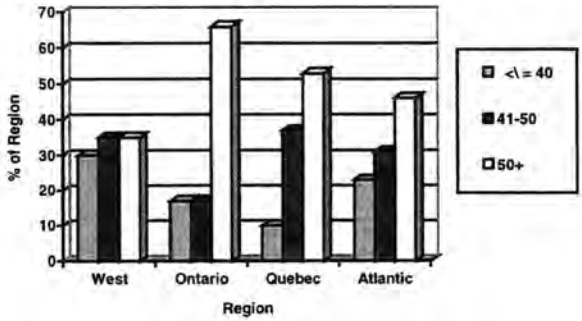
statement "Environmental issues are important in transportation education" than respondents from any other region as seen in Figure 3.

When reporting on the perceived difficulty of incorporating environmental issues and techniques, respondents from Ontario (52% in the low difficulty category) and the Atlantic (43%) region reported the least difficulty, respondents from Quebec (35% in the high difficulty category) the highest, with the West falling in between. Finally, in terms of age, respondents from Ontario and Quebec were older than those from the Atlantic or Western regions, with Ontario faculty being significantly older. Sixty seven percent of Ontario respondents were over the age of 50. In the West the figure was 35%. Figure 4 illustrates the contrast in age between regions.

Regional Summary

It would seem that Ontario with its large share of Engineering courses and high degree of faculty support for the idea of introducing environmental content into transportation courses could provide fertile ground for initiatives aimed at bridging the two solitudes we have identified. Quebec and the Atlantic region appear to exhibit a wider gulf between the approaches and attitudes associated with a

Figure 4
Faculty Age by Region
n=105



traditional transport curriculum and one that links environmental education with transportation. The West falls somewhere in between. Ontario's unique age demographics are also significant and could influence the consideration of who might lead the change to a more environmentally sustainable transportation curriculum.

Who Among University Faculty Lead in the Development of an Environmentally Sustainable Transportation Curriculum?

This section attempts to pinpoint the disciplines and demographic groups that are prone to transportation curriculum innovation and most receptive to incorporating environmental issues and content. It seeks to answer the question "Who is best positioned to lead change?"

The most obvious place to begin identifying curriculum innovators is by looking at the group of respondents indicating they have changed course material or instructional format since the last time their course was taught. Fifty five percent (63/113) of respondents indicated they had changed their course since last teaching it, 29% (32/113) reported

not changing it and 16% (18/113) taught the course for the first time[3] The type of change being made in the course did not necessarily relate to environmental or sustainability issues. In the analysis that follows, we refer to faculty who reported changing course content and/or format as "Changers" and those not changing content and format as "No-changers."

When it came to incorporating environmental issues and techniques into their curriculum, Changers were more likely to report a "High" degree of difficulty than No-changers. This could indicate that the changes currently occurring in Canadian university transport curricula have little to do with environmental issues and impacts. Both groups agreed with the statement that "Environmental issues are important in transportation education" in roughly equal numbers. Interestingly, No-changers agreed even more strongly (47% either agree or strongly agree) with the statement "I should expand the environmental content of the transportation related courses I teach" than did Changers (32% agree or strongly agree).

When it came to identifying the initiative that would be most useful in broadening and/or deepening the environmental content of transportation related courses, Changers selected course development symposia and workshops more often than No-changers (20% vs 6%). No-changers had a higher preference for new texts and readings (69%) compared to No-changers (43%), though both groups preferred the latter overall.

Variation in Incorporating Environmental Content by Course Type

Respondents could fit their course into a university's academic requirements in six different ways. These included: Transport Program Requirement, Other Program Requirement, Elective for Transportation Majors/Specialists, Elective for Other Majors /

[3] Refers to the first course recorded only. Only five out of 119 respondents didn't change materials or format of their first course, and changed materials or format in subsequent courses they recorded.

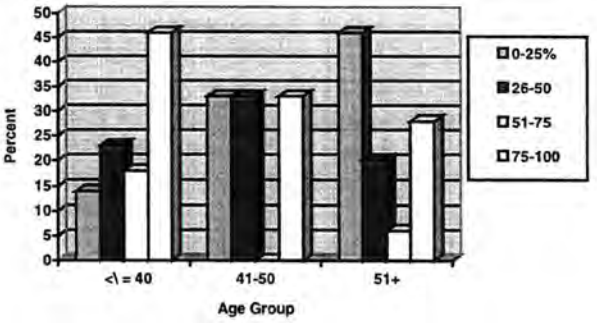
Specialists, Service Course and Professional Requirement. The one area where faculty reported difficulty in incorporating environmental material was in courses that were a "Transport program requirement." Given that such courses are central to the education of future transportation professionals, this finding highlights a potentially significant obstacle to the development of new skills that facilitate environmental sustainability.

Gender and Curriculum Development Leadership

Only eight women participated in our survey, which in itself is significant. Given our extensive outreach efforts to encourage participation, this result reveals a male-dominated workforce among Canadian transportation educators. Because of the small number of women respondents, it is not possible to draw firm conclusions about the gender gap that exists in attitudes and practices among transportation faculty. But there is some evidence that women differ significantly from their male counterparts in responses to this survey.

Five of the eight women (63%) indicated a 'low' level of difficulty incorporating environmental content into their transportation teaching, compared with 32 out of 86 (37%) men. This represents a 26% 'gender gap' on the perceived difficulty of incorporating environmental content into transportation teaching. A similar gap emerged on several other questions. Women were 25% more likely to address non-motorized transport than men, and also more likely to address inter-modal and rapid public transit, by 9% and 6% respectively. Women were 12% more likely to strongly agree with the statement "Environmental issues are important in transportation education" and addressed linkages between environmental impacts and human health issues more often (63% to 42%). Finally, more of women's transport teaching covered environmental issues: 50% of women were in the high environment content group (21-100% of teaching related to environmental issues) as opposed to just 18% for men. While these differences may be affected by the small sample of female faculty, collectively these results suggest that women may be

Figure 5
% Transportation-Related Teaching by Age Group
n=106



more willing than men to bridge the solitudes of environmental and transportation education.

Age and Curriculum Development

On the whole, faculty age and experience are positively correlated with openness to linking environmental issues and impacts with transportation teaching. No relationship was found to exist between age and transport mode covered. Further, older respondents are not incorporating any less environmental content in their transport related courses compared to younger counterparts. Those respondents with the most years (21+) of teaching at both the under graduate and graduate levels, indicated the least difficulty incorporating environmental content.

Finally, it is noteworthy that, in total, older professors teach less transport related content than do younger professors.[4] Figure 5 shows that only 28% of the professors teaching 76-100% transport related content are aged 51 and over as opposed to 46% who are 40 or less.

[4] Significant at p<.05. Gamma= -.311

This may reflect the diversity of teaching that a senior professor would take on as opposed to a younger faculty member in a specialized field. Again, this is problematic because it suggests that those faculty reporting a low level of difficulty incorporating the environment into their teaching cover less transportation-related content compared to those reporting more difficulty in incorporating environmental issues.

Curriculum Development Leadership Potential Among University Faculty

Geographers and Planners, senior professors and women faculty appear to be the most receptive to building bridges between environmental and transportation education. These groups constitute a core of environmentally conscious faculty teaching about transportation at Canadian universities who could be tapped to lead more systematic and ambitious curriculum development efforts. But engaging this core group of curriculum development leaders is-easier said than done. As we have noted, there are very few women faculty in the transportation field. Furthermore, senior professors appear to teach courses with less transport content than their junior colleagues. This relatively small number of curriculum innovators would be faced with at least three significant constraints.

First, professors currently making changes in course content and format are not particularly open to enhancing the environmental dimensions of their teaching. This resistance to integrating environmental and transport education is suggested by the lower support for environmental content development among faculty who reported changing their courses. Second, change oriented faculty also indicated greater difficulty in introducing environmental materials into their courses than faculty who have not changed their course content. And finally, faculty teaching required courses in transport programs indicated more difficulty in integrating environmental content than all other types of transportation and related courses.

What Kinds of Content Should Curriculum Development Initiatives Consider?

This section lays out what the survey has revealed about recent changes to transportation and related courses and assesses where that knowledge points to.

Materials used in presenting environmental issues and techniques

Respondents indicating a change in course material or instructional format since the last time a course was taught were invited to comment on what exactly had changed.[5] The most common change reported consisted of updating course material to keep the content current. While some of this change might have involved the elaboration of environmental issues and/or techniques, there is no reason to suppose that environmental material played a large (or even a moderate) part in such updating. The second most common form of change was very specific to the course in question, and was classified as "Other." Typical responses here would include adding material on: freeway modeling, real time approaches or I.S.O. guidelines. The third highest number of changes arose from the use of internet and related new communications media. Many of these respondents identified creation or updating of course specific web sites where assignments, learning materials, and/or notes were made available to students over the internet. Specific references to the use of new software applications were also common in this category of change.

There were only seven specific references (6% of total changes) to course modifications that added environmental and sustainability content confirming that those making change are not currently focused on introducing or expanding environmental content. Based upon current practices, it would appear that new environmental content would be most likely to be taken up by faculty in one of two ways:

[5] Respondents could indicate more than one choice per course.

First, it could be tailored to existing issues, tools, and techniques so that it would be well positioned for incorporation as an "update" of existing course material. Such an ability to fit into existing course structures would be especially important for those courses classified as transportation program requirements, where faculty reported the greatest constraint on change. Second, new environmental content could be delivered through the web and other computer based information systems. This would take advantage of the new communications and information technology being incorporated into the curriculum. Beyond these two categories of change, some respondents did provide specific recommendations on material that belonged in an environmentally sustainable transportation curriculum.

Environmental learning materials currently used in transportation courses

When respondents categorized the kinds of materials they use in presenting environmental issues and techniques from a list of nine choices, they emphasized journal articles, the professor's own research, and government documents. Respondents were also invited to include any specific reference they felt was an important part of an environmentally sustainable transportation curriculum.

The most important attribute of these citations was that no two respondents identified the same material. This divergence among the substance of environmental education resources repeats, and thus confirms, our finding in the previous curriculum content analysis. Unlike typically "mature" academic subjects, including many other aspects of either transportation (e.g. road design or demand modelling) or environmental studies (e.g. impact assessment), there is neither a textbook, nor a "canon" of common articles, book chapters, or research reports that constitute the intellectual core of learning materials. Curriculum development leaders would thus face the starting point of a "mixed bag" of educational materials currently in use.

Curriculum Content Development Options

While much change is occurring in the transport classroom, very little of that change is related to the environment and sustainability. The internet and other new technologies are rapidly making their appearance, and could provide a vehicle for future expansion of environmental content and techniques. Materials currently used to teach about the environment and transportation are a mixed bag of textbooks, (mostly foreign) videos, government documents, web sites and conference papers, suggesting a real need to boost the supply and availability of learning materials that address the environmentally sustainable transportation. One respondent commented explicitly on this need, and others alluded to it:

"I would prefer if a balanced textbook were available. It would discuss what we know about urban travel demand and ways of improving travel/accessibility options in an economically and environmentally sustainable manner."

"There does not exist a good Canadian transportation economics text...I will be switching again this year and use a US textbook."

CONCLUSIONS

This analysis of attitudes and practices among Canada's transportation faculty reveals that the low level of environmental issues and approaches in today's curriculum that was documented by our previous content analysis of course syllabi did not occur by coincidence. This limited engagement of environmental issues, impacts, and instruments in Canada's transportation courses corresponds to attitudes and practices among faculty. While many faculty are not opposed to such linkage in principle, we have found relatively few inclined to develop such linkages in their own teaching.

Left to their own devices, a small and disparate cohort of curriculum innovators appear to have developed course-specific blends of environmental and transportation education, but the likelihood of

such customized innovations spontaneously diffusing across the general population of transportation faculty must be regarded as slim. Too many faculty feel constrained by various demands – ranging from the general requirements of transportation programs to the perceived difficulty of adding environmental content to their particular course, to enable a widespread adoption of best practices without concerted effort. That effort can benefit from building upon the ideas about, approaches to, and examples of bridging environmental and transportation education that our research has uncovered. Such work can, hopefully, yield fruitful results by taking account of what has already been accomplished in curriculum innovation as well as what has stood in the way of further progress.

In regional terms, Ontario faculty offer the most promising indicators of receptivity and efficacy in curriculum development. Here, the share of environmental content in transportation courses is high, as is the perceived importance of incorporating such content into a transportation curriculum. The difficulty of incorporating environmental content is perceived to be low in Ontario, compared to other regions. Quebec faculty present the most challenging indicators, reporting low environmental content in their transportation courses, relatively low perceived importance of such content, and relatively high difficulty in incorporating such content into the transportation curriculum. The Atlantic region and the West fall in between these extremes.

The search for curriculum development leaders in any region should begin with Geographers and Planners, senior professors, and women faculty who indicate the greatest current receptivity to building bridges between the two solitudes of environmental and transportation education. But it should not end there. Engaging Engineers, Economists, and Management/Commerce faculty to support development efforts is essential to understanding and overcoming current constraints and/or preferences for "business as usual" among many transportation faculty. Clearly, new incentives and approaches will be needed to bridge the two solitudes.

Perhaps the greatest challenge facing environmentally sustainable transportation curriculum developers will be in developing the right "product(s)," that can facilitate rapid and extensive diffusion of sustainability skills. The substance and form of new learning materials that integrate environmental sustainability with transportation education will make a critical difference in their level of use across Canada. Our findings suggest that learning materials viewed as desirable and appropriate could easily be incorporated into the regular updates which faculty make to course content. Learning materials delivered via the web could be particularly attractive and rapidly incorporated into courses.

Developing such materials through partnerships among Canadian universities and/or Institutes appears attractive to our survey respondents. It remains to be seen what roles individuals will see for themselves in such a partnership. This will be one topic among many to be explored at the curriculum development workshops being organized by Canada's Centre for Sustainable Transportation. More information about the outcome of these workshops and future initiatives to develop a sustainable transportation curriculum will be available at the Centre's web site: www.cstctd.org.

MODERN ROUNDABOUTS, GLOBAL WARMING, AND EMISSIONS REDUCTIONS: STATUS OF RESEARCH, AND OPPORTUNITIES FOR NORTH AMERICA

Executive Summary

The escalating rate at which modern roundabouts are being built at North American holds a significant potential for conserving energy, reducing air pollutants and addressing global warming. Modeling and empirical studies of busy highway intersections document a substantial short term benefit from installing modern roundabouts. Motor fuel consumption and associated air pollutants are reduced, and the primary global warming gas (GHG), CO_2, is significantly cut. The modern roundabout arrived in North America in 1990 and has rapidly increased in numbers. The US and Canada urgently need a variety of strategies to return to GHG generation to levels below those of 1990 by 2012. The roundabout is one such strategy. Two decades of intersection control modeling and software development and research, establish that substantial fuel savings at busy intersections can be gained by employing roundabouts rather than traffic signals. Reduced fuel consumption–and by extension pollution emissions and GHGs–are demonstrated through analysis of empirical data and modeling reported from existing US roundabouts and those under development.

The magnitude of impact from roundabouts on Vermont statewide motor fuel use is suggested by a hypothetical installation of roundabouts in place of signals at 100 busy intersections.

According to models, this change would decrease total annual motor fuel use by approximately 8% of 1997 statewide consumption. The effect of a modest number of roundabouts on various climate action plans is probed by estimating an arbitrary number of intersections transformed to roundabouts. It is suggested that 25 roundabouts replacing existing traffic signals in the City of Burlington, Vermont would equate to over 20% of that City's goal of bringing GHG emissions to 10% below the base line 1990 level. The value and effects of increasing numbers of roundabouts in North America need to be explicitly incorporated as planners develop policies and strategies for energy conservation, global warming and air pollution.

The longer term potential for roundabouts to reduce resource use–including pollution and GHGs--by enabling higher density land uses and fostering increased transit and non-motorized modes has been documented elsewhere. Roundabouts are generally acknowledged to be superior in all aspects of performance to alternative intersection controls. Roundabouts usually recover their capital costs in a few years through savings to all users in reduced crashes and crash severity, time delay, maintenance, and fuel savings.

Introduction

This paper reviews a portion of the literature on roundabouts and pollution emissions; describes the Australian intersection model and the lack of a comparable US model; reviews applications of the Australian model to US intersections; and concludes by evaluating the impact of installing roundabouts in the City of Burlington as a strategy to implement its plan for global warming gas reduction to below1990 levels. Results of evaluation of stop delay at the first two-lane roundabout installed in the northeast at Brattleboro, Vermont, in 1999 is also presented.

Modern Roundabout History

Traffic signals, roundabouts, and stop control all fall under the transportation category of "intersection control devices." The traffic circle, the modern roundabout predecessor, as an intersection control device dates from 1905 when the first one, Columbus Circle, began to

operate in New York City. Traffic circles are large, high speed and historically, entering traffic had priority over traffic within the circle. The modern roundabout era began in 1996 with the adoption by Britain of "yield-at-entry" rule for vehicles entering a roundabout, giving vehicles in the circular travelway the right-of-way for the first time. Roundabouts evolved into relatively smaller diameters than traditional circles and now possess splitter islands that constrain speeds at entry. The "yield at entry" modification resulted in increased roundabout capacity and a sharp decline in crashes. The modern roundabout era began.

Roundabouts then spread from Britain to the European continent and British influenced countries, particularly New Zealand and Australia. The first two roundabouts were built in North America in 1990. The 19th in the US and first roundabout in the northeast was constructed at Montpelier, Vermont in 1995. There were about 100 roundabouts in the US and Canada in 1999, and 300 at the end of 2000. Observers expect US and Canadian annual development of roundabouts will reach 1,000 annually within a few years, a construction rate maintained by France for over a decade.

The Air Quality Challenges

Air Quality–Pollutant Emissions and Global Warming
Environmental air quality concerns range from neighborhood to global in scale. Global warming and the ozone layer thinning comprise two universal air-based threats to the earth's living environment. Fossil fuel pollutant emissions cause immediate and long term direct damage to human health as well as to the balance of ecology.

Polluted air and global warming continue to be major issues for the United States and Canada. US law and a regulatory framework for reducing air pollution, particularly for those below standard ("non-attainment") areas, date from the US Clean Air Act of 1970. The primary target for reduction to combat global warming, carbon dioxide (CO_2), is not an emission of concern in clean air policy and regulation. Global warming as a substantive concern emerged from

the United Nations Conference on the Environment and Development's "Earth Summit," held at Rio de Janeiro in 1991. Over 150 nations signed the UN Framework Convention on Climate Change, committing them to voluntary actions to stabilize GHGs. As developed nations, the US and Canada committed to reducing GHGs at the Third Conference of Parties to the United Nations Framework Convention on Climate Change at Kyoto in 1997. For the US and Canada this means cutting GHGs "business as usual" projections by 25-30% in the period 2008-2012. The targets for the US and Canada are 7% and 6% respectively below the 1990 levels (US Office of the President, p 21).

Air Pollution and Transportation in the United States

The transportation sector is a major source of GHGs and air pollution emissions. The US transportation sector consumes about 65% of total oil consumption, produces almost a third of US carbon emissions and substantial amounts of most air pollutants (Interlaboratory Working Group, 2000, p 6.1). "Air pollution" as used here refers to the primary pollutants that contribute to damage to health and the larger environment:

1. Smog, a combination of pollutants with direct and indirect health effects and other environmental damage, including a resultant formation of damaging ozone. Transportation emissions contribute a major portion of smog, including substantial amounts of volatile organic compounds (VOCs) and nitrogen oxides (NO_x).
2. Sulfur oxides (SO_x) which cause direct human health damage and constitute the key pollutants that cause acid rain damaging lakes, streams and forests up to hundreds of miles away. Transportation contributes a significant amount to this factor.
3. CO directly causes negative health affects and contributes to smog. Transportation produces two-thirds of CO emissions in the US, mostly from motor vehicles (Fueling Vermont's Future, I, p 3-8).
4. CO_2 that in itself comprises the major global warming gas as noted above. Each gallon of motor fuel generates 19.6 pounds of CO_2.
5. Particulates, microscopic particles, commonly called soot, can be absorbed over time in the lungs and cause respiratory disease and exacerbate other circulatory or respiratory ailments. Larger trucks and buses alone contribute to over half the soot from motor vehicles

(Rolling Smokestacks, p 2). Overall transportation contributes to less than 10% of particulate pollution.

Some progress has been made since 1970 in the US in reducing several pollutants, particularly in erasing atmospheric lead through eliminating lead in motor fuels. However, increased vehicle miles of travel continue to add NO_2 and CO_2 emissions. And, "US transportation energy consumption and greenhouse emissions are expected to grow robustly over the next few decades (US Department of Energy, 2000, Chapter 6.1)." Within this context, CO_2, CO, VOCs and NO_x are the compounds that, when empirically measured, determine whether an air shed meets (attainment) or fails (non-attainment) United States air quality standards. Particulates and CO are subject to separate standards for attainment and non-attainment. US Federal air pollution standards for the transportation sector over the past two decades have improved air quality or dramatically slowed its decline. A rapid increase in the GHG CO_2, is a clear exception.

Intersection Modeling for Energy Consumption and Emissions
The modern roundabout is acknowledged to be generally superior to other forms of intersection control in the areas of capacity, user delay, and safety. Most early roundabout research and assessment focused on capacity, delay, and effective design. Australia first developed fuel consumption models in the late 1970s and over time expanded them to include major pollutants. Only in the 1990s did the question of intersection control and emissions begin to receive the attention of North American researchers. The US Environmental Protection Administration (EPA) Mobile 5 model (Mobile) does not contain a module that effectively models intersection performance. The Mobile computer program first developed in1978, was built on sets of equations, assumptions, and adjustments to laboratory and field tests. Factors for translating fuel consumption to specific emissions are provided for each area of the country. States must use Mobile in determining air quality planning, effectiveness of different strategies, and measuring results. "Mobile" measures three pollutant emissions: CO, NO_x, and HC (US General Accounting Office, 1997, p 1). One Federal Highway Administration (FHWA) policy administrator

viewed Mobile as only providing gross estimates of transportation emissions (Savonis, 2000, p 2).

Australia Develops Models with Environmental Applications
As a result of policies emphasizing cost efficiency in selecting highway investments, Australia developed intersection fuel consumption models and software about the time the roundabout began to be used significantly there. The roundabout was added to the analytical software, SIDRA[1], and this software in the North American version operates in conformance to the highway practice manual, the US "Highway Capacity Manual." SIDRA which, in a single software, models and compares performance of all intersection control options, is the most user friendly, versatile, and comprehensive tool for transportation planners and engineering practitioners. British software is considered the most refined design software for roundabouts.

Australian concern for cost efficiency in highway investments required an analytical procedure, a micro model, to determine the relative performance of signals, stop control, and eventually roundabouts. The software, SIDRA-2, released in 1982 (Bowyer, Akcelik and Biggs, 1985, p 24) was one of the first to model vehicle fuel consumption based on a four-part cycle: cruise, deceleration, idle and acceleration to cruise.

Figure 1 provides a clear outline of fuel consumption during four phases a vehicle encounters at an intersection: cruise, deceleration, stop, and acceleration to cruise–all are scaled by time, distance, and speed (Bowyer, Akcelik and Biggs, 1985, p 3). Clearly vehicles stopping and accelerating utilize far more fuel than during cruising, and fuel used while idling in stop delay at intersections uses far less fuel than either acceleration or cruise phases.

[1] SIDRA is a generic description of a software product that has gone through several editions. Originally produced under the Australian Road Research Board, its development and distribution transferred in 1999 to Akcelik and Associates, Victoria. The current edition is aaSIDRA.

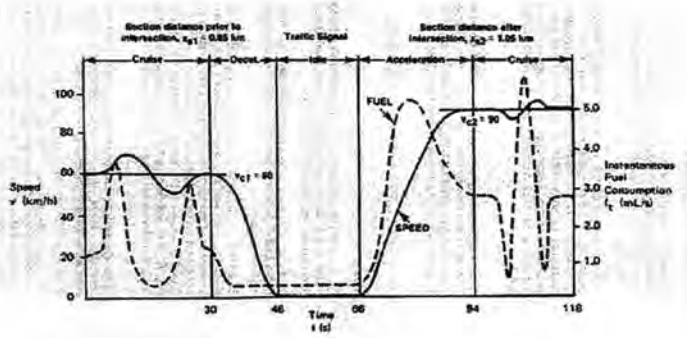


Figure 1: Speed Time Trace and Instantaneous Fuel Consumption (Bowyer, Akcelik and Biggs, 1984)

Starting in 1982 emissions, energy consumption, and cost components began to be added. SIDRA now models intersection performance for the following: pollutant emissions (HC, CO, CO_2, NO_x, and lead); delay (stopped, queuing, geometric, overall); energy (fuel) consumption, and operating cost. In order to predict performance, SIDRA models each travel lane, and both light and heavy vehicles.

Previous studies and other models have addressed intersection environmental performance. Other researchers identified differentials in emissions between roundabouts and traffic signals (Hoglund, 1993; Garder, 1998). A Swedish model, CAPCAL 2, released in 1996 also calculates performance measures, including vehicle costs and emissions, for all intersection types (Hagring, 1997). Other studies are centered on the intersection cycle of emissions at signalized intersections (Colyar, Frey, Rophail and Unal, 2000). The latter study determined that the stop and go cycle produced about twice the amount of emissions when compared to Mobile analysis predictions

Specific Intersection Research, Analyses and Assessment

Thessaloniki, Greece

A research paper (Mustafa and Vougias, 1993) tested SIDRA against known measured emissions at an intersection where a roundabout was replaced with a traffic signal. Measurements at the intersection of CO, HC and NO_x at a monitoring station showed a clear relationship between traffic levels and emission levels. The SIDRA simulation using the intersection traffic data revealed lower emissions for the roundabout for all four pollutants over a wide range of traffic volumes.

East Lansing, Michigan

A roundabout related study compared three stop control intersections and three small traffic circles at Michigan State University at East Lansing, Michigan (Savage and Al-Sahli, 1994). Traffic circles tend to be large–300 to 400 feet and more in diameter–and high speed.

The circles in this study were small, 150 feet in diameter, and similar in size to a moderate size, multi-lane roundabout. All but one approach to one intersection were two-lane divided roads with medians. All intersections experienced considerable bicycle and pedestrian traffic. Daily traffic on 14,633 to 23,710 vehicles. All six intersections were modeled along with changing the three stop control intersections to traffic circles. Using a NETSIM, a US intersection traffic model, the circles out performed the stop control in all categories, including safety, pollutant emissions (HC, CO and NO_x) fuel consumption, and delay. Historic vehicle and bicycle crashes and injuries were substantially less at the circle intersections.

Vermont Keene Turn Roundabout

The first multi-lane roundabout and one at an interstate interchange in the northeastern US, "Keene Turn," opened in Brattleboro, Vermont, October 1999. The Vermont Agency of Transportation (VAOT) (Vermont Agency of transportation, 2000) carried out stop delay

measurements at the signalized intersection in 1999 and at the roundabout in 2000. Peak hour traffic in both years was approximately 2,800 with total entering vehicles 28,000 daily in 2000. Three approaches were single-lane highways, and the I 91 Exit was two-lane. The signal did provide a dedicated left hand turning phase on all approaches.



Keene Turn Roundabout, Brattleboro, Vermont

Table 1 summarizes a.m. and peak hour data recorded. Note that during the lower a.m. traffic period, signal delay remained practically unchanged compared to the over 50% drop in delay for the roundabout. Taking an arbitrary overall delay figure midway between the a.m. and p.m. delay–26.5 seconds delay for the 28,000 average daily traffic translates to a reduction of 75,231 hours yearly of stop delay. This does not account for geometric delay or delay represented by deceleration and acceleration from cruising speed, which, together, would also tend to increase the advantage of the roundabout over a signal.

To translate stop delay to fuel usage, the parameter from the US SIDRA version (SIDRA Input Guide, 1988, p 123) of 0.41 gallon per hour applied to the 75,231 hours of stop delay translates to an annual motor fuel consumption decrease of 30,845 gallons.

Table 1
SIGNAL DELAY 1999 AND KEENE TURN ROUNDABOUT DELAY 2000 AT INTERSECTION OF US 5, I 91, AND VT 9

Peak Hour	1999 Traffic Signal		2000 Roundabout		Change
	Delay (sec)	Vehicles	Delay (sec)	Vehicles	(sec)
A.M.	44	1,216	12	1,870	-33
P.M.	46	2,764	26	2,812	-20

NOTES: 1. Video was used to measure 1999 delay. At times queues were beyond video range and therefore traffic signal delays are understated.
2. On-site methods were utilized to measure roundabout stop delay.

Maine

Maine's first roundabout replaced a stop-controlled intersection. An assessment that included field measurements of this four-leg intersection with about 13,000 daily entering vehicles, concludes there is a reduction in delay of about 5,000 to 10,000 hours annually and a reduction of emissions compared to a signal alternative (Garder, 1998, pp 7 and 43). Garder stated traffic "slowing to 15 mph and keeping that speed through a junction and then accelerating back up to a 'normal' travel speed will cause less pollution than a vehicle that is repeatedly forced into stop-and-go condition (Garder 1998, p 45)." The estimated stops per day were 3,500 with the roundabout in place compared to the 12,000 with the prior stop control.

Clearwater, Florida: Federal Transportation Air Quality Programs and Roundabouts

In 1999 the City of Clearwater, Florida, submitted one of the first proposals for FHWA funds under the Congestion Management and Air Quality (CMAQ) program administered by the State of Florida (City of Clearwater, 1999). CMAQ funds must be used in states with non-attainment air sheds for projects that improve air quality. The Clearwater "Gateway Roundabout" application (City of Clearwater, 2000) proposed a single, two-lane roundabout to replace three signalized intersections and one stop controlled intersection. EPA's Mobile (Mobile5) limitations only allowed roundabout benefits over signals for stop delay when comparing Gateway Roundabout to other CMAQ competing proposals. The emissions reduction based on stop delay were 68%, less for the roundabout compared to the signals. Using the SIDRA default figure of 0.41 gallons of motor fuel per hour of "stop delay" (idling), the annual motor fuel reduction totals for the 125,779 hours of delay was 51,569 gallons.

Since idling uses less fuel than the acceleration and deceleration phases–a factor SIDRA does model–the SIDRA results were also reported by Clearwater. SIDRA assessment predicted emissions for the stop delay period at about a fifth of the overall fuel and emission reductions that would be attained with the new roundabout. Annualized fuel savings and emissions reductions predicted by SIDRA were (City of Clearwater, p 6):

579,255	gallons of motor fuel
17.4	tons of HC
438.5	tons of CO
19.3	tons of NO_x
12,409	tons of CO_2

The Gateway Roundabout--without CMAQ funding--was built in late 1999 and significantly improved traffic operations from previous conditions.

Fort Collins, Colorado and CMAQ

Fort Collins, Colorado received a CMAQ grant and construction is scheduled for summer 2001 of a multi-lane roundabout serving as the junction of a four-lane and six-lane highway. Initially the roundabout

is to be two-lane. The single roundabout option provides the same fuel consumption and air quality benefits at grade and at half the cost compared to a grade separated facility that also employs a single roundabout for interchange. The roundabout at grade cost estimates range from $2.8-$3.1 million and the roundabout grade separated $5.4-$5.7 million (Moe, Bracke and Crown, p 26). Both options attain the same level of air quality and stop delay. In terms of air quality and delay, the emissions and delay of the no-build were so extreme on the four-lane highway as to be un-measurable. The roundabout options each resulted in a reduction of 225 hours of delay daily and 11 kilograms of CO daily compared to only the six-lane road generation of 880 hours delay and CO of 44 kilograms (Moe, Bracke and Crown, p. 27).

Montpelier US 2/US 302 Intersection

A three-leg Montpelier, Vermont, intersection study (City of Montpelier, 2000) provides a clear example of SIDRA fuel consumption in an intersection with projected traffic near the limit of the capacity of a one-lane roundabout. This study reflects a peak hour fuel saving of 92.9 gallons daily compared to a signal alternative. This can be extrapolated by a factor of ten to determine daily differential of 929 gallons, and annualized consumption reduction of 339,085 gallons reduction compared to the signal alternative. Substantial reductions were also predicted for other air pollutant emissions and CO_2.

Area and State Fuel Consumption and GHGs Potential Decrease from Roundabout Installations

The potential impact of transforming busy signalized intersections to roundabouts can be easily calculated from applying the Montpelier annual fuel savings, 330,000 gallons, to a larger set of intersections. Using this number from a one-lane intersection appears to be a fairly conservative yardstick to apply to likely group of single and multi-lane installations. For example, applying modeled reductions of the magnitude of the Montpelier US 2/302 intersection to a retrofit of 100 busy signalized intersections to roundabouts control statewide in Vermont results in a reduction of approximately 33 million gallons

annually. A motor fuel reduction of this magnitude is the equivalent of approximately 8% of the 1997 total Vermont motor fuel consumption of 401 million gallons (FHWA, 1998, p I-3). Air pollution emissions and particularly GHGs in the form of CO_2 are tied directly to fuel consumption.

Roundabouts also provide short term and long term savings that may approach or exceed immediate reductions at a specific intersection:
1. Short term-- higher performance of existing circulation in builtup areas as well as sharply improved conditions for walking and bicycling (Redington, 1999).
2. Long term--increased development densities as the result of increased access and service–for all modes but particularly for those walking and bicycling (Redington, 1999).

The emissions, GHGs, and fuel consumption benefits of these other short and long term benefits are not included in the potential benefits evaluated here in this paper. Certainly, these additional benefits need to be researched to determine their magnitude.

Global Warming--Possible Contribution of Roundabouts
Global warming strategic and implementation plans have been developed in both the US and Canada in response to the establishment of the Kyoto GHGs reduction goals for developed nations. The Canadian national plan for transportation (Transportation Climate Change Table, 1999) was generated through a collaborative effort of the Canadian and provincial transport ministries.

The City of Burlington through local initiative completed a planning process with a goal for GHGs reductions reflecting the Kyoto target. The difficulties of formulating plans and starting implementation at the national level are manifold for both the US and Canada (Butt, 2000). State, provincial and metropolitan plans in North America do not yet include roundabouts along with associated pedestrian facilities as a strategic element in transportation plans--the current Toronto strategic plan is an example (Greater Toronto Services Board, 2000). The roundabout does present another substantial element of

automobile "attrition" as the term is used by Toronto urbanologist, Jane Jacobs (Jacobs, 1961, pp. 348-349, p. 359).

Burlington "set a target for 2005 of reducing GHGs emissions in Burlington 10 percent below 1990 levels (City of Burlington, 2000, p 5). The first steps in creating the structure and initiating implementation and monitoring are now underway. The 1990 Burlington population of 37,712 (largest in Vermont) is the largest component of the Burlington metropolitan statistical area about 150,000. The area population is growing at a moderate pace, about 10-15% per decade.

At the metropolitan level, the roundabout savings at busy intersections represents a tool to meet the Burlington goal of reducing 2005 GHGs generation by 256,000 tons. Transforming 25 signalized intersections with substantial traffic volumes would, based on the Montpelier case modeled by SIDRA, amount to about 61,000 tons of CO_2. This estimates is based on assuming 250,000 annual reduction in fuel use per intersection average and 19.57 pounds of CO_2 per gallon of motor fuel. The immediate GHG reduction from the replacement of 25 traffic signals at busy intersections to roundabouts would represent about 24% GHGs reduction necessary for Burlington to return to 10% below the 1990 generation levels by 2005.

References

Akcelik and Associates (1999) aaSIDRA software (Victoria, Australia). Burlington (2000)

Climate action plan–a plan to save energy and reduce greenhouse gas emissions. (City of Burlington, Vermont).

Butt M (2000) "Getting back to basics": transportation, the environment and some recent reflections on corridor initiative in light of the monumental "November options paper." *Proceedings 35th Canadian Transportation Research Forum Conference, pp. 470-486* (CTRF, Saskatoon, Saskatchewan).

Federal Highway Administration (1998) Highway statistics 1997. p I-3 (US Government Printing Office, Washington, DC).

Garder P (1998) Little Falls, Gorham a modern roundabout (University of Maine Orono, Orono, Maine).

Greater Toronto Services Board (2000) Removing roadblocks to continued economic prosperity for the Greater Toronto Area, Ontario and Canada--a strategic transportation plan. (Greater Toronto Services Board, Toronto).

Hagring O (1997) CAPCAL 2–a new version of the SNRA capacity, delay, and VOC software. pp. 115-123. *Proc 3rd International Symposium on Intersections without Traffic Signals* (University of Idaho, Moscow, Idaho).

Hogland G (1993) Alternative intersection design–a possible way of reducing air pollutant emissions from road and street traffic. (Swedish Transport Research Board, Stockholm).

Interlaboratory Working Group (2000) Scenarios for a clean energy future. (Oak Ridge, Tennessee) http://www.ornl.gov/ORNL/Energy_Eff/CEF.htm.

Jacobs J (1961) *Life and Death of Great American Cities. (Random House, Toronto).*

Mark J and Morey C (2000) *Rolling Smokestacks.* (UCS Publications: Cambridge, Massachusetts).

Moe R, Bracke E, and Crown B (2000) Mulberry Avenue (SH 14) and Lemay Avenue intersection alternatves analysis (City of Fort Collins, Colorado).

Mustafa A and Vougias S (1993) Analysis of pollutant emissions and concentrations at urban intersections. *ITE Compendium of Technical Papers, pp 234-238.* (Institute for Transportation Engineers, Washington, DC).

Office of the President (1998) The Kyoto protocol and the President's policies to address climate change: administration economic analysis. (US Office of the President: Washington, DC).

Rhouphail N, Frey H C, Colyar J, and Unal (2001) A Vehicle emissions and traffic measures: exploratory analysis of field observations at signalized arterials. *80th Annual Meeting of the Transportation Research Board* (Washington, DC).

Savage F and Al-Sahili K Traffic circles–a viable form of intersection control? ITE Journal, September 1994 pp. 40-45.

Savonis M (2000) Strategic plan for transportation and air quality research 2000-2010. Transportation in the New Millennium: State of the Art and Future Directions (Transportation Research Board, Washington, DC).

Sides K (1999) Gateway roundabout application for congestion mitigation and air quality funding (City of Clearwater, Florida).
SIDRA 5 (1998) AARB Transport Research (Victoria, Australia)
Transportation Climate Change Table (1999) Transportation and climate change: options for action. (Canada Ministry of Transport, Ottawa).
Vermont Department of Public Service (1998) Fueling Vermont's future. Vol 1 (Montpelier, VT)
Vermont Agency of Transportation (2000) Unpublished data on 1999 and 2000 Brattleboro (US 5/T 9/I 91) intersection delay at a.m. and p.m. peaks (VAOT, Montpelier, VT).
US Department of Energy, (2000) Scenarios for a clean energy future (US Department of Energy: Washington, DC).
US General Accounting Office (1997) Air pollution–limitations of EPA's emissions model and plans to address them (US General Accounting Office: Washington, DC).

A Time Series Analysis of Vessel Traffic Service Utilization

Louis A. Le Blanc
Berry College

Nicholas A. Beadles, II
Georgia College and State University

Introduction

This paper describes a time series of more than 1,700 daily records of system utilization for the New Orleans (Louisiana) Vessel Traffic Service (NOLA-VTS). Regression time series models were constructed to gauge the impact of maritime safety technology, system closure, speed of river current and traffic level on VTS utilization. Based on five years of case history, this study represents one of the lengthiest analyses of information technology utilization. These results may have significant implications for marine policy makers.

Navigating the lower Mississippi River involves decisions that are relatively unstructured and usually relate to the passing and overtaking of vessels. A navigation decision may be unstructured because of its novelty, time constraints, or lack of knowledge.

The decision support offered by a VTS is the central collection and broadcast of traffic and related information. This reduces the uncertainty about other vessels' positions as well as their intentions. Advanced warning of critical or hazardous areas is also included in this central distribution of information by the VTS.

From this input of more accurate and complete data, vessel pilots are better able to make decisions about when and where to pass or overtake other shipping. Decisions on how to pass (face-to-face meeting of vessels) or overtake (one vessel passes another going in the same direction) are usually quickly reached but gradually implemented due to the slowness of large ships and barge tows to react to decisions to change course and/or speed.

These decisions have very permanent effects with severe consequences if they are not made in the best manner. Poor decisions cannot be reversed quickly because of the nature of vessel steerage and propulsions and may results in a serious accident. Such accidents can result in significant dollar losses as well as the loss of life.

By definition, a decision support system (DSS) is designed to enhance the effectiveness of decision making. If utilization of the DSS can be associated with the desired changes in the criterion variable, such as the frequency of accidents, then decision making has likely been improved through increased use of the DSS. Moreover, if utilization rates are affected by conditions suggesting increased use of the DSS (i.e., hazardous weather conditions, fast currents or heavy river traffic), this would also support usage or utilization rate as an appropriate DSS success factor.

Research Methodology

Regression analysis was selected as the most appropriate technique for constructing models of the factors that contribute to system utilization. The methodological approach taken was to incorporate variables in the regression equation(s) that would account for all possible sources of variation in the criterion variable (i.e., utilization) and identify those factors that contribute at a statistically significant level.

Model Variables

Four explanatory variables were identified in this study of long-term system utilization. These continuous variables, collected on a daily basis by the U.S. Coast Guard, include system utilization, river stage, and traffic level. A categorical variable, system closure, was also employed in the final model. Table 1a and Table 1b list the various descriptive statistics about the criterion and explanatory variables used in the regression analysis before and after the VTS was temporarily not in service.

Table 1a. Descriptive Statistics for Model Variables Pre-Closure (4 April 1979 – 16 March 1982)

	Utilization	Stage	Traffic
Count	1039	1039	1039
Mean	0.74	6.17	93.0
Standard Deviation	0.08	4.07	12.5
Minimum	0.26	0.24	52.0
Maximum	0.95	17.22	140.0

Table 1b. Descriptive Statistics for Model Variables Post-Closure (1 September 1982 – 18 July 1984)

	Utilization	Stage	Traffic
Count	678	678	678
Mean	0.56	8.99	56.0
Standard Deviation	0.11	5.02	15.5
Minimum	0.23	2.08	26.0
Maximum	0.94	19.71	120.0

System Utilization – The system usage rate is expressed as a percentage. This variable represents the level of participation in the New Orleans VTS at the hour of peak daily traffic. System participation, or utilization, is the percentage of total traffic that is communicating with the Vessel Traffic Center.

River Stage – The stage of the river (height above sea level) is a very critical factor for vessel safety, because river stage directly determines the velocity of the current. The hazardous conditions that often accompany the changes in the river stage (hence, river current) precipitate many vessel casualties on the lower Mississippi River.

Traffic Level – According to maritime authorities, traffic density is another major factor influencing the occurrence of accidents in the study area. Daily measures of traffic volumes include ocean-going ships, barge tow assemblies, tugs, as well as excursion craft.

System Closure – The NOLA-VTS was not operative for six months from mid-March 1982 through August 1982 due to federal budgetary constraints. A single indicator variable represents system operation before and after the shutdown. This variable measures the residual effect of the six-month absence of the NOLA-VTS on system utilization.

Analysis and Results

Regression models were constructed and evaluated to determine the best equation for NOLA-VTS utilization. Explanatory variables incorporated in the final model included measures of river traffic, system utilization lagged by one time period, a quadratic form of river stage, and a dichotomous variable for system closure.

The best equation for utilization explained over 55 percent of the variation in the daily usage rate of the vessel tracking system. All variables incorporated in the final model exhibited very high levels of statistical significance (see Table 2).

For time series regression, the Durbin-Watson statistic is highly relevant, as positive correlation of residuals that are adjacent in time is frequently a problem. This statistic is a formal test for serial correlation. This equation did not exhibit serial correlation.

Table 2. Regression Model of Utilization

Explanatory Variables	Coefficient	Standard Error	T Statistic (1708)	Probability Level
Intercept	0.534881	0.02227	24.0186	0
Traffic	-0.000500	0.00015	-3.2673	0.001107
Stage	0.006977	0.00203	3.4341	0.000609
$Stage^2$	-0.000230	0.00011	-2.0580	0.039756
Closure	-0.153670	0.00867	-17.7197	0
Utilization (-1)	0.298454	0.02282	13.0769	2.59E-37
R-Squared	.5536		F (5,1708)	423.54
Durbin-Watson Statistic	2.11		Probability Level	0.0000
Serial Correlation	-0.05596		Standard Error	.08473

Traffic – According to the regression, river traffic had a statistically significant effect on utilization. For every unit increase in traffic, the model estimated about a one-percent drop in utilization. A priori, a different sign or relationship (i.e., positive) would have been expected. The actual sign was negative, suggesting that utilization drop if traffic level increases.

River Stage – This equation includes river stage variables with a quadratic effect. Differentiation of the river stage portion of the model, which was statistically significant (probability level = 0.0398), indicates that utilization peaks when river stage reaches 14 feet.

System Closure – Perhaps the most significant result comes from an indicator variable that modeled the effects upon system utilization resulting from the closing and then re-opening of the VTS. Significant at the zero percent level, this variable estimated that system usage was more than 15 percent lower after closure because of federal budget reduction.

Lagged Utilization Rate – Another variable incorporated in this utilization model was the previous day's utilization rate. A lagged endogenous variable represents the existing momentum whereby the level of the criterion variable is expected to amount to an approximately fixed percentage lower (or higher) than the previous time period (Ostrom, 1988). A one-unit increase in the prior day's usage rate was associated with nearly a three-tenth percent (0.298) increase in the current period's level of participation by users. This lagged utilization variable was statistically significant (probability = 0).

Summary and Conclusions

The most interesting results of this study come from the statistical significance of the system closure variable for the critical factor of system utilization. With almost no probability of being a random effect, the cessation of the NOLA-VTS for six months had a residual effect of decreasing user participation by an estimated 15 percent. According to prior research (Le Blanc and Kozar, 1990), positive changes in VTS utilization usually result in a drop in vessel accidents. These findings suggest that not only are utilization levels of these systems important, but also their availability is crucial for vessels wishing to avail themselves of the service.

References

Cook, Thomas D. and Donald T. Campbell, *Quasi-Experimentation: Design & Analysis Issues for Field Settings*, Houghton Mifflin Company, Boston, MA, 1989.

Le Blanc, Louis A. and Kenneth A. Kozar, "An Empirical Investigation of the Relationship Between DSS Usage and System Performance: A Case Study of a Navigation Support System, *MIS Quarterly*, Volume 14, Number 3, 1990, pp. 263-277.

Neter, John, William Wasserman and Michael H. Kutner, *Applied Linear Regression Models*, Richard D. Irwin, Homewood, IL, 1993.

Ostrom, C. W., *Time Series Analysis: Regression Techniques*, Sage University Paper Series on Quantitative Applications in the Social Sciences, Series Number 07-009, Sage Publications, Beverly Hills, CA, 1988.

Economic Modelling of Congestion in a Transportation Network

Dr. Yuri Yevdokimov, University of New Brunswick

Introduction

Traffic congestion is a negative externality that arises due to consumer and producer activities. The economics literature on congestion focuses on externalities due to consumer activities. For instance, Rothenberg (1970) defines pure congestion as follows: *"All users are using the medium (public good) in much the same way, each is damaging service quality for both others and himself, and the ratio of self:other damage is approximately the same for all users"*. Hence, in such a framework congestion is regarded as a negative consumption

externality. Applied to transportation, such an externality results in increasing waiting time which consequently increases opportunity costs of a trip.

Passenger and freight transportation services are produced within transportation networks. As it is shown in this study, congestion arises due to a capacity constraint in the existing network - a supply side phenomenon - that eventually leads to decreasing returns to scale in the production of transportation services, or the so-called diseconomies of scale. This implies that congestion can be regarded as a negative production externality. In this study, a negative externality is directly incorporated into the production function of the transportation network, in order to derive the congestion costs.

Congestion and capacity utilization

A speed-flow diagram is an analytical starting point that is familiar to engineers . It is a forward bending line depicted in Figure 1. Speed is measured in kilometres per hour and flow is measured in vehicles per hour; this is known as traffic. In Figure 1, at a flow T_c any further

increase in the density of transportation leads to a decrease in speed and flow.

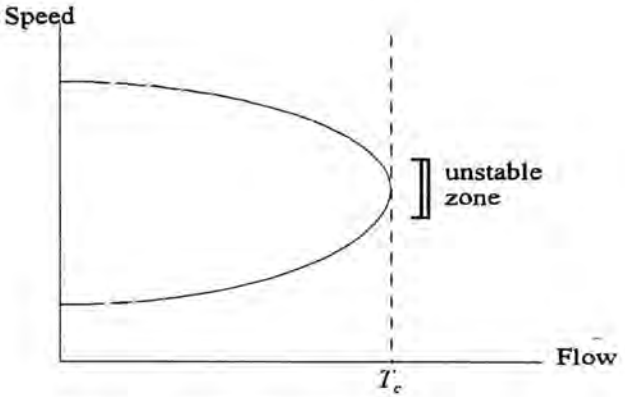


Figure 1. Speed-Flow Diagram

The last statement can be expressed in terms of the flow-density diagram presented in Figure 2, density is measured in vehicles per kilometre. In this diagram, T_c is the critical volume of transportation, or engineering capacity, of the transportation fixed facility (for example, highway). This framework has been extensively discussed in

transportation literature with regard to issues of congestion (see, for example, Neutze, 1963 and Button, 1993).

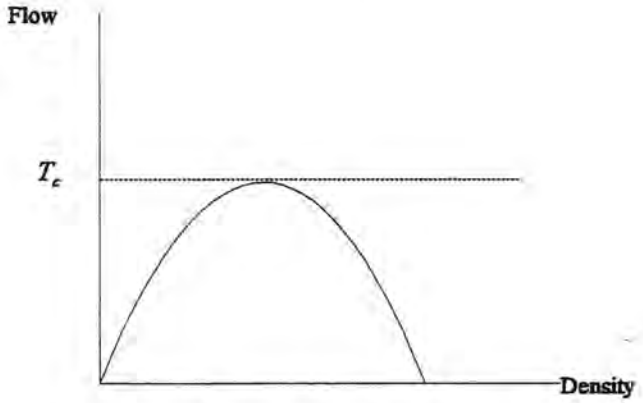


Figure 2. Flow-Density Diagram

Let us begin by redefining some variables in Figure 2. In terms of transportation network, it is more convenient to express traffic flow as the volume of transportation measured in ton-kilometres or passenger-kilometres per period of time T_t. This is a commonly accepted index of measuring the quantity of transportation (see Boyer, 1998).

The density function has been a critical element in all previous studies on congestion (see Else, 1981 and Evans, 1992). In this study, the density of transportation is defined as the number of vehicles in the network at a given point in time. Density increases as the absolute number of vehicles in the network, K_{Vt} increases, implying new vehicles. On the other hand, density in the network also increases as frequency of vehicle utilization increases, implying use of the same vehicles more frequently. Let f_t represent the frequency of vehicle utilization; f_t is initially equal to one.

Let L_{Tt} represent the quantity of labour in the network; this is equal to the number of vehicle operators multiplied by the frequency of vehicle utilization. Given all of the above, the number of vehicles in the network at any given point in time, which is the network's density d_t, can be represented by the following equation:

$$d_t = (K_{Vt} f_t)^{\alpha} L_{Tt}^{1-\alpha} \qquad (1)$$

where α is a parameter that is between 0 and 1. For instance, if there are initially 20 vehicles and 20 drivers, and frequency of vehicle utilization

is $f_0 = 1$, then equation (1) produces $20^{\alpha} \times 20^{1-\alpha} = 20$ vehicles/network. On the other hand, if the frequency of vehicle utilization in the network increases by 2 times, then the quantity of labour becomes 40 units, while number of vehicles remains at 20, and density is equal to $(20 \times 2)^{\alpha} \times 40^{1-\alpha} = 40$ vehicles/network.

It is possible to interpret the flow-density diagram in Figure 2 as the volume of transportation-density diagram, where the volume of transportation is measured in ton-kilometres (passenger-kilometres) at time t, and density is measured as the number of vehicles in the network at time t.

Let us introduce capacity utilization of the network, u_t as follows:

$$u_t = \frac{T_t}{T_c} \tag{2}$$

T_t is the volume of transportation at time t, and T_c is the critical volume of transportation. In fact, the critical volume depends on quantity of transportation fixed capital K_{Ft}. For example, a four-lane highway can handle more traffic than a two-lane highway. Let us define the critical

volume of transportation as follows:

$$T_c = bK_{Ft} \tag{3}$$

b is a scale coefficient showing the potential volume of transportation, produced by a unit of fixed capital. Taking (3) into account, equation (2) becomes:

$$u_t = \frac{T_t}{bK_{Ft}} \tag{4}$$

It is now possible to re-scale the vertical axis in Figure 2 in order to end up with the capacity utilization-density diagram. However, there is the so-called unstable zone where traffic approaches engineering capacity (see Figure 1). In order to reflect this fact, let us assume that if density d_t increases, capacity utilization u_t asymptotically approaches one. The following relationship captures properties of the network discussed above:

$$u_t = 1 - \frac{a}{d_t} \tag{5}$$

where a is some parameter reflecting the initial density in the network associated with zero capacity utilization (if $d_t = a$, then $u_t = 0$). Graphically, relationship (5) is the increasing portion of the flow-density diagram presented in Figure 2.

Equation (5) shows that $\lim_{d_t \to \infty}(1-\frac{a}{d_t}) = 1$ and

$\frac{du_t}{dd_t} = \frac{a}{d_t^2} > 0$ implying, that if density in the network increases, then the capacity utilization of the network increases and approaches one. The relationship (5) is depicted in Figure 3.

Production function of the transportation network

Let us multiply both sides of the relationship (5) by bK_{Ft}:

$$u_t bK_{Ft} = (1-\frac{a}{d_t})bK_{Ft} \qquad (6)$$

According to the definition of capacity utilization (4), the left hand side

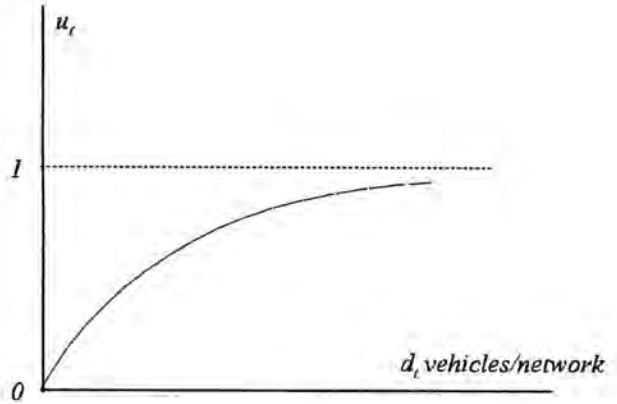


Figure 3. Capacity Utilization-Density Diagram

of the expression (6) is the current volume of transportation in the network. Taking definition of density (1) into account, the final version of (6) becomes:

$$T_t = [1 - \frac{a}{(K_{Vt} f_t)^{\alpha} L_{Tt}^{1-\alpha}}] b K_{Ft} \qquad (7)$$

Expression (7) is the production function of the transportation network. It shows the relationship between volume of transportation as the

network's output and the network's inputs - transportation fixed facilities, vehicles and labour. Moreover, the production function (7) incorporates negative externality from congestion.

Without congestion, the flow-density has a one-to-one relationship. Taking into account that capacity utilization, u, represents re-scaling of the transportation flow variable, it is obvious that

$$u_t = cd_t \tag{8}$$

where c is the re-scaling coefficient. If both sides of (8) are multiplied by bK_{Ft}, and then definitions of capacity utilization (4) and density (1) are used, expression (8) becomes

$$T_t = AK_{Ft}(K_{Vt}f_t)^{\alpha} L_{Tt}^{1-\alpha} \tag{9}$$

where $A = cb$ or it is a technological parameter.

Looking at expressions (7) and (9), one can conclude that:

- without congestion, the production function (9) exhibits increasing returns to scale;
- with congestion, the production function (7) exhibits decreasing

returns to scale;

- transportation fixed capital K_{Ft} is a complement to transportation vehicles K_{Vt} and transportation labour L_{Tt} in the production of transportation services.

In this framework, congestion is a negative externality that causes decreasing returns to scale.

In order to derive congestion costs, it is possible to generate two cost functions, one without congestion and one with congestion. The difference between the two values of costs as a function of transportation fixed capital K_{Ft} and volume of transportation T_t will give us the congestion cost function. Below a simulation procedure is presented in order to derive this function.

Simulation of congestion in transportation network

In order to simulate the consequences of a traffic increase in transportation network, the following initial conditions and parameter values were chosen: (i) initial conditions, $K_{F0} = 20$, $K_{V0} = 10$ and $L_{T0} = 20$; (ii) parameter values, $\alpha = 0.5$, $b = 2$ and $a = 7.15$. In addition, in

this simulation exercise the frequency of vehicle utilization f_t was fixed at a value of one.

Traffic in each period has been increased through a five-percent increase in the density of the network. According to the derived production function (7) and the definition of the network's density (1), such an increase implies a simultaneous five-percent increase in both, transportation vehicles K_{Vt} and labour hours L_{Tt}. Recognizing the unstable zone when traffic approaches the networks capacity, it was assumed that, if traffic reaches ninety five percent of the engineering capacity, a new increment of fixed capital, $\Delta K_{Ft} = 20$, is added.

Since capacity of the network is a function of the amount of fixed capital K_{Ft}, then if an increment of the capital is added, the capacity of the network increases, and consequently congestion is reduced.

Figure 4 presents the simulated time path of traffic. The dotted line depicts increasing traffic without congestion; the solid line reflects traffic increase with congestion.

The total production costs of transportation in the network are

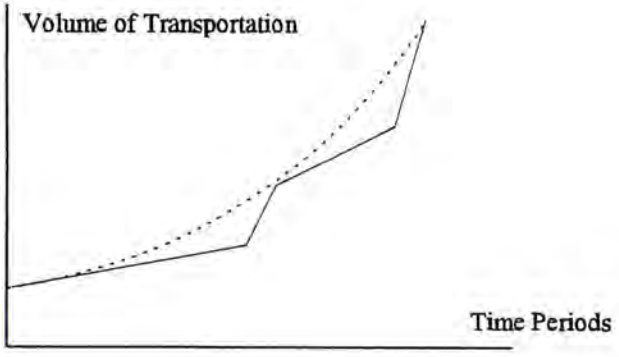


Figure 4. Volume of Transportation in a Network

defined as follows:

$$TC_t = P_F K_{Ft} + P_V K_{Vt} + P_L L_{Lt} \tag{10}$$

where P_F, P_V and P_L are the prices of fixed capital, vehicles and labour respectively. In this simulation exercise the prices are fixed at a value of one.

Figure 5 depicts the average total cost as a function of the volume of transportation in two situations: (i) with congestion (solid line), and (ii) without congestion (dotted line).

The shape of the solid line in Figure 5 reflects three important

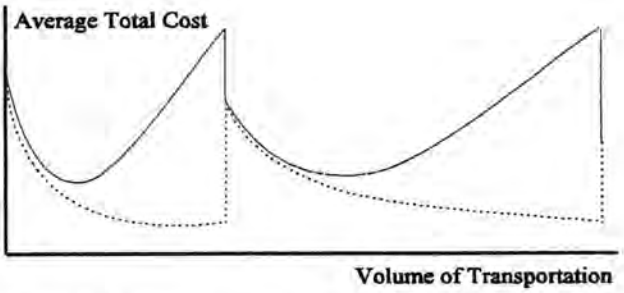


Figure 5. Average Total Cost of Transportation in Network

features: (i) it is U-shaped locally due to congestion; (ii) it is scalloped due to the lumpiness of transportation fixed capital; (iii) it is declining globally due to economies of scale.

Congestion costs were calculated as the difference between the two paths, presented in Figure 5, times the volume of transportation. Figure 5 shows that the difference between the two paths depends on volume of transportation T_t and the transportation fixed capital K_{Ft}. Hence, the following linear regression model was estimated by Ordinary Least Squares:

$$C_t = 2.44 - 0.12K_{Ft} + 0.29T_t \qquad (11)$$

where C_t is the congestion cost.

Equation (11) shows that congestion costs increase with traffic and

decrease in response to an increase in transportation fixed capital. Two elasticities were calculated according to the simulated data and equation (11): (i) the congestion cost-transportation fixed capital elasticity = -0.2; (ii) the congestion cost-traffic elasticity = 0.7.

These two elasticities imply that, in order to avoid congestion costs, it is necessary to increase transportation fixed capital in the following way: a thirty five percent increase for each ten-percent increase in traffic in the network. This is an investment into improving the transportation infrastructure.

Conclusion

Congestion that occurs in the transportation network is a negative production externality because it causes decreasing returns to scale. It was shown that congestion can be directly incorporated into a production function of the transportation network. The production function, as a function of transportation capital and labour, was derived and its properties discussed.

The proposed approach was illustrated using computer simulation. Nonetheless, it provides a methodology for estimating congestion costs. The congestion cost function, as a function of the network's traffic and infrastructure, can be directly obtained from the production function. Therefore, estimation of the production function becomes a crucial

element of the proposed approach.

The derived congestion cost function can be used as an economic tool for the justification of investment into public infrastructure. The congestion cost function provides useful economic indicators in the form of elasticities with respect to traffic and infrastructure. Furthermore, the production function and underlying cost function can be used in various microeconomic and macroeconomic applications, from cost-benefit analysis to regional economic development.

References

Button, K.J. *Transport economics. Second Edition.* Edgar Elgar Publishing Ltd., 1993

Boyer, K.D. *Principles of Transportation economics.* Addison Wesley Longman Inc., 1998

Else, P.K. A reformation of the theory of optimal taxation. *Journal of Transport Economics and Policy,* 15, 1981, 217-232

Evans, A.W. Road congestion: The diagrammatic analysis. *Journal of Political Economy,* 100, 1992, 211-217

Neutze, G.M. The external diseconomies of growth in traffic. *Economic Record,* 39, 1963, 332-345

Rothenberg, J. The economics of congestion and pollution: an integrated view. *American Economic Review, Papers and Proceedings,* 60, 1970, 114-121

The Foundation of a New Transportation Planning Model for the Manitoba Capital Region

Kean Hoe Lew, E.I.T., Traffic Engineer ISL Consultants, Edmonton
Alan Clayton, P.Eng., Professor, University of Manitoba
Doug Hurl, P.Eng., Transportation Systems Planning Engineer, City of Winnipeg

Executive Summary

This paper develops the foundation of a model to predict P.M. peak vehicle trips for the Manitoba Capital Region. This model, called the Manitoba Capital Region Transportation Planning Model (MCRTPM), is built on a Geographical Information System for Transportation (GIS-T) platform. Components of the model were calibrated using the 1996 Statistics Canada Journey-to-Work database and the matrix adjustment algorithm with newly developed traffic data from the City of Winnipeg and Manitoba Department of Highways and Government Services (MHGS). The foundation developed in this paper allows further refinement to be made to the MCRTPM in order to develop a complete model. This model is designed to facilitate transportation planning in the Manitoba Capital Region.

Introduction

This paper develops the foundation of a model to predict P.M. peak vehicle trips for the Manitoba Capital Region. This model, called the Manitoba Capital Region Transportation Planning Model (MCRTPM), is built on a Geographical Information System for Transportation (GIS-T) platform. Components of the model were calibrated using the 1996 Statistics Canada Journey-to-Work database and the matrix adjustment algorithm with newly developed traffic data from the City of Winnipeg and Manitoba Department of Highways and Government Services (MHGS).

Background and Need

The City of Winnipeg and the surrounding region are termed the "Manitoba Capital Region" (Figure 1). The 1996 Census of Canada indicates that the Capital Region has a population of approximately 700,000 people with 625,000 of them residing in Winnipeg (Capital Region Review, 1999). Like most North American metropolitan areas, the City of Winnipeg is experiencing urban sprawl characterized by the rise of suburban communities surrounding the older areas. Growth in the ex-urban communities in the Capital Region, and large activity shifts in the City of Winnipeg itself, have translated into significant changes in the traffic patterns.

To this day, transportation agencies in the region, like the City of Winnipeg, Manitoba Department of Highways and Government Services (MHGS), Manitoba rural municipalities, Winnipeg Transit, and the Winnipeg Airport Authority perform their own respective transportation planning. The lack of a regional approach to transportation planning has led to a situation in which there is insufficient understanding of the regional traffic characteristics in the region. The Capital Region Review panel highlighted this lack of understanding when they concluded that there is "inadequate information and research for assessing commuting patterns within the region" (1999). In order to increase the effectiveness of transportation planning, there is a need to develop better means to understand, structure, analyze and predict transportation demand (characterized by traffic flows) in the Manitoba Capital Region. This

can be helped through the construction of a model that links the transportation and land-use systems of the region in terms of traffic flows. This linkage provides a capability to better assess and test a variety of transportation planning alternatives. This capability in turn provides decision makers with quantitative information about the consequences of their decisions, thus allowing objective transportation systems analysis and planning in the region (Mekky, 2000).

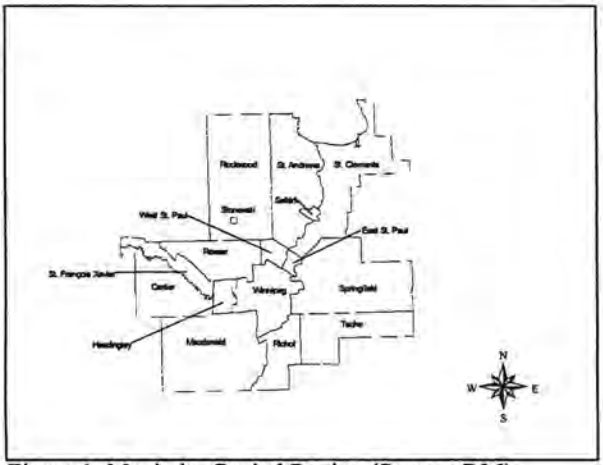


Figure 1: Manitoba Capital Region (Source: DLI)

In the United States, metropolitan areas have their respective Metropolitan Planning Organizations (MPOs). The MPOs are formed for cooperative transportation decision making for the metropolitan planning area (FHWA, 1994). MPOs are required by the Transportation Equity Act for the 21st Century (TEA-21) to develop transportation plans (FHWA, 2000). In Canada, transportation agencies (municipal, provincial, or regional) are not required to produce transportation plans by law. However, many agencies develop, maintain and improve transportation planning models in order to facilitate the development of transportation plans.

In November 1999, Statistics Canada made available the 1996 Journey-to-Work database for the Manitoba Capital Region. Furthermore, new traffic databases from the Manitoba Highways Traffic Information System (MHTIS) and the City of Winnipeg have also been made available. In addition, Geographic Information Systems for Transportation (GIS-T) technology has evolved to a level where it is readily accessible and offers an efficient means for transportation modeling. The need for a regional approach to transportation planning, the availability of the Statistics Canada database and new traffic databases, and the evolution of GIS-T technology make the development of a transportation planning model for the Manitoba Capital Region feasible.

Model Structure

The model is based on the four-step sequential modeling method minus the mode split component. Since only aggregate data are available at this time, a decision was made not to develop a mode split component for the MCRTPM. The mode split is captured by using automobile driver trips and other vehicle trips (instead of person trips). Figure 2 shows the structure of the model.

The model assumes that all trips begin within Manitoba and end in Manitoba. Trips that begin or end outside of Manitoba are not considered.

The model predicts the number of P.M. peak vehicle trips. The trips are divided into two trip types: (1) work trips from place of work (POW) to place of residence (POR) (also referred to as home based work trips), (PMPADWT); and (2) other trips (PMPVOT). For this research, the P.M. peak period is defined as the time between 3:30 P.M. and 5:30 P.M. with the highest traffic per hour.

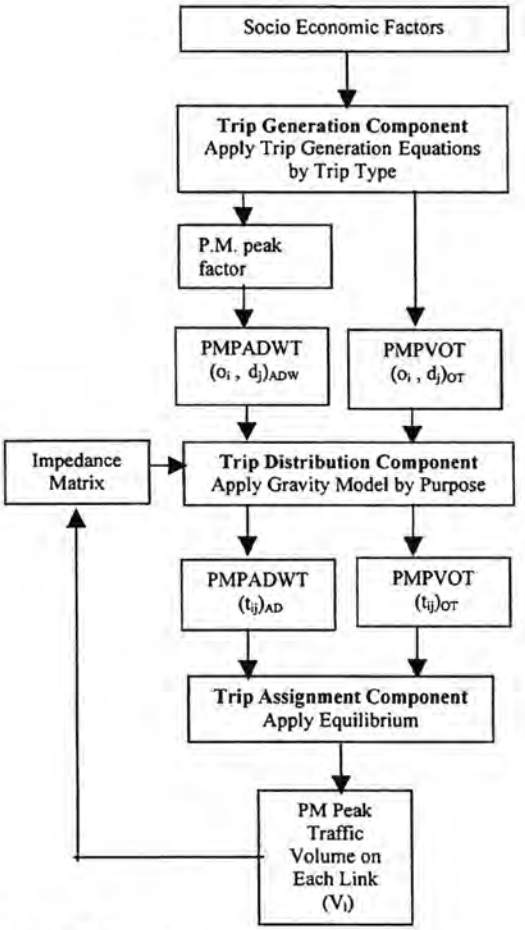


Figure 2: Model Structure

The Manitoba Capital Region Transportation Analysis Network

Prior to this research, there was no spatial layer of the road network in the Manitoba Capital Region, both MHGS and the City of Winnipeg maintained their own digital road network. This made it impossible to perform any GIS-based transportation planning analysis at a regional level. In December 1999, students from the University of Manitoba Transportation Information Group (UMTIG) worked together with the City of Winnipeg Public Works Department to create the Manitoba Capital Region Transportation Network (MCRTAN). Detailed documentation of the development of the network can be found in Han (2000). The network was constructed with the following considerations:

(a) the interchangeability of data from the City's MapInfo map which held different databases including the collision database
(b) compatibility with the MHGS provincial highway spatial layer
(c) compatibility with the City of Winnipeg regional zone system
(d) the topology of the network detailed network attribution

The network is shown in Figure 3.

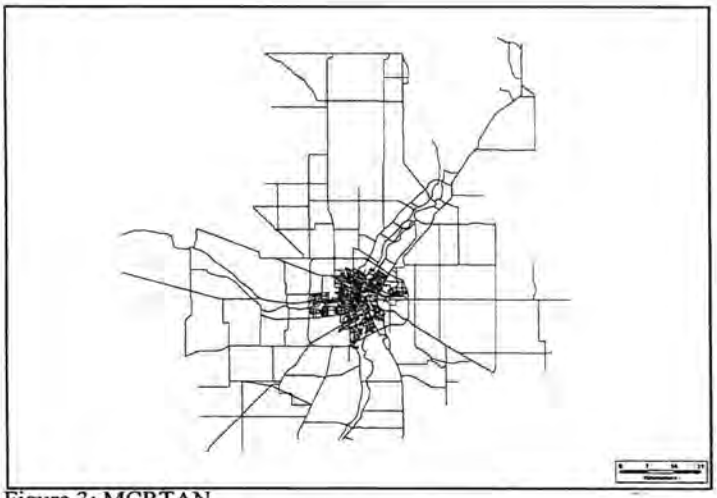


Figure 3: MCRTAN

Components of the MCRTPM
The components developed were:

(a) trip generation equations for Automobile Driver Work Trips (ADWT) using the ADWT matrix
(b) a P.M. peak automobile driver work trip (PMPADWT) matrix using the ADWT matrix and the National Cooperative Highway Research Program (NCHRP) diurnal factor.
(c) an estimated P.M. peak vehicle total trip (PMPVTT) matrix using the PMADWT matrix developed in (b), traffic data from the City of Winnipeg and MHGS, and TransCAD's matrix adjustment algorithm (which uses Nielsen's (1993) method).
(d) a P.M. peak vehicle other trip (PMPVOT) matrix using the PMPADWT developed in (b) and the PMPVTT matrix developed in (c).
(e) trip distribution equations for PMPADWT using Statistics Canada 1996 Journey-to-Work data, travel time estimates from

MCRTAN, and TransCAD's trip distribution component calibration function.

(f) trip distribution equations for PMPVOT using the PMPVOT matrix developed in (d), travel time estimates from MCRTAN, and TransCAD's trip distribution component calibration function.

(g) the trip assignment component using Bureau of Public Roads default link performance function parameters.

Conclusion

Based on the need for a regional transportation planning model, this research has developed (using the best available data) the foundation of the Manitoba Capital Region Transportation Planning Model: (1) the framework and structure of the model; (2) the Manitoba Capital Region Transportation Analysis Network (MCRTAN); (3) relevant P.M. peak traffic data; (4) the automobile driver work trip (ADWT) matrix from Statistics Canada 1996 Journey-to-Work database; and (5) components of the MCRTPM. This foundation allows further development to be made to the MCRTPM in order to develop a complete model.

Discussion and Further Work

ADWT Time Distribution

The PMPADWT is estimated from the ADWT (obtained from the 1996 Statistics Canada Journey-to-Work database) using the NCHRP diurnal factor. Sensitivity analysis has not been performed on this factor and as such the effect of this factor on the components developed in this research is unknown. The 2001 Statistics Canada Journey-to-Work Database will not contain time distribution information to help refine the diurnal factor for the Manitoba Capital Region (Hurtubise, 2000).

PMPVOT Matrix

In this research, the PMPVOT matrix was estimated using the matrix adjustment algorithm. The estimated link counts matched the

observed link counts well. However, satisfactory trip generation equations for the PMPVOT matrix could not be developed using independent variables from the National Cooperative Highway Research Program (number of households and employment by different industries) or the Winnipeg Transportation Model (employment by different industries). This means that there are two possibilities: (1) the PMPVOT matrix developed by the matrix adjustment algorithm is wrong and the trip generation characteristics of PMPVOT can be explained using the independent variables from NCHRP or the Winnipeg Transportation Model; or (2) the PMPVOT matrix developed by the matrix adjustment algorithm is correct and the trip generation characteristics cannot be explained by the independent variables examined. There are no data to validate the PMPVOT matrix developed in this research. At this time, there is very little understanding of the traffic generation and trip distribution characteristics of PMPVOT in the region. In order to develop better PMPVOT trip generation and trip distribution equations, there is a need for additional data in this area.

Trip Generation Component

This research developed trip production and attraction equations for ADWT. There are strong correlations between the dependent variables with the independent variables for the equations produced. This research also attempted to develop trip production and attraction equations for the PMPVOT. However no satisfactory equations were produced. The use of category analysis to develop trip generation equations for the PMPVOT should be investigated.

Trip Distribution Component

This research developed trip distribution equations for PMPADWT and PMPVOT. However, since the PMPVOT matrix developed in this research has not been validated, the validity of the PMPVOT trip distribution equation (which was developed using the PMPVOT matrix) is also uncertain.

Trip Assignment Component

The traffic assignment component used the BPR default parameters for link performance functions. Detailed research would have to be performed in order to determine the link performance functions which should be used for the different links in the MCRTAN.

Refinement of Centroid Connectors

The centroid connectors in MCRTAN were developed based on the three guidelines discussed in Section 4.3. Investigations should be made on the impact of different centroid connector construction on the model calibration results.

Zone System Improvement

The Winnipeg Regional Traffic Zone system is outdated and do not adequately represent the traffic generation characteristics in the Capital Region. Further work should be done to investigate and develop a new zone system for model development.

References

Capital Region Review (1999) *Final Report of the Capital Region Review Panel*, Winnipeg, Manitoba, Canada.

Mekky, A. (2000) *Analytical Transportation Planning*, Second Edition, University of Toronto Short Course Notes, Toronto, Ontario, Canada.

Federal Highway Administration (2000) *United States Code 23*, http://www.fhwa.dot.gov/environment/135.htm

Federal Highway Administration (1994) *Federal-Aid Policy Guide*, http://www.fhwa.dot.gov/legsregs/directives/fapg/cfr0450a.htm

Han, K. (2000) *Development of an Integrated Geographic Information System Platform for Transportation Applications*, MSc. Thesis Draft, University of Manitoba, Winnipeg, Manitoba, Canada.

Hurtubise, Paula (2000), *Conversation during Census Place of Work Seminar,* Missisauga, Ontario, Canada.

Nielsen, O.A. (1993) *A New Method for Estimating Trip Matrices from Traffic Counts*, Institute of Roads, Traffic, and Town Planning, The Technical University of Denmark, Paper 1993-3, Denmark

Clearing the Track: Appraising Railway Perspectives on the Issues of Regulation and Competition in the North American Rail Industry

William Drew, Canadian Wheat Board
James Nolan, Dept. of Agricultural Economics, U. of Saskatchewan

Introduction

As a consequence of pending changes to transportation legislation, the Canadian railway industry has embarked on another regulatory journey. The current legislation in the form of the Canadian Transportation Act (CTA) provides for a mandatory review of the act, to be completed by the summer of 2001. The importance of the next few months for the future of both railways and rail shippers in this country cannot be overstated.

Interpretation of prior experience with rail regulation processes indicates that Canadian railways will continue to persuade the federal government to pursue a reduced regulatory framework. The CTA interim report of December, 2000 confirms this (Gov. of Canada, 2000b). Operating links and increasing dependency on their U.S. counterparts (signaled by recent merger activity) indicates that the Canadian railways will continue to champion the adoption of U.S. style rail regulation in Canada. The issue we will discuss in this paper is the future direction of Canadian rail regulation. Foremost, we argue that the railway perspective on rail regulation, advocating increased pricing freedom without new competition (or entry) provisions, ensures further reductions in rate protection mechanisms for shippers.

Much of the evidence presented by the railway lobby favoring U.S. regulation is factually incorrect, or in many cases, simply misleading. It is our belief that the Canadian government would be well advised to seek better understanding of the position of the railways in this regulatory review process. Such understanding will be critical in a decision whether or not to repeat the U.S. experience with rail regulation.

Other evidence indicates that the U.S. rail industry is increasingly characterized by decreasing levels of competition, rising shipping rates to customers and decreasing levels of service. This situation is primarily a result of legislation and a regulatory body that has continued to insulate and protect railways from competition in bulk markets. The evidence shows that U.S. style rail regulation is not a good model for Canada. Innovative regulatory solutions need to be considered or shippers will continue to suffer and the industry itself will decline in relative importance in the overall transportation market (as defined by market share), as has occurred in the U.S.

Freight Rates Have Not Fallen

Railway industry research often asserts that freight "rates" have declined in the U.S. since the passage of the Staggers Rail Act in 1980. Railways, stakeholders and consultants often attribute this decline to the removal of regulatory restrictions at that time. Further, as part of the on-going Canadian debate on rail reform, one recent study concludes with respect to the U.S. experience that "less restrictive economic regulation facilitates improvement in railway productivity and financial performance while also reducing *rates* to shippers" (Conference Board, 2000).

The reality is that it is unclear whether rates since the passage of Staggers have declined for the majority of rail customers. Prior to 1988, the AAR published a "freight rate index", an aggregated index of rates paid by shippers to move various commodities over given distances. Since 1988, U.S. railways and the AAR are no longer mandated by regulators to report the rate index. Instead, North American railways report a proxy statistic called revenue per ton-mile (RTM). RTM is not equivalent to a simple freight rate. The computation of RTM is strongly influenced by changes in shipping distance and shipment weight. In many circumstances, it is a poor indicator of changes in actual freight rates as paid by individual shippers; RTM for a group of shipments can decrease when freight rates paid by customers remain the same or increase. Furthermore, it is well established that a proper freight rate index and RTM are not equivalent. More succinctly, RTM does not properly measure average

rate levels because it is affected by traffic composition and distance hauled (Banks et al., 1998).

In the de-regulated rail era up to the last year of the reported freight rate index (1988), the index fell by 7.9 percent, whereas RTM showed a decline of nearly 34 percent (Rastatter, 1998). Winston et al. (1990) also found that "when length of haul was accounted for, post deregulation rail rates were shown to increase while RTM decreased". Even without exact data on rates, recent research finds similar results. The US Bureau of Labor Statistics Railroad Producer Price Index reveals that rail prices increased 48% from 1980 to 1996 whereas the AAR RTM statistic indicates a decrease of 18% (Banks et al., 1998). The evidence points to the need for a consistent rate reporting scheme in order to properly assess U.S. freight rate movements in the post-Staggers era.

RTM Has Declined Under All Regulatory Regimes

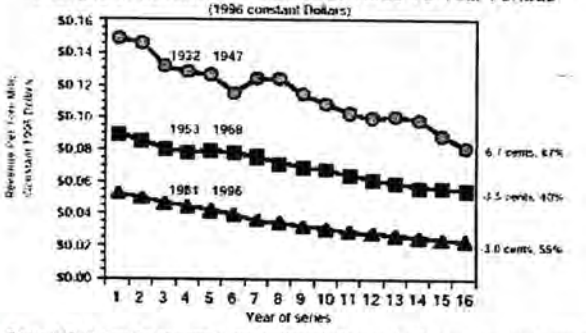
The rate debate is crucial to the future of Canadian rail policy. Are reduced freight rates a product of a favorable rail regulatory environment, an environment similar to that in the U.S. industry? In order to help answer this question, consider the following fact: RTM has declined virtually since the inception of the North American rail industry.

As alluded to earlier, a recent independent study claims that revenue per ton-mile began to experience a decline in Canada at the time of the National Transportation Act of 1967 (Conference Board, 2000). This Conference Board study also highlights a decline in RTM in the U.S. since 1980, attributed by the authors to the beneficial effects of Staggers. However if we track this data series backward in time, we find that RTM has been gradually declining for over 100 years in both Canada and the U.S., implying that the decline has occurred under various regulatory environments.

For instance, from 1870 to 1900 RTM declined by 50%.[1] And this trend continued through the passage of the U.S. Interstate Commerce Act in 1897 and the Mann Elkins Act in 1910, which gave the Interstate Commerce Commission (ICC) sweeping powers over U.S. rail rates (Mullin, 2000). Thus, it is clear that a variety of types of rail regulation have been present during the RTM decline of the last 100 years. The figure below illustrates the decline in RTM from 1932 to 1996. It seems to show that any claim about Staggers being primarily responsible for a decline in RTM is purely inaccurate (Banks et al., 1998). We conclude that the form of regulatory regime in rail appears to be irrelevant to movements of the RTM measure.

Figure 3

Changes in Revenue Per Ton-Mile, Three 16-Year Periods



Source: AAR, *Railroad Tee Year Trends*, various years; U.S. Department of Commerce, *Historical Statistics of the United States*

Why Has RTM in the U.S. Declined?

It is our contention that the decline in RTM can be traced to changes in railway technology. The earliest declines in RTM can be primarily

[1] Wilner (1997)

attributed to the standardization of rail gauge. This innovation made feasible longer haul rail business than was the norm up to that time. The next large fall in RTM in the late 1920's and early 1930's is attributable to further advances in logistics technology; one such innovation was the two-way radio (Wood and Johnson, 1996). Concurrently, the extensive paving of the U.S. road system contributed to railway comparative advantage in long-haul movements as road improvements further transferred short haul freight to the trucking industry.

In the late 1940's and early 1950's, more technology changes led to a sharp decline in RTM. These were driven by the introduction of diesel locomotives that could haul significantly higher tonnage over longer distances (Wood and Johnson, 1996). Concurrently, the construction of the US interstate system during this period shifted even more short haul traffic to highways. Finally, the full deregulation of the trucking industry, began in earnest in the U.S. in the late 1970's. This policy alone appears to have led to increases in average length of haul in rail by approximately 17% per decade since 1970 (Larson and Spraggins, 2000).

Technological advances in rail, combined with advances in other modes have significantly changed the composition of rail business. And full de-regulation in trucking brought about large-scale entry and greatly reduced shipping rates (Wood and Johnson, 1996). The effect of these changes has resulted in railway business shifting away from the movement of high value, often light weight short hauls to lower value high weight bulk traffic that must be moved longer distances. A combination of factors has significantly increased the average length of haul for U.S. railways throughout the 20th century, a fact which helped maintain the decreases in the RTM statistic throughout this period.

How Has U.S. Rail Deregulation Affected Bulk Shippers? A Focus on Agriculture

With respect to bulk goods, the price of transportation services relative to the value of the product is important (Currie, 1967). This

measure has a long history of use as an indication of the level of competition in transportation industries (Norton, 1963). Discounting differential inflation rates, if rail freight rates significantly increase faster than the corresponding increase in the price of the goods being transported, this can be an indication that market power is present in transportation.

In 1980, railway freight rates to move grain from Montana to Oregon represented between 10% and 19% of the value of the commodity at final destination. By 1990, this had increased to between 22% and 38%.[2] In 1999, freight rates in certain areas were greater than 51% of the value of the grain being transported.[3] If effective rail competition existed, rates to haul grain could not have increased in this manner, especially during a period where the world price for grain decreased. Rail competition has significant financial implications for the Canadian government. The on-going farm crisis in Western Canada has seen a renewed call for government subsidy of the agricultural sector. The Federal and two provincial governments recently provided 400 million dollars to help producers cope with increasing freight rates (Government of Canada, 2000a).

Since the passage of Staggers, there is evidence that U.S. railways have passed along much of the cost of transportation directly to their customers. Since 1981, shippers have purchased over 77% of new freight cars at a cost of $20 billion (Banks et al., 1998). The branch line rationalization process in the Prairies indicates that the same pattern is happening in Canada. This forces rail customers to truck their product over ever increasing distances to reach a rail line. These additional costs are borne by the customer (and the taxpayer with a public road system) and are not reflected in analyses of railway freight rates and/or revenue per ton-mile.

[2] Montana grain movements are captive to a single railroad.
[3] Whiteside & Associates (1999)

Rail Service in the U.S. Has Not Improved

American railways claim that another benefit of U.S. style regulation is that the level of service in rail has improved (Gallamore, 1999). But this position is not supported by recent evidence, as the rail industry publishes ambiguous performance measures that are not meaningful to the customer and are difficult to evaluate on an individual basis. For example, U.S. railways are mandated by the Surface Transportation Board (STB) to publish train velocity statistics. But this metric does not capture the important factor of total transit time for a shipment. The measure is analogous to the airline industry publishing on time schedule performance in terms of how fast their planes fly.

With respect to grain shipment service, some U.S. railways consider an order for a rail car "on time" if it is placed at the customers rail siding within a two-week placement window (Drew and Rosher, 1999). In the 1997 Burlington Northern Santa Fe annual report. BNSF reveals that in the third quarter, the railway was over 40,000 grain orders past due (BNSF, 1997).

The U.S. General Accounting Office (1999) conducted a recent study of the service issue. The GAO surveyed all rail customers and found some interesting results; 81% of shippers want increased transportation competition and 75% believed that they were being charged unreasonable rates. Even considering that some of these survey results may be explained by reaction to the merger between the Union Pacific and Southern Pacific railroads in 1995-1996 (Kwoka and White, 1998), U.S. rail shippers still appear to be a group very much dissatisfied with rail service in general.

The Issue of Revenue Adequacy

Revenue adequacy is a regulatory term defined as a situation whereby railways earn the minimum returns necessary to cover cost of capital requirements. It ostensibly measures the viability of railroads in all-important financial markets (Conference Board, 2000). Canadian railways, like their U.S. counterparts, promote the idea that they are only nominally "revenue adequate", and that this is a recent

phenomenon (Canadian Pacific, 2000a). This issue is not new; since the inception of rail, the issue of financing has permeated discussions about its existence (Burke, 1997). However, it is our contention that the goal of using this strategy seems to be to avoid pro-competitive policies that may limit railway ability to price discriminate among captive customers.

Right before implementation of Staggers, the Interstate Commerce Commission (ICC) determined that a revenue to variable cost (R/VC) ratio of 146% was sufficient to ensure revenue adequacy (Tye, 1991). Contrast this with present U.S. regulatory guidelines indicting that if a revenue to variable cost (R/VC) ratio of 180% or greater is found for a freight charge in a given market, it is only evidence (with no implications of guilt) of "market dominance". There is still a very generous range (@150-180% R/VC) within which a railway in the U.S. can be revenue adequate and yet would still not be found to possess "market dominance".

The merits of revenue adequacy as currently defined by the STB have been the subject of legal testimony (Kahn and Haas, 1991). In analyzing the market to book ratios of six Class 1 railways that the STB had deemed to be "revenue inadequate" by definition, Kahn and Haas found that all had positive book ratios[4], the lowest being 1.88 and the average being 2.41. Kahn further states that "investors collectively expect the prices these companies can be expected to be able to charge and the volume of business they can be expected to attract will be far more sufficient to produce a return in excess of the costs of capital".[5] Kahn acknowledges that the positive outlook by the investment community comes in spite of recent inflated base values of assets attributable to excessive prices paid in recent U.S. mergers and acquisitions.

---

[4] Statement of Professor Alfred G. Kahn and Jerome G. Hass on Railroad Revenue Adequacy Standards, February 1991, National Economic Research Associates

[5] Professor A.G. Kahn and J.G. Hass, statement on Railroad Revenue Adequacy Standards, February 1991, National Economic Research Associates, Page 2.

Recently, the Globe and Mail reported that the announced proposed revenue cap on grain rates might have a negative affect on CN and CP stock prices. But the article quotes an analyst: "(railroads) are very attractive investments and that is why we still recommend them as buys" (Barnes, 2000). Ultimately, we find it surprising that studies still promote the view that railway returns to capital are "fragile".[6] The debate indicates that the official measurement of financial viability for railroads must be brought in line with the kinds of indicators used in modern financial analysis.

Railways and Competition

The Canadian railways continue to claim that their business decisions are dominated by competitive processes (Rennicke, 1998). At one time in the history of North American railroading, intra-modal competition was significant. But there are indications that competition is no longer effective in the modern North American rail sector, including Canada.

In the U.S., the primary indicator of decreasing competition is increasing firm concentration in the rail industry and the geographic isolation of the largest remaining railways. Before the implementation of the Staggers Rail Act, there were over 40 Class I railroads in the United States. Currently, there are 4 Class I railways in the U.S., each dominating a particular portion of the country. The wave of rail mergers in the 1990's means that literally hundreds of shippers in the U.S. (and Canada) are now captive to only one railway for their long-haul business (Kwoka and White, 1998). There exists a very real possibility that eventually there will be just two Class I's in the near future operating throughout North America (Kaufman, 2000).

For many Canadian shippers the reality of the duopoly Class I rail environment has resulted in increasing rail rates. If we look at grain shipments, even with statutory rate regulation, recent reports indicate that Canadian Class I railways have been able to achieve contribution

[6] Conference Board of Canada (2000)

rates to capital that are in excess of the levels that would be achieved if effective competition existed in grain movement (Carlson et al., 2000).

Another potential indicator that competition is lacking in the Canadian rail sector can be found by examining the recent use of various competitive provisions for shippers in the Canada Transport Act. By construction, these competition provisions require the co-operation of a second railway, acting as a competitor, in order to work. Without the co-operation of a second railway, the CLR is especially useless for promoting rail competition.[7] In the period for which we have recent data (1996-1999), neither of these important competitive provisions have been demanded by any shipper in Canada (Canadian Transportation Agency, 2000).

A study of other captive commodities provides further evidence that rail competition is lacking. In discussing the relevance of operating ratios as a measure of rail efficiency, Boland (2000) discovered the following " ...While BNSF coal revenue yields are pulled way down by rail on rail competition for Powder River Basin Coal, most other mines are captive to a single railway. The results appear to be much higher revenue yields generated from coal shippers by CN and even higher yet by NS."[8]

Within the last two years, no fewer than six proposed bills to address the lack of rail competition in the U.S. have been discussed in the

---

[7] For cases beyond the distance set for mandatory interswitching limits, a shipper may seek an extended interswitch or a competitive line rate. Extended interswitching is applicable if the railway tracks are within 30 km of each other. The actual track distance involved in the switch from one carrier to another can be no more than 45 km. Beyond that distance, the shipper must seek a CLR. But in order to obtain a CLR, the shipper must first obtain a competing freight rate from an alternate carrier. Once this shipper/carrier agreement is made, a CLR is the authority granted to the CTA to set a transportation price so as to move the traffic over the tracks of the first carrier (that the shipper is trying to bypass) to the competing carrier. The CLR is applicable up to the interswitch point for the competing railway.

[8] K. Borland (2000)

Senate or House of Representatives. Four of the six bills call for enhanced running rights or access provisions to foster competition in the industry; another proposes to remove railway regulation from the authority of the STB, placing the industry under the authority of the Department of Justice. The scope of political activity to enact change indicates that, from the perspective of shippers, American rail regulation has evolved into a major failure.

Clarifying the Concept of Transportation Deregulation

A favourable regulatory environment for the rail industry, such as that which exists in the U.S., is often inappropriately referred to by the railways as "deregulation". But along with rate-making freedoms, full deregulation of a network industry also removes or minimizes entry barriers. Unless the franchise monopoly in a network industry is eliminated through such regulatory initiatives as open or enhanced access to the sunk or fixed asset (as in the U.S. telecommunications industry), it is incorrect to claim that U.S. style rail deregulation is really "deregulation" at all.

Considering the evidence presented here, we postulate that when Canadian railways suggest "deregulation" as a solution to industry problems in the context of Canada, they are, in effect, calling for continued regulation. Railways want pricing and service freedom (including the right to exit specific markets) without being obligated to face countervailing competitive pressures provided by the entry of new firms into various markets. The Kroeger debate made it clear that this was the type of regulation sought by the railways (Canadian Pacific, 2000b). It is most unfortunate that in the present policy debate, the Canadian railways continue to misuse the term "deregulation", possibly to influence and increase public support.

Rail Deregulation: Open Access, Freight Rates and the Case of Australia

Several countries have attempted to solve the fundamental problem that exists when a network industry can control entry into the market. Australia has in place a regulatory structure that guarantees new entrants the ability to enter the rail market place. While the

competition process is nascent and still evolving, evidence is beginning to surface that open access has allowed the Australian shipping community to gain a competitive advantage on its global competitors in the trade of certain bulk commodities, such as coal (Canadian National, 1998).

The Australian Railway Association reports in its 1999 yearbook that "a multi-user environment is emerging above rail", and "four new operators were granted access to the New South Wales network during 1997/98." And what effect has this policy had on the price of coal rail transport and volume of coal exports? The response is that "coal haulage rates when combined with the phasing out of monopoly rent in the Hunter Valley, have been reduced by as much as 17% facilitating the railing of a record tonnage of coal in NSW – a 22% increase over 1996/97." (DoTRS, 2000)

One Canadian railway has publicly stated that they responded to the Australian coal threat and accordingly reduced rates.[9] The reason that inexpensive Australian coal exists as an alternative to Canadian coal is portrayed by the railways as being a result of the closer proximity to port and less mountainous terrain in Australia. While this fact is undeniable, it is clear that open access in the rail industry has also enhanced the international competitiveness of Australian coal.

The North American railway response to growing evidence of the auspicious impact on customer rates and service as a result of open access overseas is to draw comparisons using the uninformative revenue per ton-kilometre measure. Utilizing RTM instead of actual rates gives the impression that shipping charges are less in Canada than Australia. But the distortion associated with the use of RTM in this comparison is due to the massive discrepancies in the length of rail haul for coal between Canada and Australia. In Canada the average distance from mine to port is 1,100 kilometres, compared to 135 km in New South Wales.[10] Algebraically, any comparison

[9] Dennis Apedaile, CITA Shippers Conference, Saskatoon, January 2000.
[10] Productivity Commission (1999)

between the two countries using RTM instead of an actual shipping rate is inappropriate.

What deserves highlighting in this case is the testimony to the powers of open access for the rail industry. Australian railways under open access are competing with the Australian trucking industry for 135 km length hauls! This stands in stark contrast to the situation in North America, where the vertically integrated rail industry has all but abandoned the short haul business, claiming that trucks are simply more efficient in that market segment (Wood and Johnson, 1996)

Conclusion

We have described and analyzed a series of policy positions that are being offered by the railroads as they face the latest regulatory review in Canada. Our assessment of these issues leads us to conclude that the official railroad positions on some crucial regulation and competition issues will certainly mislead and confuse those individuals unfamiliar with details about the industry.

It is apparent that the failure of the railway industry to report actual rail rates (or an index thereof) and instead report the measure revenue per ton-km is extremely misleading. Rates paid by shippers and RTM are not equivalent. Any conclusion that rail rates are lower based on revenue per ton-mile measures is inaccurate.

The rail industry obfuscates the effects of U.S. style rail regulation by drawing what we believe to be a spurious correlation between a decline in RTM and the enactment of the Staggers Act in 1980. In fact, we offer strong evidence that RTM has declined in North America regardless of regulatory framework since the last century. Furthermore we argue that the decline in RTM is primarily a function of continual increases in the average length of rail haul.

The Staggers Act in the U.S. has been particularly hard on captive shippers. The present consequence of U.S. regulatory changes has been an ever-growing group of shippers who are not satisfied with their rail service or rates. This has resulted in a renewed push via

political forums to tackle railway market dominance for these types of movements.

Considering the discussions herein, it is apparent that the continued rail efforts calling for deregulation in Canada are in reality a call for a new form of regulation. In particular, the railways continue to push for legislation protecting them from increased competition. Evidence from around the globe indicates that a practical and fair competitive framework in the rail industry is beneficial both to shippers and the rail industry itself. In Australia, open access has provided the rail industry with the incentives necessary to effectively compete with trucking on shorter haul routes.

It is important for the public to understand the reality for rail customers. As the forces of globalization continue, it is crucial that this next round of legislation provides a transportation structure allowing Canadian industry to compete fairly against other countries. Some of these countries have already addressed the issue of market dominance of rail and other network-based industries with innovative and practical solutions. We would be well advised to consider these experiments carefully before moving forward with a new regulatory framework for rail.

References

1) Australian Commonwealth Department of Transport and Regional Services (DoTRS) (2000) *Rail Facts at a Glance*, available on-line at http://www.dotrs.gov.au/land/rail/facts.htm

2) Banks, R.L. and Associates and Fieldston Company Inc. (1998) "Rail Freight Rates in the Post-Staggers Era", available on-line at http://www.railcompetition.org/library/studies/arcfinal.pdf.

3) Barnes, A. (2000) "CP shares fall as investors fret over freight rates", *Globe and Mail Business Section*, May 12, 2000.

4) Boland, K. (2000) "Probing Railway Operating Ratios", *Bridging the Gaps*. Proceedings of the 35th Annual Conference, Canadian Transportation Research Forum, 347-355.

5) Burke, J. (1997) *The Pinball Effect: How Renaissance Water Gardens Made the Carburetor Possible and Other Journeys Through Knowledge.* Little Brown and Co., New York.
6) Canadian National Railway (1998) *Canadian National Railway Yearbook, 1997.* Montreal.
7) Canadian Pacific Railway (2000a) "Railway Economics, Efficient Regulation and Mandatory Access", presentation by M. Cairns to the Dept. of Agricultural Economics, University of Saskatchewan, Saskatoon, March.
8) Canadian Pacific Railway (2000b) "Canadian Grain Handling and Transportation System", presentation by D. Apedaile to the Canadian Agricultural Economics Association, Vancouver.
9) Canadian Transportation Agency (2000) "Moving with the Times", 1999 Annual Report, Ottawa.
10) Carlson, L., J. Nolan and M. Fulton (2000) "Computing Regulated Access Rates Applicable to the Canadian Rail Industry", Working paper, Dept. of Agricultural Economics, U. of Saskatchewan.
11) Conference Board of Canada (2000) "Lower Rates and Improved Performance: Regulatory Reform of Freight Railways", Report 283-00, Regulatory Policy Alternatives Practice, Ottawa.
12) Currie, A.W. (1967) *Canadian Transportation Economics.* University of Toronto Press, Toronto.
13) Drew, W., and P. Rosher (1999) "U.S.-Canadian System Comparison: A Grain Shipper's Perspective". *Going Beyond: Moving into the New Millenium.* Proceedings of the 34th Annual Conference, Canadian Transportation Research Forum, 569-584.
14) Gallamore, R. (1999) "Regulation and Innovation: Lessons from the American Railroad Industry". Ch. 15 in *Essays in Transportation Economics and Policy*, J. Gomez-Ibanez, W.B. Tye and C. Winston, Eds. Brookings Institution Press, Washington, D.C.
15) Government of Canada (2000a) Press Release, February 24.
16) Government of Canada (2000b) Competitive Rail Access, Canada Transportation Act Review, Interim Report.
17) Kaufman, L.H. (2000) "How railroads might play the merger end-game", *Surface Reflections.* Journal of Commerce (on-line), June 20, 2000.

18) Kwoka, J.E., and L.J. White (1998) "Manifest Destiny? The Union Pacific-Southern Pacific Merger", *New York University Center for Law and Business Working Paper* #CLB-98-012.
19) Larson, P.D., and H. Barry Spraggins (2000) "The American Railroad Industry: Twenty Years After Staggers". *Transportation Quarterly*, V. 54, 31-45.
20) Mullin, W.P. (2000) "Railroad Revisionists Revisited: Stock Market Evidence from the Progressive Era", *Journal of Regulatory Economics*, V.17, 25-48.
21) Norton, H.S. (1963) *Modern Transportation Economics*. Charles E. Merrill Books, Columbus, Ohio.
22) Productivity Commission (1999) *Progress in Rail Reform*, Inquiry report No. 6, AusInfo, Canberra, Australia.
23) Rastatter, E. (1998) *Testimony of the National Industrial Transportation League, U.S. House of Representatives, April 22, 1998*:http://www.house.gov/transportation/rail/rrhearin/04-22-98/rastatter.htm
24) Rennicke, W. (1998) "Myths and Realities of Rail Access and Competition Issues". Research Report, Railway Association of Canada, Montreal.
25) Tye, W. (1991) *The Transition to De-regulation: Developing Economic Standards for Public Policies*, Quorum Books, New York, NY.
26) United States General Accounting Office (1999) "Railroad Regulation: Current Issues Associated with the Rate Relief Process", *GAO/RCED-99-46*, February, Washington, D.C.
27) Whiteside and Associates (1999) "Grain rates in the post-Staggers era", Consultant report, Washington, D.C.
28) Wilner, F. (1997) *Railroad Mergers: History, Analysis, Insight*. Simmons Boardman, New York, NY.
29) Winston, C., Corsi, T.M., Grimm, C.M., and C.A. Evans (1990) *The Economic Effects of Surface Freight Deregulation*, Brookings Institution Press, Washington, D.C.
30) Wood, D.F. and J.C. Johnson (1996) *Contemporary Transportation 5th Ed.* Prentice Hall, New Jersey.

Railway Competition and Continental Consolidation

Kells Boland
PROLOG Canada Inc.
Calgary, Alberta

Introduction

The momentum of railway market consolidation, through mergers and alliances, is growing across the North American continent. The promise of increased productivity from maintaining that momentum is countered with decreasing competitive countervail as the number of major railways dwindles from six toward two.

We are on the threshold of, in the words of the U.S. Surface Transportation Board Chairman, ". . . what will likely be the final round of restructuring of the North American railroad industry that may well result in two transcontinental railroads".[1] With ongoing continental consolidation, the competitive model for railways is now being revisited - by shippers, carriers and governments.

The purpose of this paper is to:

- highlight some of the more controversial aspects of the current Canadian railway market model; and
- set a valid context amid continental consolidation, for evolving that model into a more competitive multi-firm marketplace.

This paper concludes with a pragmatic scenario for rail-to-rail competition limited to transcontinental railways operating on shared rail plant which they would jointly own.

[1] Linda Morgan testimony before the U.S. Senate Subcommittee on Transportation, September 12, 2000.

Natural Monopoly Pricing and Productivity Sharing

Continental consolidation is increasing the rail industry's ability to most efficiently meet its full market potential at the lowest possible cost – that is as a "natural" monopoly.[2]

The railway natural monopoly provides a decreasing cost continuum as throughput increases over fixed right-of-way facilities and track systems that have relatively unlimited capacity to handle the entire foreseeable market – that is there are "global" economies of scale.[3]

To maximize the global economies of scale, railway pricing discriminates in favour of shippers with competitive alternatives (e.g., to attract trailer/container traffic) and against largely captive shippers (e.g., of bulk commodity trainload traffic) – that is discriminatory pricing is practiced where markets can be differentiated.[4]

Discriminatory pricing into differentiated markets is an economically legitimate practice to increase throughput productivity and reduce total system unit costs on natural monopoly rail plant. Productivity sharing can then provide lower rates to captive shippers and/or increase carrier profitability.

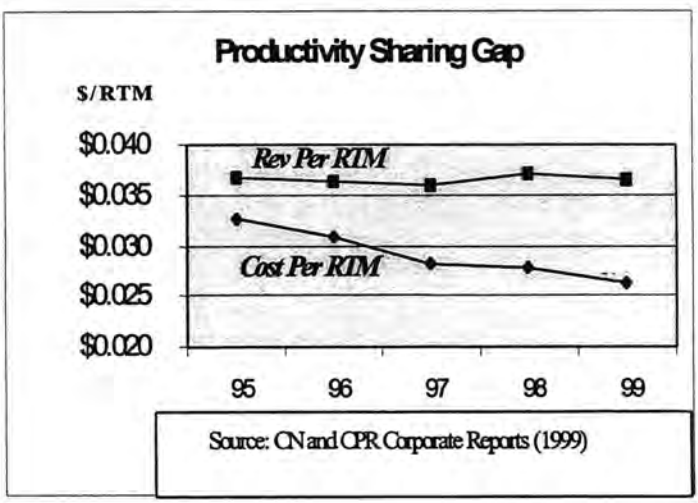
However, by definition there is no competition in the captive markets which are discriminated against. That lack of competitive countervail creates concern that productivity improvement may not be shared by shippers to the same extent that it might in a multi-firm marketplace.

---

[2] The ongoing momentum of railway mergers and shared track arrangements in North America reinforces the "natural" monopoly of consolidated rail plant which can more efficiently meet total capacity requirements than multiple competing track systems.
[3] The range over which economies of scale occur is a dynamic function of demand and capacity at successively larger, indivisible, fixed cost levels. In the railway industry, the train start, the railway tracks and the right-of-way each have successively greater fixed costs and an increasing range of scale economies. At the railway system right-of-way level of indivisibility, "local" economies of scale become "global" because of a relatively unlimited ability to achieve additional capacity at lower costs.
[4] more benignly referred to as differential pricing.

In fact under Canadian railway de-regulation there has been an increasing productivity sharing gap in terms of the growing margin of revenues above costs (see Figure 1 below).

Figure 1



Despite achieving ever-lower costs per revenue ton-mile (RTM), revenues per RTM (and by extension average rates) have been held relatively constant[5] . It would appear that the Canadian carriers have been able to dramatically improve profitability by NOT sharing productivity gains.

[5] On a discounted dollar basis, real revenues per RTM have declined somewhat, but not nearly as much as comparable expenses per RTM, and the increasing productivity gap remains the same.

Captive Shippers versus Rail Independence

The benefits of natural monopoly economies of scale have traditionally been balanced by regulatory recourse to protect captive markets against potential abuse under a discriminatory pricing model. Railway de-regulation has now removed most regulatory recourse at the same time as increasing railway consolidation has made it difficult to argue that competitive forces are effectively in play for more than a relatively small minority of rail shippers.

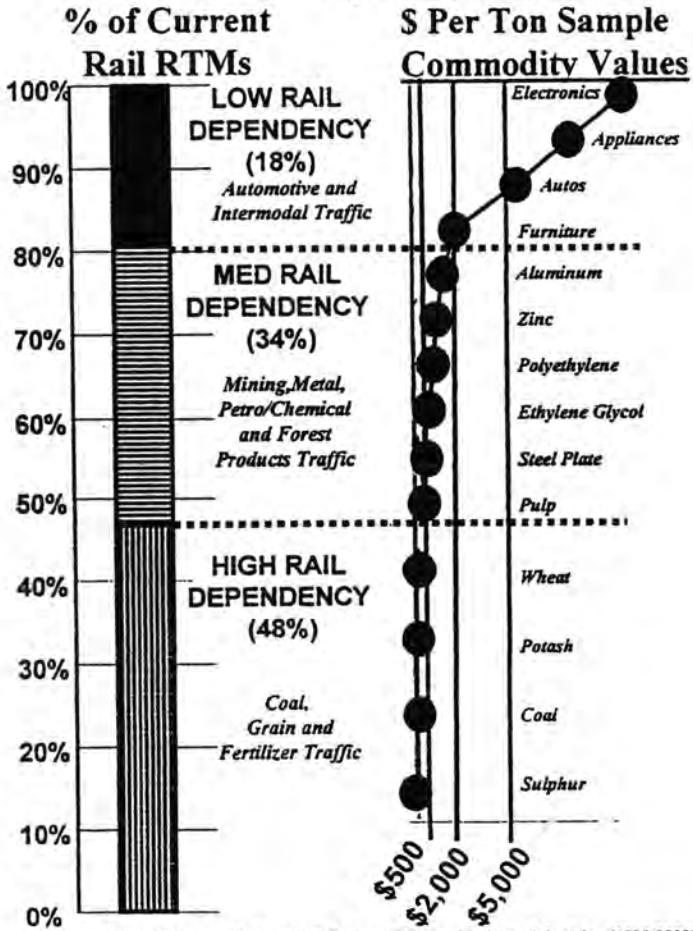
Currently less than 20% of the Canadian rail market is comprised of low rail dependent, high value intermodal/automotive traffic which can afford alternatives (see Figure 2). The balance of up to 80% of the market is lower value commodity traffic which remains economically captive - that is it cannot afford alternatives to the rail mode[6].

Direct highway competition is a practical alternative principally for the trailer and container market. Direct rail competition is increasingly irrelevant with mergers and alliances removing residual rail-to-rail competition at origins and gateways.

---

[6] The current economic force in play is cross-elasticity of demand for railway substitutes - primarily highway competition - which leads to rate discrimination against captive low value traffic. In an earlier era, railway value-of service pricing considered all traffic to be captive. At that time discrimination actually favoured low value commodities, as the high elasticity component, with the rate that would move the traffic - and high value goods, considered to have a relatively inelastic demand for transport, were charged all the traffic would bear.

Figure 2
Rail Dependency Profile
High value/low rail dependent traffic is a small percentage of total RTMs



Source: CN and CPR Corporate Reports (1999) and Industry Price Lists (1999/2000).

Within the rail mode, competition in terms of destination gateway alternatives much less multiple originating carrier options is being eliminated. The proliferation of "single line" routings which railway merger and alliance advocates advertise as significant shipper benefits also hold the shipper even more captive to that "single line" carrier.

In short, less than 20% of the rail market is independent of it. The direct competitive influence of the highway mode would only appear to require productivity sharing for high value, essentially intermodal shipments that are not dependent on rail.

For the 80% balance of lower value, higher rail dependent commodity based shipments; ongoing rail market consolidation requires some more equitable productivity sharing mechanism than the goodwill of the carriers. This can be a reversion to old methods of monopoly regulation or it can be with new approaches for rail-to-rail competition.

Vertical Separation

Potentially the most market effective - and most controversial - approach to enhance rail-to-rail competition is to "unbundle" vertically integrated train operations from track infrastructure. This is the de-regulatory approach that has been followed in the other network industries - isolating residual natural monopoly infrastructure while promoting multi-firm access on, and thereby bringing direct competition to, that infrastructure. [7]

In the railway industry, proposals to implement this approach range from limited expansion of current running rights practices to more

[7] Unbundling of telecommunications switching, electric power and natural gas transmission, under a number of different de-regulatory regimes, has ranged from complete corporate divestiture, to setting up separate subsidiary companies, to simple functional segregation on an accounting basis.

aggressive open access schemes. What they all have in common, though, is a critical requirement to set workable terms for track access – including full and fair access fees that will not undermine natural monopoly economies of scale and system cost recovery.

As outlined earlier in this paper, the natural monopoly cost structure of the railway industry is such that differential pricing which discriminates against rail dependent shippers actually protects the lowest system costs for them. This is because incremental traffic[8] is incentive priced above variable costs but below average total costs. The incremental traffic thus gained (or retained) makes a contribution to fixed overhead costs that otherwise would be incurred exclusively by rail dependent traffic and system unit costs are accordingly reduced for all traffic.

In effect, each traffic segment covers the full cost of the trains that carry that traffic. However, the common cost of shared track infrastructure is discounted to low rail dependent shippers with the difference made up by high rail dependent shippers – at a lower overall cost.

A more open access system, promoting rail-to-rail competition on the same tracks, has been perceived to jeopardize the natural monopoly differential pricing structure which facilitates full system cost recovery at the lowest possible unit costs. This is a legitimate concern that has been illegitimately applied to preclude the prospect of competitive access on shared rail plant.

In fact, the economic justification for differential pricing derives from the long run decreasing costs that prevail over any foreseeable demand requirements for the high fixed cost *Track System* - NOT high variable cost *Train Operations*.

---

[8] Incremental traffic is low rail dependency traffic that can be gained or lost

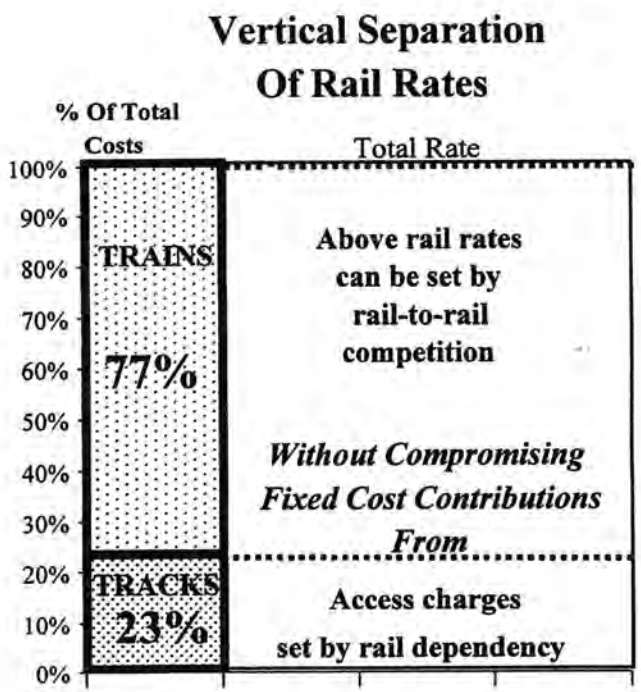
Accordingly, differential pricing to recoup full track system costs does not need to, indeed should not, be applied to the variable cost of train operations. Rail dependency based access charges with a sliding scale of contributions to the fixed cost of track system overhead are not incompatible with multi-firm competition to set the above rail rates associated with train operations.

The *only* legitimate target for differential pricing is access to the high fixed cost track system which is a relatively small component - less than a quarter - of total costs (see Figure 3). The preponderance of largely variable cost train operations is fully 77% of total costs upon which the objective market influence of multi-firm competition can be brought to bear.

While differential pricing can help achieve the greater economies of scale available from a railway natural monopoly, it simply does not follow that differential pricing must be applied to the full rail rate. Moreover, the significance of differential pricing as an impediment to competitive access has been distorted well beyond its relatively small proportion (23%) of total rail costs.

This misrepresentation is a self-perpetuating result of the vertical integration which masks allocation of revenues - to train operations versus track overhead. Competitive access, in whatever form, requires some degree of vertical separation of rates at least, if not of physical track and train operations. However, the relatively small proportion of total costs which warrant differential pricing cannot be an insurmountable obstacle – whether access fees are set by government fiat or, as envisioned below, by mutual negotiation among the carriers themselves.

Figure 3



Source: Statistics Canada Cat 52 216 XBP
and CN and CPR Corporate Reports.

A Competitive Access Scenario

Competitive access by multiple train operators to a shared track system does not need to destroy current vitality of the de-regulated, private sector Canadian rail industry. In fact it could well be accomplished by, and limited to, the two Canadian transcontinental railways.

Under such a scenario Canadian National and Canadian Pacific Railway might:

- merge all infrastructure assets into a jointly owned track system;
- establish a confidential fee structure for track system access; and
- set up train operating subsidiaries to compete on the track system.

This scenario recognizes the heavy risk capital investment in, and ongoing maintenance cost of, fixed rail infrastructure which must be funded by the transcontinental railways regardless of traffic levels. This scenario does not envision the economic inequity of risk-free, pay-as-you-go track access for short-lines, regional railways or non-track owning train operators.

By not providing access to other than the two transcontinental railways, a more purely competitive marketplace is foregone. However, the result should be a more acceptable, pragmatic approach for facilitating limited rail-to-rail competition with operational concepts currently practiced in North America.

Specifically under this scenario system-wide traffic solicitation rights would accompany mutually negotiated system-wide running rights.

North American railways are already subtly implementing such a scenario on a piecemeal basis:

- In Canada, Canadian National and Canadian Pacific share track to increase traffic on light density lines (eastern Canada) and to increase capacity on high-density lines (Fraser Canyon).
- In the western United States, Burlington Northern Santa Fe and Union Pacific Railroad have implemented joint dispatching over shared track assets initially in the Texas Gulf Coast area.
- In the eastern United States, Norfolk Southern and CSX Railroad now have mutual traffic solicitation rights in shared assets areas of the former Conrail System.

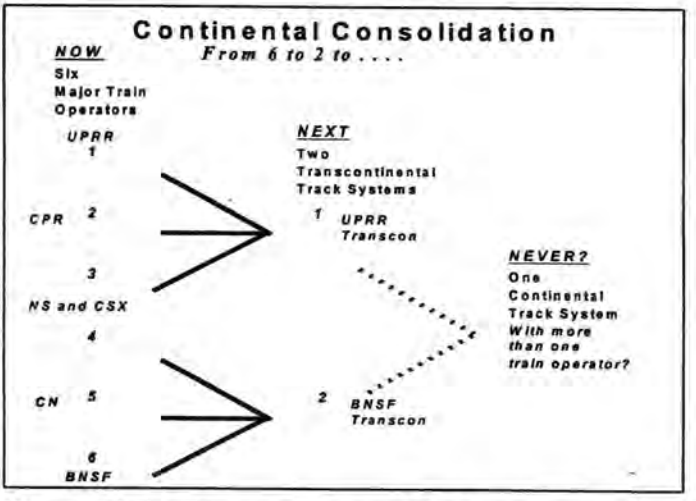
In the broader context of continental railway consolidation (see Figure 4) the six remaining North American railways are poised to further merge industry infrastructure into what amounts to common user rail plant.

There is not much technical difference between trains of several merged railways operated over the same infrastructure; and jointly dispatched train service subsidiaries operating over jointly owned infrastructure as envisioned in this scenario. However, the latter facilitates a competitive market place that the former forfeits. Most other remedies will always be artificial in that they typically impose subjective regulatory influence to replace the objective competitive marketplace.

If this scenario were implemented in Canada, it could be a model for mergers in the rest of North America. It is an expandable model that can be incrementally applied with each new merger application. That is, all future mergers could be made conditional upon setting up competing train service subsidiaries.

Ultimately, continental consolidation could lead to all six current major North American railways providing a multi-firm market on a single common user rail system.

Figure 4



However, during the next merger round(s) only two, or possibly three, train service subsidiaries would potentially compete with each other on several separately merged railway systems. The intervening transition period would avoid drastic, immediate disruption to the status quo, allowing all players a bit of breathing space to complete the inevitably difficult adjustment to a more competitive marketplace.

This scenario offers a flexible structure to gradually enhance railway competition with an increasing number of competing train operators on a decreasing number of competing track systems. The combination of track systems and permutation of train operators can be adjusted - or curtailed - as an evolutionary process.

The purpose of this process, however, is unchanging: To balance railway revenue and shipper rate requirements through the objective operation of a multi-firm marketplace rather than with the subjective judgment of regulatory authorities – or of unregulated railways.

Conclusions

Railway consolidation and cost reduction is dramatically improving throughput productivity. However, productivity sharing with shippers is decreasing.

In a deregulated environment, the objective allocation of productivity improvements between railways and shippers can only be accomplished with an effective market mechanism. However, up to 80% of rail shipments are lower value commodities that are economically dependent on that mode, with only limited practical access to the effective market leverage of truck competition.

Differential pricing is a legitimate means of maximizing the natural monopoly scale economies of the rail mode; and of fully recovering the fixed track system component of total rail costs. However, the fixed track system component is relatively small (23%) and should not significantly impede an effective market mechanism to set rates for the much larger balance (77%) of largely variable cost train operations.

The only effective market mechanism that is fair to both carrier and shipper is direct rail-to-rail competition. However, a vigorous multi-firm marketplace for rail-to-rail competition is increasingly unlikely as the six major North American railways merge toward two.

In conclusion, this paper has presented a pragmatic, incrementally expandable, Canadian model for railway competition and continental consolidation. This model can:

- accommodate ongoing rail plant consolidation;
- promote competing train operations on that plant;
- avoid potential disruption of completely open rail access; and
- keep rail plant and operations in the control of major railways.

Most importantly objective, private sector market operations are reinforced in lieu of subjective, regulatory market simulations.

RAILWAY REFORM IN AUSTRALIA: ACCESS REGIMES, CHANGES IN OWNERSHIP AND STRUCTURAL CHANGE

Derek Scrafton
Transport Systems Centre
University of South Australia (1)

INTRODUCTION

For over a century, the historical image of Australia's railways was one of government ownership, different gauges and a focus on regional ports. Although considerable change has taken place in recent years, state governments are still heavily involved in railways. The direct ownership and operating responsibilities of governments have given way to corporatised and privatised systems, but the legacy of parochialism engendered by colonial governments and different gauges can still be found in most systems.

Fifty years ago, there was an almost equal split of railways that were broad gauge (1600mm/5'3", in Victoria and South Australia), standard gauge (1435mm/4'8.5", in New South Wales and on the Commonwealth Railways) and narrow gauge (1067mm/3'6", in Queensland, Tasmania and Western Australia). After many inquiries, it was agreed the major cities should be linked by a standard gauge railway, and a standardisation program that was a mix of new track, gauge conversion and third-rail, began by linking Melbourne to the NSW system in 1962. The link from the east to Perth was opened in 1970, and the rest of the existing main lines were standardised by 1995. An extension to the northern port of Darwin is expected to be completed by 2004.

These improvements to the network, essential if railways were to be able to compete for interstate freight traffic, were paralleled by institutional changes. In 1975, the Federal Government agreed to take over the Tasmanian Government Railways and the non-metropolitan section of the South Australian Railways, and merged them with the Commonwealth Railways (2) to form Australian National Railways (3). Then in 1993 a national rail freight railway was created by excising the interstate freight business from AN and giving it operating powers, mainly over the standard gauge interstate main lines linking Brisbane, Sydney, Melbourne, Adelaide, Alice Springs and Perth.

However, these changes were not enough to shift the priorities of the State-owned railways from their bread-and-butter businesses of grain and suburban passengers (in all states) and minerals, particularly coal in NSW and Queensland (4). Increasing pressure on governments' budgets, and on rail freight rates from increasing competition from road transport (as a result of intrastate freight deregulation and improvements in road infrastructure and vehicle technology) and from competition in downstream markets, particularly export coal, had combined to establish a need for major changes in railways. But there was little change until the Council of Australian Governments agreed to implement a package of measures to extend competition policies to sectors previously exempt from the policies, including railways (5).

The *Competition Policy Reform Act 1995* amended the competitive conduct rules of the *Trade Practices Act* and extended their coverage to government business enterprises and unincorporated businesses; amended the *Prices Surveillance Act* to extend prices oversight to government business enterprises; and created new institutions (the Australian Competition & Consumer Commission and the National Competition Council) responsible for overseeing and providing advice on implementation of the NCP package.

The legislative change was accompanied by three inter-governmental agreements: the *Conduct Code Agreement*, which set out the basis for extending the coverage of the *Trade Practices Act*; the *Competition Principles Agreement,* which established principles for structural reform of public monopolies, prices oversight of GBEs, regimes to provide access to essential facilities, and review of legislation restricting competition; and the *Agreement to Implement the National Competition Policy and Related Reforms,* which provides for payments by the Federal Government to States and Territories that meet their commitments to the CPA and related reforms in the gas, electricity, water, and road transport industries.

The NCP package has proved to be a major stimulus to change in Australian railways, with new entrants, access regimes and structural change to organisations being introduced across the country. The rest of this paper summarises these changes, assesses performance to date, and highlights some current issues.

STRUCTURAL REFORM

Most governments have reformed the structure of their railways by breaking up the formerly integrated authorities horizontally and/or vertically, and in some there have been changes in ownership through franchising or privatisation. Where railways remain in government ownership, commercialisation or corporatisation models have been adopted. However, the application of these reforms varies in the intensity of their implementation, e.g. from 'true' corporatisation, requiring a GBE to act as a private company, to cosmetic models, where a Minister can still exert control, however indirectly. There is also variation in the geographic coverage of the reforms. As a result, different mixes of structural change and ownership have been applied to different networks; the following is a summary of the changes that have occurred.

Federal Involvement

Following the inter-governmental transfers of the 1970s, the Federal Government owned the Australian National Railways, operating freight services in SA, Tasmania and NT, and extending into adjacent WA and NSW, plus three long-distance passenger train services. The first step in dismantling AN was the transfer of its interstate freight operations to the newly formed National Rail in 1993, as a result of which AN lost any potential it might have had to be commercially viable. Then in 1997 the federal Government decided to sell the residual AN businesses: intrastate freight in SA (sold to Australia Southern Railroad, a subsidiary of Genesee & Wyoming), the freight operation in Tasmania (Tasrail, sold to Australian Transport Network, affiliated with Wisconsin Central), and the Indian-Pacific, Ghan and Overland passenger trains (sold to Great Southern Railway, a Serco company). Track was transferred to a new Federal authority, the Australian Rail Track Corporation.

The benefits of the sale to the Federal Government come from avoiding future revenue supplements and financial losses (6), not from the $A95.4million (7) sale proceeds. Direct sale costs totaled $A9.3million, and AN's wind-up costs were $A1billion of debt and lease obligations, $A152million outstanding contractual payments, and $A100million of redundancy payments.

Thus the former AN system now consists of two private freight operators (ASR and ATN), a government freight operator (NRC), a private passenger train operator, and a government track authority (ARTC).

National Rail Corporation is a corporatised, government-owned railway, incorporated under the Corporations Law, subject to the Trade Practices Act, and receiving no debt guarantees. The new operator received some start-up benefits from its shareholder governments

(Federal, NSW and Victoria), including cash, selected assets (transferred or leased), a greenfields labor relations agreement, and no residual AN debt (8). NRC is an above-rail operator running mainly interstate line haul freight trains, purchasing access from five track authorities. In 1997, the Federal Government indicated a wish to sell its shareholding in NRC, but complications caused by the multi-government ownership have delayed the sale, which should proceed in 2001.

The Australian Rail Track Corporation, was incorporated under the Corporations Law in 1998 to own the former AN interstate track in SA, NT, WA and NSW, and has a long-term lease of the interstate track in Victoria. The Corporation has responsibility for management of access and maintenance over the track it owns or leases, and elsewhere its function is to facilitate access to the interstate network through a 'one-stop-shop' service to train operators, marketing and selling train paths, and negotiating with the other four track owners on behalf of train operators.

New South Wales

The NSW State Rail Authority was a vertically integrated railway operating urban, intrastate and interstate passenger and freight trains. In 1996 the NSW Government passed legislation providing for an access regime for track in NSW and for the creation of four new rail agencies: Rail Access Corporation, which owns track infrastructure and is responsible for negotiating access to the NSW network and for financing investment and maintenance; Rail Services Authority, later corporatised and renamed Rail Services Australia, which provides track maintenance services to RAC under contract, and rolling-stock maintenance services (RSA has to compete with the private sector, though it has been protected to some extent by the State Government); Freight Rail Corporation (FreightCorp), which provides commercial freight services in NSW and interstate, plus freight services to some rural areas at non-commercial rates financed by a CSO contract with

the State Government; and a 'new' State Rail Authority, which provides urban, rural and interstate passenger rail services. Unlike the other agencies, SRA remains a statutory authority and receives most of its revenue from the State Government, either as CSO payments, concession fare reimbursements, revenue shortfalls or capital grants (9).

NSW therefore exhibits both horizontal and vertical separation, combined with varying degrees of commercialisation and corporatisation. After such complex changes to a major organisation, further modifications proved to be necessary. RAC and RSA were merged in 2001 into a single agency, the Rail Infrastructure Corporation (subject to direction by the Minister), following the recommendations of an accident inquiry.

Victoria

The reorganisation of rail services in Victoria was even more radical than that in NSW. The formerly vertically integrated Public Transport Corporation was restructured into five business units and the *Rail Corporations Act 1996* set up V/Line Freight and the Victorian Rail Track Access Corporation (VicTrack) as corporate bodies. Then began a process of franchising and privatising the different business units.

By 1999 the Victorian Government had sold V/Line Freight for $A163million to a syndicate headed by RailAmerica, which now operates freight trains in Victoria and interstate as Rail Australia. National Express won the 10 year franchise for V/Line Passenger (the main rural train operations), the 15 year franchise for Bayside Trains, one of two Melbourne suburban train businesses, and one of two tram franchises. There are also two private rural train operators, West Coast Railway and Hoy's, while the second suburban rail franchise went to Melbourne Transport Enterprises, a Vivendi company. The Victorian Government estimates the two metropolitan train franchises will save more than $A500million over 15 years, and the new owners have projected major increases in patronage.

Western Australia
In 2000 the Western Australian Government sold its rail freight operations (Westrail) for $A585million to the Australian Railroad Group (Genesee & Wyoming and Wesfarmers), including a 49 year track lease (10). ARG will operate trains as Australia Western Railroad, and manage the State access regime as West Rail Net. The suburban passenger train network in Perth and a few country passenger trains will remain government owned.

Queensland
Queensland Rail remains as a totally integrated railway providing passenger services, freight trains, including a major network for export coal, and infrastructure. Reform has centred on the financial arrangements between the State Government and QR, with corporatisation in 1995 under the *Government Owned Corporations Act 1993*, with the Treasurer (Minister of Finance) and Minister for Transport as shareholders. QR is subject to a tax equivalent regime, receives explicit payments for CSOs, and is obliged to develop an access regime.

South Australia
When the South Australian Railways were transferred to the Federal Governmant in 1975, the railways in and around metropolitan Adelaide remained with the SA Government as part of the State Transport Authority, together with bus and tram operations. In 1995 the STA was split into TransAdelaide (operations) and the Passenger Transport Board, a new agency for purchasing public transport services. All bus services have since been contracted out to private companies, but TransAdelaide continues to operate the single tram line and the suburban rail network. Although TransAdelaide was corporatised in 1998, the PTB contract for rail services is not contestable, so there is still scope for reform.

Thus from domination by governments at the beginning of the 1990s, the rail scene in Australia has changed to be, in 2001, a mix of private owners, franchisees, vertical and horizontal separation, and corporatised government entities. There is no simple picture, but evidence suggests that a combination of local conditions, politics, and financial stringency has determined the structural patchwork of railways that now exists across Australia. As these influences change and others arise in the next few years, so will there be further effects on the railways.

ACCESS REGIMES

Much of the structural reform of Australia's railways in the late 1990s was more extensive than would have been the case had the National Access Regime and a range of state-based access regimes not been in place. For example, new entrants now operate niche services in NSW and Victoria, FreightCorp has won a major coal haul in SA, and conversely National Rail is operating intrastate services in NSW. The most dramatic impact of the new regime is on ARTC's trans-continental main line, where several private operators operate freight trains, some of them in competition with NRC.

Under the National Access Regime, new and existing operators can, if unable to gain access by direct negotiation, request the NCC recommend the relevant Minister '*declare*' access to a particular facility (usually track). If the facility is declared, parties enter into negotiation, supported by binding arbitration, to reach acceptable terms and conditions. Alternatively, they can negotiate within the provisions of a legally binding '*undertaking*' registered with the ACCC, or a state-based regime that may or may not be certified as '*effective*' following a recommendation by the NCC.

The NCC must be satisfied that a number of criteria are met before recommending a service be declared. These criteria include a

requirement that access (or increased access) would provide competition in at least one market other than the market for the service, and that the facility is of national significance, having regard to size, trade, or the national economy. On receiving the NCC's recommendation, the Minister must then consider the matter. If the service is declared, the parties can negotiate, go to arbitration if necessary, appeal to the Australian Competition Tribunal, etc. Because the process is so protracted, it is not surprising to find no declaration applications have been pursued to completion – either the NCC has recommended no declaration, the relevant Minister has taken no action on a recommendation, or appeals have been lodged then withdrawn.

The process whereby an infrastructure owner can avoid being open to declaration by submitting an access undertaking to the ACCC for registration has similarly not been tested. There are as yet no undertakings for rail services.

The National Access Regime overrides other access regimes, including those established by state and territory governments, unless such regimes are certified as effective by the Federal Minister, following a recommendation from the NCC. If the Minister decides not to certify a regime, the state or territory Minister can appeal to the Australian Competition Tribunal. A regime has been certified covering the Tarcoola to Alice Springs and its proposed extension to Darwin. Some states have applied to have rail regimes certified, however the WA, Queensland and Victorian governments decided not to seek certification of their regimes, and the certification of the NSW access regime has expired.

The state regimes vary considerably in terms of their coverage and the provisions they contain: pricing principles, dispute resolution, transparency; scope to appeal etc. Given the different structural arrangements that prevail across the country, it is not surprising that the regimes vary to cope with different requirements for access.

There is also a variety of organisations existing in the states to administer and arbitrate the regimes. In Queensland, the state's Competition Authority (QCA) rendered, in late 2000, a draft decision QR's draft undertaking to allow access to its vertically integrated rail network. In NSW the Independent Pricing and Regulation Tribunal (or an alternative arbitrator appointed by the Tribunal) will act as arbitrator and publish its determinations. The NSW regime contains pricing principles incorporating a 'floor and ceiling' test, and establishes priority provisions for passenger trains. If a dispute arises in Victoria, the matter may be referred to the Regulator General, whose determinations can not be challenged, appealed or reviewed. In SA there are separate regimes for interstate and intrastate lines, plus one for the Alice Springs line. A specific regime does not exist in Tasmania, but Tasrail is required to negotiate with operators wishing to use its track through obligations contained in its conditions of sale.

In summary, the NCP is well supported by access provisions existing or in the course of development. The question is whether Australia has recreated its traditional gauge problems with this proliferation of access regimes.

PERFORMANCE

There were significant improvements in the productivity of Australia's railways during the 1990s. The average annual growth in TFP of *circa* 8% in the period 1990 to 1998 was greater than that of the US and Canadian Class 1 systems. However, there is still a significant gap with North American railways – Australia's level of productivity was about 66% of that of the best performing countries in 1997, including US and Canada. Some of the difference is due to factors disadvantaging Australia, such as scale of operation. Even so, when productivity is

adjusted for the effect of scale, Australia remains 30% below those best performers (11).

Freight customers have benefited from the improvement in productivity, as freight rates decreased by 30% between 1990 and 1998. This is comparable with decreases in Canada (33%) and US (27%) between 1990 and 1997. On average, freight rates in 1998 were still higher in Australia (A3.8 cents per ntk) than in Canada (2.1) and US (2.3) in 1997, partly reflecting the differences in productivity.

The Productivity Commission, using DEA, compared the productivity of Australian railways with a sample of North American railways (12). CN and BNSF achieved best practice productivity levels; compared to these two, North American railways had an average productivity level of 84% and Australian railways 47%. (Allowing for differences in locomotive power pushed Australia's average up to 52%). The best performing Australian 'railway' was AN/NRC at 63%, followed by Westrail (56%) and QR (43%). After accounting for differences in railway characteristics, to try to better determine technical efficiency comparisons, all railways in the sample were closer to best practice: North American railways had an average technical efficiency of 88%, and Australian railways 69%, led by AN/NRC at 81%.

Financial performance reflects the PC's economic analysis. During the period 1990 – 1999, the Federal GBEs (AN, NRC, ARTC) showed a net operating profit of A$0.4billion, QR A$3.7billion, and Westrail A$5.2billion. Stambrook (13) estimates total public financial support for and investment in railways in this period at *circa*A$32billion, with operating losses in NSW and Victoria accounting for almost A$7billion. Although early days, there is evidence that the reforms of the 1990s have potential to reduce this commitment from the public purse. The number of private railways rose from 6 in 1991 to 19 in 1999. The new entrants in both freight and urban passenger railways are showing signs of turning around formerly declining markets, with

commitments to investment, new services and courageous targets. In addition, private consortia are developing proposals for new trunk railways.

However the changes are not without problems and practical difficulties. Accidents in NSW have raised accountability questions between track owners, maintenance organisations, and train operators, leading to appointment of a rail 'overlord' and the RAC/RSA merger. In Victoria, the problems also relate to access: to what extent does a new private operator, who leases the track and has signed up to commitments to maintain track standards, grant access to the track and at what price? Access and vertical separation are not effective in markets where there is limited scope for more than one train operator and/or competition in downstream markets. The reforms to urban railways are aimed mainly at cost reduction, primarily labor costs (14), but subsidies for such services, whether contracts for service, provision of rolling stock, or a combination, seem inevitable.

CONCLUSION

Much has been achieved in reforming Australia's railways as a result of the implementation of National Competition Policy. Yet much more remains to be done as most government owned rail organisations are still not viable, and face increased competition from other railways, trucking, bus and coach companies, and cheap domestic air fares.

Whilst the emphasis has been on competition in and for the market, with structural reform, privatisation and access high on the agenda, it has become clear that the package of reforms for any particular railway or type of railway will be different. The mix required for a low volume regional railway facing competition from road transport will be different from that for a heavy-haul mineral railway. For urban passenger rail networks, facing strong competition from other modes of

transport, horizontal separation, with or without vertical separation may bring benefits, with contracting out or franchising creating competition for the market.

Reforms other than increasing competition are also important. Management improvements can yield benefits, as demonstrated by NRC and FreightCorp. Streamlining management of the interstate network is essential – ARTC's 'one-stop-shop' goes some way to reducing costs, but a single network manager would be better able to administer the access regime, and ensure the pricing principles are transparent and the appeal processes independent.

Competitive neutrality within the rail industry and between road and rail transport are essential elements of land transport reform in Australia. Government owned railways are able to offer lower freight rates than their private sector counterparts because the requirement to make a commercial return is not binding; existing private operators are disadvantaged, and the lack of competitive neutrality deters new private entrants and investors. More broadly, government decisions relating to investment, taxes and charges, access regimes and safety standards all disadvantage rail compared to road transport, but as long as large subsidies continue to be paid to rail transport, this issue will not be resolved. The PC therefore recommended a more commercial approach be applied to the provision of both modes of transport.

Australia is still in mid-stream in reform of its railways, and the future markets for rail will change again in the first decade of the 21st century, as they will in North America. The next obvious steps are the privatisation of National Rail and FreightCorp – if they were to be sold at the same time and/or one company was to buy both, the 1970s vision for a national freight railway might be realised. However it is achieved, the rail scene in 2010 is likely to consist of one or two freight railways, a series of urban franchises, complemented by short lines and niche operators.

Australian rail reform has been extensive and provides new investment opportunities, but the processes are those that have been in progress in North America and elsewhere for many years, applied to a handful of regional, government-owned railways down under.

Endnotes

1. This paper summarises and updates parts of the Productivity Commission Inquiry into Progress of Rail Reform (PC 1999a). The author, who was Associate Commissioner on the PC Inquiry, acknowledges the contribution of Dr John Salerian, Assistant Commissioner, in the preparation of this paper (Salerian 1999).
2. Commonwealth Railways main operations were on the Trans-Australian line linking SA and WA.
3. AN was envisaged as an embryonic Australian equivalent of CN, to be enlarged by mergers with other state railways. This concept was overtaken by other reforms.
4. The iron ore railways in WA are privately owned.
5. Although railways are subject to the general NCP provisions, they are unlike other industries such as power utilities and road transport, that are subject to inter-governmental agreements to implement sector specific reforms.
6. Similar benefits accrued to the SA and Tasmanian governments when they sold their state railways to the Federal Government in 1975.
7. In the mid 1990s A$1 was *circa* Can$0.90.
8. Nor did NRC pay AN for assets transferred.
9. In 1998, CSO and other operating subsidies were A$495million, capital grants A$356million.

10. Genesee & Wyoming thus gains a significant presence in Australia, with interests in railways in SA, NT and WA..
11. PC 1999a.
12. PC 1999b.
13. Stambrook, unpublished data. I am grateful to David Stambrook for sharing his financial analysis of Australian railways in the 1990s.
14. Employment in Australian railways dropped from *circa* 90,000 in 1986 to 36,500 in 1999; however, a large proportion of the reduction occurred outside the large cities.

References

Productivity Commission: Progress in Rail Reform. Inquiry Report #6. Ausinfo, Canberra, 1999a.

Productivity Commission: An Assessment of the Performance of Australian Railways, 1990 to 1998. Supplement to the Inquiry Report. Ausinfo, Canberra, 1999b.

Salerian, J: Australian Approach to Railway Reform, including Privatisation of Railways. Paper presented to the Regional Conference on Railway Reform and Restructuring, Hanoi, September 1999.

Stambrook, D: Australian Rail GBEs in the 1990s. Unpublished.

Abbreviations

ACCC	Australian Competition and Consumer Commission
AN	Australian National Railways Commission
ARG	Australian Railroad Group
ARTC	Australian Rail Track Corporation
ASR	Australia Southern Railroad
ATN	Australian Transport Network
BNSF	Burlington Northern and Santa Fe Railway
CN	Canadian National Railway
CPA	Competition Principles Agreement
CSO	Community Service Obligation
DEA	Data Envelopment Analysis
GBE	Government Business Enterprise
NCC	National Competition Council
NCP	National Competition Policy
NRC	National Rail Corporation
NSW	New South Wales
NT	Northern Territory
PC	Productivity Commission
PTB	Passenger Transport Board (SA)
QCA	Queensland Competition Authority
QR	Queensland Railways
RAC	Rail Access Corporation (NSW)
RSA	Rail Services Australia (NSW)
SA	South Australia
SRA	State Rail Authority (NSW)
STA	State Transport Authority (SA)
WA	Western Australia

RAIL ACCESS PRICING FOR FREIGHT
International Practice

The Canada Transportation Act Review Panel issued an interim report on December 29, 2000 regarding competitive rail access. Considerable attention has been focussed by shipper interests on the possibility of expanding the running rights provisions of the Act in order to provide new and existing railway operators the right to run over the privately-owned lines of CPR and CN for a price. The stated purpose of expanding regulated access is to increase intra-rail competition, but the access price is critical in determining the outcome of such a regulatory change.

This paper will review the international practice of providing access to rail lines for freight train operators at a regulated price. While the focus of the paper will be on the level and structure of access fees, some discussion of the industry structures and the associated terms and conditions will also be included.

I INTERNATIONAL OVERVIEW

From a survey of international practice it is apparent that a first key distinction to be made is whether there is a vertical separation of rail infrastructure from train operations – such as in the UK and parts of Australia – or whether the principal railways are vertically integrated, as in the US and Canada. In order to provide broad coverage, both the UK and the US were selected for review, together with the vertically-integrated Queensland Rail (QR) and the Australian interstate infrastructure operator Australian Rail Track Corporation (ARTC). Exhibit 1 provides a profile of the freight railways involved and a comparison with CPR and CN.

FREIGHT IN CANADA, UK, AUSTRALIA AND THE US

	Percent Freight	Route Miles	Operating Revenues (CDN$ millions)	reight Traffic Tons	MGTMs	Density MGTMs/ Route-Mile	Freight Operators
CANADA	91%						
CPR		9,771	$3,593	110,583,890	202,433,759	20,718	
CN		14,880	$5,313	135,216,142	267,972,361	18,009	
UNITED KINGDOM	8%						
Railtrack		10,352	$5,249	132,277,644	23,409,000	2,261	EWS (85%), Freightliner, Direct Rail Services
AUSTRALIA							
Queensland - QR	96%	5,959	$1,972	144,995,037	45,487,389	7,634	
Commonwealth - ARTC	93%	2,768	$76	3,802,982	18,220,178	6,584	FreightCorp, National Rail, Freight Victoria, Patrick Rail, Toll Rail, Silverton Tramway, SCT, Great Northern, Australian Southern
UNITED STATES	94%						
BNSF		25,636	$13,642	509,413,000	957,085,651	37,334	
UPSP		29,560	$14,980	552,543,000	995,686,627	33,684	
NS		17,126	$8,929	405,357,000	390,318,597	22,792	
CSX		20,914	$9,572	481,529,000	483,566,582	23,122	

US$1 = CDN$1.50 £1 = CDN$2.25 AU$1 = CDN$0.84

1 tonne = 1.1023137 tons 1 tonne -km = 0.6849691 ton-miles

The following general points about these railways may also be noted:

- The listed Australian railways are publicly owned, while all the others are privately owned;
- There are significant differences between the passenger-freight traffic mix, with the UK primarily passenger and the others freight;
- Freight traffic densities in the UK and Australia are significantly less than in North America.

II THE UNITED KINGDOM – RAILTRACK

In the UK rail industry, Railtrack is the monopoly rail infrastructure operator throughout the country, which includes the provision of electricity for traction. In addition, there are some 25 passenger franchise operators and 3 freight train operators: Freightliner, which handles marine containers, Direct Rail Services, which handles nuclear materials for its parent British Nuclear Fuels, and English, Welsh and Scottish Railway (EWS), which handles bulk freight and letter mail, and is by far the largest of the freight train operators.

The movement of freight is dominated by implicitly subsidized road haulage with rail handling just 5% of traffic. Furthermore, Railtrack itself handles predominately passenger traffic – passenger track access revenues in 2000 were £1,807 million compared with revenues of £162 million from freight operators.

The UK Rail Regulator determines the annual revenue requirements of Railtrack and the levels of efficiency to be anticipated over time. In this process – which has recently been the subject of a periodic review – the access revenue from freight is deemed sufficiently minor to be treated exogeneously. It should be noted that passenger franchises are subsidized by the government. Furthermore, freight operators can also be government-subsidized: annual freight grants of some £40 million are provided to encourage environmental

benefits associated with a modal shift from road to rail – the grants are for both freight facilities and track access.

Despite the rail network being near capacity in many areas, it is government policy to increase rail traffic significantly, and in particular to increase rail freight by 80% over the next 10 years. Overall there is a regulatory/policy climate in the UK in which there is an incentive to increase rail freight through relatively low access fees, and to have Railtrack receive offsetting compensation indirectly through the government-funding of passenger franchises.

The current approach to freight charging of access fees[1] established in 1995 is based upon individual negotiations between Railtrack and freight operators, subject to fees being between an avoidable cost (short run variable cost) floor and a stand alone cost ceiling. This approach was adopted to permit Railtrack to negotiate a share in the high value placed on transportation associated with certain types of freight – differential demand-based pricing. This resulted in a structure with a high fixed component (60%) and a low variable component, and Railtrack cost information was largely opaque to the operators.

It was intended that such an approach not provide different operators with substantially different fees, to the extent of distorting competition. However, the operators have subsequently merged into the dominant EWS with little or no rail freight competition. At this juncture it is worth emphasizing that the current approach for generating revenue or income for Railtrack through freight track access fees is **not** cost-based, although prices are expected to lie between cost-based minima and maxima. The prices being determined through negotiation implies that any economic rents

[1] The Rail Regulator has recently undertaken a review of freight charging policy, and it is the published documents and preliminary findings that are the basis for the information provided in this paper – see www.rail-reg.gov.uk

obtainable from shippers based upon demand elasticity are being shared between Railtrack and the freight operators.
The reasons provided by the Regulator for the present review of freight charging policy are as follows:

- The government has announced a policy objective to promote rail freight, and the review is timed to coincide with the periodic review of passenger access charges;
- Rail freight traffic has increased 35% in recent years and there is now a better understanding of cost causation;
- There is an interest in considering operator-based rather than traffic-based track access arrangements;
- There is a desire to negotiate longer-term agreements and new freight operators are considering entry.

Railtrack has indicated that new investment to encourage modal shift will require government financial support, and that the current freight charges provide little incentive to promote rail freight. Freight operators have argued that Railtrack is not responsive to freight needs and that no financial incentive is ever likely to be effective. Moreover, they do not want access fees to exceed avoidable costs – including a downward efficiency adjustment reflecting best international practices – in order to ensure that rail can compete with other modes. Shippers also argue for low rail access fees since rail freight is still not competitive with road in many instances.

In light of the above, the Regulator has identified the following pricing principles which have received broad support:

- Preventing monopoly abuse by Railtrack. However, Railtrack contends that inter-modal competition limits possible abuse;
- Allowing Railtrack to recover the avoidable costs of its aggregate freight operations;
- Promoting competition between freight operators;
- Increasing transparency and predictability;

- Avoiding distortions in freight markets – charges should not discourage traffic that can cover its avoidable costs, and there should be no discrimination between same-market operators;
- Providing incentives for Railtrack to develop the freight market. However, EWS considers this inappropriate.

The degree of inherent conflict in these principles, together with associated issues, has been the subject of discussion and public proceedings with the following outcomes at time of writing:

- Should a more deterministic approach to charges be adopted? Most parties have supported such a move, rejected access charges based upon the value of demand, the Regulator has concurred, but the problem of fixed charges discouraging marginal traffic has been recognized;
- Should there be greater transparency of Railtrack information – freight-specific costs and the structure of charges? Railtrack contended that transparency would inhibit negotiations of a sharing of the value of demand, but since the Regulator has opted for a more deterministic approach the point is moot. Both costs and prices will become publicly available;
- What costing methodologies should be used to determine stand alone freight costs, avoidable freight-specific costs, and anticipated efficiency gains? An independent study determined annual stand alone costs between £515 million and £798 million depending upon assumptions, very much higher than Railtrack income from freight access charges of £162 million or total freight-specific costs of £145. The detailed freight-specific costs were estimated as summarized in Exhibit 2[2]. Anticipated

[2] Note the performance costs. When a freight train delays a passenger train Railtrack must pay a full performance penalty to the passenger operator but receives only part of this cost back from the freight operator – the estimated funding gap is the performance payment. Congestion costs are those costs that Railtrack would avoid if freight were removed from the network. Freight traffic results in congestion related reactionary delays which also result in increased performance penalties to passenger operators.

EXHIBIT 2	RAILTRACK FREIGHT-SPECIFIC COSTS 2000		
	£ millions	Percent	£ per MGTM
CAPITAL COSTS			
Return on embedded assets at 8%	£1		
Enhancements	£2		
Sub-Total	£3	2%	£0.11
MAINTENANCE COSTS			
1 Shared Lines			
Track, structures and signals	£89		
Electrification assets	£1		
Sub-Total	£90	62%	£3.43
2 Freight-Only Lines			
Maintenance	£11		
Renewal	£9		
Sub-Total	£20	14%	£0.76
3 Freight-Only assets on Shared Lines	£0	0%	£0.00
Sub-Total	£110	76%	£4.20
OTHER OPERATING COSTS			
HQ and Regional Support	£6		
Freight-Only Signalling	£5		
Other	£1		
Sub-Total	£12	8%	£0.46
TRACTION ELECTRICITY	£6	4%	£0.23
PERFORMANCE COSTS			
Performance payments	£10		
Congestion costs	£4		
Sub-Total	£14	10%	£0.53
TOTAL FREIGHT-SPECIFIC COSTS	**£145**	**100%**	**£5.53**

efficiency in the range of 3-5% was settled at 3.1% over the next five years in the passenger review. Railtrack suggests no difference between freight and passenger, but EWS suggests annual freight efficiency at 11% - the issue is not yet resolved;

- What should be included in the charges and how should they be structured? It was decided that freight access charges should not include a contribution to the common costs of shared lines – they are to be allocated to passenger operators and funded implicitly by the government. While it was recognized that the fixed costs of freight-only lines should not be cross-subsidized by passenger operators, how they should be allocated is unresolved. Consideration is being given to allocation by commodity – presumably as a proxy for demand elasticity. Detailed variable costs by vehicle-type, commodity and operator

have been estimated but it is not clear whether charges are to reflect such differences or not. It is generally recognized that freight will be unable to finance freight-specific enhancements to the network, and therefore the Regulator is proposing that such enhancements be included at the time of passenger upgrades and funded by government. The passenger review will provide incentives for Railtrack to conduct track repair at night, thereby increasing the disruption of freight traffic – whether to include such impacts in the freight charges is not yet resolved;

- How should the complexities of a more deterministic approach be resolved? The Regulator has opted for commercial negotiation at the margin in relation to:
  i. The precise nature of access rights – the rights for firm schedules and/or additional, contingent or spot schedules;
  ii. The performance incentives;
  iii. The split between fixed and variable charges[3] and the allocation of volume risk between the parties.

Overall, the following conclusions can be drawn regarding future freight access charging in the UK:

- It will become more deterministic, transparent and cost-based rather than demand-based;
- The problem of the allocation of fixed and common costs discouraging marginal freight traffic has been avoided by establishing low freight access charges largely without these allocations, with Railtrack receiving compensation implicitly from the government through the passenger operators;
- The complex details have been finessed by the Regulator by retaining commercial negotiation at the margin – the scope of which is undetermined at time of writing.

[3] An independent economic analysis of the freight market has indicated that the higher the fixed freight charge the higher the freight traffic volume and vice versa.

III AUSTRALIA

Australia's rail industry hauls around one-third of the total ton-miles of surface freight, and nearly one-half of this is handled by isolated private, shipper-owned, vertically-integrated, profitable iron ore, sugar cane and mineral railways that are subject to export market competition. The balance of the rail system which was, until quite recently, in government hands is now a complex mix of government-owned and private, rail operators and infrastructure providers.

New and existing operators can gain access to rail lines and facilities through private arrangements, or through the National Access Regime introduced in 1995 as part of a National Competition Policy. The regime is complex and, in the examples considered below, the infrastructure providers are in process of developing formal Undertakings which specify in advance precise terms and conditions of access, including access prices or pricing methodologies, service standards, and extension of capacity.

The Commonwealth Productivity Commission[4] has reviewed rail access in Australia and made the following general observations:

- Regulated access may improve rail performance in markets where there is market power through the benefits of increased competition, but is unlikely to improve performance where there is significant inter-modal competition, source/market competition, or little infrastructure congestion;
- Access regimes should be flexible – encouraging optimal utilization of infrastructure but not permit cross-subsidization and not result in high transaction costs – and should be transparent, independently administered, and reflect market realities – ring-fencing of infrastructure is a necessary condition in the case of a vertically-integrated railway, while the

[4] "Progress in Rail Reform", April 13 2000, Productivity Commission, www.pc.gov.au

infrastructure provider should be discouraged from abusing market power in the case of vertical separation;

- Access arrangements are complex, slow to develop, and impose significant regulatory costs on industry participants;
- In principle, an efficient pricing and slot allocation scheme should reflect the opportunity costs across the entire network, otherwise operators may not make the best use of existing capacity and/or infrastructure providers may make inappropriate investment decisions;
- Access pricing can take one of three general forms: posted prices, negotiated prices or auctioning mechanisms;
- Posted prices do not provide a means for slot allocation, which requires a further mechanism – and usually relies on a first come-first served basis and/or "grandfathering" of existing schedules - and makes demand-based differential pricing very difficult;
- Negotiated prices could result in prices above stand alone costs, could involve a sharing of economic rents rather than increased competition, and negotiations will occur only periodically and therefore likely result in sub-optimal efficiency;
- Auctioning mechanisms address the element of competition for slots that are precluded by posted and negotiated pricing. However, they are complex, give rise to uncertainty, may be inefficient if abused by large operators, and are presently viewed as impractical[5].

3.1 Queensland Rail (QR)

QR handles 40% of all rail freight in Australia (excluding the isolated private railways) is vertically-integrated, state-owned, and is itself principally a freight railway – 87% of freight tonnage is coal, and 98% of passenger traffic is in suburban Brisbane. It is profitable but receives significant government funding as community service payments as summarized in Exhibit 3. In particular, low-volume

[5] Auctioning mechanisms for access slots are not currently used in Australia .

EXHIBIT 3: Queensland Rail 2000 Revenue (Excluding misc Other Revenue)			
	Sales Revenue	Government community services payments (CDN$ thousands)	Total
FREIGHT			
Coal and mainline freight	$869,014		$869,014
Other low-volume freight	$17,328	$1,881	$19,208
PASSENGER			
Brisbane	$157,368	$290,724	$448,092
Long distance	$47,679	$49,960	$97,639
INFRASTRUCTURE			
Network access group		$220,929	$220,929
TOTAL	$1,091,388	$563,494	$1,654,882

freight services are subsidized and there is a significant subsidy to the Network Access Group, which is the ring-fenced infrastructure provider within QR.

The Network Access Group represents some 29% of the total operating costs of QR – which includes most but not all yards, and does not include stations, maintenance facilities or terminals. It is ring-fenced to the extent of separate accounting, and will provide consistent access agreements to internal and external operators – but there have been no new entrants up to 2000, when QR has been challenged for some coal markets, and in 2000 new entrants represented only 0.8% of traffic by MGTM.

In 1999 QR submitted a draft Undertaking to the Queensland Competition Authority (QCA) and nearly two years later the QCA refused to approve the Undertaking, proposing instead 69

amendments. The following points represent the QCA position on access pricing[6]:

- Pricing principles shall include revenue adequacy for QR subject to limits on price differentiation. QCA specifically considers revenue adequacy in terms of efficient levels of assets and operations. Pricing is constrained market pricing with cross-subsidization avoided by ensuring prices exceed a lower bound of incremental cost and do not exceed a stand alone cost ceiling – determined for an efficient network provider;
- There is to be no inclusion of congestion costs at present. Performance standards are under development, but a system of financial incentives and disincentives to ensure compliance has yet to be determined;
- All operators in defined markets – based upon commodity and geography – are to be charged the same access price, subject to differences in cost and risk factors only. QR will be able to negotiate access charges with different levels of contribution to common costs, for operators serving different markets;
- Access charges may be reviewed if an applicant can demonstrate that, in the same market, QR has sold a like train path to another operator for a lower price;
- Reference tariffs for specified coal services will be used to overcome the problems with very broad limits imposed by floor and ceiling prices under a negotiated regime – namely, lack of transparency, uncertainty for operators, and negotiation costs. The tariffs will specify reference train characteristics and access prices may vary if different train characteristics requested by operators imply different cost and risk factors. The tariffs have been determined to provide QR with access revenues equal to the stand alone costs (less a productivity factor of 1.5% per annum) of the coal network in central Queensland – because the coal mines currently have no effective alternative means of transporting their product to market;

[6] The QCA five volume draft decision may be accessed at www.qca.org.au

- As indicated in Exhibit 4 for coal traffic with an average haul of 190 miles, reference tariffs have five components: an incremental maintenance charge, an incremental capacity charge, a charge for electric traction, and two allocated components. Total common costs are divided into two equal parts and the parts allocated according to MRTMs and net tons respectively;

EXHIBIT 4: Reference Tariff for Queensland Coal from Central Blackwater			
Price Component	Units	Unit Price CDN$	Contribution to Total Revenue
Incremental maintenance charge	$ per MGTM	$0.67	9%
Incremental capacity charge	$ per train path	$420	4%
Electric traction charge	$ per MGTM	$1.23	19%
Allocated component 1	$ per MRTM	$4.48	34%
Allocated component 2	$ per ton	$0.82	34%

- The QCA has reserved the right to require reference tariffs on three months notice for commodities other than coal, seeing the issue as a balance between increased transparency and the intrusion into QR's operational autonomy;

3.2 Australian Rail Track Corporation (ARTC)

ARTC is an interstate rail infrastructure provider, prohibited from operating trains, and is government-owned with a mandate to operate the business on commercially sound principles. Rail traffic over the ARTC network is predominantly intermodal freight handled by the 9 freight train operators listed in Exhibit 1 and in strong competition with highway trucking.

By the standards of the other railways in this survey it is very small with annual revenues from access charges in 2000 of CDN$76 million. ARTC makes a profit from operations but in a November

2000 preliminary draft Undertaking it states that it "is unlikely in any of its markets to be able to price at levels that will recover the full economic costs of its assets". This apparent anomaly can be understood when it is recognized that ARTC has had an 85% write-down of its CDN$1.2 billion of assets, and there is ongoing government funding of capital investment.

Access charges are posted prices, vary according to specific geographic market and level of service, and are presented in Exhibit 5 for a reference set of train characteristics[7]. These prices are subject to negotiation with the operator for required differences in the train and service characteristics.

EXHIBIT 5: ARTC Access charges July 2000 CDN$

	Specific origin-destination routes						
Distance - miles	12	46	192	233	519	530	1,245
Variable access charge per MGTM	$3.97	$4.46	$2.64	$2.85	$4.72	$2.90	$2.52
Fixed access charge per train by service							
- super premium			$625				
- premium	$36	$116	$440	$592	$2,670	$1,334	$5,028
- high	$31	$101	$394	$513	$2,313	$1,257	$4,357
- standard	$26	$86	$293	$433	$1,957	$1,094	$3,686
- low	$24	$77	$293	$394	$1,780	$1,056	$3,352

The structure of the charges is in two parts: a variable charge per MGTM and a "flagfall" or fixed charge per train for each service type and segment. ARTC has adopted equity and transparency as key elements of pricing policy and does not discriminate on the basis of

[7] Details of ARTC and its access charges may be found at www.artc.com.au

the identity of the operator, the commodity or the market being served – no demand-based differential pricing.

Prices generally exceed a floor of incremental cost – excluding depreciation and a return on assets – but may be lowered at ARTC's discretion. Overall, it appears to be government policy to provide a competitive interstate rail alternative to highways through the provision of rail access at prices that are uneconomic.

IV UNITED STATES

The four principal Class 1 US railroads are privately-owned, vertically-integrated freight railroads. Passenger services, both inter-city and commuter, are provided by other operators who have negotiated access to rail lines, but a significant majority of traffic remains freight as indicated in Exhibit 1.

Access to rail lines for freight operators – referred to as trackage rights – can be negotiated commercially or mandated by the regulator, the Surface Transportation Board (STB). Negotiated agreements do not generally provide for access in order to compete for the landlord's freight traffic, but an exception may occur in the case of a proposed merger. In order to mitigate the anti-competitive effects of a proposed merger, and smooth the subsequent approval process, a competitor may be granted access by the merging entity to its rail lines and its freight traffic in a negotiated agreement. Note that, unlike the UK and Australia, access is generally only granted for specific segments or components of track – not for the whole network, and not generally for the complete movement of traffic from origin to destination.

Trackage rights can also be mandated by the STB as part of the merger approval process[8], and the access charges can also be

[8] The STB may also mandate terminal trackage rights and the associated access charges but this does not appear to have occurred recently.

determined by the STB if the parties cannot subsequently agree. This was the situation in the case of the Union Pacific –Missouri Pacific – Western Pacific merger approved in 1982, and subsequent decisions of the ICC (the predecessor of the STB) established the methodology and made determinations regarding the trackage rights compensation that must be paid by St Louis Southwestern Railway Company (SSW) for mandated access to the MP rail line from Kansas City to St Louis. This methodology – referred to as SSW Compensation – has become the standard for regulated access charges, and was more recently applied in 1999 during the break-up of Conrail between NS and CSX, when CPR was mandated access to the CSX line east of the Hudson River between Albany and New York, and the access charges were determined by the STB.

The methodology establishes mandated trackage rights compensation in three parts:

1) The variable costs of operations, including train operations, dispatching, operating signals and highway crossings, switching and mechanical services incurred by the landlord as a result of the tenant operations. The costs to include standard additions as determined by the Uniform Rail Costing System constant cost mark-ups or flow-through option;
2) A proportionate share of infrastructure maintenance expenses, including taxes, based upon the tenant's use of the facilities – as measured by the percentage of car-mile usage;
3) The tenant's share of an interest rental component representing return on road property investment – referred to as the capitalized earnings approach;

and requires periodic adjustment over a specific period (one year after start-up, three years thereafter) to reflect changes in cost of capital, the relative number of car-miles for the landlord and tenant, and actual costs incurred in 1) and 2) above.

The methodology has been affirmed to conform to subsequent accounting principles, established by the Railroad Accounting Principles Board in 1987 to govern the determination of costs for specific regulatory purposes. In particular, as cited by the ICC:

- The Board's causality principle entails the use of incremental costs in establishing a floor for compensation[9] – avoidable costs that would not have been incurred without the movement of the tenant's trains over joint tracks;
- The variable costs identified in 1) above clearly conform;
- Similarly, an appropriate share of the costs in 2) above, constitutes an approximation of those costs that are incrementally associated with the movement of tenant traffic;
- Competitive access compensation levels may include in addition to incremental costs common fixed costs shared among the beneficiaries;
- The capital earnings approach represents a practical and reasonable measure of the specific assets required or consumed as a result of access. In particular, a return on current market value, at a current nominal cost of capital rate, reflects the relevant economic costs for future movements involving access;
- Lastly, sharing such identified costs places the parties on identical and competitive footing and ensures continued joint provision of service and adequate reinvestment.

It has been the capitalized earnings approach for determining interest rental compensation that has been the principal subject of dispute, and the methodology is illustrated in Exhibit 6 using the STB 1999 decision in the case of access by CPR east of the Hudson River over CSX. It is immediately apparent that the interest rental component is the most significant component representing more than 60% of the total access charge.

---

[9] US Railroad freight rates for a complete movement from origin to destination are generally subject to a legislated stand alone cost ceiling.

EXHIBIT 6: STB Determined Trackage Rights Fee, east of the Hudson River	(CND$)
Fair market value of CSX's portion of CR	$7,670,850,000
Road property proportion of fair market value	88.49%
Fair market value of CSX's portion of CR road property	$6,787,935,165
CSX's share of indexed 1995 CR earnings, excluding special charges	$562,806,000
CSX's share of annualized earnings from merger-related benefits	$194,070,000
Total pre-tax earnings of CSX's portion of CR	$756,876,000
Earnings multiplier	8.97
Line segment net earnings in 1995	$510,630
Inflation in 1996 and 1997	4.461%
Adjusted line segment earnings in 1997	$533,409
Value of line segment for trackage rights	$4,783,805
Pre-tax current cost of capital in 1998 for the railroad industry	15.6%
Allowable pre-tax return on line segment	$746,274
Tenant's car-miles on the line segment	1,567,112
Interest rental per car-mile: item 3)	$0.476
Variable operating costs and share of maintenance: items 1) and 2)	$0.303
Total trackage rights fee per car-mile	**$0.779**

The methodology requires: the determination of the fair market value of the landlord, multiplied by the road property proportion, to determine the fair market value of the landlord's road property; the estimation of the total pre-tax earnings of the landlord; and calculation of the earnings multiplier as the ratio of the two figures.

This earnings multiplier is then applied to the line segment earnings, multiplied by the current cost of capital, and then this figure apportioned to the tenant on a car-mile basis. Past reviews of this methodology has affirmed the following:

- Use of fair market value, rather than book or replacement value, so that the landlord is valued as a going concern;

- Valuation of the line segment based on the earnings potential of that property, not on the historical cost value shown on the balance sheet;
- Inclusion of incentive per diems as an expense.

One particular issue of recent interest is whether this methodology conforms to the Efficient Component Pricing Rule (ECPR). The ECPR establishes the price of access to a line segment to handle a given block of traffic as the sum of the incremental costs of the landlord to handle the traffic for the tenant, plus the loss of contribution to common fixed costs incurred by the landlord[10].

The SSW Compensation methodology does not conform to the ECPR. If the tenant is able to secure future contracts to serve two classes of the landlord's traffic, one demand elastic and one demand inelastic, the access fee is nevertheless the same for both classes of traffic, despite the fact that the landlord loses more contribution on the 2nd class than on the 1st class.

If the tenant serves a range of traffic equal to the average of the line segment, then on average the landlord may be adequately compensated. However, there is an obvious incentive for the tenant to "cherry-pick" the high contribution traffic, in which case the average access fee is too low. In fact, the methodology has the defect that, as the access charges are periodically reviewed, if the tenant is able to "cherry-pick" the high contribution traffic then the earnings of the line segment for the landlord decline and so the access charges for the next period are lowered accordingly. This problem has been implicitly recognized by the regulator: the relative use of the line segment between landlord and tenant is based upon a 3-year moving average which "has the ancillary effect of alleviating elasticity problems"[11].

---

[10] For details see "Efficient Access Pricing for Rail Bottlenecks", Eric Beshers, under contract with the Federal Railroad Administration.

[11] See SSW Compensation II – Trackage Rights, ICC Docket No 30,000, (1987).

V SUMMARY

The following general observations may be made:

- The development of access regimes is a complex, lengthy and costly process;
- Access fees should be set between an incremental or avoidable cost floor and an efficient stand alone cost ceiling;
- In similar markets operators should be required to pay similar access fees;
- Ongoing and evolving issues relate to performance standards and schemes of incentives and disincentives; how to handle capacity issues and congestion costs; and the management of access rights and slot allocations;
- The structure of access fees generally includes a fixed and variable component – the exception is the US;
- The level of the access fees should be such as to provide the infrastructure provider with the opportunity to be revenue adequate. The most contentious problem is how to recover the common fixed costs – demand-based differential access fees between different markets or ... ?;
    1) These costs are not charged to freight operators by Railtrack or ARTC, and the infrastructure providers receive indirect government subsidies to compensate;
    2) QR will set access fees at the stand alone cost ceiling for "captive" coal traffic, and has left access fees for the balance of minor traffic in other markets subject to negotiation for now;
    3) The US methodology sets access fees to cover the common fixed costs, to the extent that they are currently being recovered by the infrastructure provider, but sets a fixed access fee for all traffic.

It is apparent that international practice has not established a common solution to this intractable problem.

Modeling the Rationalization of Rural Road Networks: The Case of Saskatchewan

Paul Christensen (Dept. of Civil Engineering, U. of Saskatchewan)
James Nolan (Dept. of Agricultural Economics, U. of Saskatchewan)
Gordon Sparks (Dept. of Civil Engineering, U. of Saskatchewan)

Introduction

Beginning in the late 1980's, relatively few large high-throughput concrete grain elevators began to replace numerous smaller wooden elevators across the Prairies. Before 1990, the average distance traveled by trucks moving grain from farm to elevator in Saskatchewan was less than 20 km. In the past decade, the average shipping distance for grain by truck in Saskatchewan has more than doubled (Government of Saskatchewan, 1999).

More trucks use the provincial rural road network and travel greater distances as part of the grain handling and transportation system (GHTS). But from an engineering perspective, the road network in Saskatchewan was never designed to handle the truck traffic associated with a consolidated grain elevator network.

The expectation that railways would be obliged to serve the extensive wooden elevator network in perpetuity (through the Crow Rate agreement) combined with a limited tax base meant that when the rural road network in Saskatchewan was graded and paved, sacrifices in road durability were made. The consequence for today's GHTS is that the province is unique in Canada because the road network consists of three major surface types.

Saskatchewan Highways (1992) formally defines these road classes. The distinction is based on construction parameters. They are;

(a) Structural pavements (P), which range from a thin asphalt seal on a granular surface to a full depth asphalt concrete (approx. 40% of highway system).
(b) Thin membrane pavements (TMS) consist of a thin asphalt mat approximately 25 to 50 mm thick placed directly on a prepared

subgrade (approx. 40% of highway system). Unique to Saskatchewan, not designed to handle heavy traffic loads.

(c) Gravel (GR) is simply a layer of traffic gravel on a prepared subgrade.

While adequate to handle two-axle traffic under a few thousand kilograms, the latter two surface types are more prone to damage from heavier, multi-axle loads. A TMS road is especially vulnerable to such damage, and is the cause of much concern among Saskatchewan road planners.

A Decision Process for Rural Road Abandonment

We develop a basic mathematical model of rural road investment/abandonment based upon traffic flows and the cost of maintaining a given road surface type. Since the rural road network in the province will likely remain publicly funded, we assume that highway planners face the following set of decision rules for a particular road:

1) Maintain the existing road surface type because demand (traffic level) is compatible with road quality;
2) Upgrade the road segment because demand is incompatible with road quality and few viable routing alternatives exist;
3) Abandon the segment because alternatives are available. Demand is insufficient to justify costs of maintaining/upgrading.

By incorporating demand, maintenance costs and routing decisions, we hope to develop a systematic approach to the problem of rural road abandonment and make planning decisions easier and more justifiable in a political context. Since these types of problems are computationally intensive, we will develop our basic road abandonment solution algorithm using a simple network possessing characteristics roughly similar to the actual road network in the province.

It is instructive to note that most network design problems involving roadways focus on capacity issues pertinent to congested facilities.[1] A

[1] See, for instance, Friesz, T.L. and P.T. Harker [1983], Rothengatter, W. [1979] and El Dessouki, *et al* [1998]. Note that the latter two papers involve not only the discrete

search of relevant literature failed to reveal any network design models involving rural roadways. Instead, the only literature directly pertinent to the rural road network problem addressed in this paper involved *simulation modeling* of road networks—experiencing negligible congestion—in rural regions of the United States.[2]

It appears as if the continued focus on road capacity has left a void of serious investigation surrounding structural issues pertinent to rural road networks. Hopefully the research begun in this paper will help extend the application of formal network modeling to rural road problems.

A Mathematical Model of Road Investment Decisions

In order to develop an analytical means of addressing the rural road problem described above, it is necessary to first develop a mathematical formulation of the problem. This is done using graph theoretic notation. Next, we construct a formal *constrained* (or *combinatorial*) *optimization problem*. This optimization problem is coded within a spreadsheet and—given simulated data—solved using a *branch and bound* procedure embedded within a spreadsheet package (Microsoft Excel).

The Combinatorial Optimization Problem

In Figure 1, we refer to each circular vertex in the model as a *node*. The set of all nodes is denoted N. The bi-directional lines connecting each node are *arcs*. The set of all arcs is denoted A. The stylized "map", then, can be defined as a *graph*, G, containing all nodes and arcs (i.e., G = (N,A)).

---

assignment of urban road capacity improvements, but also discrete scheduling of the recommended improvements. The first paper presumes continuous capacity adjustment within freight networks (both road and rail)—applying nonlinear model formulations to determine optimal capacity adjustment.

[2] See, for instance, Hamlett and Baumel [1990] and Baumhover, McVey, Lipsman and Baumel [1990]. Interestingly, these papers examine the influence of structural changes in rural America (e.g., branch line abandonment, rural depopulation, etc.) on regional road networks. This closely mirrors the regional circumstances motivating this paper.

Although our model road map contains bi-directional lines, graph theoretic convention is that each line be separated into two, directional arcs. Any two distinct nodes i and j are connected by two unique directional arcs, (i,j) and (j,i). The road type corresponding to these arcs is consistent across direction. Conversely, the road type corresponding to arcs (j,k) and (k,j) is completely independent of the road type assigned to arcs (i,j) and (j,i).

The set of all road decisions is denoted by the index M (where $m \in M$). Within our stylized model, the set M includes: (a) *status quo* (the original road types assigned to each model arc may include TMS, gravel, or pavement), (b) decommission, and, (c) upgrade to P-class pavement. The set of unique traffic categories is termed K (where $k \in K$). Within this model, K includes: (a) passenger movements from node 5 to node 2 (Pax "A"), (b) freight movements from node 5 to node 2 (Freight "A"), and (c) freight movements from node 1 to node 6 (Freight "B").

The traffic flows assigned to each model arc are denoted by f_{ij}^{km}. This variable denotes the amount of traffic flow over arc (i,j) belonging to traffic category k and road type m. Additionally, since road type "decommission" effectively excludes traffic flows, all $f_{ij}^{k"decommission"}$ are assigned the value zero. Note that the sum of all traffic flows—across all traffic categories—in any direction may not exceed the *one-way capacity* of a two-lane rural road segment, u_{ij}.

Unit average variable costs corresponding to traffic flows are denoted c_{ij}^{km}. These unit costs account for all variable user and roadway damage costs associated with particular road types and traffic categories. We presume that these unit charges vary linearly with traffic flows (Kasilingam, 1998). Multiplying c_{ij}^{km} and f_{ij}^{km} provides a measure of total annual cost in dollars per kilometer ($/km/year). Multiplying this total cost measure by d_{ij}—the length of arc (i,j) measured in kilometers—generates the total annual variable cost corresponding to traffic flows across arc (i,j).

The supply of—and/or demand for—traffic category k at each node i is denoted by $R^k(i)$. If $R^k(i)$ is positive, then node i is a supply node and is denoted $O^k(i)$. If $R^k(i)$ is negative, then node i is a demand node and is denoted $D^k(i)$. *Transshipment* nodes feature zero net flows (i.e., $R^k(i) = 0$).

The unit capital cost F_{ij}^m (measured as \$/km) denotes either the remaining worth of an existing road segment (i.e., the arc-pair (i,j) and (j,i)) or the capital costs required to decommission or upgrade the road segment. The parameter α_m represents the annual amortization rate corresponding to road type m. To control the implementation of a decommissioning or upgrade for any given road segment, we introduce the *binary* model variable y_{ij}^m. For each possible m (i.e., *status quo*, decommission, upgrade), y_{ij}^m may be zero or one. To ensure that only one road type corresponds to each road segment, the sum of y_{ij}^m across all m is limited to one. Also, since each road segment corresponds to the arc-pair (i,j) and (j,i), y_{ij}^m must equal y_{ji}^m for all (i,j) and m within the road network.

To obtain total capital costs corresponding to each segment (measured in \$/km), we multiply y_{ij}^m and F_{ij}^m. Multiplying again by the corresponding segment length, d_{ij}, provides the total costs per segment (in dollars). We also introduce a measure of base annual maintenance cost corresponding to each road segment and road type, b_{ij}^m (measured in \$/km). Multiplying b_{ij}^m by d_{ij} provides a measure of base annual road maintenance cost (measured in dollars).

The primary constraints limiting the upgrade and decommissioning of road segments within our stylized network involve budget monies and physical resources. Available budget monies are denoted B (measured as dollars available for upgrade and decommissioning at time zero). Physical resource constraints are represented by L^m (measured as kilometers by road type at time zero).

This model is designed to determine an optimal (i.e. total cost-minimizing) network configuration. Although capital decommission and upgrade decisions are limited to time "zero" (i.e. immediate

transformation of the network), the costs corresponding to these decisions are streamed across an arbitrary planning horizon, T. For this reason, we incorporate discount and traffic growth factors within our cost computations. The variables involved in these computations include: (a) compound annual discount rate, r, (b) annual traffic growth factor, g^{kt} (≥ 1), corresponding to each traffic category and presumed constant across all t (t = 1,2,...,T), and (c) present worth factor, (P|F,r,t), equal to $1/(1+r)^t$ for any t.

Solving for a Network Optimum

We compute two categories of cost in our mathematical model: $PW_{ij}^{V}(m,0,T)$, and $PW_{ij}^{F}(m,0,T)$. The first denotes the present worth of all variable costs for arc (i,j) and road type m across our planning horizon (time 0 to time T). The latter expression applies to fixed capital (or, in the case of *status quo* road segments, remaining worth) and base maintenance costs.[3] Mathematically, we write each expression as:

$$PW_{ij}^{V}(m,0,T) = \Sigma_{s=0\text{ to }T} (P|F,r,s) \Sigma_{k\in K} c_{ij}^{km} g^{ks} f_{ij}^{km} \quad (1.1)$$

$$PW_{ij}^{F}(m,0,T) = F_{ij}^{m} [1 - (P|F,r,T) \alpha_m^{T}] + \Sigma_{s=0\text{ to }T} (P|F,r,s) b_{ij}^{m} \quad (1.2)$$

Notice that (1.1) is summed over all traffic categories, $k\in K$. Equation (1.2) explicitly accounts for depreciation of the road asset over the chosen planning horizon (α_m^{T}). Both are expressed in dollars per kilometer ($/km).

The assignment of road types within the stylized network must minimize total costs. This can be written as:

$$\text{Min}_{\{f,y\}}\ Z = \Sigma_{(i,j)\in A} d_{ij} \{\Sigma_{m=1\text{ to }M} [PW_{ij}^{V}(m,0,T) + y_{ij}^{m} PW_{ij}^{F}(m,0,T)]\} \quad (1.3)$$

The policy goal is to minimize Z with respect to both our traffic flow variables (i.e., all f's) and binary road type assignment variables (i.e., all y's). Z is the sum of all present value variable and fixed costs across all arcs contained in the network. Since $PW_{ij}^{V}(m,0,T)$ and

[3] If the status quo road type is maintained over the planning period T for any given road segment, then a depreciated *salvage value* equivalent to the road segment's remaining worth is assigned.

$PW_{ij}^{F}(m,0,T)$ are expressed as dollars per kilometer, d_{ij} must be applied across each arc to obtain the total present worth costs allocated to each road segment. The binary variables, y_{ij}^{m}, apply only to the fixed cost component of (1.3). Although y_{ij}^{m} might apply to both fixed *and* variable cost components (as traffic flow assignment must conform to the selection of road type), the influence of y_{ij}^{m} on variable costs is managed within the model constraints. The latter is necessary to maintain *linearity*.[4]

The following constraints impose basic controls over our flow and binary variables:

y_{ij}^{m} is binary, $\forall\ m \in M, (i,j) \in A$ (1.4)

$f_{ij}^{km} \geq 0, \forall\ k \in K, m \in M, (i,j) \in A$ (1.5)

$\Sigma_{k \in K} f_{ij}^{km} \leq u_{ij}, \forall\ m \in M\ (i,j) \in A$ (1.6)

$\Sigma_{m \in M} y_{ij}^{m} = 1, \forall\ (i,j) \in A$ (1.7)

$y_{ij}^{m} - y_{ji}^{m} \leq 0, \forall\ m \in M, (i,j) \in A$ (1.8)

Constraints (1.4), (1.5) and (1.6) set limits on the values assigned to binary control and flow variables. Control variables must be binary, flow variables must be greater than or equal to zero, and the sum of all flows across all traffic categories must be less than the one-way capacity of a typical two-lane rural road. Constraints (1.7) and (1.8) impose more conditions on our binary control variables. Each arc (i,j) may be represented by only one road type, and if arc (i,j) is assigned road type m, then arc (j,i) must be road type m as well.

The next set of constraints is known as the *mass balance constraints*. They are designed to ensure that the "supply" of traffic from all nodes within the network satisfies traffic "demand" at each node. The mass balance constraints can be expressed as follows:

---

[4] Branch and bound solving algorithms typically rely on some form of *LP-relaxation* (i.e., binary variables allowed to vary continuously between 0 and 1). If the binary control variable applied to variable costs within the model, both the flow and control variables would be multiplied together. The application of LP-relaxation would then result in a non-linearity (i.e., $f_{ij}^{km} y_{ij}^{m}$). This would void the use of a branch and bound algorithm.

$$\Sigma_{m\in M} (\Sigma_{j\in N} f_{ij}^{km} - \Sigma_{j\in N} f_{ji}^{km}) = \{R^k(i), \text{ if } i = O^k(i)$$
$$- R^k(i), \text{ if } i = D^k(i), \forall\, k\in K$$
$$0, \text{ otherwise}\} \qquad (1.9)$$

The sum of *outbound* traffic flows to each node i *less* the *inbound* traffic flows must equal the assigned supplies and demands at each node for each traffic category, $k\in K$.

To ensure the flow across any given arc (i,j) conforms to the assigned road type, $m\in M$, we add the following constraints:

$$f_{ij}^{km} \le y_{ij}^{m} u_{ij}, \forall\, k\in K, m\in\{M / m = \text{decommission}\}, (i,j)\in A \qquad (1.10)$$
$$f_{ij}^{km} \le (1 - y_{ij}^{m}) u_{ij}, \forall\, k\in K, m = \text{decommission}, (i,j)\in A \qquad (1.11)$$

If neither *status quo* or upgrade road types are selected, constraints (1.10)—combined with non-negativity constraints (1.5)—ensure that the traffic flows assigned to the affected arcs (i,j) are zero. But if the road type "decommission" is selected, constraints (1.11) ensure flows on affected arcs (i,j) are zero.

The final sets of constraints capture the physical and budget resource limitations imposed on the road agency responsible for "assigning" road types to each segment. These are:

$$\Sigma_{(i,j)\in A} d_{ij} y_{ij}^{m} \le L^m, \forall\, m\in\{M / m = \text{status quo}\} \qquad (1.12)$$
$$\Sigma_{m\in\{M / m = \text{status quo}\}} \Sigma_{(i,j)\in A} d_{ij} y_{ij}^{m} F_{ij}^{m} \le B \qquad (1.13)$$

(1.12) limits the total length of roadway that may be either decommissioned or upgraded across the entire network. (1.13) limits the total capital expenditures on decommission and upgrade. *Status quo* road types are not included in any of these constraints.

Minimizing the objective function (1.3) subject to constraints 1.4 through 1.13 provides the optimal network configuration. Since the model was encoded within Microsoft Excel, we used the Solver subroutine embedded within Excel. This was applied in order to find optimal network configurations across a number of model scenarios, a few of which will be discussed below.

Assumptions and Base Case Data

The following parametric assumptions were made with respect to our hypothetical road network.

1) Traffic volumes are constant across all scenarios. Assigned probabilities (p,q) alter the distribution of traffic volumes across intended destinations.
2) Traffic growth rates over our 5-year planning horizon (T) are: (a) 2.0% for Pax "A", (b) 3.0% for Freight "A", and (c) 3.5% for Freight "B". The assignments are arbitrary but apply to every scenario.
3) For each scenario, each freight "truck" is presumed to weigh approximately 55 tonnes (including a 40 tonne payload).
4) Capital costs / remaining worth and depreciation (amortization) rates are arbitrarily assigned but are similar to those values used in practice. Table 1 summarizes these values.

Table 1. Relevant Financial Data for the Road Network

Road Type	New Capital Cost/Current Salvage Value	Depreciation Rate (%/yr)
Status Quo (TMS)	25,000	85.0
Status Quo (Gravel)	20,000	95.0
Status Quo (P-Class)	75,000	95.0
Decommission	2,000	N/A
Upgrade (P-class)	150,000	95.0

5) The route traveled to the "destination" is the same route used to return to the "origin". The backhaul "damage cost factor" applied to freight vehicles is 1.60, applied to ALL scenarios.
6) For each scenario, base maintenance and damage costs are as listed in Table 2.

Table 2. Network Maintenance and Damage Costs

Road Type	Base Maintenance Costs ($/km/yr)	Damage Costs ($/tonne-km)
SQ (TMS)	750	0.08730
SQ (Gravel)	250	0.02910
SQ (P-Class)	500	0.01325
Decommission	0	0
Upgrade (P-class)	500	0.01325

7) Passenger (pax) and freight user costs (across road type) are set to reasonable values. These are shown in Table 3.
8) The "backhaul user cost factor" is presumed to be 2.00 and applies to all scenarios. Truck operating costs vary little given the horsepower and efficiency of modern trucks once they are "up to speed".
9) The one-way capacity of a two-lane rural road is 900 vehicles per hour. Studies have suggested a two-lane capacity of around 1800 to 2000 vehicles per hour but the traffic volumes in this network are well below capacity. This is applied to all scenarios.

Table 3. Passenger and Freight User Costs

Road Type	Pax Costs ($/km)	Freight Costs ($/km)
SQ (TMS)	0.39	1.21
SQ (Gravel)	0.40	1.27
SQ (P-Class)	0.35	1.10
Decommission	0.00	0.00
Upgrade (P-class)	0.35	1.10

10) The constrained capital budget is limited to $500,000. In the "unconstrained" scenario, this amount is raised to $5 million. The unconstrained scenario only uses about $1.4 million out of the available $5 million.
11) Other resource limitations in the model (i.e., L^m) are set at 15 kms for "decommission" and 10 kms for "upgrade". These are arbitrary and apply to all scenarios.
12) The discount rate is set at 10% per annum for all scenarios.

The only differences in parameters distinguishing the model scenarios described next involve budgets and the base road types assigned to each segment. Further, we do not discuss here a series of simulations that altered traffic assignment probabilities (p,q). Figures 1-4 illustrate the network investment problem. The road network contains 8 road segments (see Figure 1). Figure 2 shows the base case scenario. Arcs (1,2), (3,4) and (2,6) are thin membrane pavement surfaces covering 44 percent of the total distance in this network. Arcs (1,3), (2,4), (3,5) and (4,6) are assigned as gravel roads covering 37 percent the distance. Only segment (5,6) is assigned as pavement in the base case, representing 18 percent of the total network.

Network Design and Scenarios

We altered some base assumptions regarding the key variables. For the networks illustrated here, the important parameter settings and changes include the following:

1) Figure 2 (Base Case). $p = 0.5$, $q = 0.5$, original road types a mix of TMS, Gravel and Pavement. Total PW Cost ≈ \$11.35 million.
2) Figure 3. $p = 0.5$, $q = 0.5$, original road types a mix of TMS, Gravel and Pavement. However, arc (1,2) is now defined as a Pavement section. Total PW Cost ≈ \$9.82 million.
3) Figure 4. $p = 0.5$, $q = 0.5$, original road types a mix of TMS, Gravel and Pavement. The total roadways capital budget is unconstrained. Total PW Cost ≈ \$9.92 million. This scenario is computed for direct comparison to the scenario in Figure 2. The assumption regarding the capital budget is the only factor that separates them.

Model Results: Flow Assignments (f_{ij}^{km})

The flow assignment indices describe the traffic flow along each directional arc (i,j) by traffic category ($k \in K$). This value is reported as $\Sigma_{m \in M} f_{ij}^{km} = f_{ij}^{k}$, $\forall (i,j) \in A$. Since the road type ($m \in M$) corresponding to each road segment is made clear in each diagram, we do not list "m" on the flows reported in the figures. Further, since road type selection is mutually exclusive for each arc (i,j), all other f_{ij}^{km} will equal zero.

In general, our results show that user equilibrium appear to mirror the system-optimal results. The solution flows in this stylized network do not exclude shortest-path trip types from any user's standpoint. The model has also satisfied the "mass balance" constraints; supplies at nodes 1 and 2 match demands at nodes 2 and 6. Since we have confirmed that the optimization algorithm works in a sensible manner for users, next we interpret the key results of this paper.

Road Type Assignments (y_{ij}^{m})

The final road configuration shown in Figure 2 contains two important results. Segment (3,4) has been upgraded from TMS to pavement and segments (1,2) and (2,6) were decommissioned. Given

the network configuration and parameter set, the cost minimizing network matched transportation supplies with demands by decommissioning two longer, more expensive segments. To counter this, users received a major pavement upgrade on a short but crucial piece of road (3,4) that now serves the transportation needs of those affected by the decommissioning. Users are re-routed along segments (2,4) and (4,6). Not surprisingly, segment (3,4) absorbs a large proportion of total traffic volume in the new equilibrium.

Figure 3 starts with a similar network to that in Figure 2, except that arc (1,2) begins as a pavement section. This will clearly lower the total cost of our final network configuration; we assume that more total roadway has been built at the highest pavement grade level. The solution for this problem we call the "left-horseshoe" configuration, which includes an upgrade to one of the gravel segments on the left edge of the network (3,5) and complete decommissioning of the interior segments. The present value of total cost for this scenario is the lowest at $9.82 million.

Figure 4 shows the unconstrained budget model incorporating the same initial parameters used in Figure 2. The solution of this problem we call the "right-horseshoe" network. As in Figure 3, the interior segments are decommissioned along with gravel edge segments (1,3) and (3,5). System travel demands are met by upgrading segments (1,2) and (2,6) from TMS to pavement. The total present value of costs involved in this scenario is relatively low, around $9.92 million dollars. But more importantly, the solution to this problem involves the most direct re-routing of traffic for any of the scenarios examined, and is thus the most "user-friendly". The latter point illustrates how myopic decision making by governments with respect to roads could be costly to society in the long run. We find that the budget expenditures recommended by the model to configure the road network in this scenario were only $1.38 million (compared to $500,000), yet the long-term total cost for this scenario compares favorably to others we constructed and users were better off.

Staying mindful that this is preliminary research, we find that the road type assignments yielded by the model fully meet user needs and yet maintain the viability of the road network given the resource constraints imposed. Clearly, more sensitivity analysis and larger network models are needed to better understand the effects of changes in the resource constraints on the status of the road network equilibrium.

Conclusions

The province of Saskatchewan faces some difficult transportation decisions in the next few years as the structure of the grain economy changes. The need to properly maintain the extensive rural road network will increase over the next decade. While much of the existing road network needs to be upgraded, limited highway budgets dictate that this cannot be done for every qualifying road segment. How the government will make decisions to either upgrade or decommission the secondary road system subject to budget constraints is of primary importance.

In this paper, we attempt to provide some guidance as to how such decisions can be made in a systematic manner. Using a small stylized network, we implement the technique of network optimization in order to find a cost-minimizing network configuration that meets the needs of users and is achievable within a reasonable budget constraint.

The network configurations examined here involve a considerable amount of road decommissioning and re-routing of users to their destinations. Interestingly, the scenario with an unconstrained capital budget resulted in the most convenient network configuration for users. These findings indicate that the future of the road network in Saskatchewan will involve the consideration of such trade-offs. Officials would be well advised to recognize that the existing road network and road types will strongly impact the shape of the rural road system in Saskatchewan.

References

1) Baumhover, S.B., M.J. McVey, M. Lipsman, and C.P. Baumel (1990) "Investment Strategies for the Rural Road and Branch Rail Line Systems", *Journal of the Transportation Research Forum*, 31(1), 167-179.

2) Boyce, D.E. (1979) "Network Design for Transportation: Introduction", *Transportation Research*, 13B, 1-3.

3) El Dessouki, W., N. Rouphail, M. Beja, and S.R. Ranjithan (1998) "Multiperiod Highway Improvement and Construction Scheduling: Model Development and Application", *Transportation Research Record 1617*, Transportation Research Board, National Research Council, 96-104.

4) Friesz, T.L. and P.T. Harker (1983) "Multicriteria Spatial Price Equilibrium Network Design: Theory and Computation Results", *Transportation Research*, 17B(5), 411-426.

5) Government of Saskatchewan (1999) Parallel Process on Roads. Saskatchewan Road Impact Analysis. Regina.

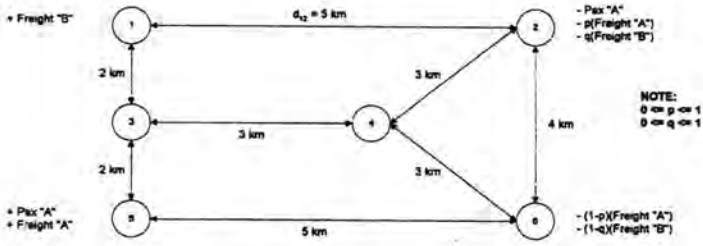
6) Hamlett, C.A., and C.P. Baumel (1990) "Rural Road Abandonment: Policy Criteria and Empirical Analysis", *American Journal of Agricultural Economics*, 72 (1), 114-120.

7) Kasilingam, R (1998) *Logistics and Transportation.* Kluwer Academic Publishers, Boston.

8) Rothengatter, W. (1979) "Application of Optimal Subset Selection to Problems of Design and Scheduling in Urban Transportation Networks", *Transportation Research*, 13B, 49-63.

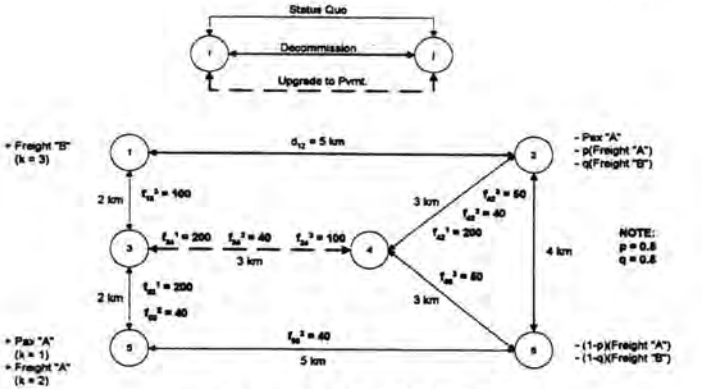
9) Saskatchewan Department of Highways and Transportation and Saskatchewan Department of Agriculture and Food. (1992). *A Description of the Grain Handling and Transportation System.* Regina: Saskatchewan Department of Highways and Transportation.

Figure 1



Presumed Supplies at each Node:
Node 1: Freight "B" = 100 ; Node 6: Pax "A" = 200 ; Node 6: Freight "A" = 80
(NOTE: Demands can be calculated according to diagram)

Figure 2



NOTE: unless otherwise specified, $f_{ij}^k = 0$, for all k and (i,j) in A.

Presumed Original Road Types (prior to modeling):
(1,2) = TMS ; (1,3) = Gravel ; (2,4) = Gravel ; (2,6) = TMS
(3,4) = TMS ; (3,5) = Gravel ; (4,6) = Gravel ; (5,6) = Pvmt.

Presumed Supplies at each Node:
Node 1: Freight "B" = 100 ; Node 6: Pax "A" = 200 ; Node 6: Freight "A" = 80
(NOTE: Demands can be calculated according to diagram)

Planning, Operations, and Maintenance of Toll Road Systems

Ata M. Khan,
Carleton University

Mohammad R. Tayyaran
Ottawa-Carleton Regional Transit Commission

ABSTRACT

This paper describes specialized aspects of planning, operations and maintenance of toll roads on a life cycle basis. Toll road development trend is examined and selected examples are provided in terms of rationale for their development. A methodological framework and constituent components are described. The maintenance part of the framework is illustrated by using a Monte Carlo Simulation Model that systematically addresses uncertainties involved in input data and output results.

INTRODUCTION

Frequently, governments are unable to pay for public road capacity and serviceability improvements that are required to keep pace with increasing user demand. Also, financial constraints are encountered in meeting the timely maintenance requirements of road infrastructure. These factors led a number of jurisdictions around the world to develop toll ring roads and some times radial roads in major urban areas.

The private sector plays a major role in building and operating toll roads. A common model followed is that of build-operate-transfer (BOT). Normally, the toll road authority has complete control over technical as well as financial aspects of the toll road business. Also, since users pay a special toll for the use of road, their expectations are high regarding service quality. Under these conditions, the planning, operation and maintenance of such roads is different than other roads in a region.

The purpose of this paper is to describe specialized aspects of planning, operations and maintenance of toll roads on a life cycle basis. For illustration purposes, examples of actual toll roads are provided.

TOLL ROAD DEVELOPMENT TREND

Rationale for Toll Roads

An examination of literature and discussions with toll road organizations suggests that there are two reasons for developing toll roads. The first reason is financial in nature. Owing to growing traffic in urban regions, the demand for road capacity keeps on increasing. As a part of strategic and tactical plans, transportation planners identify road as well as public transit improvement requirements. However, governments are frequently under financial constraints and cannot fund road improvements according to plans. Due to the absence of a dedicated fund for transportation infrastructure in many countries, transportation projects have to compete with other needs of the urban regions.

Owing to financial constraints, future strategy of some governments for enhancing the quality of road infrastructure relied on the development of toll roads. This strategy involved the privatization of some segments of the highway system (Canadian Civil Engineer 2000). Toll roads such as the Highway 407 Electronic Toll Route (ETR) in the Greater Toronto Area (GTA) could be an acceptable way

to provide the required highway infrastructure in almost any urban area (Forumation 2000). Another development in terms of commercialization of roads in Canada and U.S.A. is the outsourcing of road operations.

Toll roads offer a mechanism to raise funds required for road network capacity expansion and physical maintenance. These facilities are developed according to the transportation strategic plan for an urban region but do not impose a monetary burden on the public purse. The users pay for the high quality of service they receive. The life cycle of a toll road can be planned according to the financial and transportation engineering principles.

The second reason for considering toll roads within an urban region's transportation network is to relieve congestion at key locations. Frequently, high volumes of urban commuter traffic and the intercity traffic share the limited roadway capacity. The result is intolerable congestion-induced delays and economic losses. Traffic congestion in the Greater Toronto Area (GTA) is estimated to cost businesses in Ontario at least $2 billion a year in lost time and productivity (407 International 2001). In some urban areas, roads in the vicinity of central business district experience severe congestion due to the imbalance between demand and capacity. Under these circumstances, toll roads can offer an effective solution to the congestion problem.

Growing Trend and Selected Examples

Toll cordons have been developed and are now operated in many cities around the world. For example, toll cordons in Norway's three largest cities, namely Oslo, Bergen and Trondheim are well known (Larsen 1995). Other examples are Toronto's Highway 407 Electronic Toll Route (ETR) and toll roads in Singapore.

In the case of Norway, the reasons for building toll roads were to raise money for road capacity expansion. Also, congestion in cities of Oslo and Bergen was a problem. According to literature, it is common in Norway that the government spends less on roads than the revenue

received from taxes on car ownership and car use. The toll roads are well utilized in Norway and there is growing acceptance of paying a special charge for high quality of service. Nevertheless, it is a very common reaction of motorists in Norway to toll cordons that car owners already pay enough taxes (Larsen 1995).

In Norway, it was believed that in spite of the need to improve the road system, funds were not available for this purpose. Also, it was the understanding that toll revenues should be used for the benefit of road users in the form of road improvements. The Norwegian government was in favour of toll road initiatives in urban areas and agreed to supplement toll revenues with additional funds.

A Company, namely Fjellinjen, was established to operate the toll systems in Norway's three cities noted earlier. This company transfers the net revenues (i.e., profits) to road authorities.

It was the view of the government that a strong monopoly position and stable traffic streams would allow the toll road authority, Fjellinjen, to maintain a sound financial condition (Fjellinjen 1999).

In Oslo, the city council saw an opportunity to cross-subsidize public transit and demanded that toll revenues should also be used to benefit public transport users. The central government mediated. It was decided that toll revenues should be used for road projects and the additional government funds that were intended for Oslo should be dedicated to public transport projects. It was believed that motorists are more likely to consider tolls to be "fair", given that they pay directly for the use of a facility that is intended to improve driving conditions (Fjellinjen 1999).

The Highway 407 ETR is a 69 km long east-west expressway located just north of Toronto in the GTA. It connects major centres across southern Ontario's busiest transportation corridor. Transportation planners identified the need for a major facility in this corridor. It was also recognized that by using traditional financing methods, it would have taken decades to obtain funds for building this highway. Further

extensions are under construction which when completed will provide a total of 108 kms of high quality, transportation connection through the GTA. At present, it serves 240,000 customers on weekdays (407 International 2001).

In order to reduce the burden on taxpayers and to expedite the addition of capacity in this corridor, the Government of Ontario decided to develop the highway as a "user pay" or toll route. A toll authority, a new Crown agency, was established to design and construct the highway in an expeditious and a cost-effective manner by working in partnership with the private sector. The 407 ETR was completed in five years. The 407 International Inc. is the sole shareholder, operator and manager of the 407 ETR (407 International 2001).

FRAMEWORK FOR PLANNING, OPERATIONS AND MAINTENANCE OF TOLL ROADS: LIFECYCLE

The overall framework for planning, design, construction, operations and maintenance of toll roads on a life cycle basis is presented in Figure 1. A toll road authority can assume total responsibility for all aspects of the system. Given that this authority can apply commercial principles in managing traffic as well as the road assets, the planning, operations and maintenance methodologies applied to toll roads are different than other roads.

Planning & Concept Design Issues

Planning and concept design is the first major activity. Although the preliminary location information for a toll road is known within the overall regional transportation plan, the exact location of the facility has to be established. Also, traffic demand and composition of traffic have to be estimated in response to the level of tolls. Furthermore, the concept design is to be developed in terms of access points, exits, and the number of lanes for each segment of the facility.

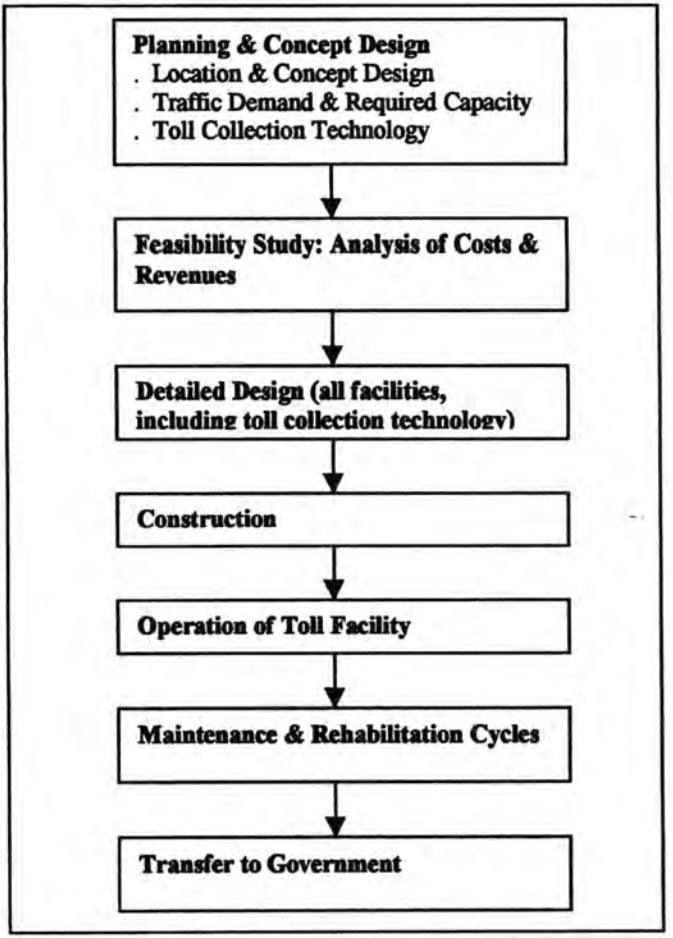


Figure 1: Life Cycle of A Toll Road System

In the concept design phase, a toll scheme may be adopted so as to improve efficiency in the transport system. For example, a toll cordon may be designed along the principles of congestion pricing (Larsen 1995). This is the case with toll road in Singapore (Land Transportation Authority 2000). In Norway, toll cordons for Oslo, Bergen and Trondheim surround the central areas of the cities and only the inbound traffic is tolled. These have played a role in relieving congestion (Larsen 1995).

At the time of planning and concept design, the toll authority has to decide on the type of toll collection technology to be employed. Although the trend is towards electronic toll collection, various details have to be worked out (e.g. total absence of toll booths, the handling of non-compliant drivers, the exact technology of toll collection, etc).

Feasibility Study

On the basis of traffic estimates, composition of traffic, and toll levels, revenue can be calculated for the analysis period (e.g. 20-40 years). Costs of construction, operation, and maintenance are to be estimated as well. A comparison of revenue and costs can establish the feasibility of the toll road project. The revenue and cost analyses can be carried out in constant dollars or in actual dollars.

In order to account for risk and uncertainty, the Monte Carlo Simulation Model can be used to find the expected net present worth and the probability of loss. Further information on this method is provided in the maintenance section of this paper.

Detailed Design & Construction

The overall toll road system as well as various components have to be designed according to the applicable criteria, standards and codes. Service quality, safety, reliability and environmental impacts are accorded special attention.

Given that a single agency is responsible for design and construction, the time required to get the facility ready for service can be minimized. The availability of funds for construction and procuring tolling technology ensures that no delays are encountered.

Toll Road Operation

The operation of a toll road encompasses the introduction phase, publicity and advertisement, traffic management, collection of tolls, and enforcement. The intelligent transportation systems (ITS) technologies play a major role in toll road operations.

Most operational tasks for an electronic toll route are automated. The examples of the 407 ERT (Toronto) and toll roads in Norway are provided here for illustration purposes.

The 407 ETR, one of the most advanced electronic toll collection systems in the world, has five main components:
. Vehicle transponders
. Vehicle classification system
. Overhead gantry mounted toll collection system
. Toll transaction processor
. Revenue management system.

The transponders are easy to install to the interior of the windshield behind the rear-view mirror. The installation process can be completed in a few minutes, without tools. Also, these are cost-efficient for regular users of the facility. A user can register (through Internet) and obtain the transponder at a number of specified outlets.

The transponders can be moved from one vehicle to another between vehicles of the same class. For heavy vehicles, transponders are mandatory. In the case of 407 ETR, over 410,000 transponders are used in the vehicles of users (407 International 2001).

For vehicles equipped with transponders, the electronic sensors located on each overhead gantry log the ETR 407 entry and exit

points. At the time of exiting from the toll facility, a green light on the transponder and four short beeps indicate that the toll transaction has been completed correctly. If a yellow light is observed and a long continuous beep is heard, it implies that the user's account needs attention. It is an advisory that a customer service representative should be contacted. If on exit, a red light is noticed and a long continuous beep is heard, it means that the account has no funds or there is a problem with that transaction.

If a vehicle enters the toll facility without a transponder, the trip is logged by using a license plate recognition system. This system is mounted on each overhead gantry and is programmed to send up to five video images to a central processing computer when the vehicle enters and exits the toll road. A bill is sent to the owner of the vehicle, consisting of the applicable toll charge plus a $2.00 Non-Transponder charge.

For 407 ETR, the toll charge for a trip depends on (a) the time of day the user enters the ETR, (b) vehicle class, (c) distance travelled, (d) and whether one has a transponder or not. The total trip charge is calculated according to the toll rate that is effective at the time a vehicle enters the facility. All transactions take place electronically and there is no need to slow down to pay a toll (407 International 2001).

In Norway, the plan is to exchange all existing electronic tags with new tags that conform to the new European Communication standard and therefore will be technically interoperable. The tags to be offered to drivers will have SMART cards as an option. Also, a plan is under consideration that is designed to exchange the current manual payment part of the tolling system with only electronic fee collection supplemented with video pictures of vehicles that enter the facility without electronic tags (Fjellinjen 2000).

Maintenance Factors

In order to maintain a road at a steady level of performance, the following infrastructure maintenance management activities are carried out (Transportation Association of Canada 1997):

- Collection and monitoring of information on current pavement conditions.
- Forecasting future conditions.
- Evaluating alternative rehabilitation and maintenance strategies.
- Prioritizing investments.

The concept of improving the serviceability of a road pavement is illustrated in Figure 2. Alternative 1 requires higher investment than Alternative 2, but lower yearly maintenance.

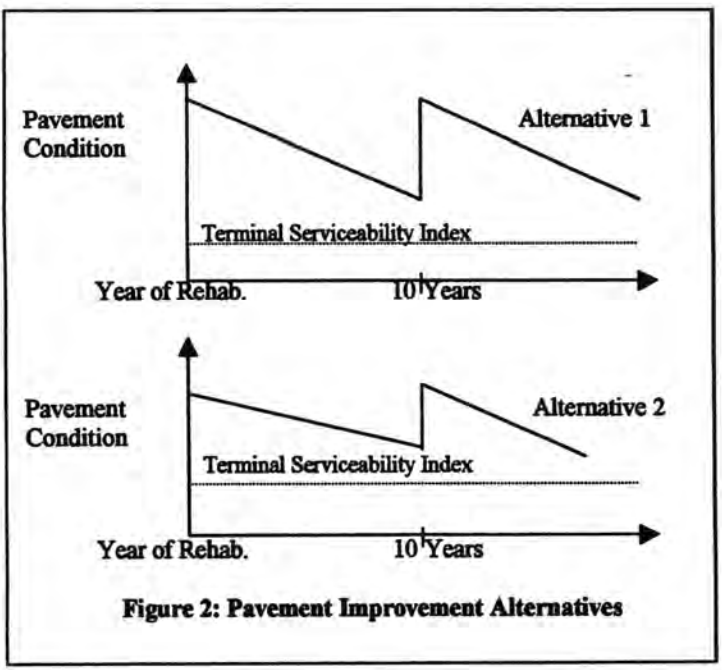


Figure 2: Pavement Improvement Alternatives

Since commercial principles of asset management can be applied to a toll road, it is feasible to maintain the infrastructure in an optimal fashion. The methodology for infrastructure management is reasonably well developed in terms of forecasting future deficiencies, identifying a list of improvement options, estimating the costs and benefits of implementing each improvement option, selection of projects and formation of infrastructure improvement program (FHWA 1994, 1999, 2000; Khan and LaFontaine 1989).

An illustrative example of a toll road maintenance/rehabilitation is provided here. Maintenance is to be studied for a 10-km electronic toll route for a 10-year period. The methodology to be developed treats risk in the form of Monte Carlo Simulation (Table 1). The inputs are to be treated as probabilistic and are assumed to be normally distributed. It is assumed that the standard deviation is 5% of mean value.

Table 1: Features of Monte Carlo Simulation Model

- Can be used to systematically address uncertainties involved in input data and output results.
- Input variables are described by using probability distributions
- Samples are randomly drawn from input distributions to calculate results
- Numerous samples may be drawn to form a distribution of results: Expected Net Present Worth, Standard Deviation, P(NPW<0)

The traffic is normally distributed. The mean traffic level on the toll road is 25 million vehicles/year and the st. deviation is 1.25 million vehicles/year (i.e., 5% of mean). For 10 kms distance, the mean vehicle-kms per year amount to 250 million.

A 10 year rehabilitation cycle is planned. Please see Figure 2. The economic analysis is to be carried out in constant dollars for a 10 year period. The discount rate is 10% (real).

Two rehabilitation alternatives are to be analyzed. The following are the inputs and results for Option1. Estimates of rehabilitation costs are $2.5 million/km in constant dollars. Therefore, the initial cost is $25 million. The standard deviation is 5% of the mean value.

On the basis of the past experience, the toll road authority expects that toll revenue of $0.20/vehicle or $5 million/year (mean value) can be allocated to the rehabilitation/maintenance account. The standard deviation is $0.01/vehicle.

The yearly maintenance expenses are expected to be $0.02/vehicle or $0.5 million/year. The st. deviation is $0.001/vehicle or $25,000/year.

The management wished to take into account the salvage value, which is estimated to be 10% of the initial investment or $2.5 million and st. deviation is $0.125 million.

At the outset, by using the mean values of initial cost, yearly maintenance, salvage value, and revenue assigned to the rehabilitation/maintenance activity, a deterministic analysis is carried out. The following results are obtained:
Net Present Worth (NPW) = $3.614 million
Equivalent Annual Worth (EAW) = $0.588
Internal Rate of Return (IRR) = 13.15%
PW Index or Return on Investment (ROR) = (NPW/Investment)x100= 14.5%.

Next, a Monte Carlo simulation was carried out with 100 runs and the following results are obtained.
NPW Mean = $3.388 million
St. Deviation = $0.231 million
P(NPW<0) = 0.07

By way of sensitivity test, instead of using st. deviation as 5% of the mean value for the random variables, 10% is assumed. Monte Carlo Simulations (100 runs) are made and the following results were obtained.
NPW Mean = $2.623 million
St. Deviation = $4.853 million
P(NPW<0) = 0.29

A second option is under consideration which has a lower initial cost but would incur higher yearly maintenance costs.

The Option 2 inputs are as follows:
Initial cost (mean) = $22.5 million
St. deviation of initial cost = $1.125 million
Revenue to be allocated for rehabilitation/maintenance activity (mean) =$0.20/vehicle
St. Deviation of revenue allocated for rehab./maintenance activity = 0.01/vehicle
Maintenance expenses (mean value) = $0.04/vehicle
St. deviation of maintenance expenses = $0.002/vehicle
Salvage value estimate is 10% of the mean value of the initial cost and standard deviation is 5% of salvage value.

For a 10 year analysis period and 10% (real) discount rate, the results of the deterministic analysis are as follows.
NPW Mean: $2.946 million
EAW: $0.479 million
IRR= 12.9%
PW Index: 13.1%

The results of the Monte Carlo Simulation Model are shown next.

NPW Mean: $ 2.539 million
St. deviation: $2.391 million
P(NPW<0) = 0.14

The Expected NPW for Option 1 is higher than for Option 2. The probability of a negative NPW is higher for Option 2 than for Option 1. The NPW and PW Index for Option 1 are higher than Option 2. Therefore, the management should select Option 1.

Given a high level of NPW, the management can afford to lower revenue allocation to rehabilitation and maintenance works and still cover the costs. A sensitivity analysis can be carried out to find the breakeven point.

Since the rehabilitation options are assumed to cover the entire toll route and funds are available for the activity, the program consists of one project, to be implemented at the optimal time.

Transfer to Government

The time frame for the operation of a toll road is normally specified at the time of planning the facility and the study of the feasibility of the scheme. The applicable public authority (i.e. government) may decide to extend the toll operation for additional years. On the other hand, a toll road may be turned into a facility without tolls. For example, the toll system in Oslo is expected to be terminated by 2007. According to the toll authority, since the main idea of the Oslo Toll Ring was to fund new infrastructure, this mission will be complete by 2007 (Fjellinjen 2000). According to another source, extending the status of toll cordon would yield further revenue for required road projects and also support the long-term urban transportation policy intended to relieve congestion (Larsen 1995).

CONCLUSIONS

1. Toll roads appear to be successful in terms raising funds for road construction and maintenance and also in reducing traffic congestion in and around major urban areas.
2. If toll roads in urban areas are to be made permanent, or operated for a considerably longer period than initially planned, these could be integrated in a long term urban transport policy designed to restraint automobile traffic and relieve congestion.
3. The concepts and methods for planning, operations and maintenance applicable to toll roads are different than those for non-toll roads. An example of the decision analysis regarding rehabilitation/maintenance is presented.
4. In the feasibility study and in the maintenance methodology, it is beneficial to treat risk by using the Monte Carlo Simulation technique.

ACKNOWLEDGEMENTS

This paper is based on research sponsored by the Natural Sciences and Engineering Research Council of Canada (NSERC). The opinions are those of the authors.

REFERENCES

407 International Inc. (2001), "407 ETR", Toronto, Webpage.

Canadian Civil Engineer (2000), "Highways: A Road Too Heavily Travelled", V.17, No.5, Sept./Oct. 2000, p.13.

Federal Highway Administration (FHWA) (1994), "Life Cycle Analysis", U.S. DOT, Publication No. FHWA-PL-94-025.

Federal Highway Administration (FHWA) (1999), "Asset Management Primer", U.S. DOT, FHWA Office of Asset Management.

Federal Highway Administration (FHWA) (2000), "Asset management: Preserving a $1 Trillion Investment", FOCUS, May 2000, U.S. Department of Transportation.

Fjellinjen AS (1999), "Annual Report 1999", Oslo, Norway.
Federal Highway Administration (FHWA) (1994), "Life Cycle Cost Analysis, Searching for Solutions, A Policy Discussion Series, U.S. DOT, Publication No. FHWA-PL-94-025.

Fjellinjen AS (2000), "Toll Road System in Oslo", Oslo, Norway.

Forumation (2000), "Bridging the Gaps at CTRF 2000, Transportation Commercialization: How is it going?" July 2000, p.1.

Khan, A.M. and P. LaFontaine (1989), "The Management of Highway Infrastructure Rehabilitation: A Forward Look", Proceedings, World Conference on Transportation Research, Vancouver, Volume 1.

Land Transportation Authority (2000), "Automation of Toll and Vehicle Entry Permit Fees Payments", Singapore (Webpage).

Larsen, Odd, I (1995), "The Toll Cordons in Norway: An Overview", Journal of Transportation Geography, V.3, No.3, pp.187-197.

Transportation Association of Canada (1997), "Pavement Design and Management Guide", Ottawa.

User Charges and Cost Recovery: The State of Idaho/American Trucking Association Case

James R. Jones, Professor
Selahattin Bekmez, Post Doctoral Research Associate
Department of Agricultural Economics and Rural Sociology
University of Idaho
P.O. Box 442334
Moscow, ID 83844-2334

Abstract:

The paper discusses the Litigation outcome of a case between the State of Idaho and the American Trucking Association regarding Idaho's weight distance truck user fee system. A study contracted by the Idaho Transportation Department to the National Institute of Advanced Transportation at the University of Idaho to analyze a replacement system based upon a registration fee structure will be discussed. Economic efficiency and equity issues will be addressed. The study is in the beginning of the first year of a three-year frame so the objective of the authors will be to stimulate discussion and explore with other participants their experience with similar issues.

User Fees and Road Financing in the U.S.

User based transportation taxes in the US were pioneered in Oregon in 1919 with a one-cent per gallon gasoline tax (Locklin, p. 657). Within twenty years every state in the Union had a gasoline tax. Before 1956 when the Highway Trust Fund (HTF) was set up as part of the system to finance the federal Interstate System, taxes on motor fuels and automobile and related products existed but they were not linked to funding for highways. The public finance principle that reigned was ***"spend where you must, and get the money where you can"*** (US Department of Transportation Federal Highway Administration Office of Policy Development, p. 31). In this discussion we will designate this as the **"Get"** principle. Another tax rule or principle is **"tax that guy behind the tree not me"** which we call the **"Not Me"** principle. Funding road construction and maintenance in this case often entails using general funding sources such as income, sales, and property general taxes. The link between the user and the party paying taxes is nonexistent or vague at best.

The Federal-Aid Highway Act of 1956 and Highway Revenue Act of that same year authorized significant funding for the Interstate System based upon a system of user taxes earmarked for that purpose

and henceforth simply referred to as taxes administered on the **"Benefit"** basis. Oregon in keeping with its tradition of user funded highway finance also pioneered a weight mile tax on heavy trucks that takes into account not just vehicle size, weight and annual miles traveled but also the number of axles. This system of taxation is based on the philosophy **"those who benefit directly from public roads should pay for them in proportion to the costs for which they are responsible"**. Economists argue that **"getting the price right"** to accomplish this principle works best when prices are associated with marginal rather than average costs and also take into account external societal costs as well as direct private costs. User fees allocated on this principle will be simply described as **"Optimally Social Priced"** fees or "PRICE."

Idaho's Weight Distance Tax on Trucks

In the early 1990's the major Idaho taxes and fees on highway users consisted of a fuel tax of 21 cents per gallon on gasoline, diesel and other special fuels, registration fees, and ton-mile or weight distance taxes on trucks with registered weights over 60,000 pounds (SYDEC, et. Al., p. 2-11). It is this latter tax on vehicles above 60,000 pounds that is the subject of a weight distance interdisciplinary (economist and engineers) study we are beginning at the University of Idaho in the National Institute of Advanced Transportation Technology (NIATT) with support from the Idaho Transportation Department. As a result of a recent court decision registration fees have assumed a larger role in Idaho's heavy vehicle (over 60,000 pounds) user tax structure by being merged in with the weight distance fee structure.

As reported in a highway cost allocation study prepared by SYDEC, Inc. in association with Cambridges Systematics, Inc. and R.D. Mingo and Associates (February. 1994) for the Idaho Transportation Department, Idaho's taxes on semis trucks were the second highest in the region behind Oregon and this was in part due to the two states both having weight distance taxes on heavy weight commercial vehicles. In 1951 Idaho implemented a weight distance tax on trucks with fees scheduled by type of operation, declared GWT, and fuel

type. The tax was calculated on a mil per mile basis. In 1957 the tax system was modified so fees were based on combined weight of maximum cargo capacity and truck weight or GVW. In the version in effect in the 1990's a ton mile tax was collected on trucks with registered weights of 60,000 pounds with rates ranging from approximately 3 to 4.5 cents per mile plus a 0.21 cents per ton mile surcharge for vehicles registering over 80,000 pounds. A separate schedule for vehicles domiciled in Idaho hauling selected wood, agricultural and mineral products provided a reduced rate of 2.245 cents per mile.

In 1999 the American Trucking Association (ATA) filed a lawsuit against the state of Idaho claiming that the two-tiered mileage-use fee schedule was an infringement on interstate commerce and was unconstitutional. The Fourth Judicial District of Idaho ruled in favor of ATA. The court ruled that one tier of the two-tiered mileage use fee schedule gave Idaho trucking firms an unfair advantage over companies based in other jurisdictions. The court ordered the state to discontinue, effective April 1, 2000, the mileage-use fee schedule known as the limited commodity schedule. The limited commodity schedule applied to the following types of transportation: Farm vehicles, and any commercial vehicle engaged in the transportation of logs, pulp wood, poles, piling, rough lumber, ores, ore concentrates, sand and gravel aggregates in bulk, livestock and vehicles used for the sole purpose of transporting milk from the farm to processing plant, having a maximum gross weight in excess of sixty thousand (60,000) pounds.

The mileage-use fee rate increased for all owners that operate vehicles partially or wholly in one of the transportation activities listed above. The rate schedule in appendix A shows the rate change per mile as of April 1, 2000 and a successor registration fee structure that became effective as of October 1, 2000. This Latter schedule, which merges a registration structure into the weight distance tax, also lowers the effective tax burden per ton mile for vehicles covering more miles than vehicles covering short distances.

In light of the chaos created by the state's legal defeat, it is probably fair to say that the state is now in the adhoc position where funding of it's road system is on the *spend where your must and get money where you can* **(GET)** foundation while various vested interest or stakeholder groups are going to frequent the principle *tax that guy behind the tree* **(NOT ME)**. The next section will address how user fees can be used as a pricing mechanism to direct road construction, improvement, and maintenance activities on the basis of economic efficiency **(PRICE).**

Several economists have argued that the best way to economize on maintaining and using an existing highway system is to charge user fees commensurate with the actual cost each user imposes on society through the user's effect on the road's condition and on the speed that others can travel (Small, Kenneth A., Clifford Winston, and Carol Evans, 1989.). McMullen pointedly states, "motor carriers need to 'bite the bullet' in the short run and pay user fees which reflect the damage they do to the roads if the industry expects to have a highway system that will meet future demands." (p. 148). Economists know well that having the user pay helps assure ***"transportation is discouraged which is not worth its cost." (Locklin, p.658.).***

In the private sector pricing mechanisms signal value and scarcity of resources. In the same sense user fees in the public sector can promote efficiency in resource allocation and discourage the wasteful use of public goods. By levying a certain charge for identifiable beneficiaries, the public authority can limit the demand for that good. In the context that those who benefit pay the cost, user charge financing is intuitively a "fair" method of financing government services including highways and roads.

On a theoretical basis, user charges are a very attractive type of taxation because the payers are the ones that use a particular public service, not the whole community. An example for this is the excise on gasoline. The income generated with this excise can be assigned to a highway trust fund earmarked to be used solely for the purpose of building new roads or improving the existing roads. Unfortunately fuel use often is a poor proxy

for benefits and costs of highway use. Besides it is probable the flat rate per gallon fuel tax will become a diminishing source of revenue to fund road construction and maintenance (See Samuel and Litman, p. 25.). Concern about global warming will discourage fuel consumption plus taxes applied to fuel consumption to cover perceived externalities of its use will likely continue to go to purposes other than road construction and maintenance.

Applying a user charge equal to the actual cost of using a road is a way to finance an existing road. With this method each user will be charged according to his effect on the road's condition. Conceptually such a charge could guide the independent decisions of each user within a framework of socially optimal pricing. The principle for optimal pricing and investment policies require capacity and durability of roads to be produced to the point where the marginal benefit from increasing investment in each dimension equals its marginal cost (Winston, 1991, p.115.). Unfortunately determining this pricing level is not always that easy due to unidentifiable social costs.

Another way to determine a pricing level is to use the Ramsey method (Boyer, 1998). Under the Ramsey method all highway user groups pay something toward the cost of the system, but those with lower demand elasticity would be charged the highest markups over marginal cost. Mathematically this can be expressed as:

$$(P_i - MC_i)/P_i = \mu/\varepsilon_i$$

where P_i is the price charged to user i; MC_i is the marginal cost caused by user i; μ is a constant whose value is determined according to target revenue to be reached, and ε_i is elasticity of demand for each user. The problem of measuring costs, specifically marginal cost (MC_i), and demand response elasticity parameters (E_i) is still daunting to say the least. But conceptually these are helpful ways to proceed.

There are two primary considerations in pricing the use of existing roads: (i) to finance road-wear cost and (ii) to finance congestion

cost. Not to demean the importance of the latter issue, road congestion in the relatively rural state of Idaho is not the issue as much as the lack of roads that are safe and suitable to accommodate heavy truck traffic. The road-wear cost includes routine and major maintenance costs caused by operators of both cars and trucks. The dominant element of maintenance cost is the periodic overlay when a pavement becomes worn. (Small, et al., 1989). It is generally recognized that wear and damage on the roads is exponentially related to axle loads. A fully loaded truck may cause approximately 10,000 times more damage than a car (Abdulghani, Arora, Balghunaim, Bayomy and Pearson-Kirk, p.18). For that reason federal highway user fees on heavy vehicles are tending to increase. Taxes on various highway fuels, excise taxes on sale of heavy trucks, taxes on tires weighing more than 40 pounds and heavy vehicle use taxes on trucks are some examples.

In many cases user fees do not cover the full cost of public services but, in some cases, very high user fees are imposed by government to cover the full cost of the program. However, the "cross-subsidy" problems may still remain. A cross-subsidy problem occurs if one party or group pays more than its share while the other party or group pays less than its share. Cross-subsidies are common in highway financing. For example, it has been argued that federal highway taxes are constructed in a way that operators of lighter trucks pay more than their fair share. In 1982, congress increased truck taxes by 55 percent, but spending increased by 69 percent. Thus, subsidies for larger trucks still remained (U.S. Congressional Budget Office, 1993).

Precision in social pricing is difficult if not impossible. Delucchi at the Institute of Transportation Studies, University of California, Davis cautions that "getting the price right" in transportation may not always be feasible because of the difficulties of estimating and implementing the right price. Some of these difficulties have already been eluded too. While he cautions embracing marginal cost pricing unreservedly he still recommends using this kind of analysis to inform political debate (p.21). Even the relationship between road damage and truck weight remains a contentious issue. Some analysts

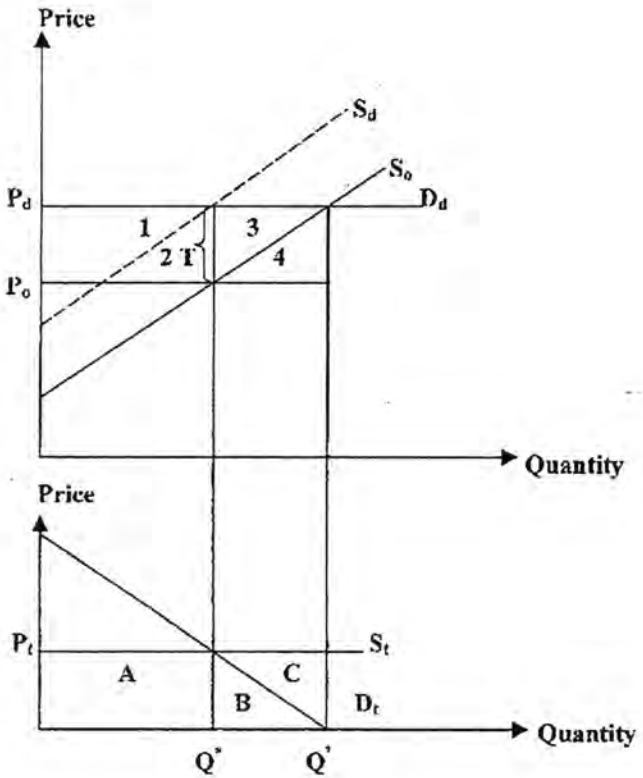
from trucking associations claim that a substantial part of the maintenance costs is due to age and weather. Thus, they claim that the extent of these maintenance costs invalidates the argument of exponential relationship between road damage and heavy axle loads. However, aging and weathering makes the roads more vulnerable to get damaged by heavy loads (Small et. al, 1989) and in the relatively recent Federal Highway Cost Allocation Study it acknowledges the need to allocate costs taking into consideration that combination trucks are responsible for over 40 percent of system preservation and new capacity costs .

A cost allocation study conducted for the state of Idaho noted that on a per mile basis heavy combination vehicle rates are 7.3 times those of autos and concluded heavy trucks were taxed unequitably (SYDEC, Inc. et. Al., p.2-17). This relation between vehicle weight and pavement damage is obviously a pivotal point of contention that cannot be ignored if an objective analysis of the state's revenue structure is to be performed.

One of the possible sources of funding of roads is from general funds such as income, sales, property and other taxes. Using user fees results in economic efficiency concerning resource allocation while using general funds to finance highways causes misallocation of public and private resources. The problem with this system from an economic efficiency framework is that it in fact operates as a subsidy that can create a misallocation of resources not just in terms of transportation expenditures but other resources as well. A partial equilibrium welfare illustration can highlight how this is so.

Assume we are transporting a commodity or product to market. Supply at the point of origin ($\mathbf{S_o}$) and demand at the market destination ($\mathbf{D_d}$) are shown in the upper panel of figure 1. The vertical distance between the price suppliers at the origin are willing to provide the commodity or product from $\mathbf{S_o}$ and the price purchasers are willing to pay from $\mathbf{D_d}$ provide the basis of deriving the demand for transportation ($\mathbf{D_t}$) in the lower panel. Transportation services are embodied in the supply function ($\mathbf{S_t}$) in the lower panel.

FIGURE 1


Price
S_d
S_o
P_d
D_d
1
2
T
3
4
P_o
Quantity
Price
P_t
S_t
A
C
B
D_t
Quantity
Q^*
Q'

In this stylized illustration we assume the producer/shipper/carrier does not move enough cargo to influence price. In Idaho which is a small population state and focusing on agricultural, forest, and mineral products that are integrated into the global economy this is a realistic assumption, although we invoke it mainly to simplify the analysis and reduce data demands later in our modeling efforts. To further simplify the illustration we assume away other transportation and marketing charges and focus exclusively on road infrastructure.

Assuming market clearing conditions are fulfilled, if a user fee is charged for the road service, Q* of the commodity will be produced and delivered. If the funds are raised from a general revenue source independent of road utilization costs, Q' of the commodity or product will be shipped. The latter will result in a net social welfare loss represented by area 4 (=C).

Nonuser taxes such as property taxes used by local government units for road improvement and maintenance can sometimes be justified on a benefit received basis. Since the value of property may be positively correlated to access and quality of roads it could be argued that property taxes are an equitable source of funding for local road and street maintenance. However there is little correlation between actual transportation mode that travels on roads to or from the property and the amount of property tax typically allocated. Specific user fees such as tolls and weight distance taxes can much more effectively target vehicle class in terms of that classes' cost responsibility.

Finally a concern that must be addressed if the state of Idaho is to find its way back to an acceptable system of funding road and highway construction, improvement, and maintenance is the amount of revenue raised and who will bear the burden of the fee or tax. To address these questions an understanding of the elasticity of (derived) demand for road service relative to supply will be required.

We of course have a lot of work ahead of us and at this point we have already made some commitment on how we will proceed. Partial

equilibrium spatial comparative static analysis and modeling will be employed in conjunction with formal and informal surveying to examine and analyze the impacts of the conversion or alternation of Idaho's weight distance tax selecting wheat shipments as a case study. A quadratic programming model developed at the University of Idaho, the *Pacific Northwest Wheat Shipment Hinterland* (PNWSHIP) model (Jones, J.R., Lu Qu, K. Casavant, and Won Koo, 1995) will be revised to simulate the effects of the tax conversion on volume and value of Idaho and foreign shipments, incidence of the tax on producers and truck operators, tax revenue generated, and the mode (truck, rail, or water) of shipment. Wheat is a convenient prototype commodity due to data availability and the fact that low value bulk wheat is likely to be one of the most sensitive cargo categories to be impacted by the tax change. The ability to state social welfare costs in dollar terms is one of the advantages of the quadratic programming modeling exercise that we will be doing. It is developed to handle multirouting of hinterland shipments as they interface with an integrated global market framework. Also modeling exercises in the study will have the ability to address the implications of different tax rate structures and levels for intermodal competitiveness of alternative modes of transportation.

Conclusion

Now the question is what is our purpose of bringing this issue to the Canadian Transportation Research Forum. The answer is that we sense the Canadians are at the cutting edge in the area of transportation economics and of course we are hoping we can learn from you. Specifically we would be interested in learning what you encountered in your experience and research related to this general area of weight distance user fees as a mechanism for funding roads. Years ago (decades actually) I taught a course in the economics of state and local taxation and I also taught a course in comparative economic systems. My colleague with me today signed on as a post doc research associate to add to his experience with public finance. While I bring very limited experience to taxation I have always found comparative analysis to be rewarding so we are looking to our

Canadian colleagues for something to compare to. Berry Prentice has already called our attention to a study conducted at the University of Manitoba Transportation Institute (Derkson and Shurvell) and we have shared some of our joint antidotal evidence on the economics of containerized shipping. We hope when we complete our study we in turn will have results that you will find useful as you look at your own funding.

With NAFTA it is imperative that economists on both sides of the border become better informed of what is happening in the area of transportation finance. "Fair" trade advocates on our side think they are already well versed in some of those different (therefore by inference unfair??) ways the Canadian's finance their transportation of exports, etc. and perhaps some of your folks have noticed some peculiar issues about how we subsidize our shallow draft and deep draft waterborne transportation-*eh*?

REFERENCES

Abdulghani, Khalid, Madan Gopal Arora, Fahd Balghunaim, Fouad Bayomy and Donald Pearson-Kirk. "Taxation Model for Road Vehicles in Saudi Arabial. *Transportation Research Record 1305.* 1991, pp 18-25.

Boyer, K.D. *Principles of Transportation Economics.* Addison-Wesley Longman, Inc., 1998.

Congress of the United States Congressional Budget Office. *Charging for Federal Services,* December, 1993.

Delucchi, Mark. "Should We Try To Get The Prices Right?'. *Access.* No. 16, Spring 2000, pp. 14-21.

Jones, J.R., Lu Qu, K. Casavant, and Won Koo. "A spatial equilibrium port cargo projection model." *Maritime Policy and Management.* Vol. 22, No.1, 1995, pp 63-80.

Locklin, Philip D. *Economics of Transportation.* Homewood, Illinois: Richard D. Irwin, Inc., 1954 (4th Edition).

McMullen, B. Starr. "The US Motor Carrier Industry at the Millennium." *Journal of the Transportation Research Forum.* Vol. 39, Number 4 (Fall, 2000).

Samuel, Peter and Todd Litman. "Optimal Level of Automobile Dependency: a Transportation Quarterly Point/Counterpoint Exchange." *Transportation Quarterly.* Vol. 55, No. 1, (Winter, 2001).

Small, Kenneth A., Clifford Winston, and Carol Evans. *Road Work: A New Highway Pricing and Investment Policy.* Washington, D.C: The Brookings Institution, 1989.

SYDEC, Inc., Cambridge Systematics, Inc. and R.D. Mingo and Associates. *Highway Cost Allocation Study*. Final Report submitted to Idaho Transportation Department, February, 1994.

U.S. Department of Transportation Federal Highway Administration. *1997 Highway Cost Allocation Study Summary Report*. August, 1997. http://www.ota.fhwa.dot.gov/hcas/summaary/97fhcas.html.

U.S. Department of Transportation Federal Highway Administration Office of Policy Development. *Highway Trust Fund Primer*. November 1998, p. 31.

Winston, Clifford. "Efficient Transportation Infrastructure Policy." *Journal of Economic Perspectives*. (Winter, 1991) pp. 113-127.

APPENDIX A

Maximum Gross Weight	Current Mill Rate Per Mile	Mill Rate Per Mile Beginning 4/1/2000	Mill Rate Per Mile Change
60,001-62,000	22.45	30.05	7.60
62,001-64,000	22.45	31.35	8.90
64,001-66,000	22.45	32.60	10.15
66,001-68,000	22.45	33.90	11.45
68,001-70,000	22.45	35.15	12.70
70,001-72,000	22.45	36.40	13.95
72,001-74,000	22.45	38.55	16.10
74,001-76,000	22.45	40.65	18.20
76,001-78,000	22.45	42.75	20.30
78,001-80,000	22.45	44.90	22.45
80,001-82,000	24.55	47.00	22.45
82,001-84,000	26.65	49.10	22.45
84,001-86,000	28.75	51.20	22.45
86,001-88,000	30.85	53.30	22.45
88,001-90,000	32.95	55.40	22.45
90,001-92,000	35.05	57.50	22.45
92,001-94,000	37.15	59.60	22.45
94,001-96,000	39.25	61.70	22.45
96,001-98,000	41.35	63.80	22.45
98,001-100,000	43.45	65.90	22.45
100,001-102,000	45.55	68.00	22.45
102,001-104,000	47.65	70.10	22.45
104,001-106,000	49.75	72.20	22.45

Source: Idaho Transportation Department; The table above shows the impact of elimination of the limited-commodity rate schedule that went in effect as of April 1,2000. The mileage-use fee was previously one schedule for all owners.

Senate Bill 1580

With passage of Senate Bill 1580 a registration schedule (see below) replaced the above milage-use fee on October 1, 2000.and the Fourth Judicial District Court approved the settlement of ATA vs. the State of Idaho. As a part of the ATA lawsuit settlement, the Legislature replaced the mileage-use fee with a registration fee schedule. The new schedule intends to replace the same amount of highway revenue generated by the mileage-use fee.

Senate Bill 1580 went into effect on October 1, 2000. After October 1, 2000, owners registering vehicles over 60,000 pounds gross weight will use the following fee schedule:

Weight of Vehicle (Pounds)	Total Miles Driven 1-7,500	Total Miles Driven 7,501-50,000	Total Miles Driven Over 50,000
60,001-62,000	$210	$ 840	$1,400
62,001-64,000	240	960	1,600
64,001-66,000	270	1,080	1,800
66,001-68,000	300	1,200	2,000
68,001-70,000	330	1,320	2,200
70,001-72,000	360	1,440	2,400
72,001-74,000	390	1,560	2,600
74,001-76,000	420	1,680	2,800
76,001-78,000	450	1,800	3,000
78,001-80,000	**480**	**1,920**	**3,200**
80,001-82,000	495	1,980	3,300
82,001-84,000	510	2,040	3,400
84,001-86,000	525	2,100	3,500
86,001-88,000	540	2,160	3,600
88,001-90,000	555	2,220	3,700
90,001-92,000	570	2,280	3,800
92,001-94,000	585	2,340	3,900
94,001-96,000	600	2,400	4,000
96,001-98,000	615	2,460	4,100
98,001-100,000	630	2,520	4,200
100,001-102,000	645	2,580	4,300
102,001-104,000	660	2,640	4,400
104,001-106,000	675	2,700	4,500

ITEM-0:
Integrated Transportation and Economy Models, Version 0

1. Introduction to ITEM Studies

Gaudry, M., Agora Jules Dupuit (AJD) du Centre de recherche sur les transports (CRT), Université de Montréal, Montréal, et Université Louis Pasteur, Strasbourg

We provide some international and local context for the development of ITEM Studies. General motivations, that echo those of the Canada Transportation Act, 1996, Section 52, include (i) the desire to account analytically and credibly for the interaction between transportation and the economy; (ii) the desire to develop Kyoto compatible policies that can be justified in front of foreign experts; (iii) the practical necessity to base transport policies on modal accounts describing clearly the direct costs or revenues—and the subsidies or surpluses—generated by various user classes. We discuss the four different types of accounts in use for these purposes. We mention new methodological developments in the valuation of externalities (pollution, human life) that is crucial to full direct costing of resources. We introduce the structure ITEM-0 developed by researchers of three universities in Montreal and Ottawa collaborating in a research project of Agora Jules Dupuit of CRT.

2. GET-0: A Computable General Equilibrium Model for Transport in Canada

Mérette, M., University of Ottawa,
Mercenier, J., Université Cergy-Pontoise et
Centre de recherche sur les transports (CRT)
Harvey, S., Université d'Ottawa
Karangwa, E., Carleton University

A multi-sector multi-region computable general equilibrium (CGE) model is built to address issues related to the linkages between transportation (of commodities) and other sectors of the economy and to analyze the inter-sectoral and inter-regional effects of government policies. The model emphasizes the inter-linkages between the different modes of transport with the rest of the economy. Transportation services offered by the different modes are used by firms in the model as input of productions and to deliver goods to the customers. There is one representative agent in each region and one representative firm by sector and region. Sectoral and regional goods are considered imperfect substitutes. Hence the Armington assumption is used to capture the inter-regional and inter-sectoral trade flows. The model is calibrated on Canadian data. Simulation scenarios are performed to analyze inter-modal, inter-sectoral and inter-regional effects generated by various government policy options. In particular fiscal and general taxation policy, as well as environmental policy regarding the transportation sector in Canada are investigated.

3. MAST-0: A Modal Accounting System for Transport, with first national and regional results

Gaudry, M., Centre de recherche sur les transports (CRT)
Laferrière, R., Centre de recherche sur les transports (CRT)
Préville, E., Centre de recherche sur les transports (CRT)
Ruest, C., Centre de recherche sur les transports (CRT)

In 1992, the Royal Commission on National Passenger Transportation developed a seminal set of direct cost and revenue network accounts for the intercity Canadian networks. These accounts presented costs, including externality costs, of modes sharing a common infrastructure. We have reproduced, corrected, parametrized and applied in more general fashion these seminal cost and revenue allocation procedures. Results for intercity road and rail by user class, submitted to the Canadian Transportation Act Review, can be found at (//www.crt.umontreal.ca/crt/agoraJulesDupuit/). These national and other results for a Province of Canada are expected to be presented, including results for the marine mode.

4. Analysis-0: Some ITEM system Tools

Gaudry, M., Centre de recherche sur les transports (CRT)
Marullo, C., Centre de recherche sur les transports (CRT)
Nikolajev, R., Digital Shape Technologies, Montreal
Tran, C. L., Centre de recherche sur les transports (CRT)

We briefly introduce two ITEM-0 utilities. The first, ESAL-0, which allows the flexible computation of Equivalent Single Axle Loadings from large samples of vehicles, such as those found in Canadian Council of Motor Transport Administrators' (CCMTA) surveys, that are not conveniently handled by EXCEL programs: this procedure was used to generate inputs for MAST-0. The second, MOM-BC, a regression procedure that allows the user to calculate for any classical regression model, even in the presence of Box-Cox transformations, the marginal rates of substitution (or Allais trade-offs) among moments of the dependent variable. The method was recently used to produce values of human life derived directly from Quebec road driver accident behaviour, for a submission requested in November 2000 by the French (Boiteux) Commission on transport investment criteria. The core results are presented, but not used in MAST-0.

The Development of Western Grain Rates

Jim Riegle, Canadian Transportation Agency

This paper presents a description of the regulatory regimes since August 1, 1984 relating to the development of rates for the movement of western grain, by rail. Following a brief historical review, the major emphasis will be directed to the current regime which impacts rate development beginning August 1, 2000.

To be subject to rate regulation, western grain movements must originate in western Canada and be:

> destined to export markets* through West Coast ports, Thunder Bay, Armstrong or Churchill, or
>
> destined to domestic eastern markets through Thunder Bay or Armstrong.
>
> * excluding US destinations through west coast ports

The grain must also move by a prescribed railway company and must qualify as an eligible grain, as identified in legislation.

Rate Development Prior to August 1, 2000

The interval August 1, 1984 to July 31, 2000 encompasses 16 crop years and involved three different rate-setting regimes. The first rate-setting regime was established under the Western Grain Transportation Act (WGTA) and lasted for 11 years. The second regime was brought in by the Budget Implementation Act, 1995 and lasted only two years; the first year under the National Transportation Act and the second year under the Canada Transportation Act (CTA).

The third rate-setting regime was brought in by the Budget Implementation Act, 1996 and lasted for three years, all under the CTA.

Despite the different rate-setting regimes over this 16-year interval, a number of elements were common to the regimes: rates were set annually in advance of the crop year and at maximum levels on a per tonne basis; rates were mileage based and consequently, for a given distance, did not vary by railway, origin or destination, routing or by commodity; and annual rate adjustments accounted for forecasted inflation (with the exception of one year, 1995-96) and the abandonment of services on Grain Dependent Branch Lines (GDBL).

The 11-year WGTA rate-setting regime had three major, unique characteristics. The first involved a shipper subsidy, referred to as the Government Commitment, which was paid by the Federal Government to the railways as grain traffic moved. The amount of the subsidy averaged about $725 million between crop years 1986-87 and 1992-93 but was scaled down to about $560 million for the WGTA's last effective crop year, 1994-95.

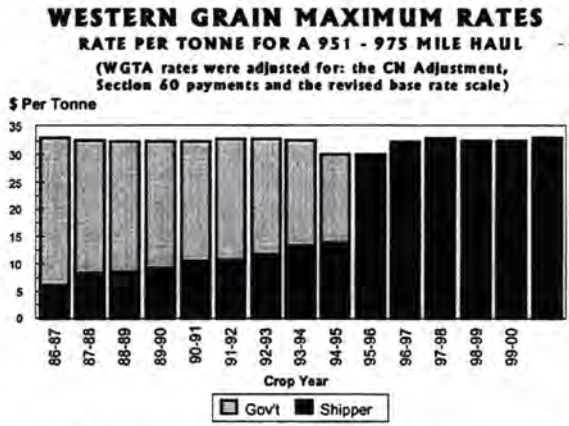
The second unique WGTA rate-setting characteristic was its Quadriennial Costing Reviews. These reviews, along with a number of component studies, determined the estimated eligible costs for western grain movements for a given calendar year. These costs included the volume-related variable costs, the grain-dependent branch line costs, and constant costs - set by the WGTA as 20% of the volume-related variable costs. The estimated eligible costs are sometimes referred to as economic costs since they include the costs of capital and allow for a fair return on equity.

The estimated eligible costs from the most recent costing review were used to develop rates for the four following crop years. Hence, the WGTA rates were cost-based rates. During this four-year interval, cost savings achieved through productivity gains accrued to the

railways, who shared a portion with shippers through lower tariff rates. But the subsequent costing review captured these productivity gains in establishing a new set of base year costs, to be used for rate development for the next four years.

The third unique WGTA rate-setting characteristic was its protective provisions. The Shipper Share Limitation Adjustment ensured that shippers were protected from their freight rates becoming too high relative to the price of grain. Another element protected shippers from increases in the railway freight cost per tonne that exceeded 6%.

As can be seen in chart below, under the WGTA rates for a 951-975 mile length of haul remained relatively stable - around $33 per tonne, until crop year 1994-95 when the impact of revised costs from the 1992 Costing Review lowered rates by about $3 per tonne.



The repeal of the WGTA had two major consequences. The first, the elimination of the government subsidy, was felt immediately as shippers suddenly paid 100% of the freight rate. For a typical grain movement, shipper rates rose by about $16 per tonne. As a partial

offset to the loss of subsidies, a $1.6 billion payout program and a $300 million adjustment fund, was implemented through Agriculture Canada. The second major consequence of the repeal of the WGTA, which drew considerable attention several years later, was the elimination of the costing reviews.

Two rate-setting regimes applied in the five years following the repeal of the WGTA. The Budget Implementation Act, 1995 imported a new regime into the NTA, for crop years 1995-96 and 1996-97. The new regime simply allowed for the annual adjustment to rates to account for inflation and the abandonment of services on GDBL's. Crop year 1995-96 was an exception, as rates were effectively frozen at 1994-95 crop year levels.

A revised rate-setting regime was imported into the CTA by the Budget Implementation Act, 1996. It applied to three crop years; 1997-98 to 1999-2000. In effect, the regime was the same as that brought in by the Budget Implementation Act, 1995. However, had the sale of 10,000 or more government-owned hopper cars occurred, the GDBL adjustment would have been eliminated, a partial productivity adjustment would have been incorporated, and an additional $0.75 per tonne charge would have applied.

In December, 1997 the Minister of Transport appointed Justice Willard Estey to conduct a forward-looking review of Canada's grain handling and transportation system. After a year's work and discussion with all facets of the industry, Justice Estey made 15 recommendations, one of which called for a revised rate-setting regime. Under the proposed regime, railways would have freedom to set their own rates but would guarantee lower revenues per tonne, and higher shipper savings, than would have resulted under the existing rate-setting regime. The Estey proposal contained what was regarded by many as a number positive factors, but overlooked the widening gap between railway costs and revenues that had resulted for grain due to the elimination of the WGTA's costing reviews.

In May, 1999, the Minister of Transport appointed Mr. Arthur Kroeger to undertake a consultation process to develop necessary legislative changes to make the system work. A number of studies were requested in support of establishing new rate-setting legislation including a CTA estimate of the 1998 costs and revenues for the movement of western grain. This study confirmed that there was a widening gap between railway costs and revenues, and indicated that the contribution to constant costs, expressed relative to variable costs, had risen to a range of 39% to 43%. Under the WGTA the contribution to constant costs was defined to be 20% for a costing review year but was generally higher - averaging about 27% , for the four years whose rates were based on a given costing review.

Mr. Kroeger ultimately recommended a rate-setting regime similar to what Justice Estey proposed, but with average, per tonne revenues subject to a 12% reduction from 1998 levels. Mr. Kroeger also recommended that steps be taken to increase competition between railways, and that the selection of specific measures be made following an assessment of alternatives by an expert panel. This assessment is currently being conducted under the CTA Review Panel.

In May, 2000 the Minister of Transport presented legislative amendments to the grain handling and transportation system to Parliament via Bill C-34. By August 1, 2000 a number of changes had taken effect concerning grain handling, including changes to the CTA's rate-setting regime discussed below.

Rate Development Commencing August 1, 2000

Overview:

As of August 1, 2000 the Agency no longer establishes rates for the movement of western grain. Prescribed railway companies, currently CN and CPR, set their own rates for grain movements subject to two conditions:

i) a single car rate from a branch line origin for the movement of the same type of grain under substantially similar conditions for a substantially similar distance from its nearest main line origin, can not be more than 3% higher than the corresponding main line rate, and,

ii) a prescribed railway company's revenues, as determined by the Agency, for the movement of grain in a crop year may not exceed the company's maximum revenue entitlement for that year, as determined under subsection 151(1) of the CTA.

It has been observed that, as of February, 2001, the railways have not exercised a number of their freedoms under the new regime. For the most part, rates are still mileage based hence, for the same distance, do not vary by origin or destination or by routing; rates do not vary by commodity; and single car branch line rates are the same as similar-condition mainline rates - thereby satisfying condition i) above. However, the railways now charge for a number of services that were not billed before, and have tightened the terms and conditions under which other charges apply.

The Revenue Cap:

The main thrust of the new rate regime stems from condition ii) above, whereby railway revenues for a given crop year are now subject to a maximum revenue entitlement, referred to as the "revenue cap". The revenue cap is derived by the formula given below.

Currently, it applies to CN and CPR since they are currently the only "prescribed railway companies". In order to determine the CN and CP revenue caps for a given crop year, railway-specific statistics are put into the formula. The only exception relates to input F, the volume-related composite price index, which is common to both CN and CP. A general explanation of the formula follows.

$$\text{Revenue Cap} = [A/B + ((C - D) \times \$0.022)] \times E \times F$$

The revenue cap formula has four components. The first component, A/B, takes the company's base year revenues for the movement of grain (A) and divides it by the number of tonnes of grain moved by the company in the base year (B). The result is the company's average base year revenue per tonne. Values for A and B, provided in Subsections 151 (2) and (3) of the CTA, are as follows:

CN: A = $348,000,000 and B = 12,437,000 tonnes;
CPR: A = $362,900,000 and B = 13,894,000 tonnes.

Upon substitution of these values the average base year revenue per tonne is $27.98 for CN and $26.12 for CPR.

The second component, (C-D) x $0.022, adjusts the above figure (A/B) in order to take into account the difference in the average length of haul (in miles) for a given crop year compared to the base year average length of haul. C is the average length of haul for the given crop year and is determined by the Agency after the crop year has ended. D is the average length of haul for the base year, set at 1045 miles and 897 miles for CN and CPR respectively, as given in subsections 151 (2) and (3) of the CTA. The $0.22 figure is the coefficient of change in economic costs, when plotted against miles. If the average length of haul for the given crop year is equal to the average length of haul for the base year, the adjustment is nil.

The third component, E, multiplies the adjusted revenue per tonne

figure for the given crop year by the number of tonnes moved by the company in the given crop year. The result is a revenue cap figure, before an allowance for inflation. When this figure is multiplied by the fourth component, F, the volume-related composite price index, the final revenue cap figure is derived for the given crop year.

Note that paragraph 151(4)(a) of the *CTA* sets the value for F to be unity for crop year 2000-2001. This does not indicate zero inflation for crop year 2000-2001, but reflects the fact that the Agency's estimated 4.5% inflationary increase for crop year 2000-2001 had already been captured during the development of the base year revenue statistics.

While the revenue cap sets a ceiling on a railway's grain revenues for a given crop year, it may also be considered to be a cap on the average per tonne revenue for a given crop year. For example, for crop year 2000-2001, assuming that each railway's average length of haul is the same as for the base year, CN and CPR's average revenue per tonne must not exceed $27.98 nor $26.12 respectively.

There has been some confusion as to the actual percentage reduction to grain revenues implied by the above revenue cap figures. In this regard, a combined railway average revenue cap figure of $27.00 per tonne is usually quoted. The confusion results from the fact that there are two benchmarks that serve for comparison. The first benchmark, 1998 grain data, was used by Mr. Kroeger when he proposed that crop year 2000-2001 revenues per tonne be 12% below 1998 levels.

The second benchmark related to the projected grain revenue per tonne for crop year 2000-2001 that would have resulted in the absence of reform. This benchmark arose since the Agency, pursuant to legislation at that time, had already set rates for the 2000-2001 crop year - which increased by 4.5% from the previous crop year's level. Using the second benchmark, and the data that went into legislation, the crop year 2000-2001 revenue cap figures reflect an 18% reduction

in per tonne revenues. However, using Mr. Kroeger's benchmark, the legislated 2000-2001 revenue cap figures reflect a 14.3% per tonne reduction compared to 1998 levels. The Federal Government accepted Mr. Kroeger's recommendation but used base year statistics derived from updated Agency cost and revenue estimates for grain. Consequently, the original 12% reduction became a 14.3% reduction.

The timing for establishing the railway revenue caps is rather interesting. Using the upcoming 2001-2002 crop year as an example, the Agency must establish one component, the Volume-Related Composite Price Index, by April 30th preceding the crop year. Thus, by April 30th of this year both shippers and railways will know the inflationary allowance built into the 2001-2002 revenue caps. But the actual revenue caps will not be set until December of 2002. The reason being that two other inputs required in the revenue cap formula - the tonnage moved and the length of haul for a given crop year, can only be derived after the crop year ends.

Grain Revenues:

As noted earlier, subsection 150(1) of the CTA states that a prescribed railway company's revenues, as determined by the Agency, for the movement of grain in a crop year may not exceed the company's maximum revenue entitlement {i.e. revenue cap} for that year. This section deals with the determination of railway company revenue for the movement of western grain.

Section 150 of the CTA provides a partial list of items that affects the determination of "initial" grain revenues, and a partial list of items that may, or may not, reduce grain revenues. In order to establish a clearer picture of what items affect railway grain revenue or constitute a revenue reduction, Agency staff provided the western grain industry with more comprehensive listings, as shown in Tables 1 & 2 below.

TABLE 1: INITIAL REVENUES	
INCLUDE	EXCLUDE
REVENUE FROM THE TRANSPORTATION OF FREIGHT - generated through application of rates contained in published tariffs or in confidential contracts	INCENTIVES, REBATES or any other similar reduction paid or allowed by the railway company. {CTA, Paragraph 150(3)(a)}
Amounts received by the railway company for ENSURING CAR SUPPLY	Amounts EARNED by the railway company... characterized as a PERFORMANCE PENALTY or DEMURRAGE OR STORAGE {CTA, Paragraph 150(3)(b)}
Amounts received by the railway company for PROVIDING PREMIUM SERVICE	Compensation for RUNNING RIGHTS {CTA Paragraph 150(3)(c)}
Amounts received by the railway company for PERFORMING INTERSWITCHING or EXCHANGE switching	
ADDITIONAL SWITCHING, in certain situations	ADDITIONAL SWITCHING, in certain situations
OTHER ITEMS, as they arise, if deemed to be appropriate.	OTHER ITEMS, as they arise, if deemed to be appropriate.

The table listings appeared in an Agency report used to solicit feedback from western grain participants as to the interpretation of various items (i.e. words, phrases or concepts) relating to revenue. At first glance, the interpretations appear to be straightforward. However, more than 30 issues are being addressed over the interpretations of the table items.

The Agency Decision will not be released for several weeks (i.e. until mid March of 2001), but one example of a complex interpretation follows. It relates to "amounts received by the railway company for ENSURING CAR SUPPLY" which is listed as a revenue inclusion in Table 1. This item was listed in order to capture railway revenue from a U.S.-style "bid-car" system where a successfully-bid amount is paid in order to ensure the receipt of cars. But what about the railways' newly-designed Advance Ordering Systems (AOS) wherein an amount is either set - or is bid as a penalty, which applies only if the shipper does not take order of, or ship with, the contracted cars? Meanwhile, if the railway company does not supply the car as per the agreement, the railway company pays a penalty to the shipper. Are net amounts received by the railway companies under the AOS revenues, penalties, or a combination of the two?

The designation is very important. If the net amounts under the AOS (i.e. railway receipts less payouts) are declared by the Agency to be revenue, then, keeping grain revenue below the revenue cap will require the railways to provide a similar amount of reduction to other grain revenues - all other things being equal. Alternatively, if the net AOS revenues are declared to qualify as penalties, their net revenues will not constitute revenue for revenue cap purposes, as revenue for other grain movements is unaffected.

TABLE 2: REDUCTIONS TO REVENUES	
REDUCE REVENUES	DO NOT REDUCE REVENUES
INDUSTRIAL DEVELOPMENT FUNDS "...if the Agency determines that it was reasonable for a prescribed railway company to make a contribution for the development of grain-related facilities to a grain handling undertaking that is not owned by the company, the company's revenue for the movement of grain in a crop year shall be reduced by any amount that the Agency determines constitutes the amortized amount of the contribution by the company in the crop year." {Subsection 150(5)}	"...amounts paid or allowed as DISPATCH by the company for loading or unloading grain before the expiry of the period agreed on for loading or unloading the grain." {CTA, Subsection 150(4)}
	Amounts paid by the railway company to municipalities or districts (at $10,000/mile) resulting from the discontinuance of operations of a grain-dependent branch line listed in Schedule 1 of the CTA.
Amounts paid or allowed by the railway company for INTERSWITCHING or EXCHANGE switching performed by another railway.	Amounts paid by the railway company as a PERFORMANCE PENALTY
OTHER ITEMS, as they arise, if deemed to be appropriate.	OTHER ITEMS, as they arise, if deemed to be appropriate

The most complex item listed in Table 2, as a reduction to revenue, pertains to Industrial Development Funds (IDF). The amortized amount of IDF contributions applicable to a given crop year was proposed to consist of:

> i) contributions made during the crop year which are fully applicable to the given crop year, plus
>
> ii) the amortized amount of contributions made during the given crop year which are to be amortized over a number of years, beginning with the given crop year, plus

iii) any amortized amounts for past years contributions which are amortized over a number of years, and which carry forward into the given crop year.

Several western grain parties have questioned the inclusion of IDF as a grain revenue reduction. But its inclusion as a revenue reduction is appropriate since the Bill C-34 reform package, which brought in an average base year revenue cap of about $27.00 per tonne, was derived under the assumption that IDF would be recognized as a reduction to revenue. Had IDF contributions not been part of the reform package, the average railway company revenue cap would have been higher than $27.00 per tonne.

The Federal Government recognized the importance of IDF when it included IDF in the grain reform package. Most IDF relate to the construction of new sidings at new high-throughput elevators, or the extension of sidings at existing elevators. In both cases, larger blocks of cars are handled creating efficiency gains and lower shipper rates.

The reason that IDF reductions are complex to determine is that their subsection 150(5) definition requires interpretations for: reasonableness, prescribed railway company, contribution, development, grain-related facility, grain handling undertaking, and company ownership. IDF reductions also require determination of an appropriate amortization period and cost of capital rate.

Revenue Cap Compliance

In accordance with the CTA's subsection 150(2), if a prescribed railway company's revenues, as determined by the Agency, for the movement of grain in a crop year exceeds the company's maximum revenue entitlement {i.e. revenue cap} for that year, the company shall pay out the excess amount, and any penalty that may be specified in the regulations, in accordance with the regulations.

The authority to make regulations specifying the penalty and respecting the paying out of the penalty and the excess amount is provided in the CTA's section 152. These regulations currently do not exist, but are expected by April, 2001.

Final Comments

The rate development regimes that apply to the movement of western grain have changed considerably over the past two decades. Under the recent Bill C-34 changes, the Canadian Transportation Agency's setting of maximum rates has given way to the setting of railway-specific maximum revenue entitlements, while the railways have been given greater freedom and flexibility in setting rates.

There are a number of interesting questions about the current rate development regime whose answers might affect future policies regarding western grain rate regulation. A few are listed below:

i) will the railways try to maximize their grain revenue under the revenue cap regime or will grain revenue fall below their caps?

ii) will the railways experience added competitive pressures from short line railways through the use of running rights, thereby moving towards a more fully commercial rate regime for western grain? and,

iii) will the results of the CTA Review Panel dealing with improved competitive access allow grain rates to be set in a fully commercial environment like any other commodity?

Regional Railway and Grain System Project

Roy Atkinson[1], Javier Caceres[2], Gary Housch[3], Darrell Richards[4]

Project Summary and Objectives

The project objectives are:

- to sustain communities, good jobs and enhance regional economies by providing a not-for-profit grain collection service to farmers, after covering the fixed cost of investment.
- to utilize and maximize the usage of Canadian infrastructure for the benefit of all Canadians.
- to engage the provincial and municipal governments in an intermodal planning process in order to make the best use of existing road and rail infrastructure to minimize the total cost of transportation

The concept is to develop a new form of enterprise for prairie branch lines that increases the participation of workers, communities and farmers in a regional grain collection system within the reality of the deregulated global marketplace and within a proposed commercial agreement with CNR.

---

[1] Founding President of the National Farmers Union
[2] AgriTrade & Transport
[3] Brotherhood of Maintenance Way Employees
[4] AgriTrade & Transport

Farmers cannot stop global restructuring, but they can develop an alternative to increase the market power of farmers and minimize costs within the deregulated environment that has emerged. The ultimate goal is to create a vertically integrated grain system under farmer and community control that can effectively compete with transnational grain companies within the rules of the deregulated global economy.

Lessons learned from developing this new form of worker enterprise can be a model for other situations to minimize the costs and capture the benefits of industrial restructuring affecting workers and communities.

The Challenges in Current Trends

Western farmers have lost control and ownership of the grain marketing, handling and transportation systems. Elevators and branch lines have been closed. Foreign transnational companies have increased their market power by taking over or integrating with cooperative elevator companies, building inland terminals, and setting the conditions for elimination of the Canadian Wheat Board. These corporations are increasing control of both farm inputs and outputs.

Farmer costs are increasing. Between 1995 and 2001, about 3,000 kilometres of rail line will be discontinued. There were 1,400 grain elevators in 1995, but this is shrinking rapidly. Eighty-nine prairie inland terminals planned or built can handle 84 per cent of the grain, leading to the closure of many rail lines and elevators.

The following tables illustrate the impact of rail and elevator closure on road costs. However, road costs are external to the decisions made to close elevators and rail lines.

Impacts of Abandonment -- Saskatchewan Case Study[5]

Prospective Rail branch lines abandoned (kms)	991
Roads with increased truck traffic (kms)	6,608
Roads with reduced truck traffic (kms)	3,584
Upgrade gravel roads to pavement (kms)	64
Upgrade gravel roads to pavement ($)	$10.7 million
Upgrade thinly paved to structural pavement (kms)	270
Upgrade thinly paved to structural pavement ($)	$45 million
Truck haul increased (annual tonne-kms)	27.8 million
Annual interest Cost on upgrades (15.4%)	$8.74 million*
Interest Cost per tonne-km for road upgrading	31.4 cents*
Interest Cost per tonne (15 km haul)	$4.71*
Interest Cost per tonne (75 km haul)	$23.55*

* Transport Concepts calculation based on Saskatchewan data of $56.9 million road upgrading costs, 15.4% annual cost of capital, and 27.8 million additional tonne-kilometres of truck haul

Source	Road Upgrade Cost	Rail kilometres Abandoned	Cost per Rail kilometre Abandoned
Sask Govn't 1985	$75 million	1771	$42,349
Sask Govn't 1989	$56.9 million	991	$57,417
Sask Average			$50,000
Our Extrapolation	$125 million	2500	$50,000
Our Extrapolation	$475 million	9500	$50,000

[5]ADI, Clifton Associates, E.M. Ludwick & Associates, Final Report, A Methodology for Measuring the Impact of Rail Branch Line Abandonment Provincial & Municipal Roads in Saskatchewan, February, 1989.

If the system consolidates to a system of inland terminals, the additional annual road construction and maintenance costs would be $86 million (low road upgrading costs) to $293 million (high road upgrading costs, compared to the 1995 configuration. If road upgrading costs are included, the total cost to ship grain 75 kilometres by commercial truck is $7 per tonne more than by rail branch line. [6]

Between 1995 and 1999, there were 2,355 kms of rail line and 165 elevator locations closed. Truck kilometres of hauling would double if the system consolidates to 186 prairie elevators with storage capacity over 10,000 tonnes. Another 8.7 million tonnes of grain per year is trucked an average 685 kilommetres to processing plants. This includes 4.2 million tonnes of feed grain trucked an average 1,067 kilometres, generally from Saskatchewan to Alberta and Manitoba feedlots.[7]

Farm to Elevator Trucking

	94/95 (924 locations)	98/99 (760 locations)	2002 (713 locations)	2002 (679 locations)	2002 (186 locations)
tonne-kms	701,260,178	727,957,222	756,099,873	776,553,714	1,318,074,165
avg haul kms	21.3	22.1	22.9	23.5	40.0
truck kms	54900000	57000000	46200000	47400000	80500000

Source: Trimac Consulting for Transport Canada

[6] Transport Concepts, Impact of Prairie Branch Line Abandonment, May 1995

[7] Trimac Consulting Services Ltd., Truck Traffic Changes in Grain Transportation, Sept., 1999, ES

The amount of grain truck haul is 17 times higher today than it was 25 years ago and it may double again if elevator consolidation continues. The magnitude of road impacts on Western Provinces is between $600 million and $1 billion, with Saskatchewan likely incurring the majority. Grain accounts for about one-half of the truck tonnage on thinly paved roads. Because of peaking in deliveries, grain accounts for the majority of damage on these roads.[8]

The present value of impact on Saskatchewan provincial highways to 1998 was $185 million. Consolidation to 60 delivery points will increase this to $500 million. These costs do not include impacts on municipal roads. These costs do not include grain hauled between elevators or to processing facilities. These movements far exceed the farm to elevator movement of grain.[9]

Trucking of Grain to Processing Plants in Western Canada

Item	Tonnes	Avg km	tonne-kms	Annual Loads	truck-kms
wheat flour	676,250	424	286,560,500	15,912	6,742,600
canola crushing	2,518,750	308	776,691,000	59,265	18,275,082
malt barley	602,500	310	186,570,500	14,176	4,389,894
distilling (rye)	137,500	411	56,527,500	3,235	1,330,059
durum milling	127,500	344	43,874,500	3,000	1,032,341
oats processing	380,500	207	78,747,000	8,953	1,852,871
feed grains	4,225,700	1,067	4,510,835,500	99,428	106,137,306
total	8,668,700	685	5,939,806,500	203,969	139,760,153

Source: Trimac Consulting for Transport Canada

[8] Government of Saskatchewan, Submission to the Estey Grain Review Phase II, Impact of Grain Truck Hauls on Saskatchewan Roads, Sept, 1998

[9] ibid

Conventional Solutions Not Working

Farmers, communities and labour have lost control of their livelihood as the government relinquishes its responsibilities under the pressure of globalization. Isolated responses by farmers, communities, and workers will not stop the race to the bottom. They should look for ways to find market power within the reality of the deregulated global market. How? Bring some transport/grain infrastructure under local control and capture the benefits from different segments of the grain chain.

Private short lines have had success in reviving some prairie branch lines, but these are generally in the north where there are other stable non-grain commodities. Examples include the Hudson Bay Railway, the Carleton Trail Railway and RaiLink Mackenzie Northern. Governments contributed $50 million to help the Hudson Bay Railway with its initial rail and port infrastructure.

However, the private short line model has been less successful for light density grain dependent lines. For example, Southern Rails Coop and the Central Western Railway have had to abandon lines as a result of grain elevator closure. In Saskatchewan, the Red Coat Road and Rail and the Great Western Railway are facing the closure of elevators affecting their operations.

It is not feasible to have a self-contained operation on a line with irregular train service requirements. The overhead cost of crews and locomotives is prohibitive. Therefore, regional economies of scale are required, for example to pool equipment and crews region-wide. The Hall Royal Commission report of 1977 recommended a Prairie Rail Authority to manage the maintenance and operation of grain branch lines on a system wide basis.

Also, because of the rapid pace of elevator closures, the branch line solution must be vertically integrated with a grain company willing to attract grain to the branch line. The main elevator companies have a financial interest in maximizing the grain flows through their concrete elevators and are closing elevators on branch lines. The traditional cooperatives have been privatized or are aligning with the practices of the transnational grain companies. Thus, a new locally controlled enterprise of grain collection and handling is required.

The Prairie Alliance For the Future (PAFF)

The Prairie Alliance for the Future is a non-profit, non-partisan, organization of people from Western Canadian rural communities working to regain democratic control of their destiny; to increase power for farmers in the marketplace; and to minimize costs for producers and taxpayers by making the best use of existing grain handling and transportation infrastructure under a new system approach.

Solutions to Position Farmers, Workers and Communities in the New Order

Farmers cannot stop global restructuring, but they can develop an alternative to increase the market power of farmers and minimize costs within the deregulated environment that has emerged. The ultimate goal is to create a vertically integrated grain system under farmer and community control that can effectively compete with transnational grain companies within the rules of the deregulated global economy.

Principles

The grass roots alternative is based on the following principles:

- Regain market power for farmers and other shippers and provide a real option for the marketing, handling and transportation of grain as an integrated farmer-controlled system
- Intermodal planning making the best use of existing road and rail infrastructure on the prairies to minimize the total cost of transportation
- Sustain communities, jobs and regional economies
- Provide not-for-profit service to farmers, after covering the cost of fixed investment
- Farmers, communities and labour to be partners in the development and operation of a regional grain transportation system

Agreement with Canadian National

Labour and farmer representatives of PAFF have been working since 1998 to create an integrated regional grain collection system to reduce total costs. On May 8, 2000, CN responded positively to the proposal made by the Brotherhood of Maintenance Way Employees (BMWE) to explore these options.

On January 25, 2001, the Brotherhood of Maintenance Way Employees signed a memorandum of understanding with Canadian National Railway to enable BMWE to develop a new regional grain collection system in cooperation with farmers and communities. The system would initially involve leasing 1,600 kms of CNR line. More lines may be transferred to the new cooperative entity in the future. The Canadian Pacific has also expressed interest.

Without detracting from the success of several private short lines, the PAFF proposal has several advantages for a regional network of grain dependent branch lines:

- CN has agreed to work with us on a region-wide branch line solution by developing a commercial lease agreement compatible with NAFTA trade rules.
- Provide service at cost to farmers, after covering the fixed cost of investment
- Generate system synergies by coordinating branch lines with grain handling
- Obtain regional economies of scale and attract grain to its lines by working together with local communities and farmers for mutual economic benefit
- Increase competition at the *system* level by creating an alternative to the transnational model taking over the grain industry
- Capture downstream benefits (blending, margins, volume discounts) for farmers through a vertically integrated grain system
- Joint planning of road and rail infrastructure for a least-cost intermodal system

Preserving and reviving these grain dependent branch lines has significant financial benefits to taxpayers. For example, the Government of Saskatchewan estimates that road upgrading costs $50,000 per kilometre of rail line abandoned. At this level, the 1,636 kilometres preserved in the initial network agreement with CN would avoid $80 million in road upgrading costs. At 10 per cent interest, the annualized saving would be $8 million.

There are interesting dynamics created by linking users to the economics of light density branch lines. Railways have low marginal costs and average costs that decline with volume of traffic. This is even more the case with light density branch lines. Thus, every carload added to the coop regional system reduces branch line rail costs per tonne for every shipper already using the system. There is a community of interest in maximizing the use of the rail line and attracting new traffic. The local users also benefit via lower road costs paid for by local governments. At the same time, the costs and benefits trade-offs of changing service can be compared and internalized for local decision.

Local control and marketing, combined with region wide intermodal coordination, opens up the possibility of attracting traffic to the rail system other than the traditional grain for export. About eight million tonnes of grain is currently trucked an average 685 kilometres for consumption/processing on the prairies. That is equivalent to 100,000 carloads. The shipment of inbound supplies should also be investigated.

Alternatives are required to the traditional way grain is collected and handled for transportation. For example, the Western Rail Coalition is contributing innovation through its FARMCO pilot project to test the use of information technology to improve returns to farmers by reducing handling costs and maximizing the value of grain through blending.

The elevator companies have abandoned many of their customers by closing country elevators and forcing grain into their concrete inland terminals. Many farmers have turned to producer cars as a way to regain some control of the system. Thus, we appear to be headed to two extremes: concrete centralized elevators on main lines and trackside loading to producer cars on branch lines. We also need some moderately-sized handling facilities that are more decentralized.

In addition, there is a need to meet the transportation requirements of specialty crops such as pulses. Total exports of pulses have risen to three million tonnes, with almost half as bagged containerized food shipments. These crops require hopper cars, box cars, trucks, intermodal vans or containers, depending on the circumstance and customer.

Where Do We Go From Here?

PAFF has identified four components to a vertically integrated system:

- Producer component
- Transportation component
- Grain gathering component
- Grain marketing component

PAFF developed a producer component that participated actively in the election of farmers as Directors of the Canadian Wheat Board, and has done the foot work in introducing its proposal for a farmer-controlled system.

On January 25, 2001, CN and BMWE signed a memorandum of understanding setting out a process to develop a commercial agreement whereby BMWE can lease branch lines. Producers in partnership with labour, communities and CN can engage progressive organizations and governments to start working on an intermodal least-cost system making the best use of existing infrastructure. Thus, we are dealing with the transportation component. But this is just one, albeit important, piece of the puzzle. We need to attract grain to these lines.

In order to deal with the grain gathering component, we will need grain gathering infrastructure and systems (e.g. elevators) . To deal with the grain marketing component, we will need to find an alternative for farmers who want an option to the transnational corporations that are incrementally taking over the grain companies and the role of the Canadian Wheat Board.

Mechanism

The mechanism to achieve this goal of a vertically integrated system could involve a *regional federation of agricultural cooperatives* to increase the market power of farmers in conjunction with the interest of labour and communities. This federation could:

- Operate a regional system of branch lines as part of an intermodal system making the best use of existing road and rail infrastructure

- Serve farmers and farm communities with supplies of inputs and operate a farmer-owned elevator/grain gathering system as well as related value-added and processing activities

- Own an international marketing agency that competes with transnational grain companies in a deregulated global market

Lessons Learned to Date

There is a great deal of mistrust among players in the western grain industry. New solutions are required that bypass long-standing divisions and entrenched cultural identities.

Farmers and communities are grasping for pieces of infrastructure such as elevators or rail lines because they are losing control of the system. The focus of the main railway companies is on adapting to continental competition under NAFTA. The *elevator* companies are in transition to transnational ownership and restructuring to become *grain companies* as the government incrementally dismantles the Canadian Wheat Board. The challenge is to find the common interest in these trends.

An important first step is to get those involved to recognize the reality of globalization and the changing role of government. Under trade agreements, government can no longer adopt domestic policies such as forced access to rail infrastructure or commodity specific subsidies. It will also be forced to deregulate the Wheat Board, the last State trading entity in the world. The transnational grain companies will acquire control of the Canadian elevator companies, and this is well underway.

Once we recognize that the old system is crumbling behind the facade, it is then possible to look for common interest in designing a new system. CNR wants to haul grain on the main line on commercial terms in a deregulated market. Farmers and communities want to control the use of branch lines to minimize *system* costs. Workers want good jobs. Common interest meets in the vehicle of a commercial agreement with CNR for the use of the lines.

Conclusion

PAFF has been able to open an opportunity to develop and implement a grass roots alternative in the restructuring of the grain system. This enables us to move the debate beyond the rigid traditional, and often adversarial, relationship among stakeholders to identify creative solutions in our common interest.

Appendix – List of Initial Proposed Coop Lines

Mantario Conquest Network (480 kms)

Elrose (Conquest Jct – Glidden)
Conquest (Conrose Jct – Beechy)
Matador (Matador Jct – Kyle)
Mantario (Mantario Jct – Glidden)
White Bear (Eston Jct – Lacadena)

Robin Hood Turtleford Network (340 kms)

Robin Hood (Denholm – Glaslyn/Spiritwood)
Turtleford (N. Battleford – St. Walburg/Paradise Hill)

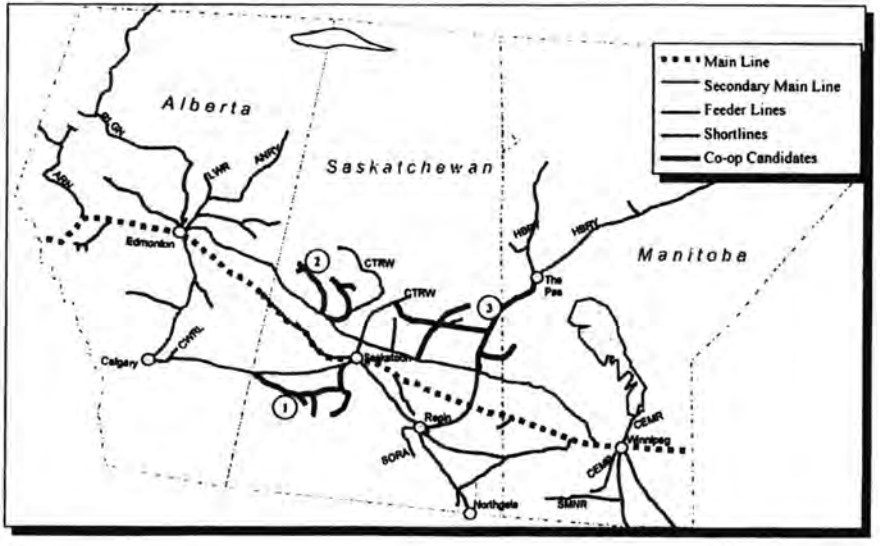
Tisdale Turnberry Network (816 kms)

Tisdale (Birch Hills – Hudson Bay)
Brooksby (Naisberry Jct – Crooked River)
St. Brieux (St. Brieux Jct – Humboldt Jct)
Turnberry (Hudson Bay – The Pas)
Preeceville (Preeceville - Swan River - Minitonas)
Assiniboine (Canora – Hudson Bay)

Total 1,636 kilometres

Co-op Line Candidates

- 3 line groupings considered as candidates for Co-op concept
  - ➔1,016 miles in total



*Grain Car Canada: Electronic Railcar Market Allocation System**

Dr. Barry E. Prentice, Director,
Tamara Thomson, Research Associate,
Transport Institute, University of Manitoba

The Canadian grain handling and transportation (GHT) industry is imperfectly competitive. A monoposonist controls the sale of the largest single crop, a duopoly provides rail transportation services, an oligopoly provides grain storage and handling, plus merchandising of major non-regulated crops, while grain production and agricultural trucking are atomistically competitive. .. Farm concerns about market power and workable competition in the GHT industry have generated over 100 years of government intervention. Since the mid-1970s, however, the government of Canada has been slowly reducing its role as an active participant and would, ostensibly, like to treat the GHT industry as it does other sectors of the economy.

The grain industry is adjusting to a less regulated and unsubsidized environment, with railway branchlines being abandoned and high throughput elevators (HTEs) being built. While the physical infrastructure is rapidly changing, the institutional setting by contrast, seems to be frozen in a bygone era. The unique market structure of the GHT industry makes it difficult to strike the right competitive balance. No single solution exists for such a complex problem, but improving the competitive environment could help inch the GHT industry in the right direction.

In the case of the GHT industry, an essential element necessary for market access is the railcar. If access to railcars became more competitive, so should the entire market. In particular, a highly fluid market in railcars could improve the indirect competition among

the railways and their corresponding supply chains. The economic viability of grain shippers is directly related to their having access to enough railcars to move all the grain they purchase. If the cars are not tied to one or other of the carriers, but could be disposed between the two carriers by shippers, the competitive environment is greatly enhanced.

Railcar supply is central to two major policy decisions that confront the logistics system of the western Canadian grain supply chain. The government of Canada must decide on the fate of its 13,000 hopper cars, and the industry must develop a new framework to determine shipper access to the railcar supply. The federal government has administered its hopper car fleet according to political rather than financial considerations. We recommend a new institutional relationship for ownership of the government fleet that would be efficient and promote competition, hereafter called Grain Car Canada (GCC).[i]

GCC would preserve the government-owned hopper cars for the betterment of farmers and society in general. Although a publicly accountable, not-for-profit organization, GCC would operate a railcar leasing pool on a purely commercial basis. GCC would be precluded from using railcars to pursue other mandates such as social and political goals dealing with branchlines and elevators, or economic development goals, such as sustaining municipal tax bases of communities or negotiating freight rates. Access to the railcars would be transparent, open, and determined by competition among the shippers, without bias due to size or location.

For the moment, the future of government owned railcars has faded from the center stage of grain transportation policy. But it is likely to return to the spotlight as the expiration date of the railways' right of first refusal to purchase the cars draws near. Removing the cars from the railways' control, and giving shippers access through GCC will remove any bias that may currently exist. The GCC board, comprised of shippers and farmers, would develop

rules to be non-discriminatory and ensure that the cars are freely available to all.

GCC would encourage indirect competition between the railways by allowing railcars to move freely between the two rail systems. Inefficient rail service would not be tolerated by shippers who would be free to choose the most capable carrier. Grain companies utilizing GCC railcars would have the freedom to employ their assets (railcar and elevator) in the delivery system in the most effective way to meet the needs of the CWB and the farmer.

Indirect Competition

Governments hesitation to deregulate the grain handling and transportation system is linked to their inability to guarantee farmers that the railways would not abuse the freedom they would have under deregulation to set their own freight rates. The challenge is to create a grain handling environment in which highly concentrated industries will act as if they are in direct competition. Much of the history of Canada's prairie grain policy has been the struggle to find such an institutional structure.

The concept of rail competition, in the Kroeger process, became hinged on the so-called open access recommendation of the Estey report, which would change the conditions under which the railways let other operators use their network. Although no evidence exists to prove that open access would have the desired effect on rail competition, it became the *quid pro quo* for greater deregulation of the grain industry that determined the government's policy course.

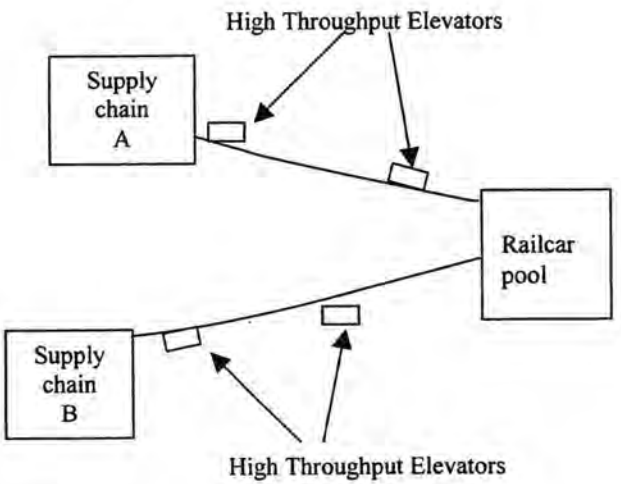
This paper takes no position on the desirability or feasibility of an open access regime for the railways. GCC would encourage the railways to engage in more competition and provide better service, regardless of the protection given to the rail network owner. GCC would enhance indirect competition between the railways by pitting one supply chain against the other. A policy of indirect competition

requires that the key entry point to the market be made highly contestable.

In the case of grain transportation, the key point of competition is the railcar. Like the gate at the airport, access to the railcar fleet (and whoever controls that access) will profoundly affect competition within the industry. If the railways have no control over which shipper gets access to a railcar and cannot restrict the transfer of railcars from their supply chain to a competing supply chain, they will be able to maintain their market share, only indirectly.

Figure 1 illustrates the concept of indirect competition between competing supply chains. The supply chains indicate the two primary rail lines that cross the Prairies, supply chain A being Canadian National and supply chain B being Canadian Pacific. Each line has high throughput elevators capable of handling unit trains.

Figure 1. Competing Supply Chains in Western Canada



Under the administered system, government-owned railcars are allocated to the individual lines and grain companies on the basis of performance, and sales are dependent upon delivery. Each railway has a fixed share of the car pool, which is rebalanced at five-year intervals. This process limits the competition for cars between the two supply chains.

The proposed GCC allocation method would allow railcars to flow to the location of the shippers in response to the highest bid. The auctioning process instills competition between the two supply chains (railways) as well as among shippers situated on the different rail lines. A rational shipper cannot afford to bid more for a car than it will yield in revenues. If one railway decides to charge higher rates, or its service deteriorates, shippers located on its track will not be able to bid as high as shippers located on the competing rail line. As railcars begin to migrate to the more efficient line (indirect competition), the other rail line is forced to become more competitive to maintain its market share.[ii]

The concept of indirect competition can be observed in many parts of the economy. Shopping centers for example, do not compete directly for consumers; instead, through their store leases and general appearance, they vie intensely for traffic. Similarly, the CTA Review Panel (2000) recognizes that forest product loading stations create indirect competition between the railways.

Electronic Market for Railcars

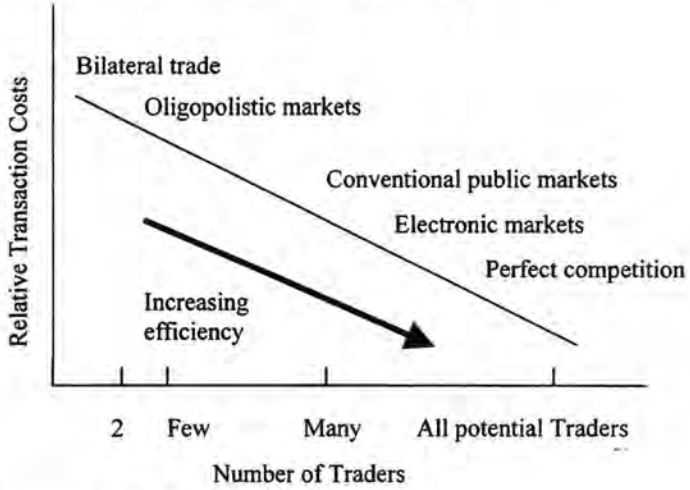
Technological advances in telecommunication and information processing make it economic to operate auction markets over the Internet. Electronic markets reduction in transaction costs produce private and public benefits normally associated with the theoretical model of perfect competition. Transaction cost economics defines a continuum that begins with a single supplier firm and ends with a perfect market institutional structure. Commercial relationships may be placed along the continuum, as shown in Figure 2. E-commerce reduces transaction costs to the point where

commercial examples approach the theoretical extremes of the market model and the integrated firm (Mulligan, 2000).

Movement down the continuum represents the reduction in relative transaction costs. In markets with only a few traders, buyers and sellers must spend considerable effort searching for an acceptable match and negotiating a price. In conventional public markets, there are more buyers and sellers making it easier to find a match, but the price may be subject to "local" conditions. In electronic markets, buyers and sellers can easily search for the best match without any geographic constraints, and can observe all the available prices simultaneously.

In electronic markets, individuals are atomistic in their ability to influence price that is the sum of the individual actions determine prices. Atomistic access to railcars means that the relationship between the shipper and the railway does not depend on "entitlements" or business volume. A shipper that lowers handling costs, or finds the best export prices for each and every freight load has a way to express the value it puts on the use of the railcars it needs to deliver grain.

Figure 2. Industry Specific Relative Transaction Costs



Economics of Railcar Supply and Demand

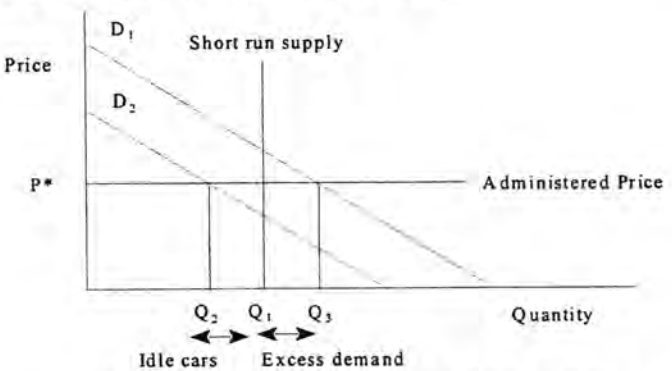
Most forms of transport experience regular peaks in demand for their services. The difficulty is to determine a pricing mechanism to ensure that infrastructure is used optimally, and fixed costs are recovered. Hopper car rationing is inevitable in the export of Canadian grain because of the seasonal imbalance in freight traffic. Over 60 percent of shipments travel during the five-month period from September to January. At the peak in shipping demand, all the available equipment is pressed into service, whereas in the off-season, surplus equipment is idle. This dynamic means that an operator can never afford to purchase enough cars to serve more than the average peak volume. An operator choosing to maximize profit may instead supply fewer cars during peak periods, collect monopoly rents, and reduce the opportunity cost of idle equipment in the off-season.

When the demand for a good or service is not uniform over

time, the analysis of peak-load pricing may be used. The market for covered hopper cars in western Canada is a case of the peak-load demand problem. Public intervention is common in markets that exhibit a peak-load demand, such as public transit services. Government intervention can range from the administered rationing of publicly owned equipment to the use of public utility peak-load pricing systems.

Figure 3 presents an administered car-rationing model. The demand for grain transportation is separated into peak demand and non-peak demand curves, D_1 and D_2 respectively. During the peak season shipping demand (D_1), capacity is rationed to shippers by an administered system of rules and entitlements. The supply of rail service becomes perfectly inelastic as the system reaches capacity at Q_1.[iii] At P*, the administered price,[iv] shippers in the peak period would like Q_3 cars. The excess demand, $Q_3 - Q_1$, is a source of shipper complaints and pressure to alleviate the perceived car shortage. During the non-peak period shipping demand (D_2), the administered price is such that all shippers get access to all the cars they desire, and $Q_1 - Q_2$ cars remain idle.

Figure 3. Administered Car-Rationing Model



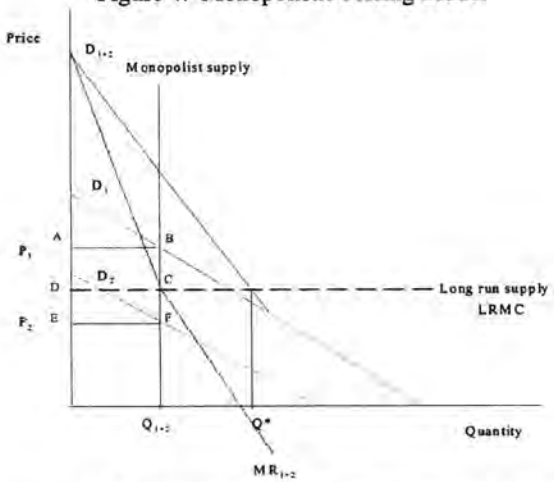
A price-based system of allocation can relieve car shortages because the prices rise to ration the demand. Prices can also help to

smooth out peaks and troughs in demand. If users know that they can get lower prices by waiting, some will try to delay shipments to the lower priced period. The problem of the peak-load market is who gets to determine the supply.

Figure 4 presents the case of a monopolistic supplier. To set prices for the peak and non-peak periods, the demand curves are added vertically allocating costs over the full year. In this illustration, D_1 represents the peak demand period from September to January. D_2 represents the non-peak demand period from February to August. A monopolist supplier can maximize its profits by setting the supply of cars so that the LRMC equal the combined marginal revenues of the two markets. The combined marginal revenue (MR_{1+2}) is determined by the vertical addition of the peak and non-peak marginal revenues to obtain the total marginal revenue curve. The monopolist sets the supply of cars at Q_{1+2}, and charges prices P_1 during the peak period and P_2 during the non-peak period.

Figure 4 shows that during the peak demand period (D_1), the price rises to P_1, which is well above the LRMC. The non-peak price falls to P_2, which in this case leaves no cars idle.[v] A monopoly supplier has an opportunity to capture supra-normal profits by undersupplying the railcar market. The net economic rent is ABCD (the revenues earned in the peak period above the LRMC), less economic losses DCFE (the revenues earned in the non-peak period that are below the LRMC).

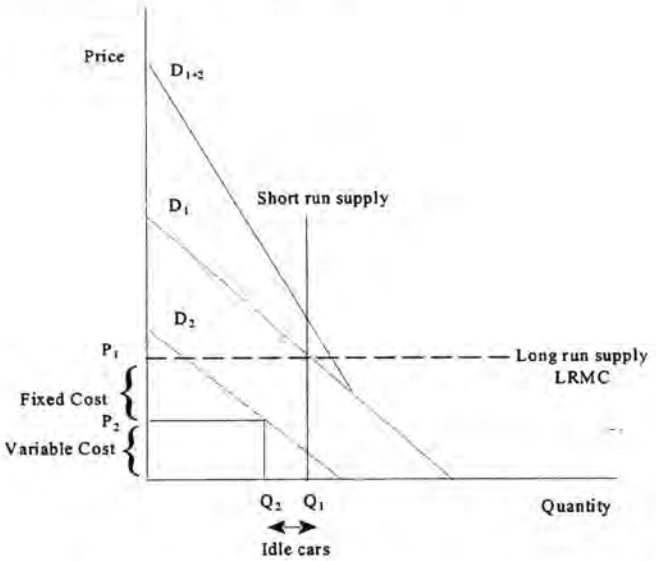
Figure 4. Monopolistic Pricing Model



The monopolist car supply reduces social welfare. The socially optimal car supply Q*, is where total demand D_{1+2} equals the LRMC.[vi] A private operator, if given unfettered railcar ownership, can aggressively build commercial relationships to capture the benefits of demand fluctuations. At the same time, no administered system without perfect information on shipper freight and railway capacity can allocate railcars as efficiently as a price system. An alternative used by governments to generate the benefits of market pricing, without transferring those benefits to a private monopoly, is to create a not-for-profit agency that can operate as a public utility.

Public utility systems such as those for electricity and public transport commonly use peak-load pricing. Electricity and transport alike cannot be stored to reconcile systematic changes in demand with smooth even production. Reconciliation may only be accomplished in these systems through pricing. Systems that use prices to determine service priority can yield benefits of efficiency, equity, and stability. Figure 5 presents the ideal pricing system for the peak-load market problem.

Figure 5. Optimal Peak Load Pricing Model



In a perfectly adjusted peak-load market, users in the non-peak demand period pay less (P_2) than the peak demand users. In this case the peak demand users bear the fixed costs of the total supply at P_1.[vii] The extra cars, $Q_1 . Q_2$, are needed only during the peak demand period, and otherwise are idle. If the size of the railcar fleet is optimally supplied, the short run supply will cut the peak demand curve at the same point as the long run supply. The combined revenues of the peak and non-peak demand users will equal the LRMC of the service. The railcar supplier can capture no economic rents.

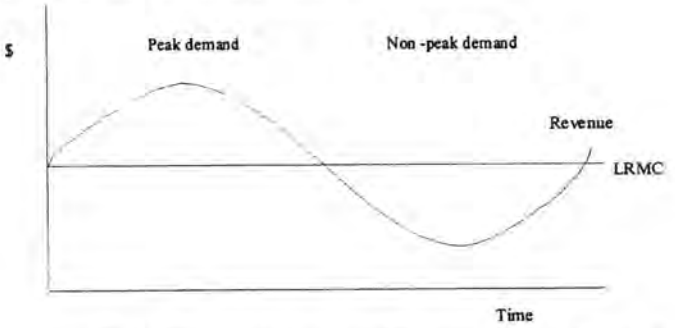
The Case for Grain Car Canada

From a producer's perspective, an optimal railcar supply

involves a tradeoff between service cost and fleet size. Year-to-year fluctuations in the peak market demand occur with the size of the harvest. The dynamic nature of the demand requires that the owner of the car supply react to the risk of having too many cars. GCC must have a decision rule that determines the long run supply as well as the short run allocation.

We envision GCC as having two prime functions: it would operate an electronic auction market to provide unbiased access to the railcars, and it would use the proceeds to maintain a railcar fleet for western Canadian grain farmers. Figure 6 shows how the auctioning of railcars by GCC could result in a self-sustaining market. During the peak demand period from September to January, leasing revenues would exceed the long run marginal cost LRMC.[viii] At peak-demand, the car fleet would be used at capacity and supplied to the shippers prepared to pay the most. During the non-peak demand period, it is anticipated that some cars would sit idle. Shippers placing a call for cars at that time would not pay the full costs of usage. The LRMC is balanced between these two extreme periods, and the supply would approach the optimal peak-load allocation.

Figure 6. Self-sustaining Railcar Market



The self-sustaining nature of the market could be achieved if GCC followed a simple decision rule. The revenue obtained by GCC

through the market mechanism would guide it toward the optimal fleet size. If at the end of any fiscal year, GCC had a positive balance, it would add new cars to the fleet. By contrast, if costs exceeded the revenues of the auction, it would not add new cars. Because of depreciation, this rule would have the effect of a net reduction. Consequently, bids placed in the auction market would provide the signal for GCC to either expand or contract the car numbers to reach the optimal fleet size.[ix]

In practice, the optimal fleet size of the GCC fleet depends on the car supply response of other parties, especially the railways. Given the railways' common carrier obligation, GCC is likely to be the residual supplier of grain cars. In addition, efficiency improvements, in particular high throughput elevators, could reduce the total number of cars required. Ultimately, the GCC fleet size would be determined by market forces.

Concluding Comments

The design of a less regulated and more competitive grain handling and transportation system has been an unsolved policy issue in western Canada since the late 1950s. Successive governments have attempted to find a balance between the social concerns of the farm population and the commercial necessities of the grain supply chain. Car supply disputes stem from the natural seasonal fluctuations in the demand for grain transportation. This is aggravated by year-to-year fluctuations in the size of the harvest. In a system as complex as the logistics of grain transportation, no single institution can solve all problems. The creation of GCC could, however, move the entire industry toward a more stable and in the long term efficient equilibrium.

The deregulation of the Canadian grain transportation and handling system must include an evolution away from the existing institutional structure. The concentration of market power in the existing institutions undermines the belief of farmers that unfettered competition is likely to work in their interest. Deregulation is

necessary because the supply chain for grain must adapt to an unsubsidized environment and reduce its costs. The existence of 13,000 government-owned railcars is a residual of a regulatory system that represents an implicit subsidy to producers. In the long term, a decision must be made to replace this fleet and to engage a system of railcar access that serves the pubic interest by limiting railway market power.

We recommend GCC as the not-for profit corporation to take over the ownership of the government fleet. GCC would be an attractive policy option if its only merit were as a means of sustaining the fleet of government-owned cars, while providing unbiased access to shippers. The added benefit of GCC, however, would be its impact on competition. First, as an alternative market-based source of railcars, GCC would reassure farmers that a commercialized system can be competitive, while railway power is kept at a minimum. Any shipper that could obtain a railcar from GCC would pay no more for a railway-owned car unless it were provided with additional services. GCC need not provide all the railcars, only a fleet sufficient to give shippers an option.

Second, GCC could enhance indirect competition between the rail lines and among the grain companies. The ability of the cars to flow between the railways to the location of the highest bidder would encourage the railways to compete, providing rates and service that would attract shipments to their lines. The electronic market also would reduce the market power of the large grain companies to influence the carriers. E-commerce reduces the advantages of size, while anonymity guarantees that all parties compete from a common base.

References

Canadian Wheat Board, *Transportation Reform: Report to Producers,* August 2000.

Canada Transportation Act Review Panel Interim Report, *Competitive Rail Access,* December 2000.

Earl, Paul, *Mac Runciman: A Life in the Grain Trade,* University of Manitoba Press, 2000.

Estey, Willard, *Grain Handling and Transportation Review – Final Report,* December 1998.

Hayward, Brian, *Movement of Grain from a SEO and CAPG Perspective*, In the First Annual Fields on Wheels Agribusiness Logistics Conference Proceedings, p. 88-98, University of Manitoba Transport Institute, November 1997.

Kroeger, Arthur, *Stakeholders Report and Working Group Appendices A, B, and C,* September 1999.

Mulligan, John, *Electronic Railcar Market Allocation System: A Conceptual Model for Increasing Competition in Western Canadian Grain Handling and Transportation,* M.Sc. Thesis, Department of Agricultural Economics, University of Manitoba, January 2000.

Parson, Graham, *Have Truck will Travel – Some Effects of Trucking on Railway and Elevator Competition.* In the Third Annual Fields on Wheels Agribusiness Logistics Conference Proceedings, p. 3-12, University of Manitoba Transport Institute, November 1998.

Prentice, Barry, *The Potential Electronic Market in Grain Hopper Cars.* In the Third Annual Fields on Wheels Agribusiness Logistics Conference Proceedings, p. 36-39, University of Manitoba Transport Institute, November 1998.

Transport Canada, *Disposal of the Government Hopper Cars.* September 1999.

---

* This report is based on an earlier study completed under a general contract between the Department of Manitoba Highways and Government Services and the University of Manitoba. The authors of this study have worked independently and opinions expressed, therefore are their own, and do not necessarily reflect the opinions of the

Province. The authors would also like to thank four anonymous reviews and the many helpful comments and criticisms received on earlier drafts of this report.

[i] Similar works on this subject refer to Grain Car Canada as a Railcar Authority. It is the authors understanding that the word authority brings with it thoughts of regulation and rules. The goal of this study is to bring forward an alternative to past restrictive policies placed on the Canadian grain handling and transportation system.

[ii] Some observers might argue that the railways each have "captive markets" that will not force them to compete. While this may have been true during the early development of the Prairies, the advent of semi-trucks and B-trains leaves few areas without competitive delivery points. Parson (1998) provides a thorough analysis of the new competitive setting in UMTI's 3rd Annual Field on Wheels Conference Proceedings.

[iii] As peak capacity is approached some actions could be undertaken to alleviate constraints, such as appropriate selection of origins (closer to port, on mainlines, favouring larger elevators), and controlling actual commodities shipped. This does not change the analysis, in part because who is going to give up their right for the improvement of the whole? Moreover, the constraint is still likely to be binding, i.e. Q_1 will still fall to the left of Q_3.

[iv] In western Canada, the government owned cars are provided free of charge, but access to the cars is bundled with a freight rate that represents the administered price.

[v] Whether the monopolist has idle cars in the non-peak season depends on the nature of demand. As long as price is above variable costs cars will be supplied, otherwise the monopolist is better to leave them idle.

[vi] Technically, this is the point where the marginal social cost (LRMC) is equal to the marginal social benefit that is the combined demand. At any other point, society will suffer a dead weight loss of efficiency.

[vii] If the seasonality were less pronounced, the non-peak users would also share some part of the fixed costs. In general, if this were to occur, no cars would remain idle in the non-peak period.

[viii] Technically this is referred to as "first degree price discrimination." Each buyer is forced to reveal his maximum willingness to pay. An efficient result occurs because the entire consumer surplus is appropriated by GCC. None of the welfare loss normally associated with monopolies occurs because GCC through the auction can perfectly price discriminate.

[ix] Given the long service life (40 years) of railcars, any expansion or contraction would be very gradual.

Identifying High-Hazard Locations for Pedestrian and Bicycle Crashes

Richard A. Raub, Senior Research Scientist
Northwestern University Center for Public Safety
405 Church Street, Evanston, IL 60208 USA

Abstract

Frequency of occurrence, rather than use of exposure-based data, traditionally has served as the basis for identifying and alleviating crashes involving pedestrians and bicycles. For most smaller cities, the use of frequency has been sufficient. On the other hand, where many crashes are being examined, or the city covers a large geographical area, an exposure-based measure may be more useful. However, a difficulty results from the lack of adequate bases for measuring pedestrian- and bicycle-crash exposure. Neither the number of vehicles using the roadway nor the pedestrian or bicycle volume has shown adequate relationship to crashes involving these groups. Moreover, pedestrian and bicycle volumes are not gathered except for isolated locations within a city. This research explores a method of combining an exposure measure based on population, crashes, and geographic information systems (GIS) to map locations. It uses population which is readily available from the U. S. Census as the basis for computing rates. These rates then are assigned to roadway segments with the aid of GIS tools. This process allows persons involved with highway safety to view a map and rapidly pinpoint locations which may be experiencing an abnormal level of crashes. The process of converting pedestrian and bicycle crashes to rates based on population and then displaying them on a map is tested for Evanston, Illinois. Analysis shows how the mapped data can be interpreted. The process appears to hold promise for use, particularly in larger cities. Several issues including the applicability of population for generating rates are discussed.

Identifying High-Hazard Locations for Pedestrian and Bicycle Crashes

Richard A. Raub, Senior Research Scientist
Northwestern University Center for Public Safety
405 Church Street, Evanston, IL 60208 USA

Abstract

Frequency of occurrence, rather than use of exposure-based data, traditionally has served as the basis for identifying and alleviating crashes involving pedestrians and bicycles. For most smaller cities, the use of frequency has been sufficient. On the other hand, where many crashes are being examined, or the city covers a large geographical area, an exposure-based measure may be more useful. However, a difficulty results from the lack of adequate bases for measuring pedestrian- and bicycle-crash exposure. Neither the number of vehicles using the roadway nor the pedestrian or bicycle volume has shown adequate relationship to crashes involving these groups. Moreover, pedestrian and bicycle volumes are not gathered except for isolated locations within a city. This research explores a method of combining an exposure measure based on population, crashes, and geographic information systems (GIS) to map locations. It uses population which is readily available from the U. S. Census as the basis for computing rates. These rates then are assigned to roadway segments with the aid of GIS tools. This process allows persons involved with highway safety to view a map and rapidly pinpoint locations which may be experiencing an abnormal level of crashes. The process of converting pedestrian and bicycle crashes to rates based on population and then displaying them on a map is tested for Chicago, Illinois. Analysis shows how the mapped data can be interpreted. The process appears to hold promise for use. Several issues including the applicability of population for generating rates are discussed.

The Role of Inclement Weather in Heavy Truck Accident Causation: Implications for Driver Training, LCV's and Safety Programs

by
Jake Kosior[1], Shannon Summerfield[2]

ABSTRACT

Inclement weather is a natural conseqence of living in Canada. The role of hostile environmental conditions in heavy truck accident causation appears to be under appreciated. For single vehicle incidents, nearly 41 percent of heavy truck accidents in Manitoba occurred with poor roads and windy conditions as opposed to 32 percent for passenger vehicles. However, heavy truck incidents per Million Vehicle Kilometers Travelled (MVKT's) are at least 50 percent that of passenger vehicles. Also, Longer Combination Vehicles (LCV's) did not appear over represented in accident statistics. The reason is that motor carriers practice self imposed "graduated licensing" by using the most experienced drivers with LCV's. In addition, mechanical failure and driver fatigue together account for less than 10 percent of accidents. Collisions with wildlife pose a greater threat to the motoring public than heavy trucks.

These findings suggest that greater emphasis should be placed on advanced driver training, easing LCV policies and shifting financial resources for advanced weather warning and technologies to extract the next level of productivity gains for motor carriers.

---

[1] Senior Research Associate, University of Manitoba Transport Institute
[2] Research Affiliate, University of Manitoba Transport Institute

1.0 Introduction

Motor truck accidents in Manitoba have declined from 87 Heavy Truck Accidents (HTA's) per 100 Million Vehicle Kilometers Traveled (MVKT's) in 1987 to 76 HTA's per 100 MVKT's in 1997, a 13 percent decrease. This occurred in a transition period of economic reform for not only the transportation sector, but for interprovincial and international trade. The incidence of heavy truck accidents has decreased while the number of long haul truck movements in North America has increased. The reduction in heavy truck collisions is due to policies addressing driver standards, training, hours of service and vehicle safety and operating codes. Vehicle technology such as "self adjusting" air brakes and anti-jackknife devices have also contributed to a reduction in heavy truck accidents.

Policy and technological initiatives have improved heavy vehicle safety on our highways, but these gains may have plateaued. A natural consequence of increased trade – and increased trucking activity – is the corresponding increase in total truck accidents. Further research is required to ascertain where the next level of gains in trucking productivity by reducing accidents can be made.

With the ever-increasing number of heavy commercial trucks using the public highway network in Canada, their involvement in collisions continues to be a significant public safety issue. The increased number of trucks competing for operating space on the highways system will probably result in overrepresentation in collisions resulting in serious injuries or death. Not only are the numbers of commercial trucks increasing, but also truck size, weight, and permissible configurations are changing as provincial jurisdictions move toward harmonized regulations.

For the U.S. experience, heavy truck accident rates are at the lowest level since data have been collected. However, there are still approximately 5,000 fatalities annually in accidents involving trucks in the United States. Of these, only about 700 fatalities are to the

occupants of the truck. Public fear of trucks seems to grow as highways become more crowded, causing political leaders to call for more controls. The USDOT has called for a 50 percent reduction in fatal accidents by 2010. Reaching such goals will require many new techniques applied to trucks, drivers, fleets and highways.

2.0 The Role of Adverse Weather in Heavy Truck Accident Causation

The role of weather in heavy truck accident causation may be greater than previously accounted for in previous studies. Seasonality in heavy truck, passenger and other vehicle accidents shows a correlation with adverse weather conditions during winter months. Determining trends in accident occurrences has application for a variety of planning purposes.

Khasnabis and Lyoo showed that monthly accident data over several years (figure 1) has a strong seasonal component, which would have otherwise been masked if the monthly data were combined to make

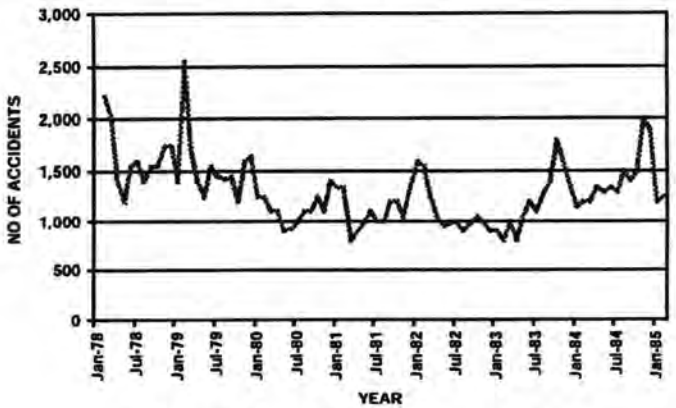


Figure 1: Time Trend in Heavy Truck Accidents in Michigan – 1978 to 1985

yearly counts. This factor further justified the use of monthly data in their forecasting model to capture seasonal trends. From January 1978 to January 1985, the data shows peaking during winter months, the only exception being January 1983. A general decrease was observed between January 1980 and January 1983 when heavy truck accidents began to increase. This may be correlated with the economic downturn observed in the early 1980's.

However, when the data base reflects seasonal peaking (as in figure 1), the feature of autocorrelation, although no necessarily reflecting dependence, was effective in fitting a time series model. In their case study, the model appears to indicate a statistically significant correspondence between the observed data and the forecast values. This could be partly because of the seasonal peaking of the accident observations, indicating some degree of seasonal correlation, which may not necessarily be construed as dependence. What the *Khasnabis and Lyoo* study does not reveal are the underlying combination of factors that result in seasonal peaking of accidents.

An in-depth heavy truck accident study conducted in Ontario in 1981 (*Billings et al.*) revealed that inclement weather had a significant factor in accident causation. The study concluded that of 474 accidents, 88 percent were attributed to the actions of the driver. Another 14 percent were jacknife accidents of which 63 occurred on wet roads. Of these accidents, 9 percent involved unladen vehicles with others partially loaded. Only 53 cases were caused by mechanical failure, load loss or other non-driver related factors.

The Ontario study further revealed that of the principal singular reasons for heavy truck accidents, 44 percent were by drivers speeding too fast for prevailing conditions (figure 2). Mechanical failure and load loss together accounted for about 26 percent of accident causal factors while the remaining 30 percent were also driver related.

What the Ontario study revealed was that a focused effort to make drivers aware of the environmental envelope in which they operate is

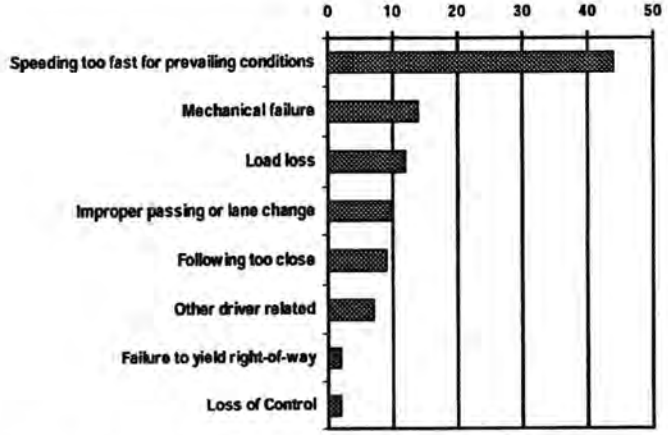


Figure 2: Percent Distribution of Heavy Truck Accident Causal Factors (Source: Billings et al., 1982)

perhaps the greatest contributor to accident reduction. Driver awareness influences driver actions for unforeseen events. Advanced defensive driver training that takes into account inclement weather evasive maneuvers may play an important role in increasing highway safety.

3.0 The UMTI Inclement Weather Heavy Truck Safety Study

Literature on heavy truck accident statistical studies reveal that while great strides have been made during the past two decades in attempting to understand the underlying nature of causal factors, considerable gains can still be made by understanding the role of weather in accident causation. This necessitates a different approach to investigate large databases without falling into the traps and limitations of standard statistical analysis techniques.

The approach is a simple one, 1) Weather data from five years will be combined with heavy truck accident data to form a master accident database with weather elements, 2) statistical techniques will be used to guide the analysis, 3) cluster analysis will be used to group data into various "trees" to uncover combinations of simultaneously occurring variables.

4.0 Building the Master *MA*nitoba *C*ommerical-Truck *A*ccident *D*atabase (MACAD)

All articulated vehicle accidents from 1994 to 1999 were extracted from the master provincial database[3] and placed in a new file. The second task was to append weather data from the Environment Canada database to each accident record. The purpose for this was to provide actual measured weather data at the time of the accident rather than rely solely on encoded data from police reports. The date, time and highway control section numbers were used to correlate the accident to weather station data. Weather data was calculated using either triangulation or bipolar methods, or used directly depending on accident proximity to recording stations. Figure 3 shows the process to assemble MACAD. In total, there were 1,231 commercial articulated vehicle accidents in rural Manitoba extracted from the master database from 1994 to 1999.

4.1 Data Limitations

The provincial master accident database has several shortcomings that are not uncommon to other accident databases – that is subjectivity. Data collected at the scene is characteristically compartmentalized according to the available choices of data codes. While this expedites accident reporting by police and provides broader coverage of incidents, it does not provide the depth of an exhaustive analysis. For example, the measurement of exact wind

[3] Urban cases and non-articulated trucks were excluded. However, articulated truck accidents occurring on Winnipeg's perimeter highways were included.

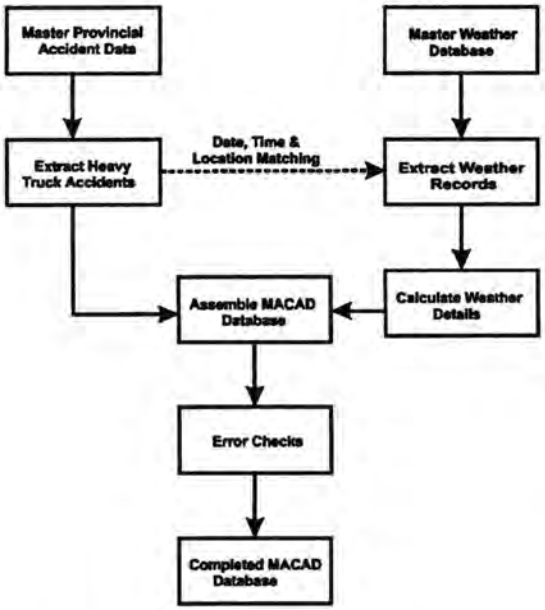


Figure 3:Flowchart showing MACAD Database preparation process.

speed and gust versus "high winds", or measurement of road friction versus "icy road".

The second shortcoming is in non-police reported accidents whereby data may be relayed to officers several days later. Memories fade and most notably in the case of single vehicle cases, causal factors of the incidents may be deflected to other than the actual reasons[4].

[4] An example may be a driver that fell asleep at the wheel or was over the legal limit for alcohol and reported the incident the next day, laying blame on another vehicle or animal crossing the road.

While data from the provincial master weather database is extracted from instrument measurements at weather stations and is not disputed, accidents are often between stations and weather data is therefore averaged for a region. While this method still does not provide the degree of data accuracy desired, it provides better estimates of the actual precipitation, temperature, wind speeds and direction than coded data. The weather dataset did not contain wind gust factor as it not recorded by automated means. Wind gust is suggested as a more influential factor in articulated truck (and passenger car) accidents than wind strength. More extensive analysis would have to be employed to assess wind as a factor in heavy truck accidents.

However, despite these limitations, sheer data volume mitigates these factors to some extent. This is consistent with Level I studies (statistical reviews) and is used to derive trends and indicators for accident causal factors.

5.0 Heavy Truck Accidents in Manitoba 1994-99

Figure 4 shows the date of occurrence of articulated heavy truck accidents versus wind speed for all incidents in Manitoba from 1994 to 1999. The red circles in figure 4 indicate that a disproportionate number of incidents with high winds correspond with winter months. The data further shows data "spikes" (the straight lines) at certain times. This appears obvious at first, that high-sided vehicles would not perform well under high wind conditions, particularly with icy roads and other contributing factors. However, the objective is to understand the interaction of these factors and to mitigate their role in accident causation. This cursory view of the data shows that a link between accident occurrence and inclement weather exists. Exactly how the factors influence each other and to what degree can be elusive.

While some colleagues in accident safety research would disregard the comparison between passenger vehicles and heavy trucks due to

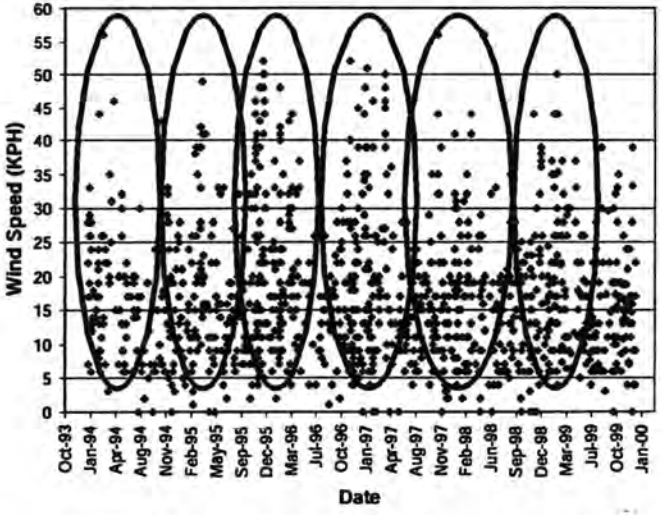


Figure 4: Wind Speed and Temporal Distribution of Commercial Vehicle Accidents in Manitoba: 1994 to 1999

inherent performance characteristics of different drivers and vehicles, the comparison points exactly to that observation. High sided vehicles such as heavy trucks should not perform as well as passenger vehicles in inclement weather, but truck drivers should be able to compensate for this difference because of their greater road experience than passenger vehicle drivers. The validity and degree of this statement remains to be tested. This additional comparison provides another perspective on the performance of heavy vehicles in adverse weather rather than a stand alone analysis.

6.0 Single vehicle heavy truck and passenger vehicle accidents with poor road conditions

This paper will provide a portion of the cluster analysis from the main report done for Transport Canada to illustrate the technique. Space limitations for proceedings do not allow all "trees" to be presented. A

cluster analysis between heavy trucks and passenger vehicles is presented in figures 5 and 6 for icy/snow covered road with low winds. Low winds are conditions below 25 kph. When early analysis began, it had become apparent that two distinct groups had emerged for both classes of vehicles, accidents occurring under poor and good visibility conditions. Poor visibility was anything that impaired a driver's vision – nighttime driving, rain and snow for example. Good visibility was daytime driving under no precipitation conditions so as the driver had maximum vision. This outcome appears to confirm *Mensah and Hauer's* claim that different conditions produce different accident patterns, hence the need to disaggregate data.

Referencing figures 5 and 6, it is appears that heavy trucks have a greater proportion of accidents under poor road conditions compared to passenger vehicles. Nearly 41 percent of heavy truck accident cases, as opposed to 23 percent of passenger vehicle cases, occurred under winter (icy/snow covered) and spring/fall (mud/wet/slush) road surface conditions. Moving further down the trees, another distinction is found at wind speeds. For passenger vehicles, 71 percent of icy/snow covered road accidents occur with low winds, with about 29 percent under moderate winds. For heavy trucks 63 percent of accidents with icy/snow covered roads occurred under low winds with 37 percent under moderate winds. For heavy trucks, not only is the proportion of the accident population occurring under poor road conditions, but also the number accidents with higher wind values is greater.

However, for low winds at least, the distinction ends with visibility conditions. For both passenger vehicles and heavy trucks, the split between poor and good visibility conditions is the same at 69 and 31 percent respectively. The lowest level reserved for primary causal factors indicate that similar patterns exist for both heavy trucks and passenger cars with low winds. The code NOS stands for "Not Otherwise Specified". This means that the accident report form listed poor road, wind or visibility conditions (or a combination) as the primary factors for the accident. Unless there was a distinctive driver (alcohol, fell asleep), non-driver (animal on the road), mechanical

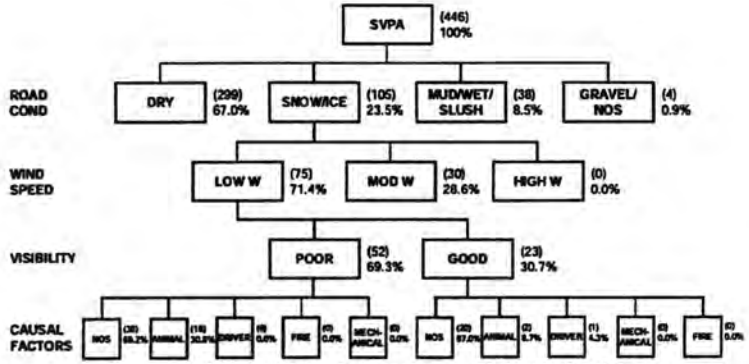


Figure 5: Single Vehicle Passenger Vehicle Accidents in Manitoba with Icy Roads 1994-99

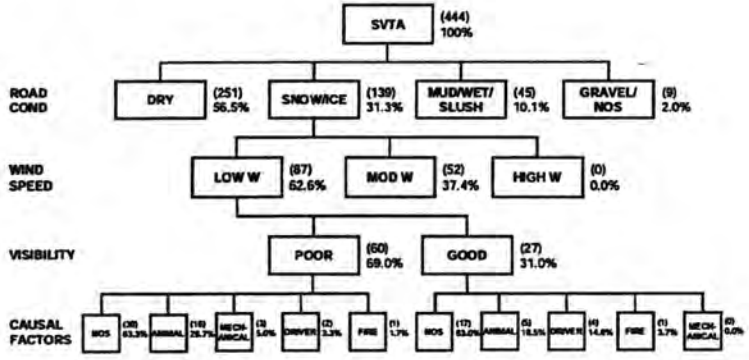


Figure 6: Single Vehicle Heavy Truck Accidents in Manitoba with Icy Roads 1994-99

failure or vehicle fire that precipitated the accident, it was left as NOS.

Constables listed NOS under poor roads, low winds with poor visibility at 60 to 70 percent of all accident cases for both heavy

trucks and passenger vehicles. An interesting observation was noted in the causal factors branch under poor visibility. While NOS coded accidents accounted for approximately 60 to 70 percent of incidents, "animal" accounted for another 25 to 30 percent for both vehicle classes. Poor visibility conditions do not provide sufficient reaction time for drivers. The result is often a hasty (and wrong) maneuver or braking action. This is particularly acute when nocturnal animals dart onto roadways at night while foraging[5].

Passenger vehicles had no mechanical, fire or driver factors under the "poor road, low wind and poor visibility". Heavy trucks had only three cases of mechanical failure, one fire and one driver related causal factor responsible for accidents. This represents only 1.3 percent of all accidents in the population and only 10 percent of all accidents occurring with "poor road, low wind and poor visibility". It appears that under these conditions at least, factors other than the weather conditions do not account for a significant proportion of accident causation.

Under the "poor road surface, low wind and good visibility" branch, the proportion of causal factors changes modestly for both categories of vehicles. For passenger vehicles, NOS accounts for the lion's share at 87 percent (20 of 23 cases) of causal factors. Only 2 cases involved animal action with one driver related case. For heavy trucks, a similar pattern exists with a decline in the amount of "animal" related incidents (5 cases or 18.5 percent). Relative to the "poor surface and visibility and low wind" branch, the number of mechanical failure and driver related incidents was nearly identical under good visibility as opposed to poor visibility for heavy trucks.

---

[5] In Manitoba's forested areas there are high incidents of vehicle collisions with large animals such as deer, elk or moose. Due to their size, these animals can inflict considerable damage to passenger vehicles while creating a dangerous situation for truck drivers. In Australia, trucks are outfitted with a robust grill called a "roo bar" to prevent damage to tractors when in collision with kangaroos or other large animals.

However, conclusions cannot be made due to the low number of cases in both categories except perhaps for animal actions. The higher number of "animal" collision incidents under "poor visibility" - and by extension night conditions - clearly indicates that drivers of all vehicles must pay extra attention while driving in the evening.

7.0 The accident experience of Longer Combination Vehicles (LCV's) in inclement weather conditions

Table 1 provides a summary of single trailer versus Longer Combination Vehicle (LCV's) single vehicle heavy truck accidents under icy roads with low and moderate wind conditions (moderate is between 25 to 50 kph). Cases with specialty trailers (flatdecks, tankers, etc.) were excluded since they are not impacted to the same degree by gusting winds as van trailers due to their shape. NOS cases were excluded because the proportion of specialty trailers in relation to van trailers is an unknown, despite the fact that known cases could be used as a proxy. However, since there were only 19 cases of NOS trailer types this would not significantly diminish the value of the analysis.

Table 1 shows that for low winds combined with icy roads, NOS cases account for 63 percent of all accidents. Not only is NOS the largest category from an aggregate perspective, but only for low wind, single trailer NOS incidents do "other" factors account for a greater number, "animal action" being the singlemost reason[6].

Surprisingly, only 7 percent of cases were apportioned to driver errors or condition (inattention, fatigue, etc.). Low wind conditions do not appear to be a significant contributor to incidents, however, they may exacerbate the severity of an incident that may not have otherwise occurred. The number of incidents with empty versus loaded trailers

---

[6] Since NOS cases are speculated to be mostly "run-off-road" cases with minimal damage, this "line item" in table 1 is suspect since it is inconsistent with all other data presented. Drivers may have deflected the blame onto "animal" action from other accident causal factors.

for low wind conditions appears to coincide with the empty backhaul rate for Manitoba carriers (15 to 25 percent empty) depending upon the region.

Table 1: Summary of single trailer versus LCV heavy truck accidents for single vehicle cases with snow/ice covered roads under all visibility conditions (specialty and NOS trailer types excluded)

			Causal Factors		
Winds	Trailer type	Loading	nos	Other	Driver
Low	Single	Empty	3	2	1
Low	Single	Full	21	6	3
Low	Single	Nos	9	10	1
Low	LCV's	Empty	1		
Low	LCV's	Full	4		
Low	LCV's	Nos	6	3	
Sub-total (low winds)			44 (63%)	21 (30%)	5 (7%)
Moderate	Single	Empty	7		
Moderate	Single	Full	15		
Moderate	Single	Nos	17	1	
Moderate	LCV's	Empty	1		
Moderate	LCV's	Full	2		
Moderate	LCV's	Nos	3		
Sub-total (moderate winds)			45 (98%)	1 (2%)	0
Note: "Other" factors are mechanical, fire and animal action					

The number of incidents involving LCV's with low winds and icy roads do not appear to be disproportionate to singles. LCV's account for 10 to 25 percent of truck traffic in Manitoba depending on the season. However, for LCV's 11 of the 14 cases were NOS while for single trailer incidents there were 33 of 56 cases with NOS. While this may appear significant, the low number of LCV cases does not point definitively at this stage to this class of vehicle being more unsafe in the broad sense than single trailer units.

For all cases involving moderate winds, 45 of 46 cases were classed NOS. While it is understandable that animals would not venture forth under more hostile winter conditions, there is a notable absence of mechanical, fire and driver related incidents. One plausible explanation for the absence of driver related factors is that drivers may be more alert under inclement weather conditions, whereas they may be prone to fatigue and inattention during favorable weather. However, the absence of mechanical and fire related cases is perplexing. Except perhaps for vehicle fire in warm weather (due to overheating) cold weather conditions should cause some mechanical problems (frozen brake lines). Non-weather related mechanical problems should occur in equal proportions throughout the year, but this does not present itself in the statistics.

Doubles showed a decline from 11 cases with low winds to 6 cases with moderate winds. This appears to run contrary to the premise that LCV's would be more unstable under windy, icy road conditions than single trailer rigs, and therefore should be over represented in this category of accidents. However, there are two methods that most motor carriers use to deal with this issue. The first is the use of more experienced drivers on LCV's, and the second is instructing all drivers to adjust for conditions or cease operation until the adverse weather has passed. Use of either or both tactics will reduce the probability of an inclement weather accident.

8.0 Summary - Implications of the study for LCV policy, driver training and safety programs

The study findings point to several issues and opportunities to improve motor carrier safety and hence transport efficiency.

The first and foremost is the apparent higher percentage of heavy trucks in inclement weather statistics relative to passenger vehicles. This does not imply that heavy trucks are inherently unsafe, rather, heavy trucks (including other high sides vehicles such as motor homes) require extra precaution under poor driving conditions.

This raises the question of the adequacy of current driver training programs in addressing inclement weather driving skills. Although standardized national driver training programs have modules on dealing with inclement weather, information from this project and subsequent research could enhance course material. The second issue regarding training is the need for advanced driver schools to test classroom theory and skills in controlled settings (truck bashing school). Motor carriers would weigh the cost of sending drivers to advanced schools against the expected benefit. Insurers would need to be involved to promote such schools. Given that the average cost of a heavy truck accident is $100,000 U.S., (*Savage and Moses*) preventing only 5 heavy truck accidents annually in Manitoba could provide the funding base for advanced training.

The second, and at first a surprising outcome was that LCV's don't appear over represented in the accident statistics. However, in conversation with motor carriers, the most experienced drivers are usually assigned to LCV's. This impacts accidents differentially. That is to say, if drivers of equal capability were assigned to all configurations, LCV's should be over represented in statistics. This industry self imposed "graduated licensing" reduces the probability of an accident occurrence and demonstrates potential for productivity gains by improving driver skills.

The use of LCV's can provide not only opportunities to the motor carrier industry, but to shippers and the public by;

1) Reducing the total number of trucks on the road will reduce truck exposure to passenger vehicles, traffic congestion and emissions.
2) Shippers gain by increasing loads in a trip and (possibly) rate reductions. This has enormous benefit for agricultural exports.
3) Motor carriers could afford to pay higher wages to more experienced drivers and thereby ease driver shortages. The reduction in equipment expenditures and insurance exposure is also a benefit.

The deployment of LCV's has been an ongoing debate in most jurisdictions for decades. While infrastructure issues such as bridge and pavement impacts are a main concern of highway officials, efforts by organized public groups to thwart increased use of LCV's have met with some success. However, this may be more of a perception management issue than an engineering one. Advanced computer tools can analyze various mechanical and operational scenarios to ascertain risk.

The third and perhaps mute point are provincial and national heavy truck safety programs that focus on drivers and mechanical condition of the vehicle. Analysis revealed that mechanical failure and driver fatigue is not as significant as it once was in heavy truck accident causation. While these programs have achieved substantial reductions in driver and vehicle related problems in the past, it is suggested that advance weather warning devices, in-cab technology and high performance driver training would be a better avenue to pursue.

9.0 References

Moses, L. and Savage, I.,"*A Cost Benefit Analysis of U.S. Motor Carrier Safety Programmes*", Journal of Transport Economics and Policy, January 1997.

M.E. Wolkowicz and A.M. Billing, *"Commercial Vehicle Accident Survey: An On-the-scene Study"* Ontario Ministry of Transportation and Communications, Report No. RR-235, 1982.

Khasnabis N. and Lyoo S., *"Use of Time Series Analysis to Forecast Truck Accidents."* Transportation Research Record No 1249 Truck Transportation and Safety Issues. 1989.

Mensah A. and Hauer E., *"Two Problems of Averaging Arising in the Estimation of the Relationship Between Accidents and Traffic Flow"* Transportation Research Record No 1635 Highway Safety Modeling, Analysis and Design. 1998.

Simulation of Heavy Trucks in Inclement Weather

Shannon Summerfield[1], Jake Kosior[2]

Abstract

High winds, slippery roads and empty trailers pose a hazardous situation for heavy truck drivers. Unfortunately, not all carriers and drivers fully understand the mechanical dynamics of articulated vehicles under hostile environmental conditions. Also, different vehicles under varying scenarios do not always react the same.

To explore the effects of poor weather conditions on trucks, computer simulations were used to analyze a heavy truck under different road surface conditions, wind speeds, and trailer loadings.

Pulsating wind gusts pose a particular hazard for heavy trucks by a phonomenon called the "child on a swing" effect. That is to say, the trailer motion and winds coincide to create mechanical resonance. This effect can occur with lower winds than previuosly believed but can be corrected by simply slowing down.

While national driver training programs address inclement weather in the classroom, information derived from computer simulation can be utilized by carriers and policy makers alike to assess safety of heavy trucks under various operating scenarios and thereby extract additional effiency gains.

---

[1] Research Affiliate, University of Manitoba Transport Institute
[2] Senior Research Associate, University of Manitoba Transport Institute

1.0 Introduction

It is postulated that high sided vehicles (such as trucks) are inherently unstable under inclement weather conditions, - most notably gusting winds, empty trailers and icy roads. However, not all carriers and drivers are aware of the mechanical dynamics of different vehicle combinations and adjust for the prevailing conditions.

Computer simulation is an economical means to test various operational scenarios. Some of the effects of inclement weather conditions on heavy truck stability are explained and solutions for the prevention of accidents caused by inclement weather are provided.

2.0 Approach

A computer simulation program was created in Excel using Runge-Kutta numerial methods to assess the degree of vehicle instability under different wind speeds, loading and road surface conditions. Vehicle instability was analyzed by monitoring the trailer angle (a.k.a. yaw angle) changing with time. It is suggested that when the trailer angle exceeds 5 degrees in either direction, then the truck becomes unstable and an accident is most likely to occur. Therefore, to determine the effects of inclement weather on heavy trucks, numerous simulations were performed under diverse conditions with attention focused on the yaw angle, especially noting when the angle exceeded 5 degrees.

To calculate the yaw angle θ shown in Figure 1, four moments applied on the trailer first needed to be determined first to build the spreadsheet. A moment causes rotation of an object about a point and is equal to a force acting upon an object multiplied by the perpendicular distance to the point of rotation. With a heavy truck, these four moments combined cause a rotation of the trailer about the kingpin, hence rotating the trailer and creating a yaw angle.

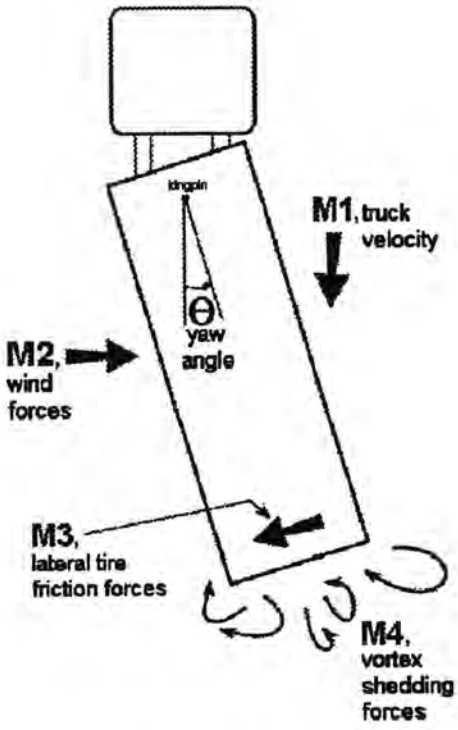


Figure 1: Moments Causing Trailer Yaw

Figure 1 shows the four moments applied upon the trailer. The first moment, M1, is due to the forward velocity of the heavy truck. The second moment is due to the wind forces acting upon the trailer (M2). The third moment is due to the lateral tire friction forces of the trailer (M3). Finally, the fourth moment (M4) applied to the trailer is the vortex shedding of the air behind the trailer (a.k.a. wind turbulence at the rear of the trailer). After assessing these moments and

establishing the mathematical equations, the spreadsheet was constructed.

3.0 Wind Gust Effects

To best simulate a real-life situation where the heavy truck experiences wind gusting, the simulation program was designed to vary the wind speeds randomly with time. Also, analyzing a heavy truck under fluctuating winds will produce much higher yaw angles than a steady wind. In sailing, for instance, when a steady wind is applied, the boat tips to a certain angle and then does not move. However, if the wind is gusting, the boat rocks and can tip over.

This same analogy occurs with the trailer of a heavy truck. When a wind gusts and then recedes, the trailer rocks from one side to the other, rotating about the kingpin instead of rocking back and forth like a boat. Gusting winds can be undesirable because if they are applied at the right time, resonance can occur. If the wind gust is applied just when the trailer is about to switch directions, the amount the trailer will move is amplified. If this occurs more than once in a row, the yaw angle can increase to unstable conditions. This is analogous to a "child on a swing". If you keep pushing a child on a swing just as the child is at the maximum height, very little force is required to maintain motion. This occurrence, called resonance, can render a heavy truck unstable when the motion of a trailer is sychronized with wind gusting.

4.0 Analysis

To determine the effects of inclement weather on a heavy truck, a 48 foot (14.63 m) trailer was analyzed in empty and full loading conditions, along with different wind speeds, and five different types of road surface conditions (from slick ice to dry) shown in Figure 2.

The wind gust range chosen for this analysis was 20 kph. The reason this was chosen is because it is a commonly occurring wind gust range for weather in Manitoba as indicated from Environment Canada

Wind Speed	
Full	*Empty*
20-40 kph	10-30 kph
40-60 kph	20-40 kph
60-80 kph	30-50 kph
80-100 kph	40-60 kph
100-120 kph	50-70 kph

X

Road Surface Condition
Slick Ice
Ice
Snow
Wet
Dry

X

Truck Speeds
100 kph
80 kph
60 kph
40 kph
20 kph

= 250

Figure 2: The different combinations analyzed on a 48 ft. trailer

weather data. Also, higher wind speeds for full loading conditions were analyzed and compared to empty conditions in order to ascertain when truck instability occurred.

5.0 Results

Fundamentally, the largest angle a trailer moves over a time period is a good indicator of its stability in a particular situation. Also, the longer the time period the yaw angle is calculated, the more accurate the results will be. Each program simulation run is 20 seconds. Therefore, to increase the time period considered for each combination, the program was run 100 times for each combination, giving a time period evaluation of 2000 seconds or 33 minutes, 20 seconds. The largest angle that occurred in this time frame was recorded and used in final tabulations.

5.1 "Child on a swing" effect

The yaw angle data collected from the simulation runs proved to be useful in demonstrating resonance of the trailer. The graphs in Figures 3 and 4 show the yaw angle of the trailer varying over time as calculated by the simulation program at wind speeds from 30-50 kph with an empty trailer on slick ice. Comparing the graphs, both occur under the same conditions, but the difference between the two is one shows the trailer experiencing resonance (Figure 3) and the other

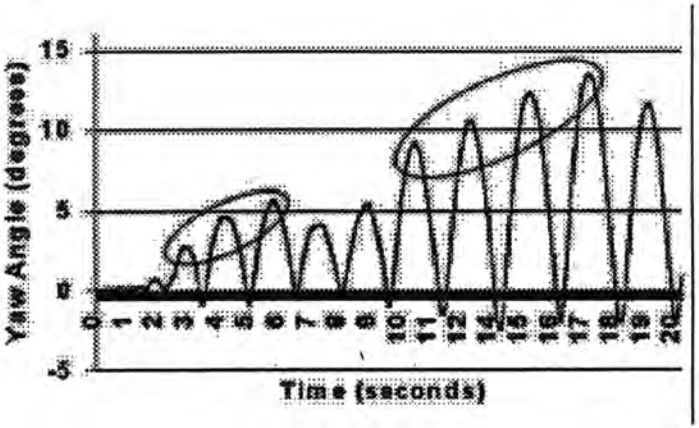


Figure 3: Resonance of the trailer yaw angle resulting from pulsating wind gusts

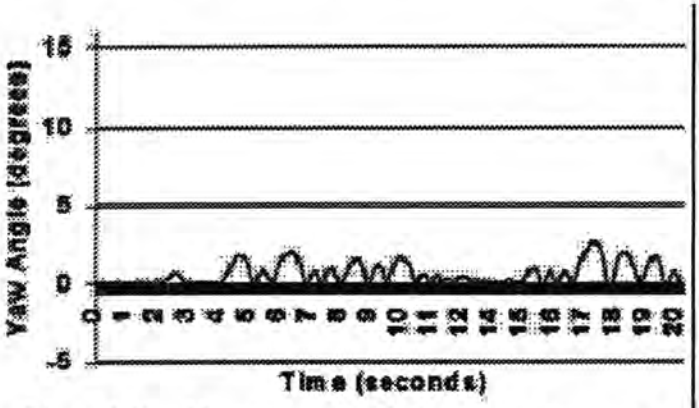


Figure 4: Resulting yaw angle with time and no resonance

shows the yaw angle changing with time with no resonance (Figure 4).

The two ellipses on the graph in Figure 3 show instances where resonance occurs. In these instances, the wind gusts are being applied just when the trailer is switching directions. The graph in Figure 4 shows the yaw angle changing with time under the same conditions as Figure 3. Except this time, no resonance occurs and as a result much lower yaw angles are achieved. In both situations, the size of the wind gusts may be equal, but it is when the gust is applied that makes the difference. If the wind gust is applied just when the trailer is switching directions, this amplifies the effects of the gusts and can result in resonance. This shows that wind gust can behave in such a manner to become synchronized with the trailer motion and increase the probability of an accident.

Therefore, wind gusts and resonance play a major role in limiting the maximum wind speeds a heavy truck can safely drive through. As the wind speeds increase, the more likely the effects of resonance will cause the yaw angle to exceed the safety maximum (5 degrees).

Using maximum safety criteria and the results obtained from the simulations, the table in Figure 5 shows the maximum wind speeds a 48-foot trailer under various conditions could drive at. Remember, these maximum speeds were determined because resonance occurred in the time frame the simulation was run. Chances are, if the simulation was run for a longer time, these numbers would be lower.

	shear ice	ice	snow	wet	dry
empty	20-40	40-60	> 50-70	> 50-70	> 50-70
full	40-60	60-80	> 100-120	> 100-120	> 100-120

Figure 5: Maximum wind speeds (kph) a heavy truck can handle safely with a 48 foot trailer under certain loading and road surface conditions

5.2 Trailer Loading Conditions

The results also showed that increasing the trailer load capacity reduces the movement of the trailer. Figure 6 depicts a graph comparing the maximum yaw angle that occurs with empty and fully loaded trailers at various wind speeds on ice. As can be seen, the yaw angle of an empty trailer increases considerably more with increased wind speed than a loaded trailer put under the same conditions. Similar results occurred for the other vehicle speeds and road surface conditions analyzed. The reason this occurs is adding weight to the trailer increases the friction between the tires and the road surface, hence decreasing the amount the trailer moves in the wind. This is important to understand because it shows that empty trailers under windy conditions are more susceptible to accidents due to inclement weather than loaded trailers.

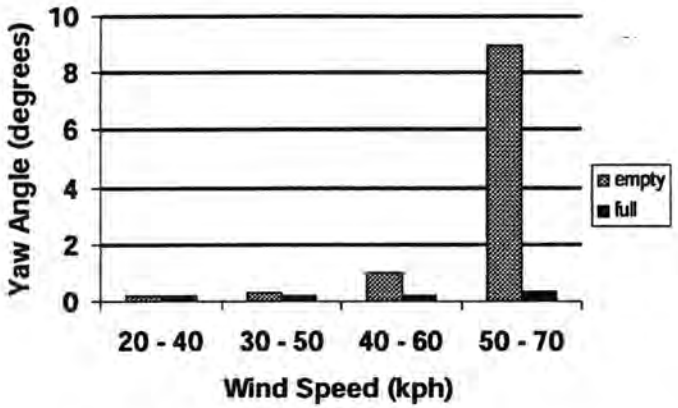


Figure 6: Load Comparison of a heavy truck travelling 100 kph on ice

5.3 Road Surface Condition

Another factor the results showed that is related to friction is the road surface condition.

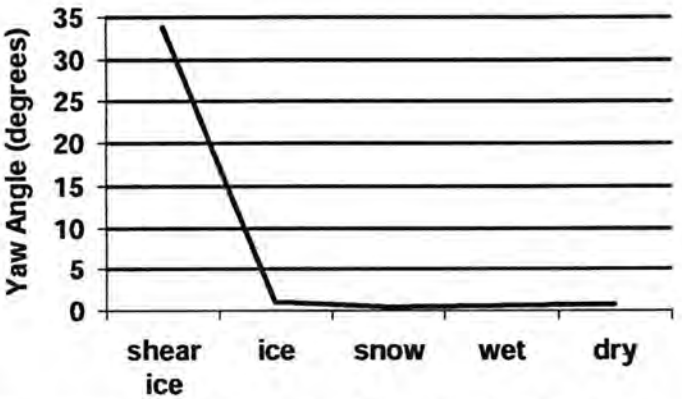


Figure 7: Yaw Angle Changing With Road Surface Condition

Figure 7 shows the yaw angle decreasing with increased friction from the road surface condition (empty trailer, 40-60 kph wind, 100kph truck speed). Similar results occurred with all the other conditions analyzed. Hence, a conclusion can be made that the stability of the truck decreases when the road surface condition goes in the following order of road surface condition: dry, wet, snow, ice, to slick ice.

6.0 Summary

Analysis of a heavy truck under various wind speeds, road surface and loading conditions using a simulation program demonstrated how resonance of wind gusts plays a large role in determining the maximum wind speed a heavy truck can safely drive in. Essentially the higher the wind gusts, the more likely resonance is to occur. This leads to vehicle instability and most likely an accident.

It was also shown that heavy trucks can handle higher wind speeds when the friction between the tires and the road is increased either by increasing the trailer load or driving on a dryer road surface.

7.0 Prevention of Accidents in the Future

Helping drivers understand about the possible negative consequences of driving in inclement weather may incline them to take precautions in these conditions and slow their vehicles down or perhaps cease operations. While this is currently discussed in driver training programs, information gleened from computer simulations could provide visual confirmation to drivers about consequences of certain actions under various conditions.

On the technological side, devices measuring crosswinds and/or yaw angle on the heavy truck are available. Yaw angle devices in the cab notify the driver about impending instability. These devices could be an excellent indicator in telling the driver when to slow down the vehicle. The crosswind measuring device could be used to determine a maximum speed for the truck.

While these suggestions could be an expensive proposition for carriers, the cost must be weighed against safety improvments and hence productivity gains. Furthermore, the cost of installing and maintaining these measuring devices should also be considered by insurers since they benefit society has a whole.

8.0 References

M.L. James, G.M. Smith, and J.C. Wolford, *Applied Numerical Methods for Digital Computation 2nd Edition*, 1977, New York: Harper and Row Publishers.

J.M. Kosior, "*Wind Driven Instability of Tractor-Trailer Combinations*", Univeristy of Manitoba Masters Thesis, September 1989.

The A to DC of Order Fulfillment in E-Tailing

Ignatius V. Hsu, B.Com University of British Columbia

ORDER FULFILLMENT FEVER

The battle for dominance over electronic retailing will take place in an area that few consumers are aware of, the distribution center. The distribution center (DC) is where orders are fulfilled, and this place can be the difference between the success and failure of an electronic retailer's supply chain.

This report deals specifically with the new role of warehouses and DCs as e-commerce order fulfillment centers. It will identify the major issues surrounding the bricks and mortar of the DC and discuss a number of models that are suitable to various electronic retailers. These issues and models will provide a background and framework to guide an e-tailer through the process involved in operating an effective DC. Finally, to serve as a working example of order fulfillment systems, a case study of Toysrus.com is presented. Christmas of 1999 was an unsuccessful season for a number of e-tailers, including Toysrus.com, and the cause was insufficient order fulfillment.[1] Electronic retailers must understand and carefully plan their order fulfillment needs and identify their core competencies, and they must do so in a rapidly changing climate.

"IN E-COMMERCE, THERE'S A VALUE PROPOSITION, YOUR DISTRIBUTION CENTER BECOMES YOUR STORE."[2]

The shopping experience of consumers is changing. Traditionally, a customer fulfilled his own orders at the store by selecting the items from the display racks, purchasing them, then

---

Since this paper's writing, Toysrus.com has changed its operations by arranging a co-branding agreement with Amazon.com. This only underscores the velocity of electronic retailing business. See endnote 33.

delivering the goods by carrying them back home. In online shopping, the online e-tail warehouse completes the order fulfillment process and delivers the product to the customer's home.

Order fulfillment in logistics terms is getting the right product to the right customer at the right time at the right price in the right condition, and the DC holds the key to this. "Fulfillment planning must consider the entire planning process from manufacturing, through distribution and transportation – within a single integrated model."[3]

In referring to order fulfillment here, the term is narrowed to encompass the processes of picking, packing, conveying, sorting and labeling of items at the warehouse. Order fulfillment is already the cornerstone of the Internet strategy for many e-tailers such as Amazon. Businesses must examine their options to optimize the delivery of goods after they are ordered.

The Old and the New: Clicks & Bricks

Online retailers can be classified as either legacies or pure plays. The majority of traditional retail businesses are known as legacies, they operate networks of stores served by regional and national warehouses. Internet start-ups also known as pure plays do not have these established networks. There are benefits to both types of companies and neither has a clear advantage in customer service.

Legacy businesses such as Wal-Mart experienced difficulties in applying their old retailing methods to their e-commerce operations. They faced the major distribution challenge of switching from delivering to retail outlets by the pallet load, to delivering to the customer's home by individual item. At times, 40 percent of Toysrus.com's customers order a single stock keeping unit (SKU).[4]

The e-tailing delivery process involves split case distribution, which means opening cases of consumer goods and storing them at the DC, then picking individual items to fulfill each customer's order. The operational complexity is far greater in split case distribution than in traditional distribution, since order fulfillment is added to the DC's operations. DCs will also need to facilitate the transfer of a large number of small packages to delivery trucks.

Legacy businesses often have the advantage of using their stores for reverse logistics. Their stores can accept returns from customers with convenience that a pure play cannot match.[5] For example, Toysrus.com used its offline retail chain to provide reverse logistics.[6]

Outsiders Outsource Fulfillment

A major issue before an e-tailer can begin taking customers is whether or not to outsource its order fulfillment process to a third party. In order to make this an effective decision, the business must analyze its strategy and determine what its core competencies are. Some e-tailers will decide that their core competencies lie outside order fulfillment, or will outsource for particular business reasons, such as lacking the volume to justify owning a warehouse, or for small to medium sized businesses, having inadequate capital financing. Outsourcing shifts costs from being fixed to being variable, hence reducing risk.

For the delivery of the product itself, the clear majority of online stores rely on a third party. There is usually no competitive advantage gained by using a large and expensive private fleet of planes, trucks and personnel to deliver the product on a national level, and there is sufficient competition and operational coverage within the package delivery industry that outsourcing is usually selected.

There are two main models for outsourcing DC operations, third party operations (third party fulfillment centers) and partner fulfillment operations.

Third Party Operations & Third Party Fulfillment Centers

Third party operations (3PO) and third party fulfillment centers (3PFs) are similar terms used to describe the outsourcing of DC operations and most or all other aspects of order fulfillment.

Though there are some advantages in relying on 3POs, not surprisingly, there are significant drawbacks as well. By combining the needs of several customers, 3POs have higher asset utilization than small businesses operating exclusive DCs do. However, 3POs provide services for a number of clients, and they may be unable to maintain a high level of service for all their clients. Taking this risk

may lead to serious implications during peak periods such as the time sensitive Christmas holiday season.[7]

As a precaution, businesses deciding to use a 3PO should not completely rely on the third party, but should devote staff to manage its relations with the provider. As Jack Horst notes:

> very few providers have developed the experience to do it [operating a DC] better and less expensive over the long run. Organizations considering outsourcing should define their service requirements and consider dedicating their own management resources to oversee their outsourced business.[8]

Partner Fulfillment Operations

Sometimes instead of outsourcing to a 3P0, a company can rely on its partners to perform its order fulfillment, while retaining the delivery function. Partner fulfillment eliminates the e-tailer's need to store inventory. Though there are no inventory carrying costs related to this model, there are high delivery costs associated with sourcing from multiple partners that can make the business uncompetitive.[9] On one hand, similar to 3POs, it does not oversee the fulfillment component and can concentrate on its core competency. On the other hand, the e-tailer is still accountable to its customers for order accuracy and stock-outs on ordered products, a vexing issue that it needs to deal with from time to time.

A number of businesses initially used partner fulfillment and subsequently switched to another model more suitable to their needs. Peapod, a grocery e-tailer that used this model, has since dismantled its partner fulfillment operations and has decided to bring fulfillment in-house by using its own DCs.[10] Barnes and Noble also tried partner fulfillment for its e-tailing division. The legacy book retailer relied on virtual inventories, where its suppliers – the publishers and wholesalers – fulfilled and delivered its customer orders. Soon after, Barnes and Noble realized that it needed to retain control of customer orders by owing its inventory to fulfill online orders.[11] In other words, Barnes and Noble decided that order fulfillment was a crucial element of its operations and was its core competency.

Insiders Fulfill In-House

Like Barnes and Noble, an e-tailer will often find that order fulfillment is its core competency. Customer satisfaction depends on the ability of a business to deliver the product, and considering that repeat customers are necessary for its continued success, order fulfillment often becomes too risky to outsource.

While relying on United Parcel Service (UPS) and the U.S. Postal Service for actual deliveries, Amazon feels strongly about controlling its picking and packing, "we control the customer [satisfaction] when we fulfill the orders in-house."[12] Indeed, 66 percent of Amazon's sales go to repeat customers.[13]

There are three different types of in-house order fulfillment models, build-to-order, distribution delivery centers and dedicated fulfillment centers.

Build-to-Order

Build-to-order (BTO) is the assembly of manufactured components specifically for a customer's order (often used for mass customization). It requires tight and often complex coordination through the supply chain to bring the components in as needed. This makes the business more reliant on its suppliers. It helps reduce the level of inventory, while increasing the value added through disintermediation, that is, bypassing several layers of the supply chain to sell directly to the consumer.[14] Lower inventory costs reduce obsolescence costs by using fewer supplies faster, and disintermediation minimizes obsolescence by cutting the wholesalers and retailers out of the pipeline and by shortening the lead-time.

BTO is the optimal model to consider for evolving technology products, where obsolescence costs are highest. It requires a degree of specialization that makes this model impractical for most e-tailers that have no prior manufacturing expertise.[15] Dell, a personal computer manufacturer, produces computers on a BTO basis, and ships the product to the consumer within a short period of time.[16]

Distribution Delivery Centers

The second option legacy retailers have is to establish distribution DCs that fulfill and deliver online orders from stores

and/or DCs. Legacy retailers can quickly establish online operations without significant capital investments by opting to combine the retail distribution channel with split case distribution and fulfillment for the e-tail channel. As described previously, e-tail warehouse operations including fulfillment are more complex than traditional warehouse operations. The operational complexity of managing both channels together will likely increase from maintaining separate channel operations. As volume grows, the company will need to reconsider the model's suitability.[17]

One anonymous distributor used this distribution delivery center model and encountered difficulties in preserving operations. First, it fulfilled and delivered Internet orders through its stores. As volume grew, it included its regional DCs with its e-tailing operations, using the closest store or DC to complete the fulfillment and delivery. Now it is struggling to integrate e-tailing with its existing distribution system.[18]

A second option within the delivery center model is to only provide online ordering and fulfillment, which requires the consumer to visit the retail outlet. While this saves on delivery and split case distribution costs, the operational complexities will also increase from managing e-tail inventory at every store. This could increase inventory-carrying costs and could confuse store employees that are unfamiliar with picking and packing procedures. In addition, there may be scheduling conflicts between facilitating consumers who shop during business hours and taking care of the fulfillment operations,[19] possibly resulting in delayed fulfillment and a customer that arrives before the order is ready. The Best Buy electronics chain is trying this model of distribution and uses its stores to handle reverse logistics.

Dedicated Fulfillment Centers

E-tailers will often discover that operating a specialized, dedicated center will bring economies of scale and keep warehousing operations from becoming overly complicated by simultaneous offline and online operations. The lead-time from order placement to delivery can be shortened through extensive automation and use of specialized equipment for handling a high volume of small items.

When using a dedicated fulfillment center, a large capital investment is incurred in purchasing the machinery. This requires a high volume of business to bring down fixed operating costs. Committing to a high volume operation is difficult since most e-tailers cannot estimate demand until after beginning operations, this uncertainty is compounded by the wider geographic coverage of the e-tailing DCs compared to traditional retailing.

The system may also be inflexible in supporting a large number of SKUs.[20] This has not proven to be true in many cases, including Toysrus.com, Amazon and Priority Fulfillment Service, which all have a high degree of flexibility built into their warehouses.

The discussion will now move from buying or producing the fulfillment service experience to staking the grounds of the DC.

Location: "Welcome to Memphis, America's Distribution Center"[21]

If the option of keeping order fulfillment in-house is made, the e-tailer must select a strategic location. Delivery sensitive businesses should consider locating near the hubs of overnight couriers, including Memphis – the central hub of Fed Ex, and Louisville – the central hub of UPS. Less time sensitive businesses can afford to locate outside large urban centers and deliver more packages by truck.[22] Locating on the outskirts of urban centers lowers rent while providing access to major markets. There are also benefits of locating near a main supplier, reducing both delivery times to its warehouse and inventory carrying costs. Amazon's original warehouse is in Seattle, where America's largest book wholesaler is located.[23]

Other issues to consider are the number of warehouses needed and the economic climate. Having one central DC instead of multiple ones precludes the complications arising from the division of demand between the facilities, inventory consolidation and expansion effects and extra facility costs. The use of multiple facilities will require analysis of the size of each one, which DC supplies a particular customer, which products to stock at each location, how to manage inventory of each product, and so forth. All these factors will affect the overall customer service level. Economic factors range

from labor costs to taxes and the long-term economic climate. Geography, economy, the size and the number of the facilities are all important to the success of location decisions. A range of the simple and complex software is available to assist in the decision-making.[24]

The type of distribution center that is used is a decision that needs to be considered side by side with the choice of location. These two dimensions are pertinent to a successful planning phase.

Build or Buy: Planning for the Future

The traditional cycle of developing a DC, either using an existing facility of erecting a new one, and installing the equipment, took as long as eighteen months. The velocity of e-commerce business is so rapid that this cycle has been cut to only three or four moths. As e-commerce consultant Patrick Sedlak phrased it, "Put something in operation in three or four months. Make it close, and then I'll worry about it because I don't know what the business will look like in 18 months."[25] This may explain why all five of Amazon's new warehouses were leased instead of constructed from the ground up.[26]

While expediency allows for many opportunities, it also allows for an equal number of mistakes. The most amount of time placed in implementing a system should be invested in the planning stage. Rushing the process may cause costly mistakes that could have a ripple effect through the business's operations.

DISTRIBUTION CENTERS THROUGH THE MICROSCOPE

There has been discussion of whether or not an in-house DC is necessary, the different distribution models, location decisions and the project time frame. This section aims to address reliability, accuracy, speed and co-ordination of fulfillment by examining the anatomy of the DC.

The Warehouse Management System

The warehouse management system (WMS) is an information technology that can potentially coordinate and centralize warehouse operations and improve order fulfillment. The system

tracks the location and stock of all SKUs stored in the DC and plans the activities of the personnel. In a word, it orchestrates.

Many current WMSs are cumbersome. Once an order is placed through the Web site, a facsimile is generated and the employee manually keys the entry into the system. These steps, when repeated hundreds to thousands of times each day, breed inefficiency and human error. The explanation for this archaic procedure is that the Internet system is not compatible with the traditional system. Some estimate that 60 percent of e-commerce orders are manually keyed by DC personnel into the WMS of legacy businesses.[27]

Recent studies show two major problems that an integrated WMS can amend: an average 90 percent picking accuracy and an average eight-day order-to-delivery-cycle.[28] An integrated WMS eliminates human errors and reduces the lead-time of orders. Automation of the ordering process eliminates mistakes made in data entry, potentially improving the DC's order accuracy to 99 percent or higher. The system can efficiently and accurately route workers to pick areas, and can verify orders by radio frequency (RF) based hand held computers. Quicker fulfillment will also lead to shorter order-to-delivery times; for example, Amazon's adoption of RF technology improved its efficiency by 70 percent.[29]

A WMS system that integrates the entire procurement and order fulfillment cycle within the DC can provide much more visibility to actual product inventory and order status of parcels, to both businesses and their customers (see Chart 1). Using the Internet, customers may check the status of their orders, regardless if it is backordered, being picked at the DC, or en route to their home or office via courier.

One can put forward the argument of integrating the WMS of the e-tailer with the WMS of the suppliers to allow total visibility through the pipeline and the supply chain.[30] Interconnectivity can enable real time tracking of orders independent of the physical location of the products. For instance, an out-of-stock item may be arriving at the warehouse the next morning or it may be held up in a Taiwanese factory and would take an extra fortnight. This information is invaluable to a customer considering order cancellation. In reality, this is already common practice in electronic

data interchange (EDI) networks, where corporate clients and suppliers share a standard and interconnected system for placing and receiving orders.

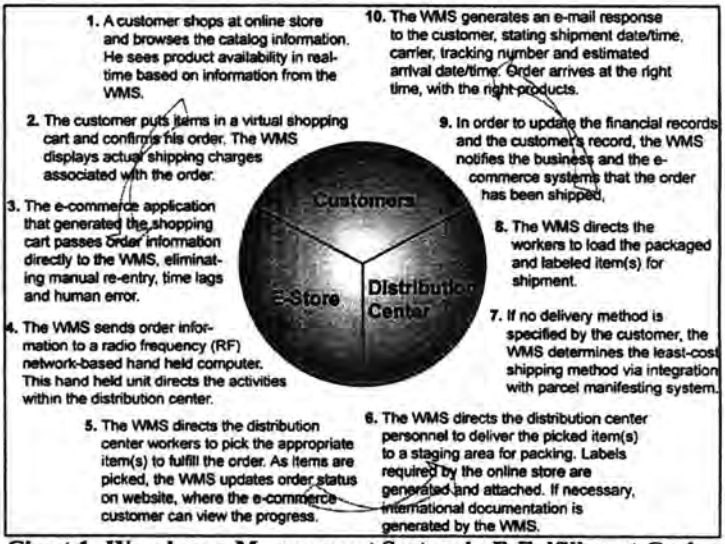


Chart 1: Warehouse Management System in E-Fulfillment Cycles
Source: Material Handling Management. Jan. 2000

One major difficulty of integrating a legacy business's e-tail operations into the retail WMS is the difference in inventory tracking needs. The pallet loads that are scanned and tracked in retail distribution are in contrast to the item by item tracking required in e-tail distribution. The task of merging the two systems is complex, and will require configurable WMS software that can shift between the different needs.[31]

The following section looks inside the DC at its network of vessels – the conveyors and sorters that are used to transport products.

Materials Handling Equipment: Conveyors & Sorters

Materials handling equipment, particularly conveyors and sorters, enable a DC to quickly and efficiently process a large volume of orders. The conveyors only need to convey relatively lightweight cartons and totes (plastic/paper containers used to hold items through processing) as opposed to the case and pallet conveyors found in traditional DCs. However, they must handle large volumes of individual items. It is usually a requirement that the automated system is versatile enough to handle items of disparate dimensions and weights. The ideal conveyor is high speed and can carry light-to-medium weight packages. Some carry totes, some carry small items and prevent them from slipping through the conveying surface, and others provide spacing between the cartons or totes.

Sorters are increasingly important to e-tail DC operations. Where conveyors intersect, sorters redirect items from one conveyor to another. With an appropriate WMS in place, various types of sorters can divert items to "pick zones," to "order accumulation," or to chutes where they are packed. Sorters can reduce delivery costs by automatically directing items to designated dock doors for carriers that offer the best shipping rate for each parcel.[32]

A hybrid combination of conveyors and sorters will achieve optimal results, bringing the right item to the right place and at the right time. Larger operations will require more automation and a wider mix of conveyors and sorter systems than smaller ones. Catalog companies have overcome the many challenges of consumer direct distribution, and have served as effective examples to e-tailers.

The Toysrus.com case study will exemplify the WMS cycle.

TOYSRUS.COM CASE: LESSONS FROM CHRISTMAS PAST[33]

Setting up Shop

The legacy retailer Toys R Us re-engineered its online operations to prepare for the 1999 winter holiday season. It decided not to fulfill online orders through its existing stores or DC infrastructure, choosing instead to separate the distribution channels by operating a dedicated fulfillment center for the e-tail operations.

To reduce setup time in Amazon style, a warehouse in Memphis was acquired from the apparel catalog company Proteam,

which already had a direct-to-consumer system in place similar to that of a typical e-tailer. The versatility and flexibility of the facility's equipment allowed for a relatively effortless transition from sports apparel and equipment to the storage and transport of toys.[34] By locating the warehouse in Memphis, Federal Express's national hub, Toyrus.com could process new orders and ship them the same day until 9 p.m.[35]

Despite the preparation, Toysrus.com was unable to meet the 355 percent climb in sales over the previous holiday season. Just days before Christmas, the e-tailer realized it was unable to fulfill some orders. It notified by e-mail and compensated these dissatisfied customers with 100-dollar gift certificates.[36]

A major problem Toysrus.com and other e-tailers faced was a lack of adequate forecasting tools. At its retail division, it had sophisticated forecasting at the store level based on regional demographics and reasonably predictable historic demand trends. Conversely, its online arm served the entire nation from one DC, leading to unpredictable demand, with unforeseen trends having an exponentially larger impact on its operations. To add to this, nearly two thirds of toy sales are made in the last quarter of the year, giving Toysrus.com minimal room to maneuver with.[37] When it combined inaccurate forecasting with aggressive marketing, the number one toy retailer lost control of its online fulfillment. It learnt its lesson and optimized its operations during 2000, with order accuracy just one tenth of a percentage point below perfect.

Machines that Move: The ABCs of DCs

A common logistics measure is the 80/20 rule, it predicts simply that the top 20 percent of products account for 80 percent of business volume. The sharp difference in popularity of products necessitates the use of special equipment for handling the fast moving 80/20 items and the slower moving items. According to David Maloney, an associate editor of Modern Materials Handling, the Toysrus.com DC "Is designed to hold very deep inventory in the most popular toy lines... the Toysrus.com facility consists of several major areas choreographed to interact depending on volume, SKU, and product need."[38]

The ABC product classification scheme is commonly used to divide products. Products are divided into three categories: 'A' for fast moving products (i.e. 80/20 products), 'B' for medium moving products, and 'C' for small slow moving products. 'A' products are placed on flow racks, that is, racks with rollers tilted to one side so products 'flow' to the rack's edge. The top 300 toys at toysrus.com were stored in flow racks for easy inventory replenishment. An order could quickly be filled from the racks if only toys from this area were requested in an order. The 'B' and 'C' items were placed in active pick zones. For 'C' products, Toysrus.com also used static racks.

Oversize items are items that cannot be conveyed easily, they are placed on special racks, usually accessible by forklift. At Toysrus.com, large items such as bikes were held in the single item shelving area and were shipped separate from the rest of the order. Oversized 'A' items were stored in pallet form on flow racks and were also shipped separately.[39]

Warehouse Management Systems

The Toysrus.com WMS operated in a similar fashion to the process outlined in Chart 1. The Toysrus.com customer could view real time product availability before placing items in a virtual cart and make a credit card purchase. The WMS processed the order, then, via the RF network, proceeded to relay the pick list to the handheld computers of the DC personnel.

The WMS could direct the pickers to fulfill the order in two ways: batch picking that made up three-quarters of the picks, and pick by order that composed the remaining quarter. In a batch pick, the orders were split up by product location within the complex. An employee received a printout of items from several orders that were in the particular zone. Once all the items were picked, they were placed in a tote and sent by conveyor to the corssbelt packing sorter; the sorter subsequently divided the batch into their individual orders.

Toys being toys, it was inevitable that some of the orders contained items of odd shapes, weights or other packaging properties that precluded automated sorting. In a pick by order, these items were picked and placed into a separate tote, one for each order, and

conveyed to the crossbelt sorter. The process was similar to batch picking, only the tote was designated for a single order.[40]

Upon completion, the orders were packed and labeled, then a maze of conveyors and sorters led them to one of the shipping chutes. Depending on the delivery option selected, the parcel would reach an U.S. customer anywhere from 1 to 10 business days. Meanwhile, a unique barcode attached to the package was traced, and the location and inventory status were updated on the WMS, which in turn updated the Web site database for the customer to view.

CONCLUDING REMARKS

The report has identified a framework that electronic retailers can use when establishing their fulfillment network. An overview of the major issues were provided, from the planning stage, with a discussion on outsourcing and the main warehouse models, to the implementation stage, with discussions on materials handling equipment and warehouse management systems.

To weave the concepts together, the experiences of Toysrus.com were given in a case study. The dedicated distribution model adopted by the toy e-tailer seemed to be a very popular and fairly successful e-commerce model. The case exemplified how a warehouse management system operated in tandem with the material handling system.

Toysrus.com's DC operations were similar to its catalog company counterparts. This is not to say that converting a catalog operation into an online one is a recipe for success, or experienced legacy retailers the size of Toys R Us would not have faltered in their online ventures. The point is to understand the framework, and to apply it with a clear understanding of the core competency of the business and the necessary investments involved in online retailing. The rules of the game are still taking shape, so flexibility and adaptability are requirements. One need only look at Toysrus.com's recent co-branding arrangement with its former rival, Amazon.

Electronic retailers are expanding across the nation, the question now is "can they turn a profit?" With the assistance of proper design, operation and control of the distribution center, there is potential.

ENDNOTES

[1] David Maloney, "E-Biz R Us," Modern Materials Handling, Feb. 2000: 44-51.
[2] Steve Alexander, "E-commerce Distribution," Computer World, 20 Mar. 2000: 58.
[3] Fred R. Ricker and Ravi Kalakota, "Order Fulfillment: the Hidden E-Commerce Success," Supply Chain Management Review, Fall 1999: 60-68.
[4] David Maloney, "E-Biz R Us" 49.
[5] Steve Alexander 58; and "E-Commerce Survey," Economist, 26 Feb. 2000: 15-27.
[6] After its co-branding agreement with Amazon in the summer of 2000, reverse logistics fell under Amazon's postal return system. See Molly Prior, "TRU.com Hopes Amazon Can Help Replenish Image," DSN Retailing Today, 5 Aug. 2000: 1, 46; and Cheryl Rosen, "Ready for the Returns?" InformationWeek, 8 Jan. 2001: 22-24.
[7] "E-Commerce Survey," Economist 32.
[8] "Go for the Gold with Logistics X-Training," Sporting Goods Business, 1 Feb. 2000: 19-21.
[9] Fred R. Ricker and Ravi Kalakota 65.
[10] "Peapod Will Open San Francisco DC; Ends Safeway Deal," Frozen Food Age, Apr. 1999: 44.
[11] Tom Andel, "Logistics @ Barnesandnoble.com" Material Handling Management, E-Commerce: Fulfillment. Jan.2000: 39.
[12] James Aaron Cooke, "Clicks and Mortar," Logistics Management and Distribution Report, Jan. 2000: 36-39.
[13] "E-Commerce Survey," Economist 24.
[14] Fred R. Ricker and Ravi Kalakota 64-66.
[15] Note that volume-gaining products such as knockdown furniture are the exception, where simple assembly techniques can be completed at the distribution center before product delivery (or in the case of Ikea, the consumer completes assembly). Here, the idea behind delaying assembly until the last possible moment is to reduce warehousing costs.
[16] "E-Commerce Survey," Economist 6-15.
[17] Fred R. Ricker and Ravi Kalakota 60-68.
[18] Gary Forger, "Distribution only as Good as its Weakest Link," Modern Materials Handling, June 1999: 55-59.
[19] Fred R. Ricker and Ravi Kalakota 64-66.
[20] Fred R. Ricker and Ravi Kalakota 64-66.
[21] The sign at the Memphis International Airport.

[22] Jim Schriner, "E-Commerce: Prepare for Location Decisions," Industry Week: Fulfillment, 6 Dec. 1999: 20.
[23] "E-Commerce Survey," Economist 6-15.
[24] Ronald H. Ballou, Business Logistics Management, 4th ed. (New Jersey: Prentice Hall, 1999) 483-524.
[25] Patrick Sedlak, "E-Commerce: Speed at all Costs," Material Handling Management, E-Commerce: Fulfillment, 6 Dec. 1999: 40.
[26] James Aaron Cooke 36-39.
[27] Gary Forger 55-59.
[28] Christopher Trunk, "Help Wanted: WMS Solves E-Store Problems," Material Handling Management, Jan. 2000: 44-46.
[29] James Aaron Cooke 36-39.
[30] Gary Forger 55-59.
[31] Gary Forger 55-59.
[32] David Maloney, "E-Conveyors," Modern Materials Handling, Jan. 2000 49-55.
[33] Shortly after this paper's writing, Toysrus.com announced it would co-brand its site with the pure play Amazon. Under the terms of the deal, they would concentrate on their core competencies. Toysrus.com would handle marketing and procurement, and Amazon would handle order fulfillment from its seven distribution centers. The deal would transfer fulfillment duties to Amazon, but also help both online operations become profitable faster. The three Toysrus.com DCs were to be sold (apart from Memphis, Toyrus.com also acquired 2 additional warehouses between April and August 2000, when the deal was announced). See Molly Prior, "TRU.com," 1, 46; & "Toysrus.com adds two new DCs," Warehouse Management, May 2000: 14.
[34] David Maloney, "E-Biz R Us" 45-47. (Note that there are conflicting accounts, Carol Sliwa, "Holiday Lessons," Computerworld, 10 Jul. 2000: 42.)
[35] David Maloney, E-Biz R Us" 47.
[36] David Maloney, "E-Biz R Us" 44-51; and Carol Sliwa 42. The Federal Trade Commission fined Toysrus.com $350,000 for its fulfillment failure that led to violations of three sections of the "Mail and Telephone Order Rule" during the 1999 holiday season. See Federal Trade Commission, "Seven Internet Retailers Settle FTC Charges Over Shipping Delays During 1999 Holiday Season," 26 July 2000 <www.ftc.gov/opa/2000/07/toolate.htm>.
[37] Molly Prior, "FAO Schwartz Upgrades Site, Gears up for the Holiday E-tail Battle," DSN Retailing Today, 4 Sep. 2000: 6, 46.
[38] David Maloney, "E-Biz R Us" 45.
[39] David Maloney, "E-Biz R Us" 45-48.
[40] David Maloney, "E-Biz R Us" 45-48.

The Civil Aviation Industry: A Discussion of Air Pollution Problems

By Samantha Sawyer
Bachelor of Arts May 2000, Major in Economics
University of British Columbia

Introduction

The civilian aviation industry has expanded from a highly specialized, expensive mode of travel to a relatively affordable means of long-distance transport for a large number of people. Commercial aircraft, because of their size and high frequency of use, are the main source of emissions even though they make up less than half than all the aircraft in operation (FAA, 1999 p. 2-139).

Air transportation's negative environmental impacts on air quality, water quality, and noise levels have long been recognized. New concern has arisen over high altitude emissions and their intensifying global warming and ozone depleting effects. Recent forecasts of huge growth in the aviation industry suggest the pollution problem will only worsen.(TDC, 1996, p.vii).The required emission reductions can be encouraged by emission standards and taxes but their application is hampered by legal barriers, cost considerations, and international obstacles. Consequently, in the face of uncertainty, there are even potential benefits from delaying regulation.

Aviation Pollutants

The public debate on the aviation industry's contribution to air pollution and climate change is loud and raging with due cause. The main pollutants generated by the air transport industry, particulates (PM-10), carbon monoxide (CO), volatile organic compounds (VOC's), unburnt hydrocarbons (HC), and oxides of nitrogen (NOx), have drastic impacts on both the environment and personal health (Transportation Development Centre (TDC), 1996 p. 28).

Particulates, (PM-10) are fine particles less than 10 micrometers in diameter, like dust, fog or fumes that form as a result of incomplete combustion (Draper and Pernigotti, p.5). They can cause lung problems, cardiovascular disease and weaken the immune system (Draper and Pernigotti, p. C-5). CO, VOC's and PM-10 occur too at higher levels during aircraft taxiing, when engines are on lower power settings (Draper and Pernigotti, 1999). NOx refer to two types of compounds, nitrous oxide NO, and nitrogen dioxide, NO_2. These emissions are created during high-temperature combustion processes, like takeoff and climb-out (Draper and Pernigotti, 1999) and are the primary emission of aircraft by weight (Emmanuel, 1993, p. 59). In the presence of sunlight, at low levels in the atmosphere, NOx, VOC's and HC, together can form ozone, O_3, an irritant gas causing smog (TDC, 1996, p.28). The long-range health effects from this gas can cause throat, eye, and nose irritation, as well as coughing and breathing problems related to reduced lung capacity. (Draper and Pernigotti, 1999).

Aviation emissions also contribute to global warming. Greenhouse gases, CO_2 and H_2O, are produced in fuel-combustion (TDC, 1996, p.24). Scientists warn that if the percentage of greenhouse gases in the atmosphere becomes too high then drastic climate changes will occurr (TDC, 1996, p. 24). According to the Transportation Development Centre of Canada, between 2.5 and 3% of world carbon dioxide emissions from fossil fuel combustion are

attributed to air travel alone. This represents roughly 1.25 to 1.5% of global warming from human activities (TDC, 1996, p.24).[1]

New concern has arisen that *cruising level* aircraft-emissions can have intensive greenhouse effects. NOx emissions above cloud level undergo reaction processes that form global-warming compounds (TDC, 1996, p.28). Water vapour from engines that form clouds of ice crystals, called condensation trails, are thought to add to global warming (TDC, 1996, p.28). Dietrich Brockhagn, a physicist studying air transport, claims that for each kilogram of fuel burned in aviation there is 3 times the heat effect associated with ground transport (Ambrogi, 1997). Brockhagn maintains that aviation has a high pollution factor because it injects pollutants 10 to 30 km into the upper troposphere and lower stratosphere (Ambrogi, 1997). This viewpoint is in accordance with statistics from a report, entitled "Indicators of Environmental Impacts of Transportation," that claim NOx emissions at high levels have 50 times the global warming effect that NOx emissions have closer to the earth's surface (1996, p.162).

Another concern, in addition to global warming, is ozone depletion. At above 12 km, NOx can have the opposite smog-producing effect that it has at lower levels, and actually reduce the amount of ozone in the atmosphere (TDC, 1996, p.28). Problems associated with ozone depletion include, environmental degradation and health deterioration due to increased ultraviolet radiation. Ultraviolet rays are responsible for skin cancer and cataracts, and if the intensity of rays increases, there will be a higher frequency of these types of afflictions (TDC, 1996, p.26). The ozone problem is expected to reduce crop yields and damage aquatic ecosystems (Block, 1990, p.182).

---

[1]. This contribution to global warming must be put in perspective. Manmade emissions, of which aircraft produce roughly 3%, are only 3% of the world total CO2 in existence. Therefore aviation accounts for just 3% of 3%, of CO2 in the atmosphere (Ambrogi, 1997).

New Technologies vs. New Procedures

There are two approaches to reducing emissions of these harmful pollutants. Entirely new technology with lower emitting engines can be developed or, alternatively, the existing equipment can be used more efficiently to reduce pollution (Environment and Energy, 1996). The former approach, inventing new designs for engines, has reduced the emission of HC's, CO_2, and NOx significantly since 1970 levels (TDC, 1996, p.29). The new engines require less fuel per kilometre of distance, so there is less fuel consumption per flight, thereby reducing emissions. The accuracy of the technology has also improved. For example, the amount of fuel burn-off during take off can be more precisely predicted, so the exact amount of fuel needed for each flight is known and less fuel will be wasted (TDC, 1996, p.25).

There are drawbacks to the new engine designs of the last decade. Noise pollution reduction-technology, with the installation of a heavy "hush kit" to silence the engine, actually decreases fuel efficiency because it makes the aircraft expend more energy in take-off and climb out modes (Ambrogi, 1997). Furthermore, major improvements in engine design have resulted in a higher percentage of NOx emissions. The HC and CO reduction technology actually produces more NOx emissions on a percentage basis because of the new high-temperature combustion chamber (TDC, 1996, p.29). According to a study by the International Air Transport Association, NOx compose 12 % more of emissions in 1990 as compared to 1982 levels (Michal, 1993). However, due to the increased fuel efficiency, there are less overall emissions of NOx (TDC, 1996, p. 29). Recent studies have completed high altitude testing in the hopes that a lower NOx emission jet engine combustor can be designed (Environment and Energy, 1996). A resulting 40-50% decrease in NOx emissions are forecast by the year 2010 (TDC, 1996, p.26).

Operational procedures and air traffic management can also be developed to reduce emissions. First, flight paths are criticized for their indirect routing that unnecessarily increase flight time and emissions (Parker, 1999). They could be redesigned to minimize distance, however this would probably be a very costly activity because of the required changes to air traffic control operations.

Second, to reduce tarmac and runway emissions, pilots are encouraged to taxi and idle with the least use of engines possible (Draper and Pernigotti, 1997, p. 45). Pilots are also encouraged to derate in-flight engine power in appropriate weather conditions to lower NOx emissions. (Draper and Pernigotti, 1997, p. 46)

Lastly, the "reverse thrust" technique in aircraft landing, where engines are used to slow the aircraft quickly, has been discouraged because of the increased NOx emissions produced (Draper and Pernigotti, 1997, p. 46). The authors of "Air Quality Procedures for Civilian Airports and Air Force Bases" note that this alteration has several drawbacks. First, reverse thrust on the runway can actually reduce the HC emissions associated with taxiing. The technique reduces each airplane's average time on the runway by decreasing the number of missed runway turnoffs and speeding up each air plane exit off the runway. Reverse thrust also lowers the maintenance costs of brake repair and replacement (Draper and Pernigotti, 1997, p.46).

A study, released June 1999, by the Intergovernmental Panel on Climate change, based in Geneva, claims that improvements in operational and technological advancements have the potential to reduce emissions by 18% (Parker, 1999). There is thought that these measures may not have the suggested pollution reducing impact after all. Great environmental concern surrounds the predicted growth of the aviation industry. Rising GDP and trade, in combination with the reduced costs of travel, suggest a continued rate of high growth for aviation into the next century (TDC, 1996, p.7). Researchers of the Geneva-based study claim air travel will increase roughly 5% annually until 2015 (Parker, 1999).

Congestion at airports is already a major emission-producing problem. Swiss Air, estimated the excess fuel consumed due to congestion-related delays as roughly 1% of its total fuel consumption. This represented an extra 170 tonnes of NOx emissions, as a result (TDC, 1996, p.28). Air traffic delays at London's Heathrow and Gatwick is said to cause the use of an extra 60, 000 tonnes of fuel for British Airways. Half of this quantity is attributed to holding delays on the runway prior to take off and half is required as extra fuel in case of delays for permission to land.

With demand for aviation set to grow, can the development of new technologies and operational procedures stabilize emissions, much less reduce them? The study on climate change concluded that the outlined changes will not likely offset the increased emissions resulting from the expected expansion of the industry and recommended emission standards as the "initial answer"(Parker, 1999).

Emission Standards

With emission standards, airlines would face hefty fines for aircraft that surpassed allowable NOx or CO_2 emissions. Airlines would then have strong incentives to develop pollution reduction technology and to invest in it when it became available. This would compel aviation firms to recognize the so-far-ignored social costs of pollution (Draper and Pernigotti, 1997). However, environmental standards pose problems in their market application. There are significant legal barriers, inefficiency considerations, and international obstacles that make the implementation of emission standards and taxes very difficult.

First, airport charges on polluting aircraft are difficult to impose because of legal tactics taken by airlines to dismantle them, as found in Switzerland and Sweden (Feldman, 1998). Switzerland, the first to try to raise levies on polluting aircraft, were charged with discriminatory practices by IATA. The levies were never fully implemented because IATA obtained a temporary injunction. Likewise, in Sweden, the proposed emissions tax was stopped because they supposedly violated the "Treaty of Rome" obligations. The mandatory emission standards necessitate the use of pollution reducing technologies, no matter what the cost to firms in the industry (Block, 1990, p.186). Therefore they generate huge protest from aviation interests who fear diminished profit because of the expensive new technology that will be required.

Second, emission standards can generate gross inefficiencies. Once firms have to reduce pollution to a specific level, there is no further incentive to strive for lower emissions (Siebert, 1998, p. 131) Two suggestions to deal with this inefficiency, an emissions tax and emissions trading, are explained.

An emissions tax can be set in addition to an emission standard, whereby firms polluting below the standard pay less tax the higher the level of their abatement. This gives firms the incentive to reduce their emissions even after they meet the regulated standard (Siebert, 1998, p. 147). A second option is the sale of transferable pollution licences that give airlines rights to emit pollutants (Parker, 1999). The number of licenses is restricted to achieve a quantity of pollution that is deemed tolerable for the environment (Siebert, 1998, p. 138). If some airlines are able to reduce emissions, they can sell their pollution rights to other firms for which the costs of pollution abatement are very high (Waters, 1999, XI-6).

With emissions trading, firms that can reduce emissions cheaply will do so, and firms that can not, purchase extra permits. This method is thought to lead to optimum prices for emission rights where the amount of resources directed towards pollution abatement is minimized but a sufficient amount of emissions are reduced (Siebert, 1998, p. 143). These tactics, suggested by the ICAO study, in theory, would eventually lead to a decreasing marginal cost of abatement in the industry and lessen the inefficiency associated with the setting of emission standards.

The third obstacle in the way of emission standards resides in the nature of international relations. The hidden characteristics problem, where each country inaccurately reports its current emission levels, makes regulation across national borders very difficult (Campbell, 1995, p. 8). If there is a tax on emissions above a set quota, then countries have the incentive to report a higher than actual emission rate. Another problem related to international emission standards, is lack of enforcing power. Avi Gil, director of the environment for Airports Council International, says,"In the absence of an international regime which has some teeth to it, we are starting to see countries going their own way." (Ambrogi, 1997) We have seen that Switzerland and Sweden both made an attempt to impose their own emission standards. Agreements and laws between nations cannot be strictly enforced because there are no binding consequences to shirking (Waters, 1999, X1-9).

Despite the recognized obstacles, there is a big push for the International Civil Aviation Organization, to enforce global standards

(Feldman, 1998). New airborne sampling methods were used to track North American pollutant movement patterns, in the 1997 North Atlantic Regional Experiment (Scott, 1998, p.56). Findings indicated that a country's air pollution eventually becomes its neighbour's air pollution (Scott, 1998, p. 56). The issuing of emission controls on an international scale is deemed essential due to the "global consequences" of pollution from aviation (TDC, 1996, 58).

Lastly, there are valid concerns about immediately implementing any sort of emission standards or environmental taxes due to uncertainty. The lack of completely accurate knowledge about the greenhouse and ozone depletion phenomena means that delaying policy implementation may actually be beneficial. It is possible that new information may be learned that will make pollution reduction less costly (Block, 1990, p.181). This results from the quasi-option value, the expected value of new information (Shaffer, 1999), associated with the possible policy or, in this case, emission regulations. The expected value of new information depends on the probability of there being new information and the cost savings of the alternative options, made possible by the information (Shaffer, 1999). This may be a factor influencing the proceedings of various conferences throughout the 1990's. The conferences agreed on a need for emissions standards, but failed to set any (Ambrogi, 1997).

Conclusion

There are significant challenges facing the aviation industry. Environmental progress has been realized in terms of aircraft engine-emission reductions, however, with the expected boom in demand for air travel and the consequent congestion, this may cease. The report, "The Greening of Aviation" advocates an approach to growth that takes environmental quality into consideration.

> "We have entered an age in which industrial growth must be considered in the light of global sustainable development. That is, development must progress in a manner that does not exceed the future capacity of the environment to sustain itself and its human populations"(TDC, 1996, p. 57)

Whether this feat is possible depends on the industry's ability to resolve the technology trade-offs, reduce the operational inefficiencies and overcome the policy implementation obstacles.

Bibliography

Ambrogi, Stephano. "Action stations on the Environment Front". *Air Finance Journal.* (198) (1997): 74-76.

Block, Walter, ed. *Economics and the Environment: A reconciliation.* Canada: The Fraser Instititute, 1990.

Campbell, Donald E. *Incentives: Motivation and the Economics of Information.* Cambridge:Cambridge University Press, 1995.

Crayston, John. Civil aviation and the environment. *Industry and the Environment.* 16(1-2) (1993): 51-54.

Draper, Webb and Augustine Pernigotti, and Liang Plante. "Air Quality Procedures for Civilian Airports and Air Force Bases" Bureau of Transport Statistics April 1997. Online www.bts.gov/ntl/data pdf file: [aq-aq-hdbk] Nov. 26, 1999.

Duraiappah, Anantha K. *Global Warming and Economic Development: A Holistic Approach to International Policy Co-operation and Co-ordination.* Netherlands: Kluwer Academic Publishers, 1993.

Federal Aviation Association. *Environment and Energy.* Online. www.nasdocs.faa.gov/nasiHTML/960307/RED96/9red9.html Nov. 26, 1999.

Federal Aviation Administration. *1999 National Aviation Research Plan.* Online. http://nasdocs.faa.gov/nasiHTML/Red/026env.pdf Dec.1, 1999

Feldman, Joan. "We are all green now". *Air Transport world.* 35(2) (1998): 70-75.

Forrest, Anne S., Thomas M. Selden and James E. Lockhart. "Analyzing the Reductions in US Air Pollution Emissions: 1970 to 1990" *Land Economics.* 75(1) (1999): 1-21.

Indicators of Environmental Impacts of Transportation: Highway, Rail, Aviation and Maritime Transport. United States Environmental Protection Agency. Oct. 1996
http://www.bts.gov/ntl/data Pdf file [indicall/pdf] Dec. 1, 1996.

Levinson, David, David Gillen, Adib Kanafani and Jean-Michel Mathley. "Full Cost of Intercity Transportation: Comparison of high speed rail air and highway transportation in California" University of Berkely June 1996. Online. http://www.bts.gov/ntl/data
Pdf file: [ucb-its-rr-96-3.pdf] Nov. 26, 1999.

Michal, Emmanuel. "The International Air Transport Association and the Environment" *Industry and the Environment.* 16(1-2) (1993): 59.

Parker, John. "Aviation tied to pollution" *Traffic World.* Vol. 258 no. 11 (1999): 43.

Scott, William B. "Pollutants tracked Over North Atlantic" *Aviation Week and Space Technology.* 148(16) (1998): 56.

Shaffer, Marvin. "Econ 370 Online Notes" September 1999. Online. http://WEB.ARTS.UBC.CA./econ/Shaffer/CLASSN~1.htm Oct. 13, 1999.

Siebert, Horst. *Economics of the Environment: Theory and Policy.* Berlin: Springer-Verlag, 1998.

Somerville, Hugh. "Airlines, aviation and the environment-the British Airways programme. *Industry and the Environment.* 16(1-2) (1993): 54-59.

Transportation Development Centre of Canada.(TDC) *The Greening of Aviation.* January 1996.

Waters, W. G. "Externalities and Market Failure: Environmental Issues and Transportation". Econ 480 Lecture Note Package, University of British Columbia, Vancouver, Nov.22 1999.

A THEORETICAL ANALYSIS OF LOW-COST ENTRY IN THE AIR TRANSPORT INDUSTRY

By:
Bonnie-Lynn Robertson
The University of British Columbia

EXECUTIVE SUMMARY

The last three decades have seen a dramatic move towards air transport deregulation in many nations. One common element in this new era is the proliferation of low-cost airlines. Therefore, academic research on low-cost competition has public policy implications worldwide.

Evidence from work by Whinston and Collins, (1992); Dresner and Windle (1995); Richards, (1996); Dresner, Windle, and Lin, (1996); and Dresner and Windle, (1998) indicates that low-cost competition greatly influences fares and traffic levels. These empirical authors have not, however, included theoretical support for their results. Using the theory of new industrial organization, Levine (1987) has provided a theoretical background. However, the concepts of new industrial organization; information costs, the principle-agent problem, and transaction costs lack traditional industrial organizational modelling.

This paper draws from the two versions of industrial organization, new and traditional, to explain the empirical results. It draws upon new industrial organization to describe the lack of contestability in the airline industry. In addition, it uses concepts from both new and traditional industrial organization to outline the motivation for the structure and product differentiation seen today. Both theories have, also, been used to explain the competitive influences that have resulted in the price and traffic fluctuations found in the empirical data. In short, this paper offers a thorough theoretical analysis of low-cost airline competition. The motivation is to give a complete understanding of low-cost airline competition.

A Theoretical Analysis of Low-Cost Entry in the Air Transport Industry

In 1978, the United States liberalized domestic airline competition. Canada followed in 1988. The European Union completed its deregulation in 1997 (Graham, 1997b) and other markets are following. One common result has been the proliferation of low-cost airlines such as Southwest Airlines in the U.S., WestJet in Canada, and EasyJet in Europe. Liberalization leads to low-cost airlines, and low-cost airlines revolutionize air travel. Since liberalization is spreading worldwide, this research has policy implications everywhere.

A sea change in academic thought has questioned the efficiency of regulation. The profitability of low-cost carriers in the less-regulated California and Texas markets began this inquiry. By the late 1970's, analysts thought deregulated airlines would approximate a competitive equilibrium.[1] However, initial results were inconsistent with this prediction. Mergers resulted in a small number of competitors, a characteristic of imperfect competition. It was suggested that the number of competitors did not matter, as long as markets could be entered freely (Levine, 1987).[2] This contestability model was accepted by most analysts until it was challenged by Michael Levine in 1987. Using new industrial organization theory, Levine (1987) explained the industry's deviations from perfect contestability. Subsequent empirical research has confirmed that air transport markets do not conform to the contestability model. Further research on low-cost airline competition by Whinston and Collins (1992) and Richards (1996) support this view.

Additional empirical work concluded that low-cost competition reduces fares by 34%-69% (Whinston and Collins, 1992; Dresner and Windle, 1995)[3] and increases traffic by 74%-300% (Dresner and Windle, 1995). It follows that low-cost competition is assumed to improve consumer welfare (Dresner and Windle, 1995).

Even the threat of low-cost entry lowers fares (Whinston and Collins, 1992 and Richards, 1996). In addition, spillover effects from

low-cost competition are found to be present on non-competitive routes at the airport entered and on competitive routes at neighbouring airports (Dresner, Windle, and Lin, 1996). Moreover, an incumbent competing with a low-cost carrier at its hub lowers fares on entered routes terminating at the hub. It, also, reduces fares on one-stop routes that flow through the hub if low-cost competition is on one, or both, of the spokes. However, the incumbent does not raise fares on non-competitive routes to, or through, the hub (Dresner and Windle, 1998). Exit is, also, preceded by a fall in prices of 27 percent and a decrease in concentration of 17 percent. One year after exit, fares were not found to have risen (Dresner and Windle, 1995). These results suggest that consumer gain may be greater than the direct effect associated with route entry (Dresner and Windle, 1998). Consequently, Dresner, Windle and Lin (1996) suggest that public policy should encourage the expansion of low-cost airlines.

These empirical authors have not, however, included theoretical support for their results. Using the theory of new industrial organization Levine (1987) has provided a theoretical background. However, the concepts of new industrial organization, information costs, the principle-agent problem, and transaction costs lack traditional industrial organizational modelling. This paper draws from the two versions of industrial organization, new and traditional, to explain the empirical results. The motivation is to give a complete theoretical understanding of low-cost airline competition. As most research has focused on the U.S. domestic industry, the examples given will be American. Specifically, many will be of Southwest Airlines. Because most of today's entrants emulate Southwest, these examples have applications for future low-cost competition. Moreover, the results and theories have worldwide implications.

The paper will proceed as follows. Section 1 will review the evidence refuting perfect contestability in the air transport industry and will draw upon the new theory of industrial organization for explanation. Section 2 will use both traditional and new industrial organization to describe the motivations for entry and product choice in the industry. Using new industrial organization, section 3 will outline the forces behind low-cost competition. Next, Section 4 will explain the

competitive results found in the empirical works using traditional industrial organization. A conclusion will complete the paper.

Section 1: Contestability and New Industrial Organization

The theory of contestability proposes that, under certain conditions, a natural monopoly can reach an efficient equilibrium without regulation (Baumol, Panzar, and Willig, 1982). The defining feature of contestability is the presence of costless entry and exit at an efficient scale. Conditions for this to exist are: (1) equal access to technology and economies of scale; (2) the absence of sunk costs so that a firm can enter and exit a market costlessly; and (3) the sustainability of prices i.e., a set of prices exists that will support the operation of at least one firm (Levine, 1987). Although the air transport industry is not a natural monopoly, contestability was thought to apply to city-pair markets that involved only a few carriers.

If markets are contestable, the threat of hit-and-run entry will force firms to lower prices to their competitive level and, accordingly, receive no economic profit (Tirole, 1987). Initially, analysts felt that airline markets were contestable because airplanes were "capital on wings" and could facilitate costless entry and exit. If correct, the level of market concentration would not affect fares (Richards, 1996). However, preliminary research does not support the contestability model of airline competition (Levine, 1987). Moreover, the low-cost entry papers of Whinston and Collins (1992) and Richards (1996) demonstrate the lack of contestability in the following ways.

Whinston and Collins (1992) evaluated stock price reactions to the announcement of low-cost entry into airport-pair markets. They found that incumbents in the markets entered only felt losses. These losses provide evidence of the degree of profits, or sunk costs, in these markets and, as such, support the view that air transport markets are not perfectly contestable (Whinston and Collins, 1992). In contrast, Richards (1996) analyzed the actual, potential, or absence of entry by a low-cost carrier. Her findings show that incumbent pricing strategies differ depending on whether the low-cost airline is an actual, a potential, or not a potential competitor on a route. Again, this differing pricing strategy is not consistent with contestability (Richards, 1996).

Levine (1987) has explained the deviations from contestability using factors not readily considered in traditional industrial organization. After deregulation, holdover firms were disadvantaged (in comparison to new entrants) because of retained high costs such as committed costs, expensive labor contracts, purchase debts, and long-term leases, that were incurred during regulation. To survive with above- market costs and yield normal returns, holdover firms had to generate rents. To do so, they pursued revenue-generating strategies that were not possible for smaller entrants. At the core of these strategies are the new industrial organization ideas of information costs, the principle-agent problem, and transaction costs. These ideas led to such deviations from contestability as mergers, vertical integration, hub and spoke networks, complex fare structures, frequent flyer plans, travel agents' incentives, computerized reservation systems, slots and gate controls, predation tactics, and new entrant casualties (Levine, 1987).

Air transportation is greatly affected by the economies of information. Specifically, airlines incur large costs in developing and communicating information about routes, schedules, services, and prices. The quasi-public good nature of information (once produced it may be supplied at almost no incremental cost) can greatly affect airline competition. Furthermore, information costs and economies of scope and scale exist in generating a reputation as a fierce competitor among actual or potential entrants. Costs are, also, involved in monitoring the "principle-agent problem," by that, in observing the actions of parties (travel agents) whose interests differ from those they are representing (the traveler). Moreover, there are economies of scale and scope in the exploitation of principle-agent effects. In addition, production indivisibilities exist in informing travel agents and in providing competitive, frequent service on an adequate number of routes. Because a single passenger may demand trips for both business and pleasure, it follows that airlines offering business and leisure travel in many city-pairs are best prepared to exploit information and principle-agent effects (Levine, 1987).

Similarly, brand recognition is essential to succeed in the airline business. Carriers providing many products in different markets under the same name, however, complicate recognition.[4] Consequently, the

need for brand recognition creates significant economies of scale and scope in the development of product and brand information. To gain market identity, new entrants must commit aircraft for trial period and advertising/marketing costs to inform consumers of their service. Generally, these information costs make it necessary for new entrants to announce entry before the start of operation. For example, Southwest announced its entry into the Baltimare/Washington market two months before it began service (Ott, 1993).[5] Once announced, these information costs are sunk; they cannot be recouped. As a result, exit is not costless which contradicts a primary condition of contestability (Levine, 1987).

Transaction and information costs are, also, important to airlines as they influence access to capital, a necessity in the uncertain and cyclical industry. Moreover, access to capital helps marketing strategies, airport access, and the ability to withstand money losing, competitive tactics. Between carriers emerging from the regulation era, there was extreme inequality in the amount of capital accumulated. Inequality was even more pronounced, though, between holdover airlines and new entrants (Levine, 1987). As such, although holdovers were disadvantaged by higher costs, they were advantaged by easier access to capital.

A complete description of how each of these factors contributed to the deviations from contestability is beyond the scope of this paper. The descriptions here simply suffice to present the influences that contributed to the lack of contestability found in the low-cost entry research. Only those factors deemed most importance will be expanded upon in the subsequent analysis.

Section 2: The Motivations for Entry and Product Choice

Both new and traditional industrial organization theories can be used to describe the entry and product choices that characterize the deregulated (U.S.) airline industry. Economists agree that asymmetrical information is present in the industry. Although a carrier may know its own costs and fare/capacity allocations, it seldom does of its rivals. Central reservation systems (that have evolved since deregulation) have facilitated this lack of transparency by allowing hidden capacity controls of fare levels (Levine, 1987). In addition, low-cost entrants do not

exclusively enter monopoly markets. This is illustrated by the response of both Continental Airways and USAir to Southwest's entrance into the Baltimore/Washington - Cleveland market (Ott, 1993). Consequently, a model of Bertrand competition with asymmetrical information, as opposed to entry into monopoly markets, is appropriate to show the existence of entry incentives in the deregulated industry. Spulber developed such a model (Spulber, 1995) which, although for homogeneous goods, is applicable to airlines if one assumes that two carriers operating between points A and point B offer the homogeneous product of transportation.

The standard Bertrand model of competition with perfect information and two identical firms concludes that price will equal marginal cost, the firms will equally share the market, and both will make zero profits. However, if all firms have constant marginal cost, but one has an absolute cost advantage, it prices at the marginal cost of the next-to-lowest cost firm and captures the entire market.[6] This creates a discontinuity that is eliminated when asymmetrical information about rivals costs is present (Spulber, 1995).

Spulber (1995) has shown that, with imperfect information, a unique symmetrical equilibrium exists with each firm in the industry pricing below its monopoly price. Moreover, the expected profit is positive for all costs. This profit creates an incentive for entry and price competition.[7] However, as the number of firms increases, the expected likelihood of winning the market decreases. Therefore, the number of firms affects the expected profits through both the likelihood of winning and the equilibrium pricing strategy. In the limit, the equilibrium pricing strategy converges to average cost pricing (Spulber, 1995). Prices below monopoly levels and average cost pricing in the limit support Dresner, Windle, and Lin's recommendation that public policy should encourage low-cost entry in the airline industry (Dresner, Windle, and Lin, 1996).

Spulber's (1995) extends his analysis to construct a consistent entry model where equilibrium price, profit, and market structure are dependant on the sunk costs of entry. Market structure is decreasing in these sunk costs, which is expected. This is consistent with Levine's (1987) explanation that airline entry is more successful in markets with the lowest, non-recoverable information costs. Sunk information costs

have played a key role in the development of both hub and spoke and linear networks, as will be illustrated (Levine, 1987).[8]

Since information costs exhibit economies of scope, a hub incumbent can add a spoke at a much lower cost than a new carrier trying to enter the same market. For the hubbing carrier, continuous advertising allows it to combine information about a new service with that of its other promotions at virtually no incremental cost. Moreover, the information effects create brand recognition that can be easily extended to new markets (Levine, 1987). Therefore, expansion of routes from hubs and the use of multiple hubs are advantageous for carriers large enough to do so. To date, successful adoption of hub and spoke networks has been dominated by the pre-deregulation, trunk carriers.

Similarly, economies of information benefit regional and niche airlines; however, they influenced these carriers' development of linear networks. For example, Southwest Airlines, originally a regional airline expands by developing routes to or from points it already serves.[9] As Southwest's experience and reputation are already known at one endpoint, these traits can easily be transferred to the new market. Southwest is able to limit its information costs by advertising in only regional sub markets. The resulting linear strategy has, clearly, benefited Southwest and is emulated by most of today's low-cost entrants (Beddoe, 1998).

Filling seats created by entry requires the communication of fares, services, and seat availability to potential customers. Both information costs and indivisibilities in the public media influence the type of service offered, full or discount. Economies of information make it cheaper for national carriers to offer more than one product, such as first and coach service, to a large market. These same economies encourage smaller carriers to target niche groups. Furthermore, indivisibilites in the public media, specifically in large metropolitan newspapers and television, make it more efficient for large airlines to offer multiple products to people in many geographical areas. Conversely, focusing on communications sub markets is more practical for smaller entrants. These factors have contributed to holdover carriers providing full service flights and low-cost entrants offering no-frills service. Evidence of these alternate strategies is given in Dresner, Windle, and Lin (1996).

Here, the authors' state that Southwest enters shorter routes with higher levels of market concentration than pre deregulation carriers (Dresner, Windle, and Lin, 1996). By servicing shorter routes at low fares, Southwest targets a niche group of travelers who would otherwise drive (Beddoe, 1998).

Clearly, product differentiation is inherent in the post deregulation airline industry. Hub and spoke and linear networks provide different services while full and discount carriers offer heterogeneous products. A simple location / spatial entry model can be used to explain this product differentiation.

Consider the linear city model proposed by Hotelling (1929). In this model, a linear city of length one forms a line with consumers uniformly distributed with density one along the interval. Two firms may enter the city and each is allowed to choose only one product (hub versus linear service or full versus discount product). The product choice is characterized by the firms choice of location on the line. It is assumed that prices can be adjusted faster than products, which is consistent with the airline industry. Consequently, firms choose their location anticipating that it will affect price competition. Thus, a two-stage game is defined with firms first choosing their location simultaneously, and second choosing their prices simultaneously[10]. In the first period, using quadratic transportation costs (again consistent with airlines), the equilibrium results in the two firms locating at opposite ends of the city (d'Aspremont et al., 1979 in Tirole, 1988). This is analogous to maximum product differentiation. In the second period, after both firms have maximized with respect to price, the equilibrium continues to exhibit maximum product differentiation (Tirole, 1988).

Although simplistic, this model formulates a theoretical foundation for the product differentiation found in post deregulation airline competition. However, the symmetry of location and price choice eliminates any strategic aspects of product positioning. Two issues arise when sequential entry is considered; what is the equilibrium pattern of location (product choice) and what is the optimal timing of entry. Both issues may deter entry or affect a rival's choice of product (Tirole, 1988). These competitive effects of low-cost entry will be examined in the next section.

Section 3: The New Industrial Organization Forces Behind Low-Cost Competition

Initially following deregulation, holdover firms were unsure how to react to competition from low-cost airlines. Still trying to adjust to this new form of competition, the 1980-1981 recession forced holdovers to keep seats full. They responded by becoming sophisticated managers of their information intensive environments. This led to the development of frequent flyer plans, complex fare structures, and elaborate inventory management systems. These features facilitate the competitive strategies used by incumbents facing low-cost entry (Levine, 1987). In turn, they contribute to the empirical results found in the low-cost entry research.

Frequent flyer plans exploit the principle-agent problem and develop loyalty in consumers. Because awards are distributed non-linearly, airlines with generous frequent flyer plans can attract larger shares of the high revenue traffic. Moreover, the flexibility of plans allows incumbents to award large bonuses on entered routes or on those with the potential for entry. Frequent flyer plans are now such a competitive tool that even Southwest has one (Brandy, 1999).

Complex fare structures have evolved due to the existence of production indivisibilites at the city-pair level, principle-agent effects, and information costs. Although Southwest only has one class of service, it has six fare levels on its flights (Brandy, 1999). Conversely, the now extinct People Express, offered simple peak/off-peak fares at rock bottom prices (Levine, 1987).[11] Production indivisibilites make it necessary to sell seats on the same flight at different prices. Principle-agent effects and information costs make it more efficient to have different fares for the same services in different markets. They, also, make it more beneficial to offer different fares to consumers with different needs in the same market. The resulting complex fare structure allows incumbents to match fares with low-cost competitors without matching fare conditions. Moreover, this structure enables firms to generate maximum rents by segmenting traffic and charging accordingly. To gain maximum revenue, competitors must offer discount seats (Levine, 1987). This may be one motivation for incumbents to reduce fares in response to low-cost entry.

Furthermore, elaborate inventory systems permit carriers to maximize revenue by changing the levels, characteristics, and allotment of fares on a flight almost instantaneously.[12] In addition, such systems ease rapid fare response to entry. These abilities benefit those airlines like American and United that control central reservations systems. Efficient market segmentation improves a carrier's ability to respond to low-cost entry and price competition. Therefore, the complex fare structure and inventory management systems allow price discriminating incumbents to quickly match entry-level fares. However, they need only match on some of the available seats, keeping the others free for high yield passengers. This gives incumbents a revenue advantage over their low-cost competitors (Levine, 1987).

Another competitive element in low-cost competition is hub and spoke networks. Economies of information and indivisibilities in production and mass media limit hub entry by small, low-cost carriers. Therefore, entry into markets from a hub is better done by a carrier already hubbing at the city or by an airline with a hub at the other endpoint. In contrast, flow-through traffic is susceptible to entry from both point-to-point and other hub carriers.[13] Since hubbing airlines need this traffic to sustain their networks, competition on flow routes is intense. Furthermore, only part of the incumbent's revenue is susceptible to entry at a spoke. This, coupled with production indivisibilities that favor the hubbing carrier, allows it to generate the revenue necessary to maintain frequent flights. In turn, it can use this frequency to attract customers which is one factor that explains the increased traffic (in spite of reduced fares) found in the empirical papers.

Recall that People Express entered hub markets such as Newark (Whinston and Collins, 1992) and Atlanta (Dresner and Windle, 1998). It is no longer in operation. It follows that entry into major markets has been cited as a major flaw in People Express's growth strategy (Whinston and Collins, 1992).

Slot-controlled airports is, also, a competitive variable that contributes to the pattern of low-cost entry. It is mentioned briefly here because low-cost carriers in both the U.S. and Europe enter mostly secondary, non slot-controlled airports (Beveridge, 1998). Examples of this are Southwest at BWI and EasyJet at Stanstead. Nevertheless, it is

assumed that the barriers created by slot control influence entering alternate airports.[14] Note that these barriers are at the airport, not at the market of entry. Therefore, slot control cannot be attributed to the failure of contestability in air transportation (Levine, 1987).

Low-cost entry at secondary airports implies that slot-controlled airports are not subject to the same low-cost competition as alternate airports. Not surprisingly, the slot-controlled variable in most econometric models has a positive effect on yields.[15] Only in Richards (1996) was this coefficient insignificant, although the sign was positive. Because Richards focused on Southwest's entered or potential entry routes, and given that the mean of the slot-controlled variable in the pooled equation was .202, the insignificance of the slot-controlled coefficient is understandable. Therefore, it is inferred that econometric models that do not control for the lack of low-cost competition at slot-controlled airports upwardly bias the impact of this coefficient on yields.

Consequently, to compete with low-cost entrants, incumbents use complex fare structures, inventory management systems, and hub and spoke networks, while exploiting principle-agent effects and production indivisibilities (Levine, 1987). By doing so, they build a top dog strategy (Tirole, 1988).[16] To look strong and aggressive, incumbents match or beat an entrant's low fare. However, they place restrictions on this fare to fence out high yield traffic while attracting the leisure traveler. Meanwhile, incumbents ensure that capacity is available to accommodate any increased traffic. They also match business fares, but include extra benefits to retain the loyalty of travel agents and frequent flyers. Moreover, incumbents add frequency to "sandwich" an entrant's departures between their own. The purpose of these tactics is to eliminate all incentives for a traveler to choose the low-cost carrier (Levine, 1987). Ensuring seat availability and "sandwiching" departures are other contributors to the increased traffic evident in the empirical papers.

Airlines that have successfully exploited the factors outlined here are those that survive today. Large holdover carriers prosper by offering full-service products through hub and spoke networks while low-cost carriers compete by providing no-frills service on linear systems. The following section uses traditional industrial organization to illustrate how

an incumbent's top dog tactics can result in the outcomes found in the low-cost airline research.

Section 4: Traditional Industral Organization Explains the Emprircal Results

Fares fall and traffic rises on the routes entered by a low-cost carrier (Whinston and Collins, 1992 and Dresner, Windle, and Lin, 1996). The magnitude of these changes is greatest if Southwest is the competitor (Dresner and Windle, 1992). Fares, also, decline due to potential low-cost entry (Whinston and Collins, 1992 and Richards, 1996). Moreover, spillover effects are evident on non-competitive routes at the airport entered and on competitive routes at neighboring airports (Dresner, Windle, and Lin, 1996). At hub airports, an incumbent airline competing with a low-cost carrier lowers fares on competitive routes ending at the hub and on routes that flow through the hub. It does not, however, raise fares on non-competitive routes at, or through, the hub (Dresner and Windle, 1997). Prior to exit, fares, also, drop and concentration declines (Dresner and Windle, 1995). Standard economic theory predicts that increased competition reduces prices, thereby, benefiting consumers. This aggregate view is too simplistic for this analysis. Instead, various economic models will be used to describe the results found in the data.

Airlines operate in many markets but low-cost entry may occur in as few as one of an incumbent's markets. Armstrong and Vickers (1993) model how prices respond to direct competition. They examine the effects of price discrimination policy in a model where a dominant incumbent faces an endogenous degree of competition in one of its two markets. Undoubtedly, complex fare structures and elaborate inventory systems facilitate price discrimination in air transportation. Therefore, the Armstrong and Vickers model of entry without regulation is applicable to post deregulation competition.

Consider an established firm that operates in two markets. It has constant unit costs and its demand in the two markets is independent and symmetric. This is analogous to a full-service incumbent operating two, non-stop routes terminating at a hub (i.e., with no flow through demand) or operating in two unrelated markets. For the entrant, entry into one

route is prohibitively expensive. Consequently, it chooses to enter the second market. The authors assume that the entrant is a price taker, which is realistic if the entrant is small relative to the incumbent.[17] They conclude that, as the scale of entry increases, the incumbent's residual demand curve in the entered market becomes more elastic. As a result, its optimal price response function is decreasing in the scale of entry (Armstrong and Vickers, 1993).

This theory supports the finding of lower fares on routes after direct low-cost entry. Moreover, it explains the coefficients found in the Dresner, Windle, and Lin (1996) regressions. First, if one assumes that entry by Southwest is on a larger scale than that by other low-cost carriers, the theory supports the result that yields fell by 53 percent when Southwest competes on a route and by only 37 percent if entry is by another low-cost carrier.[18] Second, consider markets entered as opposed to routes.[19] The increasing size of the coefficients as entry increases concurs with the Armstrong and Vickers (1993) proposition. Therefore, these authors have presented a theoretical description for the fare response following direct entry found in the empirical research.

However, in the Whinston and Collins (1992), Richards (1996), and Dresner, Windle, and Lin (1996) papers, fares fell when entry was announced or threatened. This can be explained by considering sunk costs as a barrier to entry. An important aspect of sunk costs is their commitment value. A firm that invests heavily in sunk costs today signals that it will remain in the market. A strong signal is essential for an incumbent to acquire the reputation as an aggressive competitor. Using the sequential entry model of Stackelberg (1934), Tirole (1988) has shown how investment in sunk costs and first-mover advantage allow established firms to restrict competition. This model defines investment in capital as the sunk cost. However, other forms of capital could be increased capacity or development of a clientele (Tirole, 1988). Lower fares are associated with both of these factors.[20]

Consider a two firm industry with the incumbent choosing its level of capital. The potential entrant observes this level and, then, chooses its level. These levels are fixed after chosen.[21] The firms' profit functions are identical and dependent on both levels of capital. Each firm dislikes capital accumulation by the other, and each firm's marginal value of

capital decreases with the other's capital level. Therefore, capital levels are strategic substitutes. The model, also, assumes that there are no fixed costs of entry (Tirole, 1988). These properties are not foreign to airlines. In the two-stage game, the incumbent must anticipate the reaction of the entrant to the incumbent's level of capital. By substituting the entrant's reaction function into the incumbent's profit function, a "perfect" Nash equilibrium is determined. Despite identical profit functions, the incumbent obtains more profit than the entrant does by limiting the size of entry (Tirole, 1988). Consequently, the incumbent has an incentive to invest in capital to restrict entry. It follows that fare reductions before low-cost entry have been used by incumbents to signal competitive commitment and to lower the scale of entry once it has been announced.

Spillover effects onto non-competitive routes at the airport entered, onto competitive routes at neighboring airports (Dresner, Windle, and Lin, 1996), and onto hub flow through routes (Dresner and Windle, 1998) can be explained using the theories of limit-pricing and predation. In the previous analysis, it was seen that an established firm may lower its fares to build a reputation for toughness. By doing so, it attempts to block potential entry or induce exit (Tirole, 1988). It may, also, use these tactics to deter further expansion of a low-cost entrant.

An incumbent facing entry has the choice of signaling pre entry (limit pricing) or post entry (predatory pricing) (LeBlanc, 1992). The basis of limit pricing is that the incumbent wants to signal that it is a low cost type to raise doubts about the entrant's viability in the market (Tirole, 1987). Although analysts believe that airline incumbents have higher costs than new entrants (Levine, 1987), imperfect information makes the level of their cost structure unknown to the low-cost carrier. Conversely, there may be two motivations for predation: (1) to induce the rival to exit and (2) to signal that entry is not profitable (Tirole, 1988). As such, limit pricing and predation are viable responses to low-cost competition. The Milgrom-Roberts model of limit pricing presents two conclusions. The first offers another reason for lower fares on routes facing direct competition. The second explains the spillover effects onto routes where neither entry nor its announcement has occurred.

The Milgrom-Roberts model involves two firms and covers two periods. The incumbent is a monopoly at date one and chooses its first

period price.[22] This price is visible to the entrant.[23] The established firm has the choice of charging a high or a low price. Charging a high price will signal that it is a high-cost type. Yet, charging a low-price can signal that it has either high or low costs. Given this, in period two, the entrant decides whether to enter (Tirole, 1987).

Due to asymmetrical information, the incumbent knows its costs but the entrant does not. Nevertheless, the entrant has an *a priori* belief about the probability of the incumbent being a high or a low cost type. It is assumed that the entrant's decision is influenced by its belief about the incumbent's cost structure. Since the incumbent wishes to remain a monopoly, it wants to convey that it has low costs by charging a low price. If it did not, it would be inviting entry. Knowing this, a rational entrant may not infer that a low price signals the correct costs of the incumbent. Furthermore, the incumbent realizes that the entrant knows about its incentive, and so on. The correct way to analyze this dynamic game of incomplete information is to solve for a "perfect Bayesian equilibrium" (Tirole, 1987).[24]

In the game, two potential equilibrium exist; a separating and a pooling equilibrium. In a separating equilibrium, the incumbent's first period price fully reveals its cost. However, this is not applicable to the airline industry because multiple fares and inventory systems reveal little about a firm's costs. In the pooling equilibrium, the first-period price is independent of the cost structure. Thus, no information is given about the incumbent's costs and the entrant's posterior beliefs remain the same as its *a priori* beliefs. Milgrom and Roberts show that a pooling equilibrium exists with both firms charging the low-cost monopoly price (Tirole, 1988). This first conclusion concurs with the Armstrong and Vickers (1993) prediction of fare reductions on entered routes. However, it holds only if one assumes that low-cost airlines set fares equal to the low-cost monopoly price.

The conclusion is not robust in explaining the differential impact on yields of the carrier coefficients in the Dresner and Windle (1995) regression. Here, fares dropped by 69 percent on routes entered by Southwest but by no more than 25 percent on those entered by established carriers (Dresner and Windle, 1995). Although a pooling equilibrium still exists, the causality of the game is reversed. By that, the

incumbents appear to be reacting to fares set by the low-cost entrants.

The second conclusion of this model is that the incumbent manipulates its price so as not to reveal its cost, and the high-cost type engages in limit pricing to deter entry (Tirole, 1988). Airlines' use of inventory management systems ensures that fares will not signal the true cost of a flight. This is analogous to the manipulation described above. Limit pricing, in turn, explains the lower fares found on non-competitive routes at the entered airport in Dresner, Windle, and Lin, (1996). Moreover, it gives cause for the reduced fares on non-competitive routes at the entered airport, and for the lower fares on competing routes at neighboring airports (Dresner, Windle, and Lin, 1996).

Predation, on the other hand, involves a reduction in price in the short run, possibly below cost, to drive firms out of the market. In the Milgrom-Roberts model, predatory prices can be above or below marginal cost (Tirole, 1987). The motivation is for the incumbent to regain monopoly profits or for it to strengthen its top dog reputation (Tirole, 1987). Initially, economists felt that predation would not be be an effective strategy in post deregulation airline competition. They thought that constant economies of scale at the firm level and mobility of capital by way of airplanes would allow a predatory victim to exit a market, yet, re-enter once fares increased. However, conventional views ignored the impact of information impediments, frequent flyer plans, complex fare structures, capacity controls, and hub-and-spoke networks. These factors have made predatory practices feasible in low-cost airline competition (Levine, 1987).

Evidence of predation appears in the time series analysis by Dresner and Windle (1995). Here, fares fell by 27 percent in the three quarters before exit and traffic rose by 17 percent (Dresner and Windle, 1995). This result suggests that both predatory pricing and the "sandwiching" of departures were used to cause exit.

Another example arises in a recent U.S. Department of Justice antitrust lawsuit against American Airlines.[25] The suit accuses American Airlines of slashing fares and saturating contested routes to drive three carriers out of its Dallas-Fort Worth hub. Two have since gone out of business. According to the plaintiff, in 1996 the fare from Dallas to Wichita (with low-cost competition) was $55 (U.S.). Since low-cost exit,

it has risen to $107 (U.S.) (Kuitenbrouwer, 1999). This increase contrasts the time series analysis of Dresner and Windle (1995) where fare increases were not found one year after exit. Possibly a one-year duration is not long enough to detect predatory pricing. In the American Airlines case, the higher fare is reported three years after exit. Or perhaps an incumbent may try to strengthen its aggressive reputation by not raising fares. Or, maybe, it may not want to invite entry by increasing fares (Tirole, 1987). Consequently, the failure of fares to rise in Dresner and Windle (1995) does not negate the existence of predation.

In the previous two examples, the data do not include the incumbents' frequent flyer points offered on the entered routes. Generous bonuses may be used to lure an entrant's prospective customers, either before or after entry, thus reducing the entrant's demand in the new market. As such, frequent flyer plans may be another predation tool used to precipitate exit.

Moreover, economies of scope and scale exist in building a top dog reputation. Economies of scope exist because larger firms who use predation and limit pricing a few times gain a reputation as fierce respondents to entry. Economies of scale allow larger firms to use this reputation more effectively than smaller entrants. (Levine, 1987). This may shed light on the prevalence of limit pricing and predation in low-cost entry competition.

Nevertheless, these tactics do not describe the evidence that carriers do not raise fares on non-competitive routes that flow through a hub (Dresner and Windle, 1998). Two influences explain this result. First, recall that hubbing carriers need flow traffic for the hub to be efficient. Therefore, competition for this one-stop traffic will be intense Levine (1987). Standard economic theory states that increased competition will reduce fares. Therefore, theory predicts that fares will not rise; in fact they should fall. Furthermore, entry theory posits that entry will occur where the probability of the greatest profits is maximized. Accordingly, an incumbent would not want to increase fares on non-competitive hub flows because these are the routes most easily entered by a low-cost airline already serving the hub. Therefore, the economic theories of competition and entry support the finding that incumbents do not raise fares on these hub flow routes.

The models in this section explain many of the findings in the empirical papers. The Armstong and Vickers (1993) model of entry without price regulation gives one reason for reduced fares on routes facing direct low-cost competition. Another, although not as robust, is presented in the model of limit pricing. Tirole (1988) has adapted the sequential entry model of Stackelburg (1934) to show that sunk costs and first mover advantage induce fares to drop before low-cost entry. Limit pricing explains why fares fall on non-competitive routes at the airport entered and on competitive routes at neighboring airports. Predation gives cause for fares to fall on direct and flow through routes at hubs. Moreover, building a top dog reputation is a strong motivator for the use of limit pricing and predation. The failure for prices to rise on non-competitive routes to, or through, hubs is described using standard pricing and entry theory. In sum, traditional industrial organization theory has been used to describe low-cost airline competition.

Conclusion

Low-cost airline competition has, and will, continue to impact the aviation industry worldwide. Examples from the U.S. domestic industry have been used to characterize the industry. Today, most large, trunk airlines offer full service through hub and spoke networks. Low-cost carriers using predominantly linear route networks to provide competition. This pattern is already evident in Canada and the European Union, though to varying degrees. Contestability has been revoked as a model of deregulated airline competition (Levine, 1987; Whinston and Collins, 1992; and Richards, 1996). Evidence from the research of Whinston and Collins, (1992); Dresner and Windle, (1995); Richards, (1996); Dresner, Windle, and Lin, (1996); and Dresner and Windle (1998) indicates that low-cost competition greatly influences fares and traffic levels. Therefore, it is concluded that low-cost competition improves consumer welfare and should be encouraged (Dresner and Windle, 1995).

This paper has drawn upon new industrial organization theory to explain the lack of contestability in the industry. In addition, it has used concepts from both new and traditional industrial organization to outline

the motivation for the structure and product differentiation seen today. Both theories have, also, been used to describe the competitive influences that have resulted in the price and traffic fluctuations found in the empirical data. In short, this paper offers a thorough analysis of low-cost airline competition.

---

[1]A competitive equilibrium is a set of prices such that each consumer is choosing their most preferred affordable bundle, and has demand equal to supply in all markets (Varian,1993).

[2]Contestibility theory says that the benefits of pure competition can be achieved in a market with few competitors if the threat of entry provides an effective substitute for actual competition (Levine, 1987).

[3]The 69 percent coefficient is for competition by Southwest Airlines. Unless otherwise stated, all regression coefficients are significant at the 5 percent level.

[44] American Airlines offers two classes of service, First and Coach, on one aircraft while Southwest Airlines offers one class of service but at six different fare levels (Brandy, 1999).

[5]Southwest entered Baltimore-Washington International Airport (BWI), a competing airport to Washington-National and Washington-Dulles Airports (Dresner, Windle, and Lin, 1996).

[6]Constant economies of scale characterize air transportation (Levine, 1987).

[7]It is the *expected*, profit that creates the entry incentive, even with positive set up costs. Information costs related to airline entry could be regarded as positive set up costs.

[8]Hub and spoke development has also been influenced by economies of scope in operations and production indivisibilities in providing frequent jet service in markets whose demand does not support such frequency (Levine, 1987).

[9] Southwest served Cleveland and Chicago before it began the BWI-Cleveland and BWI-Chicago routes (Ott, 1993).

[10]Sequential entry may better represent the present air transport industry, but simultaneous entry and pricing may well characterize behaviour directly following deregulation in the U.S.

[11]This may have been one factor leading to its demise.

[12] Therefore, discount seats availability on flights is usually not visible to competitors (Levine, 1987). This is one factor contributing to imperfect information the air transport industry.

[13]Flow through traffic must transfer at the hub because its route does not support non-stop service by the hub carrier. These flows may support less frequent point-to-point service, though (Levine, 1987).

[14]Other influences may be cheaper landing fees and ground handling charges (Levine, 1987).

[15]For evidence, see Whinston and Morrison (1995), Dresner and Windle (1995), Dresner, Windle and Lin (1996), and Windle and Dresner (1997)

[16]A top dog strategy is defined as being big and strong to look tough or aggressive (Tirole, 1988).

[17]Apart from Southwest, low-cost carriers are small in relation to the dominant trunk carriers in the U.S.

[18]This is a realistic assumption given that Southwest was the seventh largest U.S. airline in terms of enplanements in 1993. No other low-cost carrier was in the top six (Morrison and Winston, 1995).

[19]Dresner, Windle, and Lin (1996) have coefficients for entry into more than one competitive route. If the incumbent operates on Los Angeles-San Fransisco but the entrant enters Los Angeles-Oakland, the coefficient is defined as one low-cost carrier on a competitive route. Assuming that these are the same market, the Armstrong and Vickers theory applies.

[20]Pricing theory states that increased capacity will result in reduced prices. Furthermore, advertising investment will have a greater impact if lower fares (enhanced by generous frequent flyer bonuses) are announced. Therefore, lower fares coupled with frequent flyer plans may be used to attract and keep the leisure passenger targeted by the low-cost entrant. As such, the incumbent has a higher probability of capturing the entrants' prospective clientele.

[21]This assumption may not be robust in the airline industry where investment in capacity and information costs is not fixed in a market. However, investment in aircraft may be considered fixed in the short run.

[22]Even though low-cost entrants do not only enter monopoly markets, it is realistic to assume that some markets entered are monopolized.

[23]Recall that fares are visible but the capacity at each fare level is not (Levine, 1987).

[24]A Baysian-Nash equilibrium is a set of strategies for each player, based on the belief of the type of the other player, that maximizes the expected value of each type of player (Varian, 1992).

[25]The suit was filed in May, 1999 by Vanguard Airlines of Kansas City.

References Cited

"American Airlines Renowned For Playing Hardball." Kuitenbrouwer, P. Financial Post, 26 August 1999, C3

Armstrong, M., and Vickers, J. "Price Discrimination, Competition and Regulation." *The Journal Of Industrial Economics*, Vol. 16, December 1993: pp. 335 - 359.

Beddoe, Clive J. "Success in the Western Skies, WestJet Prospers with a Low Cost Focus." *Business Review Journal*, 1998 - 1999: pp. 6 - 8.

Beveridge, Dirk. "Fare wars heating up in Europe." *The Augusta Chronicle Online: business@augusta*. Http:augustachronicle.com, June 2, 1998.

Brandy. Southwest Airlines Reservation Agent. Interview by author, August 1999.

Dresner M. E., Lin J.C. and Windle R. J. "The Impact of Low-Cost Carriers on Airport

and Route Competition," *Journal of Transport Economics and Policy*, September 1996: pp.309-398.

Dresner M. E., and Windle, R.J. "The Short and Long Run Effects of Entry on U.S. Domestic Air Routes." *Transportation Journal*, Winter 95: pp. 14-25.

Dresner M. E., and Windle, R.J. "Competitive Responses to Low Cost Carrier Entry."*Journal of Transport Economics and Policy*, 1998.

d'Aspremont, C., J. Gabszewicz, and J.-F. Thisse. "On Hotelling's Stability in Competition." In The Theory of Industrial Organization. by J. Tirole, 277 - 300, Boston: MIT Press: 277 - 300, (1988).

Graham, B. "Air transport liberalization in the European Union: An assessment."*Regional Studies*, Vol. 3, No. 4. October 1997: pp. 807 - 812.

Hotelling, H. "Stability in Competition." In The Theory of Industrial Organization. by J. Tirole: 277 - 300, Boston: MIT Press, (1988).

LeBlanc, G. "Signalling strength: limit pricing and predatory pricing." The RAND Journal of Economics, Vol. 23, No. 4. Winter 1992: pp. 493 - 506.

Levine, M. E. "Airline Competition in Deregulated Markets: Theory, Firm Strategy, and Public Policy." *The Yale Journal on Regulation*, Vol. 4, 1997: pp. 393 - 480.

Morrison, S. A., and C. Winston. The Evolution of the Airline Industry. Washington, D.C.: The Brookings Institution, (1995). Ott, James, "Southwest Enters East Coast Market," *Aviation Week & Space Technology*, July 19, 1993, pp. 26-27.

Ott, James, "BWI Airport Braces for Southwest Entry," *Aviation Week & Space Technology*, September 6, 1993: pp. 37-38.

Richards, K. "The Effect of Southwest Airlines on U.S. Airline Markets." *Research in Transportation Economics*, Vol. 4, 1996: pp. 33-47.

Spulber, D. F. "Bertrand Competition When Rival's Costs Are Unknown." *The Journal Of Industrial Economics*, Vol. 18, March 1995: pp. 1 - 11.

Tirole, J. The Theory of Industrial Organization. Boston, MIT Press. (1988): pp. 277 - 494.

Varian, Hal, R. Microeconomic Analysis. 3rd. ed. New York, N. Y.: W.W. Norton & Company, 1992.

von Stackelberg, H.. "Marktform und Gleichgewicht." In The Theory of Industrial Organization. by J. Tirole: 277 - 300, Boston: MIT Press, (1988): pp. 314 - 322.

Whinston, M. and Collins, S.C. "Entry and Competitive Structure in Deregulated Airline

Markets: An Event Study Analysis of People Express," *RAND Journal of Economics,* Vol. 23, No. 4. Winter 1992: pp. 445 - 462.



ISSN#
1183-2770

DESIGNED AND PRINTED BY PRINTING SERVICES • UNIVERSITY OF SASKATCHEWAN • CUPE 1975
COVER PHOTO: LIGHT AND SHADOW IN THE CARINA NEBULA
IMAGE CREDIT: NASA, THE HUBBLE HERITAGE TEAM (AURA/STSSci)
"Copyright Association of Universities for Research in Astronomy Inc. (AURA), all rights reserved."