



TRANSPORT RESEARCH CENTRE

THE WIDER ECONOMIC BENEFITS OF TRANSPORT

MACRO-, MESO- AND MICRO-ECONOMIC TRANSPORT PLANNING AND INVESTMENT TOOLS

ROUND TABLE

140

OECD

International Transport Forum





THE WIDER ECONOMIC BENEFITS OF TRANSPORT

MACRO-, MESO- AND MICRO-ECONOMIC TRANSPORT PLANNING AND INVESTMENT TOOLS

ROUND TABLE

140

TRANSPORT RESEARCH CENTRE

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where the governments of 30 democracies work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The Commission of the European Communities takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

This work is published on the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the Organisation or of the governments of its member countries.

Also available in French under the title:

BÉNÉFICES ÉCONOMIQUES ÉLARGIS DU SECTEUR DES TRANSPORTS

INSTRUMENTS D'INVESTISSEMENT ET D'ÉVALUATION MACRO-, MÉSO ET MICRO-ÉCONOMIQUES

Corrigenda to OECD publications may be found on line at: www.oecd.org/publishing/corrigenda.

© OECD/ITF 2008

OECD freely authorises the use, including the photocopy, of this material for private, non-commercial purposes. Permission to photocopy portions of this material for any public use or commercial purpose may be obtained from the Copyright Clearance Center (CCC) at *info@copyright.com* or the Centre français d'exploitation du droit de copie (CFC) *contact@cfcopies.com*. All copies must retain the copyright and other proprietary notices in their original forms. All requests for other public or commercial uses of this material or for translation rights should be submitted to *rights@oecd.org*.

INTERNATIONAL TRANSPORT FORUM

The International Transport Forum is an inter-governmental body within the OECD family. The Forum is a global platform for transport policy makers and stakeholders. Its objective is to serve political leaders and a larger public in developing a better understanding of the role of transport in economic growth and the role of transport policy in addressing the social and environmental dimensions of sustainable development. The Forum organises a Conference for Ministers and leading figures from civil society each May in Leipzig, Germany.

The International Transport Forum was created under a Declaration issued by the Council of Ministers of the ECMT (European Conference of Ministers of Transport) at its Ministerial Session in May 2006 under the legal authority of the Protocol of the ECMT, signed in Brussels on 17 October 1953, and legal instruments of the OECD. The Forum's Secretariat is located in Paris.

The members of the Forum are: Albania, Armenia, Australia, Austria, Azerbaijan, Belarus, Belgium, Bosnia-Herzegovina, Bulgaria, Canada, Croatia, the Czech Republic, Denmark, Estonia, Finland, France, FYROM, Georgia, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Mexico, Moldova, Montenegro, Netherlands, New Zealand, Norway, Poland, Portugal, Romania, Russia, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, Ukraine, the United Kingdom and the United States.

The OECD and the International Transport Forum established a Joint Transport Research Centre in 2004. The Centre conducts co-operative research programmes addressing all modes of transport to support policy making in Member countries and contribute to the Ministerial sessions of the International Transport Forum.

Further information about the International Transport Forum is available on Internet at the following address:
www.internationaltransportforum.org

TABLE OF CONTENTS

SUMMARY OF DISCUSSIONS 7

INTRODUCTORY REPORTS:

Recent Evolution of Research into the Wider Economic Benefits of Transport Infrastructure Investments, by Roger VICKERMAN (United Kingdom) 29
1. Introduction 33
2. The Purpose of Infrastructure Studies 34
3. Macro-Level Evaluation of Infrastructure 36
4. Market Level Evaluation of Infrastructure 39
5. Micro-Level Evaluation of Infrastructure 42
6. Conclusions and Implications 44

The Wider Economic Benefits of Transportation, by T.R. LAKSHMANAN (United States) 51
1. Introduction and Overview 55
2. Macroeconomic Modeling of Economic Impacts of Transport Infrastructure 55
3. Lessons from Economic History 60
4. The Wider Economic Benefits of Transport: An Overview 62
5. Concluding Comments 64

Wider Economic Benefits of Investments in Transport Infrastructure, by Jeffrey P. COHEN (United States) 69
1. Introduction 74
2. Motivation 74
3. General Background 76
4. Spatial Econometrics 79
5. Applications 83
6. Conclusions and Future Work 88

Agglomeration Economies and Transport Investment, by Daniel J. GRAHAM (United Kingdom) 93
1. Introduction 98
2. Agglomeration economies and transport investment 98
3. Estimating agglomeration economies 103
4. Results 105
5. Conclusions 108

Transport Infrastructure Inside and Across Urban Regions: Models and Assessment Methods, by Börje JOHANSSON (Sweden) 117
1. Networks and The Spatial Organisation of Economies 122
2. Transport Networks and Agglomeration Economies 125
3. Transport Infrastructure and New Growth Theory 127
4. Networks and Accessibility 132
5. Empirical Results from Accessibility-Based Studies 137
6. Conclusions and Remarks 144

The Broader Benefits of Transportation Infrastucture, by Ian SUE WING, William P. ANDERSON and T.R. LAKSHMANAN (United States) 149
1. Introduction 154
2. Context: The Broader Economic Impacts of Infrastructure Investment 155
3. Conventional Methods of Impact Assessment 157
4. A Review of General Equilibrium Analyses of Congestion 158
5. A Hybrid Meso-Macro Approach 162
6. Discussion and Summary 169
7. Appendix: Implementational Details 171

Progress and Challenges in The Application of Economic Analysis for Transport Policy and Decision Making, ***Concluding Comments for the Research Roundtable on Infrastructure Planning and Assessment Tools,*** **by Glen E. WEISBROD and Brian Baird ALSTADT (United States)** 181
1. Introduction: Research Directions and Policy Assessment Needs 186
2. What Do We Mean by "Wider" Effects? 187
3. Classification of Predictive Transport Economic Models 187
4. Modeling Implications of Recent Research 191
5. Methodological Enhancements Needed for Policy Evaluation 193

LIST OF PARTICIPANTS 199

SUMMARY OF DISCUSSIONS

ASSESSING THE ECONOMIC EFFECTS OF TRANSPORT INFRASTRUCTURE INVESTMENT: INSIGHTS AND CHALLENGES

SUMMARY

EXECUTIVE SUMMARY 14

1. INTRODUCTION 15

2. RECENT RESEARCH ON WIDER ECONOMIC EFFECTS 15

2.1. Setting the stage 16
2.2. Empirical work on wider benefits 17
2.3. Comprehensive modeling frameworks 19
2.4. Progress with and challenges for applied economic project appraisal 20

3. THE PRACTICE OF TRANSPORT PROJECT APPRAISAL 21

4. WHAT KIND OF APPRAISAL FOR TRANSPORT INFRASTRUCTURE IS BEST? 22

NOTES 24

BIBLIOGRAPHY 25

Boston, January 2008

ABSTRACT

This paper summarizes and organizes presentations and discussions of the Round Table on Macro-, Meso and Micro Infrastructure Planning and Assessment Tools, which took place at Boston University, on 25 and 26 October 2007. The goal of the meeting was to investigate how recent research on direct and wider economic impacts of investment in transport infrastructure can be used to improve the practice of transport project appraisal. While the potential importance of "wider benefits" is clear, it is less obvious that attempts to quantify them should be part of all project appraisals. Timely availability of results of simpler approaches might improve the quality of decision-making just as much. And when wider impacts are part of the appraisal, their quantification should follow consistent procedures. Policy-oriented research should focus on these procedures, not on producing general results, as the latter are thought to be irrelevant to policy, to the extent they exist.

EXECUTIVE SUMMARY

This Round Table evaluated the relevance of research on the wider economic impacts of investments in transport infrastructure for the practice of project appraisal. Wider impacts are those not captured in standard cost-benefit analysis, including effects relating to returns to scale, agglomeration, thickening of labor markets, and market power, as well as firms' and households' behavioral adaptations to changes in transport costs.

Macroeconomic analysis of the effects of investment in transport infrastructure, in the Aschauer tradition, suggests that there are modest wider economic benefits from such investments. Recent, more disaggregated work that focuses on the impact of infrastructure investments on markets at the local level, and particularly labour markets, confirms that there are wider economic impacts. It also confirms that the sign and size of these wider effects differs strongly across projects. Results for one project therefore cannot simply be transferred to other projects. There is thus little prospect of developing simple rules of thumb to factor wider impacts into routine project appraisal. Undertaking more sophisticated analysis on a routine basis is hampered both by shortcomings in the availability of the data needed and in the analytical frameworks that might be used.

Accepting that wider impacts are potentially important, what recommendations can be made for improvements in the appraisal of transport infrastructure? Manuals for transport project appraisal can include guidelines for extensions of standard cost-benefit analysis with valuations of wider effects in a methodologically consistent fashion. Research should focus on the development of sound and practical frameworks, not on a search for widely applicable results.

In constructing such frameworks, it is useful to relate the range of the analysis to the size of the project. For smaller projects, an ambitious analysis that includes wider impacts would be too costly and probably yield results too late to affect decisions. The most practical approach for small projects is therefore to work on the assumption that there are no wider economic benefits. The risk of excluding real wider benefits or costs exists, but there was considerable agreement that this is outweighed by avoiding the risk of introducing double-counting of benefits and avoiding delays in project evaluation. For large projects and for the evaluation of investment programs, more sophisticated analyses may well be justified. But also in these cases it useful to keep in mind that the provision of information early in the decision-making process has a larger impact than information that becomes available only further down the line – even if that information is based on a more comprehensive analysis.

Another way to increase the policy impact of economic appraisal is to improve the analysis of direct impacts. Standard cost-benefit analysis does cover these impacts but the results are not always presented in a form that is easily understood by policy-makers. Economic modeling, for example along the lines of the applied general equilibrium tradition, can help outline how direct benefits are transmitted through markets and transferred between economic agents like households and firms. It might be possible to supplement the economic indicators typically presented in project appraisal summaries with a description of the expected economic effects of an investment on the basis of such modeling.

1. INTRODUCTION

This paper summarizes the Round Table's presentations and discussions, draws conclusions where possible, and points out where opinions differ. It is divided in three main sections. First, the presentations and discussions provided an overview of the advances, promises, and pitfalls of current research on the economic impacts of investments in transport infrastructure. A first recurring theme was that advances in the analysis of "wider impacts" were acknowledged, but their transferability across projects was questioned, so there are "no simple rules" for generalizing results. Moreover, routine analysis is difficult because of shortcomings both in data availability and in the analytical framework. This theme is developed in some detail in section two. A second recurring issue was the major differences in the approach to transport project appraisal between countries. The impact of economic appraisal on policy decisions varies greatly from one region to another and this has consequences for the way wider economic impacts might be taken into account. These issues are addressed in section three. Building on the insights from sections two and three, section four tackles the key question of the Round Table: given the current state of research and the practice of transport project appraisal, what recommendations – if any – can be made for improvements in the appraisal of transport infrastructure? A broadly accepted position was that simple rules of thumb, for example taking the form of multipliers to capture wider economic benefits, are to be avoided. Instead, recommendations might be integrated in manuals for transport project appraisal, allowing extensions of standard cost-benefit analysis with valuations of wider effects in a methodologically consistent fashion. The focus for researchers ought to be on the development of sound and practical frameworks, not on a search for widely applicable results.

2. RECENT RESEARCH ON WIDER ECONOMIC EFFECTS

This section covers the main topics addressed in the presentations and discussions. It follows the program of the Round Table, as shown in Box 1.

2.1. Setting the stage

The core purpose of the Round Table was to investigate how emerging insights from research on the direct and wider benefits of investments in transport infrastructure may inform the practice of the appraisal of transport project infrastructure. In his opening statement, T.R. Lakshmanan's sketched the challenges for the research community. Macroscopic approaches to estimating the effects on productivity of public capital in general, and of transport infrastructure in particular, produce a wide range of results. In order to understand this diversity of results, the mechanisms that generate the economic impacts need to be uncovered. An explicit framework that captures the linkages between (changes in) the provision of infrastructure and economic impacts is also a required if the analysis of wider impacts is to be relevant to the practice of project appraisal. This is because macroscopic approaches do not directly relate to the policy levers that are of central concern in economic analysis to support decision-making on transport projects.

Box 1 Programme of the Round Table

Setting the stage (Section 2.1)

Opening statement:	T.R. Lakshmanan
Presentation:	Roger Vickerman
Discussant:	Peter Mackie

Empirical work on wider benefits (Section 2.2)

Presentation:	Jeffrey Cohen
Discussant:	Yossi Berechman
Presentation:	Dan Graham
Discussant:	Andrew Haughwout

Comprehensive modeling frameworks (Section 2.3)

Presentation:	Börje Johansson
Discussant:	Ulrich Blum
Presentation:	Ian Sue Wing
Discussant:	Bruno De Borger

Progress with and challenges for applied economic project appraisal (Section 2.4)

Presentation:	Glen Weisbrod

Various strands of research contribute to a more microeconomic understanding of the effects of transport infrastructure investments, but progress is uneven: much has been done on the analysis of increasing returns to scale and on agglomeration effects, but less attention given to improving knowledge of the dynamic effects of innovation and technical diffusion.

Roger Vickerman developed these themes, by providing a classification of research on the (wider) economic benefits of transport infrastructure investments, and an assessment of their usefulness to the question at hand: how does this research help us make better decisions on infrastructure investments? The main insights are as follows:

- Macro-studies, in the Aschauer tradition, focus on overall impacts. The literature is prone to methodological problems, especially in pinning down the direction of causality, and it is based on insufficiently detailed representations of transport infrastructure to be of direct use in project appraisal.[1] Furthermore, as emphasized by Peter Mackie, there is potential confusion over whether measurements of the economic benefits of infrastructure concern **wider** benefits (i.e. those not captured in standard cost-benefit analysis, which considers effects in transport markets alone), or whether they refer to the ultimate **incidence of direct effects** (that is: the equilibrium allocation that would result from a project without considering wider effects).[2]

- Substantial work has been done at the meso-level, here defined as work that makes transport and other market interactions explicit.[3] Some contributions, like the general equilibrium framework proposed by Sue Wing et al., mainly serve to clarify how changes in transport costs as perceived

by network users translate into costs and benefits, and the distribution thereof, for various economic agents (like households and firms). But, with standard applied general equilibrium assumptions of constant returns to scale and perfect competition, the approach sheds no light on wider benefits associated with returns to scale, agglomeration, thickening of labor markets, or the weakening of market power, or the limiting effect of market power on benefits from better infrastructure. Such wider benefits are addressed in narrower market-studies, of which Dan Graham's is an example.

- Microscopic approaches, that aim to capture the effects of changes in transport conditions on the internal reorganization of firms and households, are scarce. This is not surprising, given that these types of responses are difficult to integrate in microeconomic frameworks that focus on market interactions, but it is unfortunate, as there is evidence that households and firms do re-organize in response to changes like, for example, the congestion charge in London, or the opening of high-speed rail links in Western Europe.

- Also scarce are ex post studies. The results of those that have been done do not provide strong support for the existence of wider economic benefits from transport infrastructure investments.

In summary, recent research suggest that if project appraisal is to go beyond standard cost benefit analysis and wishes to include wider economic effects, it should distinguish between direct user benefits and effects on productivity, agglomeration, competition, and on the labor market. In addition, when spatial spillovers are large (irrespective of whether they include wider benefits or only direct benefits), one should expect different levels of jurisdiction to arrive at different evaluations. Understanding spatial spillovers hence is of clear relevance to policy.

2.2. Empirical work on wider benefits

The presentations by Jeffrey Cohen and by Dan Graham illustrate the current state of econometric work on spatial spillovers and agglomeration effects. A common feature of the econometric work is that empirical specifications are explicitly motivated by a microeconomic framework. This is desirable, as it makes clear which interactions are included in the analysis and which ones are not, allowing a consistent and transparent discussion of the results. Of course, making behavioral assumptions implies the possibility that the assumptions are wrong, leading to misspecification. Two examples of this problem were discussed:

- The estimation of spatial spillovers rests on assumptions of cost minization and the treatment of transport as a costly input. The validity of these assumptions was challenged.

- The assumed direction of causation is critical. Most studies assume growth is caused by infrastructure. But as wealthier economies may choose to spend more on infrastructure, infrastructure may follow growth as well.

While these limitations need to keep in mind when interpreting results, it is clear that empirical analysis requires an explicit framework in order to make sense of data, and that such a framework will always contain restrictive assumptions. Refinements of the specification, on the basis of improved theoretical understanding, will lead to more robust results. And more flexible statistical techniques to deal with error terms, e.g. non-monotonic forms of spatial autocorrelation, will increase the practical relevance of such econometric work.

Despite the methodological limitations, the empirical work generates several relevant insights. First, **spatial spillovers** of investment in public capital are real in the sense that firms' variable costs in one

jurisdiction depend on infrastructure provision in other jurisdictions. These effects can be large, and they differ strongly between transport modes, as well as being dependent on local conditions ("the starting point"). This was illustrated in a review of some applications. A study for the US (Cohen and Morrison Paul, 2004) finds that higher highway capital in one State slightly reduces variable costs in neighboring States, while a Spanish study (Moreno et al., 2004) finds evidence of cost increases. A study of port infrastructure at the level of US States (Cohen and Monaco, 2007) finds that higher port capital stocks in one State increase variable costs in neighboring States. For US airports, however, States with many flights to States with a lot of airport capital have lower variable costs.

While information on spatial spillovers is of obvious interest to policy-makers, questions were raised regarding the extent to which the framework used is relevant to the appraisal of individual projects. Some arguments to support this skepticism are as follows:

- Although plausible hypotheses were formulated, there is no explicit explanation for the large diversity in results. This makes it impossible to separate out the impact of local conditions, and this strongly limits the transferability of results from one case to another.

- The presence of substantial spatial autocorrelation in many studies can be seen as an indicator of the extent of our ignorance, as imposing a structure of spatial autocorrelation on the errors essentially is a statistical technique that helps us deal with incomplete understanding of, or data on, relevant economic interactions.

- Public capital is measured as (the value of) the stock, while project appraisal is about changes in (the physical level of) the stock of infrastructure.[4]

Second, the empirical work on **agglomeration economies** shows that they exist and they can reasonably be measured (although there are obviously caveats here as well, some of which are discussed below). The concept of agglomeration is made operational by constructing an index of the amount of economic activity that is accessible to a firm at its location ("economic or effective density"). Effective density is treated as an input in a (translog) production function, so allowing estimation of agglomeration economies. Agglomeration economies vary strongly among industries; an application for the UK finds they are rather small for manufacturing industries (e.g. the elasticity of productivity with respect to effective density is 0.08 for manufacturing) and large for service-oriented activities (e.g. an elasticity around 0.22 for business services, and around 0.24 for banking, finance and insurance).

Accessibility clearly depends on available transport infrastructure, amongst other factors, so an empirical link between infrastructure and agglomeration can be established. Such an exercise was carried out for the CrossRail project in London, suggesting that this project's (local) benefits increase by about 20% when agglomeration economies are accounted for. The same exercise for a bus subsidy in South Yorkshire (also in the UK) increases direct benefits by some 3%.[5]

Questions were raised regarding the interaction between agglomeration effects and traffic congestion. Agglomeration economies may become exhausted and can be outweighed by congestion effects; the analysis for a Dutch project indeed found "negative agglomeration effects" (Oosterhaven and Broersma, 2007). Empirically separating agglomeration from congestion is difficult but useful (and some work on the issue is available, e.g. Graham, 2006). Analyses of the interaction between agglomeration, location decisions, and transport costs in polycentric contexts shows that lower transport costs may induce firms to move out of the center, as cheaper transport reduces returns to density. Location decisions are, however, ignored in much of the empirical work. It was also noted that congestion pricing can stimulate agglomeration economies if it succeeds in allocating roadspace to activities that benefit most from agglomerations; this is an element of the debate on road pricing in New York city. A related point is that technological developments affect the trade-off between congestion and agglomeration economies. For example, improved information

technology may reduce firms' need to locate close to other firms or two workers (Blum and Dudley, 1999 and 2002).

As with the discussion on spatial spillovers, there were concerns that the modeling of agglomeration effects is too much of a "black box" to be truly useful to project appraisal. A better "microscopic" understanding of the mechanisms that generate the benefits of agglomeration would be very valuable. Such mechanisms include production effects, ease of deliveries, and access to diverse inputs. But there are dispersion economies too. For example, a good highway system allows just-in-time deliveries. Manufacturers exploit this opportunity by dispersing the production of automobiles over several locations so as to avoid upward wage pressure associated with producing in a single location. Opening up the black box is challenging. Many sources of agglomeration effects are empirically equivalent, at least with the sort of data currently available, meaning that econometric identification of the various sources presents a major challenge.

Lastly, it was agreed that the work on spatial spillovers suggests that care should be taken with local estimates of agglomeration effects. For example, the work on the Crossrail link in London found agglomeration benefits that increase the benefits identified in standard cost-benefit analysis by some 20%. But it is not clear to what extent these additional benefits are offset by losses in other jurisdictions.

2.3. Comprehensive modeling frameworks

Börje Johansson and Ian Sue Wing presented analytical frameworks that aim to embed the analysis of economic effects of changes in transport infrastructure in a context that is broader than the narrow transport focus of standard cost-benefit analysis. Johansson's approach is rooted in spatial economics combined with a standard discrete choice travel model. Although the conceptual framework is somewhat different from the static neo-classical microeconomic framework that underlies cost-benefit analysis and its extensions, it leads to empirical strategies that aim to integrate wider economic effects that are similar to the ones identified above (agglomeration effects, in particular). The work of Sue Wing et al. is firmly rooted in neo-classical economics, as it integrates a network representation of space with a standard computable general equilibrium framework. In its present form, the general equilibrium model focuses on making interactions between markets explicit. Agglomeration effects are not included as such, but it appears that such extensions pose no particular conceptual problems.

The Johansson approach emphasizes that transport networks generate a spatial structure, and the particular spatial structure may entail agglomeration economies. The central concept to describe spatial structure is that of a functional urban region, which corresponds to the distance that can be travelled within an hour or so (implying that times and distance matter). The framework is operationalized by constructing measures of how (improvements in) transport networks lead to (improvements of) accessibility. Households desire access to jobs, services, and to the wage sum (as a measure of economic opportunities). Firms demand access to labor and to specific skills, and they are better off when labor and production factors are more abundant (more accessible). Empirical results suggest that central cities respond primarily to internal accessibility, and all urban areas benefit from intra-regional accessibility. It is emphasized in the empirical work that infrastructure should be measured in physical characteristics, not capital values, and that studies based on panel data produce more robust results than those relying on only cross-sectional or time series data. Although not stressed in Johansson's contribution, it may be added that an accessibility measure based on a discrete choice model allows calculating log-sum welfare measures of changes in transport networks.

The discussion centered on whether focusing on accessibility as an objective or as a measure of network performance is valid. There was wide agreement that performance measures refer to intermediate variables and that they should not be seen as policy goals in themselves. A comprehensive welfare measure provides better policy guidance than narrow performance indicators. For example, accessibility is large when households live in skyscrapers, but welfare may be low. Similarly, road congestion is avoided by banning cars, but welfare

may decline. Nevertheless, careful analysis of the likely impacts of changes in infrastructure is a prerequisite for good cost-benefit analysis.

The general equilibrium model introduced by Ian Sue Wing represents a meso-level approach, in that it makes explicit the interactions among the many markets that are affected by changes in transport infrastructure. It does not tackle the issue of how better infrastructure relates to long run economic development, nor to other "non-linearities" like agglomeration effects. A sizeable, though not huge, literature on general equilibrium effects of a variety of transport policies exists. Most of this work has an analytical emphasis, and the numerical results that are available are on too high a level of aggregation to be directly relevant to project appraisal. The model proposed by Sue Wing improves on the available "maquette models" of the interaction between transport conditions and input and output markets, by integrating a detailed representation of transport networks with the economic model.

The approach is useful for at least three reasons. First, it shows how benefits from better infrastructure are transmitted between markets; the final equilibrium allocation shows how costs and benefits are distributed across economic agents, and this information is useful to policy-makers. Second, compared to existing spatial general equilibrium tools, the particular network representation allows investigating the effects of localized network improvements on the overall economy, which is useful as it fits well with the nature of many transport infrastructure projects. Third, on a methodological level, the framework can be used to analyze the impact of spatial aggregation on modeling results, an issue which is known to matter – in the sense that model results depend on the level of aggregation – but which is poorly understood. Whether operational implementation of such a framework is sufficiently easy and reliable to provide routine policy support remains to be seen. In other words, it is not clear whether general equilibrium modeling will be able to make the transition from a research tool to a standard policy support tool for transport.

2.4. Progress with and challenges for applied economic project appraisal

Glen Weisbrod extracted common themes and policy messages from the papers and from the discussions. His focus was on the application of economic analysis for transport decision making. One key message is that the match between research on wider economic benefits and policy makers' needs is far from perfect. The level at which effects are measured and the tools that are used, with the associated lack of replicability and transferability, reflect a preoccupation with pure research interests; there is no strong correspondence between research and the policy levers available to decision makers. This mismatch carries some risks. First, research may be misused when it is taken out of context. Second, interest groups, in particular from the business community, become increasingly dissatisfied with economic appraisal because it ignores wider issues of core interest to them.

A prime example of such issues is the impact of infrastructure on productivity and competitiveness, measured through conditions of market access, connectivity, and reliability. The state of research on these and other issues, as exemplified by the various presentations, suggests strongly that standard cost-benefit analysis does not capture many of the effects of central concern to interest groups and policy-makers. But the research does not provide a set of operational tools for including them in project appraisal. In particular there is a lack of attention from research for microscopic, intra-agent processes, and their connection to transport infrastructure. In contrast, business-led studies have adopted a case study approach where the wider issues take center stage. The impacts addressed in these case studies concern the effects of infrastructure on market access, connectivity and reliability. And the focus in dealing with these effects is on the recognition of non-linear and threshold effects related to market size.

The dissatisfaction of at least some users with the state of transport project appraisal poses a challenge, but at the same time, should not come as a surprise. The research community is aware of many shortcomings: standard cost-benefit work misses wider effects, which are known to be real and potentially large. The understanding of

some, but not all, of those wider effects can be integrated in the static framework underlying cost-benefit analysis. But there are no general, hard and fast rules for project appraisal. In addressing the challenge, participants cautioned against the generalization from ad hoc case study work. A central characteristic of economic project appraisal is that it consistently applies a consistent methodology. Research can provide such a framework, and include any direct or indirect impact to the extent that tools and data for quantifying them are available. This means that project appraisal cannot be tailored to politicians' or interest groups' concerns, nor should it be. Instead, it is just one imperfect input into an equally imperfect decision-making process. Section 4 develops ideas on approaches to project appraisal in more detail.

3. THE PRACTICE OF TRANSPORT PROJECT APPRAISAL

The previous section reviewed research on the economic impacts of transport infrastructure, and clarified how such impacts are or are not captured in standard cost-benefit analysis. Several participants emphasized that we ought also to look for improvements in the actual practice of project appraisal, where it needs to be recognized that the current practice often falls short of ideal cost-benefit analysis.

In the United States, cost-benefit analysis – in the sense of a formal comprehensive welfare economic valuation – is not systematically applied to transport infrastructure investment projects. Most cost benefit appraisals undertaken are for road projects in rural areas.[6] In these cases, safety benefits are frequently larger than the time savings benefits. Because funding is generally apportioned or allocated by type of project (e.g., resurfacing, capacity expansion, safety, etc.), the analysis focuses on cost effectiveness. Similarly, although documentation of environmental consideration is a legal requirement for federally funded transport investments economic analysis is sometimes done within this context. Cost-benefit analysis is occasionally incorporated in this documentation process. It was noted that because this documentation process occurs prior to the completion of project design, costs sometimes change and the cost-benefit analyses are rarely revised when new cost information about a project becomes available (although new information on environmental impact would occasion a supplement to the documentation process).

One further reason (in addition to the use of cost effectiveness noted above) suggested for this relative paucity of cost-benefit analysis is that overall net benefits are not of prime interest in the decision-making process. Instead, decision-makers are, for example, strongly interested in a project's distributional impacts. Spatial distribution gets particular attention, given the spatial structure of politicians' constituencies. The question as to whether inclusion of distributional impacts in project appraisal – which poses no conceptual problems and for which analysis tools are increasingly available – would lead to wider implementations, was left open. A second possible reason is that the policy practice in the US is to allocate funding geographically even within States as well as allocating funding to different goals, such as pavement maintenance, congestion and safety. There is therefore less reason for a systemic "all projects" benefit-cost analysis. The question was asked whether an imperfect cost-benefit analysis is necessarily useful. But it was also pointed out that fragmentation of the analysis increases the risk of double-counting of benefits.

Cost-benefit analysis is applied more systematically in Northern European countries, although there too it is only one input in the decision-making process. In the United Kingdom, which employs CBA systematically, the results are presented to decision makers in a summary appraisal form, side by side with the results of EIA and multicriteria analysis to reflect the relevance of factors that cannot easily be monetized. Financial and environmental indicators are presented together with a description of how they and the project relate to the governments equity and other policy goals, on a single page. The strength of this system is transparency,

but it tends to leave the political decision makers out of the discussions involved in the economic appraisal. It was noted that the US tradition, structured around EIA, appears to be much more successful in engaging political decision makers in discussions of the economic as well as environmental aspects of projects from an early stage.

For other European countries, there was an impression that cost-benefit analysis is often carried out simply because it is a legal requirement, and it takes place late in the decision-making process, casting doubt on whether it has a strong impact on decision-making.

A potential explanation for the mixed success of cost-benefit analysis in affecting decisions is that there is a disconnection between policy-makers' objectives and the objectives implicit in cost-benefit analysis (e.g. maximizing surplus).[7] Policy-makers may wish to increase densities in cities, or they may aim to boost employment, or they may focus on accessibility or similar performance measures. Although such intermediate objectives don't necessarily clash with surplus-maximization, the connection between them is not always clear. Several suggestions were made to improve the match between what policy-makers are interested in and what cost-benefit analysis provides. First, researchers can increase efforts to arrive at an accurate analysis of a project's impacts. Second, going beyond impact assessments, cost-benefit analysis should be used to avoid serious mistakes, i.e. it should guard against projects that constitute a major waste of resources. Arguably, it has been relatively successful in doing so. Third, researchers could gear their analysis more carefully towards policy-makers concerns, rather than to their own research agenda. On this point, however, it was emphasized that this should not lead to the abandonment of the basic principles of cost-benefit analysis, which are those of welfare economics, as information on a project's impact on efficiency and on economic surplus is a valuable input into the decision-making process. Otherwise said, project appraisal can inform decision-makers on intermediate objectives, but should go further and provide an overall assessment.

4. WHAT KIND OF APPRAISAL FOR TRANSPORT INFRASTRUCTURE IS BEST?

The macroscopic analysis of the economic effects of investment in transport infrastructure suggests that there are modest wider economic benefits from such investments. But different projects show different scales of wider effects, and sometimes negative effects. Care also needs to be taken to avoid double-counting. While the macroscopic literature helps debunk the crowding out argument, it is not of direct relevance to project appraisal. Meso- and microscopic methods seem promising, as they provide ways to extend and improve cost-benefit analysis. But which specific improvements can we suggest? Round Table participants arrived at some common ground, along the following lines.

Standard cost-benefit analysis focuses on a project's direct effects, i.e. it restricts attention to changes in transport users' economic surplus. A first question of interest to policy-makers is how these direct transport benefits translate into (regional) economic benefits, or more bluntly, do time savings really translate into tangible gains. Using terminology introduced by Peter Mackie, in his comments on Roger Vickerman's paper, this is the "**alchemy** question". If there are no wider economic benefits, cost-benefit analysis provides a complete answer, but it does not come in a form that is easily understood by policy-makers. Economic modeling, for example along the lines of applied general equilibrium tradition, can help outline how direct benefits are transmitted through markets and transferred between economic agents like households and firms.

The second question on policy-makers' minds is the "**additionality** question": are there any wider, additional economic effects (benefits or costs) attached to a project? It is useful to distinguish between static

wider impacts and dynamic wider impacts. By static effects we mean productivity impacts, external economies (e.g. increasing returns to scale, agglomeration, thicker markets) and diseconomies (e.g. congestion). On a conceptual level, static effects are easily captured in the framework of standard cost-benefit analysis. On an operational level this is more difficult, but there is progress. By dynamic effects we mean adaptations to changes in transport conditions that take place at the microscopic level, e.g. within households or firms. One example is the ability of spouses to take jobs at greater distances from home, with the housing location determined by school choice rather than employment opportunities. Such dynamic effects clearly matter, because they affect economic welfare, but they are difficult to capture in the static framework of cost-benefit analysis, and little progress has been made to date.

On the advice that could be given to policy makers on the existence or otherwise of wider economic benefits additional to those captured by standard cost-benefit analysis, the position emerging from the discussion was one of caution. While the economic profession's understanding of wider economic benefits is improving, it is insufficient to provide a strong basis for *routine* project assessment. There are several explanations for this situation: limited availability and low quality of data, incomplete theoretical understanding of directions of causality, and econometric issues of identification.

Given this state of the research on additional effects, it seems impractical to recommend the inclusion of wider economic impacts in routine project assessment. The risk of excluding real wider benefits or costs exists, but there was considerable agreement that this is outweighed by avoiding the risk of introducing double-counting benefits. For large projects, and especially for the assessment of investment programs, a more ambitious analysis that addresses wider impacts may very well be justified. The recognition of wider effects in the evaluation of entire programs is particularly important, as the interactions between various parts of the program are likely to be underemphasized (or ignored entirely) in a typical cost-benefit analysis.

It is clear that wider benefits are important for some projects, and that an operational understanding of these effects improves decisions on transport infrastructure investments. There is thus a strong case for continued research and development of empirical and analytical frameworks, including operational general equilibrium models.

A particularly strong warning was made against the adoption of "hard and fast rules", like average multiplier effects, to account for additional benefits. Examples were given of projects where the additional benefits are negative, because of congestion effects that outweigh agglomeration effects (Elhorst *et al*., 2004). Furthermore, discussions of the econometric work on agglomeration effects and on spatial spillovers made it clear that results are strongly context-dependent, and no transferability should be expected. While complexity should not be sought for its own sake, researchers should resist policy-makers calls for comprehensive, simple, and transparent decision making rules to capture wider economic benefits; such rules are inappropriate and may produce highly undesirable outcomes.

A constructive way forward would be for the research community to agree on a practical framework for applied project appraisal. Such a framework may contain guidelines on which effects to include and how to measure them, and can be accompanied by a typology of projects that indicates how broad-ranging the analysis should be for each type.[8] So, while there should be a single framework, the complexity of the method can be adapted to the size of the project: for small projects, the main issue is to get results quickly, so that a less sophisticated approach is preferable; for large projects, more sophisticated analyses may be justified. Even for such big projects, however, it useful to keep in mind that the provision of information early in the decision-making process has a larger impact than information that becomes available further down the line – even if that information is based on a more comprehensive analysis.

Focusing on timely availability of appraisals has its downsides: new information may emerge, and this obviously can affect results. One way of dealing with this is to see appraisal as an ongoing process, where the analysis is updated as relevant information becomes available. Alternatively, the ex ante analysis may contain a quantification of risk, e.g. by specifying several scenarios and attaching probabilities to them.

NOTES

1. It is worth pointing out that Aschauer's work was not originally intended to inform the practice of project appraisal, but rather addressed the issue of whether public investment crowds out private initiative.

2. This problem arises with macro-studies, but also with meso- and micro- studies, and its importance will be highlighted throughout much of this paper.

3. The definition of meso-approaches in these conclusions differs from that used in Vickerman's paper, in that we put general equilibrium work under the meso-approach and not the macro-approach. We do so because general equilibrium models make market interactions explicit, even while they possibly focus on aggregate outcomes. Also, our classification fits better with the meso-scope of the paper by Sue Wing et al. This classification, however, has no bearing on the substance of any of the arguments made.

4. On a technical note, it was pointed out that using the size stock instead of changes may help address endogeneity problems.

5. The costs-benefit analysis for South East airport developments in the UK does not include any measure of wider benefits. The reason is that there is no empirical basis for quantifying then (presentation and comments by David Thompson, UK Department for Transport, at the Workshop on Competition in Transport Markets, ZEW, Mannheim, Germany, 25 November 2007).

6. It was mentioned that many ex post analyses are available for such projects. Cf. http://www.fhwa.dot.gov/planning/econdev/ and http://www.fhwa.dot.gov/hep10/corbor/border/laredo/fhwastatement.htm.

7. We abstract here from the problem mentioned earlier, that economic analysis is often carried out at too high a level of aggregation, so does not speak directly to the policy instruments available to policy makers.

8. Not all participants were convinced that such a single framework is desirable. Some advocated the use of different partial models at different stages of the planning process, or suggested limiting the analyst's role to implementing standard cost-benefit analysis while leaving all other dimensions of the decision to politician's discretion.

BIBLIOGRAPHY

Blum, U. and L. Dudley, 1999, The Two Germanies: Information Technology and Economic Divergence, 1949-1989, *Journal of Institutional and Theoretical Economics* Vol. 155, No. 4 (710-737).

Blum, U. and L. Dudley, 2002, Transport and Economic Development, in: Transport and Economic Development, *European Conference of Ministers of Transport* 199, OECD, Paris (51 – 79) [also in French: Transport et Développement Économique].

Cohen, J.P., 2007, Wider economic benefits of investments in transport infrastructure, JTRC Discussion Paper 07-13

Cohen, J.P. and Morrison Paul, C.J., 2004, Public Infrastructure Investment, Interstate Spatial Spillovers, and Manufacturing Costs, *Review of Economics and Statistics* 86: 551-560.

Cohen, J.P. and K. Monaco, 2007, Ports and Highways Infrastructure: An Analysis of Intra- and Inter-state Spillovers, manuscript.

Elhorst, J.P., J. Oosterhaven and W.E. Rom, 2004, *Integral cost-benefit analysis of Maglev technology under market imperfections*. SOM Report 04C22, University of Groningen (forthcoming in Journal of Transportation and Land-Use).

Graham, D.J., 2007, Agglomeration economies and transport investment, JTRC Discussion Paper 07-xx

Johansson, B., 2007, Transport infrastructure inside and across urban regions: models and assessment methods, JTRC Discussion Paper 07-12

Moreno, R., E. Lopez-Bazo, E. Vaya, M. Artis, 2004, External Effects and Costs of Production, Chapter 14 in *Advances in Spatial Econometrics: Methodology, Tools, and Applications* (L. Anselin, 1981, R.J.G.M. Florax, and S.J. Rey, eds.), Berlin: Springer.

Oosterhaven, J. and L. Broersma, 2007, Sector Structure and Cluster Economies: A Decomposition of Regional Labour Productivity. *Regional Studies* 41/5: 639-59.

Sue Wing, I., W.P Anderson, and T.R. Laksmanan, 2007, The broader benefits of transportation infrastructure, JTRC Discussion Paper 07-10

Vickerman, R., 2007, Recent evolution of research into the wider economic benefits of transport infrastructure investments, JTRC Discussion Paper 07-9

Weisbrod, Glen E., and Brian B. Alstadt, 2007, Progress and challenges in the application of economic analysis for transport policy and decision making - Concluding comments for the research round table on infrastructure planning and assessment tools, JTRC Discussion Paper 07-14

INTRODUCTORY REPORTS

RECENT EVOLUTION OF RESEARCH INTO THE WIDER ECONOMIC BENEFITS OF TRANSPORT INFRASTRUCTURE INVESTMENTS

Roger VICKERMAN[1]
Centre for European, Regional and Transport Economics
University of Kent
Canterbury
United Kingdom

SUMMARY

1. INTRODUCTION 33

2. THE PURPOSE OF INFRASTRUCTURE STUDIES 34

3. MACRO-LEVEL EVALUATION OF INFRASTRUCTURE 36

3.1. Measurement 36
3.2. Output 36
3.3. Productivity 37
3.4. Employment 37
3.5. Alternative models 37
3.6. Land use transport interaction models 38
3.7. Computable general equilibrium models 38
3.8. Ex-post studies 39

4. MARKET LEVEL EVALUATION OF INFRASTRUCTURE 39

4.1. Competition effects 40
4.2. Agglomeration effects 41
4.3. Labour market effects 41
4.4. Implications for appraisal 42

5. MICRO-LEVEL EVALUATION OF INFRASTRUCTURE 42

5.1. Labour market effects 43
5.2. Business organisation effects 43

6. CONCLUSIONS AND IMPLICATIONS 44

NOTES 46

BIBLIOGRAPHY 47

Canterbury, September 2007

1. INTRODUCTION

The debate on whether there are wider economic benefits from transport infrastructure investments continues to cause debate and controversy. This debate occurs both between analysts seeking to find a robust method for identifying and measuring the size of such benefits and between policy makers seeking to justify or refute the need for a particular investment. It is timely to review progress on arriving at a consensus view of the contribution of infrastructure to the wider economy which is consistent with best practice in appraisal. This paper will review progress and try to bring out some common themes for discussion.

The main aim of this paper is to bring together the various alternative methodological approaches to this problem which differs not just in the detail of the analysis, but more significantly in the scale at which the analysis is undertaken. It is argued that it is of particular importance to understand the way in which changes in the provision of transport affect microeconomic decisions, including those within firms and households, and to understand the operation of markets as well as to model the resultant flows and their macroeconomic consequences.

By wider economic benefits we mean all economic benefits which are not captured in the direct user benefits of the type which are normally analysed in a well constructed transport cost-benefit analysis after allowing for environmental and other directly imposed external costs. Such benefits are typically thought of as being positive, but logically they can also be negative implying that the direct user net benefits could over-estimate the value of a project. The traditional transport appraisal approach assumes that a well-specified cost-benefit analysis will capture all the economic impact of a transport infrastructure investment since users will be willing to pay exactly the economic value of the transport to them. Any attempt to add on wider economic benefits would thus represent double-counting. On the other hand macroeconomic studies have shown strong positive links between the aggregate level of infrastructure investment and economic performance as measured by GDP or productivity growth or employment. If it is the case that increased investment leads to faster growth then this needs to be identified and included in demand forecasts. Are these positions consistent, and if not can they be reconciled?

There are two main avenues of debate to effect such a reconciliation. One relates to the assumptions made about the nature of competition and returns to scale. This argues that when the traditional assumption of constant returns to scale in perfectly competitive markets is relaxed there will be agglomeration effects which generate wider benefits not captured in the user benefits. The second argues that the non-marginal nature of many large scale investments results in traditional forecasting approaches failing to capture the changes in behaviour of transport users.

The intention of this paper is to explore the linkages between these different approaches to identify the relationship between the different levels of analysis in order to develop a way towards a more synthetic approach which can capture best practice. However, it will be stressed that the purpose of any analysis always needs to be made clear in order to avoid inconsistencies between the appraisal of individual projects and overall evaluation of policy towards networks.

There will be a brief review of the objectives of infrastructure studies followed by a summary of the key issues which emerge from the various types of study in order to identify common themes and differences. This will lead to an attempt to synthesise the key issues and identify priorities for further work.

2. THE PURPOSE OF INFRASTRUCTURE STUDIES

One of the major areas of confusion between the different types of study into the wider economic impacts of infrastructure is that different studies have clearly different objectives. These need to be understood before any attempt is made to reconcile the results or to apply the results of one type of study to a different case. At the lowest level comes the investment appraisal of an individual link in a network. At the highest level comes the attempt to relate overall macroeconomic performance to aggregate investment in infrastructure and hence to the stock of infrastructure. The difficulty is knowing whether an elasticity obtained from the macro-study is in any way applicable to a single investment decision.

Investment appraisal is where the critical decisions are taken about transport infrastructure. The majority of individual decisions are about link improvements. These have historically been determined by methods which rely almost exclusively on the identification of user benefits, heavily dominated by user time savings, relief of congestion and reduction in accidents, but also allowing for environmental impacts. But investment decisions based on cost-benefit type procedures depend critically on the accurate measurement of future demands which in turn require correct modelling of the responses of users to the new investment (see Vickerman, 2007a, b for recent discussions of this issue). This is the problem related to a move from the traditional assumption of fixed trip matrices in which new transport investments would simply lead to a reassignment of traffic in a network rather than a revision of travel patterns.

Allowing for generated or induced traffic is a two-edged sword: failing to allow for it can lead to the sort of underinvestment which produces more congestion rather than less and hence overall benefits less than the estimated user benefits (see Venables, 1999, for a discussion of the theoretical basis of the problem); grasping it can lead to the optimism bias often used as a basis for justifying projects which might otherwise not appear to generate sufficient user benefits (as shown by Flyvbjerg *et al*., 2003).

One of the major problems with the traditional investment appraisal exercise is that it is seen as a purely transport exercise which ignores the interactions between transport and all the activities which use transport. It ignores in other words the market situation in which transport is located and how it interacts with the locations of economic activity, residences, workplaces, sources of inputs, markets for outputs etc. This is why a market based approach is essential to understand the way in which a particular transport investment serves particular markets. The traditional theoretical approach to appraisal relied on the well-known results of Dodgson (1973) and Jara-Diaz (1986) that, assuming that all other markets were in perfect competition such that price equalled marginal cost, the user benefits would exactly equal the total benefits because the full value of transport to all users would be exactly measured. Jara-Diaz demonstrated how these results might differ if the state of competition differed in the regions linked by the transport improvement, but the simplest solution was always to ignore the problems of imperfect competition.

This may not be unreasonable for the typical appraisal of a link in a network which makes only minor changes to overall accessibility, but where there is a need to appraise a fundamental change in a network, or indeed a network in its totality the dimension of the problem changes and the market situation cannot be ignored. The temptation has been to look for simple adjustments to the user benefit result – a wider economic benefits multiplier – which would enable the aggregation of a set of unspecified wider benefits. This multiplier may be thought to be related to the price-cost mark-up associated with imperfect competition,

the greater the imperfection in competition the more pro-competitive would be a transport improvement and hence the greater the uncaptured benefits from a simple user benefits appraisal. Such an approach needs to consider that the degrees of imperfection will differ between sectors of the economy and that in some cases poor transport may serve as a useful trade barrier protecting a region from the more aggressive competition gained by a larger region enjoying greater scale economies in its production. Improved transport in such situations may be globally beneficial but to individual regions could cost both output and employment. This emphasises clearly the need for a careful definition of the geographical scale of the region of interest from a transport improvement, bringing it again closer to an integration with the various spatial markets relying on that transport.

Scale economies are closely related to the existence of imperfect competition. The existence of scale economies implies the greater concentration of economic activity and this has clear implications for the most efficient transport network. As scale economies increase the barrier to trade posed by transport costs can more readily be overcome and hence there is a tendency towards concentration. If transport costs fall beyond a certain level then the advantages of concentration for the individual activity may become less and dispersion may arise. However, this ignores the interaction between the activities and it is these agglomeration forces which will tend to dominate and preserve the concentration. In this case improved transport is no longer unambiguously pro-competitive. The role of transport in agglomeration has been explored thoroughly in the new economic geography approaches to the spatial economy (see for example Fujita *et al*., 1999; Fujita and Thisse, 2002) and the implications for transport appraisal considered in Venables and Gasiorek (1999). This has implications at two levels. One is the way inter-regional transport can accelerate the agglomeration of one region at the expense of another, the second is the way intra-regional transport can reinforce that process. The clearest example of the latter is the way improved commuter transport can help keep down the real unit cost of labour to firms whilst maintaining the real wage differential to workers encouraging them to remain in the agglomeration.

Market-based approaches only go so far in helping our understanding of the impact of transport on the economy since like the macro studies they work on the basis of average propensities and elasticities. But in order to understand the real impact of new investments in transport we need evidence of how these changes actually affect activities at the micro-level, that of the individual firm and household. This is not just about these agents' market behaviour, where they buy and sell, where they live and work, but how their activities are organised internally. Hence we need to examine how firms will reorganise their operations to reflect reduced transport costs, will they concentrate all activities into a single location or will they use existing locations, but functionally specialise between these locations? Similarly for households, not only will individuals be able to use improved transport to enlarge their own labour market search, but different members of the household may be able to match a wider range of potential job offers thus enabling the household to reallocate activities between household members or its optimal location.

How does this relate to the overall contribution of transport to an economy's macroeconomic performance? Is this a simple aggregation of the impact of the individual links, or is it a problem of a different dimension? There are two aspects to this which again take us back to the basic question of what it is we want to measure. The first is geographical scale. Whilst it is not suggested that all transport improvements are likely to be a zero-sum game as far as individual regions are concerned, most will have some redistributional effects between regions, either relatively in that they benefit some regions more than others or absolutely in that there are gainers and losers (but the former could compensate the latter and still leave an overall gain). The second is the need to consider the mechanism by which it is believed that the transport improvement works through the economy. One approach would to treat it simply as an adjustment to the price of a key input which leads to changes in the relative prices of activities according to their transport content, and the impact on competition – this takes transport simply as a derived demand. The other views transport as a substitutable input to activities such that it has an impact on total factor productivity. The simple aggregate production function approaches to transport infrastructure fail to make the relevant mechanism clear leading to some of the problems of interpretation of the results from such studies.

3. MACRO-LEVEL EVALUATION OF INFRASTRUCTURE[2]

Much has been written on the macro- evaluation of infrastructure in terms of its impact on productivity and growth, typically using some form of production function approach. There has been a considerable output of empirical studies which aim to test the impact of infrastructure at both national and regional level. The main issues to emerge from this are the problems of measurement and the difficulty of making firm statements about the impact.

3.1. Measurement

The first problem with most macro studies is that rely on a very aggregate view of transport infrastructure, typically just using the volume of investment or stock of infrastructure capital as the variable which impacts on output. The problem with measuring infrastructure by its capital value is that this is likely to be a much less accurate measure of the services provided by that infrastructure than is the equivalent value of private capital. This is for two reasons: infrastructure has high asset specificity (zero opportunity cost); and is much less likely to be provided under conditions representing a free market in which the price paid is indicative of the marginal productivity of the asset. For this reason many studies prefer to use physical measures of infrastructure such as lane kilometres or track kilometres (usually expressed as a density per square kilometre to standardise for differences in region or country size). This is closer to incorporating a clearer measure of the level of service provided by the infrastructure.

The second issue is the what is being measured, output, productivity or employment. To some extent this depends on the purpose of the study. Studies concerned about the role of infrastructure in growth or convergence will use output measures such as GDP or GDP/capita. Technically, to ensure consistency with the normal Solow growth model premise, convergence studies should be based on a productivity measure of GDP/worker to allow for less than full employment. For political reasons there has obviously been a lot of interest in the employment impacts of infrastructure since this is a way of selling expensive publicly funded projects to an electorate on the promise of more jobs. Each of these approaches implies a very different underlying process of infrastructure impacts.

3.2. Output

Output-based models imply infrastructure working essentially as any other factor of production; regional economies with more infrastructure will have more output, the logic of this argument actually tends to derive from the opposite – that the lack of infrastructure would act as a constraint on output. Regions with denser infrastructure are presumed to have a more efficient transport system which will enhance the productivity of other factors of production, especially private capital, and this will generate the growth bonus which formed the basis of Aschauer's (1989) argument in the work which sparked the current round of interest in the role of infrastructure.[3]

The problem with such an approach is that it takes no account of the way on which infrastructure is used by the activities within the economy in question. A given volume of infrastructure can be either adequate or inadequate for the needs of the economy depending on, for example, the sectoral structure of the economy

or the physical configuration of the infrastructure. It is clear that in a purely aggregate demand view of the production function, the scale of infrastructure spending will affect output and growth simply because of its scale. Since construction expenditure has a particularly rapid pass-through and usually generates a relatively large employment multiplier, infrastructure investment is a good means of providing a short-run boost to the economy as long as it does not crowd out other more productive investments. This was the motivation of the Aschauer work, to deny the argument that public infrastructure would be a less good use of available investment funds than expenditure on private capital. But this is not helpful evidence for use in planning or appraising infrastructure and, as many subsequent studies have argued, may confuse the direction of causality. This was the essence of Gramlich's (1994) review article which focused on the importance of identifying the specificity of particular infrastructures.

3.3. Productivity

The spillovers in productivity have become the focus of more recent work, not only between sectors but also spatially, following the work of Holtz-Eakin (1994) and Holtz-Eakin and Schwartz (1995). This is not always overtly spatial, except in the limited sense of inter-state comparisons (see Pereira and Andraz, 2004, for an example) although a more detailed study using county data does come closer to examining the more local complementarities in network developments (Boarnet, 1998).[4] The issue of the endogeneity of infrastructure and overall output leads to a consideration of the appropriate leads and lags. It is clear that there may be a lags in both directions; the time taken for output growth to generate the demands which can justify new infrastructure and the time for activities to adjust to a new level of transport provision. However, there is also the possibility of a leading response in which the promise of major new infrastructure stimulates investment as firms try and obtain a first-mover advantage to exploit new opportunities. All of this adds to the potential econometric confusion which even the most sophisticated techniques find it difficult to unravel.

3.4. Employment

The alternative use of employment data addresses a slightly different problem. The underlying assumption is essentially one of fixed input coefficients so that the impact on employment is directly related to that on output. As Jiwattanakulpaisarn (2007) shows, the impact of infrastructure on jobs is not universally positive (especially when taking into account different types of road) and this, along with other evidence, may cast some doubt on the wisdom of policy makers pushing for infrastructure expansion. The problem here is that effective infrastructure which reduces transport costs will induce the substitution of cheaper transport for more expensive, less mobile inputs. Land is one obvious substitute – the justification for just-in-time production techniques saving on inventories – but labour, especially less skilled labour, may also be a casualty as it too may be less mobile. Furthermore, the improved infrastructure increases the competition from more mobile labour from outside the region which may take up any increase in jobs resulting from the higher level of activity. Hence the improved infrastructure is good for the local economy in terms of growth but may be bad for the employment prospects of (some) local residents. This reinforces the need to look at more disaggregated models which allow for the differences between both infrastructure type, sectors and employment structure.

3.5. Alternative models

Fully aggregate econometric models have not been found to be appropriate for this and most work has been done using advanced land use-transport interaction (LUTI) models or more recently spatial computable general equilibrium (SCGE) models. These can capture more specific spatial impacts of infrastructure, but tend to be limited by their data requirements and/or their need to make highly simplifying assumptions about the operation of the various markets or the spatial coverage of the impacts. A number of studies have carried

out ex ante evaluations both of network developments as a whole and of specific infrastructure improvements. Many of these have been as part of European Union funded projects to look at the potential impact of the development of the Trans-European Networks (TENs).

3.6. Land use transport interaction models

LUTI models have been used by urban planners for some time as extended travel demand models which allow for the interaction of transport and land use (Simmonds, 1999). More recently LUTI models have been extended to deal with regional and inter-regional impacts of transport development (Wegener and Bökemann, 1998; Bröcker *et al*., 2004). These models vary in the precise way they operate but essentially comprise a series of linked detailed models covering travel/transport, production and GDP, labour markets and population and land use. At the heart of the model is the transport sector. Changes in accessibility which change the cost of transport, impact on both production and the labour market. The production sector is typically modelled through a set of input-output relationships which define the need for transport to move goods into and out of a defined spatial area. This includes the need for labour inputs which interacts with the available labour force (and hence local population) to determine commuting and migration patterns. Land use acts as a constraint on the development of the economy since production and the resident labour force have minimum requirements for land.

The main problem with LUTI models arises from the assumptions implicit in each of these constituent models. Hence input-output models are often static in nature, dependent on existing patterns of behaviour and are solved by ensuring that equilibrium is reached in each relevant market. Similarly the links between population, labour force and labour demand also depend on assuming that existing patterns of behaviour do not change, when the evidence from major changes in the transport network is that behaviour can actually change quite significantly. Furthermore, the models make assumptions about the land-use requirements which do not allow for changing capital and labour intensities and tend to treat different sectors equally. LUTI models assume perfectly competitive markets in which the market outcome is a valid measure of the welfare change.

3.7. Computable general equilibrium models

CGE models, by their nature, also assume equilibrium and are based on the fundamental input-output relationships in the economy, but in this case they allow for more interaction between constituent markets in order to achieve a general equilibrium of all sectors through a process of numerical iteration. The key difference is that CGE models have at their core the possibility of assuming that consumers display preferences over differentiated goods which are produced by imperfectly competitive firms (Bröcker, 2000, 2004: Bröcker *et al*., 2004). Because of this use of a utility function CGE models can make a direct estimate of the welfare effects resulting from a change.

Bröcker's CG-Europe model generates three important results. First, despite significant changes in transport costs and accessibility occasioned by the development of the TENs, the impact on welfare is relatively modest (equivalent typically to less than 2 per cent of regional GDP). Secondly, the network as a whole has positive impacts on some regions and negative impacts on others. Thirdly, specific investments have differential impacts both on the specific regions they serve and in the added value they bring to the European economy as a whole.

More specific project applications include an evaluation of the regional impacts of highway developments in Japan (Miyagi, 1998, 2001) and to evaluate the impact of a high-speed rail link between the Randstad and the Northern Netherlands (Oosterhaven and Elhorst, 2003; Elhorst *et al*., 2004). The Dutch RAEM model focuses not just on the output and welfare implications, but also very specifically on the labour market

since the improvement to transport will not just affect the location of employment but also the residential location decision. This introduces further difficulties because it requires not just a balance of production and consumption in the goods markets, with a potential response through migration to long-term imbalances, but a period by period balancing of labour markets demands and supplies, zone by zone. Furthermore, once the key beneficiaries are passengers rather than goods some of the simplifying assumptions used in the typical CGE structure become less plausible. For example, the use of 'iceberg' transport costs, in which the cost of transport of a good is subsumed into the value of the goods moved such that they are worth less at the destination than at the origin by the amount of the cost of transport, is inappropriate for passengers. Similarly the assumption of constant costs of transport per unit of distance is even less appropriate for passenger transport.

Nevertheless the application of a CGE model to this project has produced an interesting set of results. The wider benefits are shown to vary significantly as a result of the precise nature of the project and the region studied (especially core-periphery differences), and constitute a higher proportion of direct benefits than earlier studies suggested, of the order of 30-40 per cent. These wider benefits are higher than theoretical simulation models have suggested; SACTRA (1999) suggested that a figure of 10 to 20 per cent was a likely range, following the conclusion by Venables and Gasiorek (1999) that 30 per cent was a likely to be exceeded in only a few cases. (It is worth noting however that in the earlier version Oosterhaven and Elhorst had produced a figure of 83 per cent). What is clear from Elhorst *et al*. (2004) is that the degree of detail in the modelling of labour market responses may be crucial here.

But CGE models do still have major drawbacks: assumptions about equilibrium, the need for large data inputs from existing sources and the 'black box' nature of large models all limit their usefulness and ease of application. Thus far CGE models have tended to be used for cases where there are thought to be significant non-transport impacts; their use as part of the regular appraisal of minor transport projects might be difficult to justify. SACTRA (1999) strongly recommended that the UK Government should invest further in this approach. Following an assessment by RAND Europe (Gunn, 2004), the Department for Transport (2005) has issued a discussion document suggesting how this could be achieved.

3.8. Ex-post studies

Most of the empirical evidence quoted above relates to ex ante studies of potential future projects. There have been relatively few in depth ex post studies of the revealed impacts of completed projects. One of the difficulties is that of identifying the specific impacts of a project over the timescale necessary to allow for these to be revealed. However, one of the relatively few ex post studies indicates a much lower level of impact than ex ante studies. Hay *et al*. (2004) have shown how a very significant project, the Channel Tunnel, has not produced significant wider benefits over its first ten years of operation, at least on the regional economies close to the tunnel. In fact it is suggested that any wider benefits are so dispersed and so long term as not to be easily detectable.

4. MARKET LEVEL EVALUATION OF INFRASTRUCTURE

The previous section has identified in several places the importance of disaggregation in order to identify the particular needs of individual sectors and activities. We have already noted the extent to which the labour market is likely to play a major role in this process. Disaggregation by space is also an essential element of a fuller understanding of the impact which infrastructure investment will have. This emphasis on

markets becomes important once we move out of the comfort zone of the perfect competition assumption. In a world of increasing returns and imperfect competition we need a more subtle evaluation of the role of infrastructure.

The theoretical justification for this approach is provided by the new economic geography or new spatial economics. The principal result of this approach is to demonstrate that agglomeration can take place and continue without a process of self-balance setting in. Transport costs play a key role in this process. However, the nature of this approach is that the impact of any particular reduction of transport costs cannot be determined a priori. It will depend on the initial level of transport costs, the degree of agglomeration already present, the size of each market, the extent of scale economies and of the backward and forward linkages within that market (Fujita *et al*., 1999; Fujita and Thisse, 2002).

What becomes relevant here is the extent of the mark-up over marginal cost in the transport-using industries. In perfectly competitive sectors there is no mark-up and hence any changes in transport costs will have to be passed on directly to the final activity, so the extent of the impact on the wider economy is dependent on the elasticity of demand for that final activity. Since the amount of transport demanded depends directly on the demand for the final activity the direct user benefits capture all the economic benefits. As mark-ups increase there is in effect a wedge driven between the market for the transport-using activity and the transport associated with it. Any reduction in transport costs from new infrastructure does not need to be passed directly on to the customers of the final activity, but firms can use the opportunity to increase or reduce the mark-up. Reducing the mark-up by passing on more than the reduction in transport costs could be a way of increasing a firm's market area and gaining market advantage over firms in a more competitive market. On the other hand firms may use the fall in transport costs to increase the mark-up, for example to invest so as to reduce other costs and gain from potential scale economies. It is also possible that the net impact can be negative. If the mark-up is negative, for example where there are industries with significant subsidies, such as in economically lagging regions, then the direct user benefits may over-estimate the total economic benefit. Hence the ultimate impact from any infrastructure project is likely to be unpredictable, both in terms of magnitude and sign.

There are three main elements to the total economic impact. First is the impact on competition in the affected regions, secondly there is the impact on the ability to gain benefits from the change in market power through agglomeration, and thirdly is the impact on the linkages and in particular on backward linkages such as the labour market. Once these have been assessed we have to identify how to include them in a full cost-benefit framework.

4.1. Competition effects

The impact on competition is ambiguous. In perfectly competitive markets, as we have seen, the impact of increased competition is essentially neutral and should be adequately captured by the direct user benefits. In imperfectly competitive markets, the direct effect of any increased competition resulting directly from lower transport costs is also likely to be essentially neutral in its impact. It is traditionally argued that monopoly power is derived from the effective barriers to competition provided by higher transport costs so that reductions in such barriers are pro-competitive, reducing monopoly mark-ups and hence there is a wider benefit resulting from the reduction of prices. On the other hand such competitive pressures if they do exist may also drive firms out of the market and the effect of lower transport costs is to reduce the number of firms able to compete in the market in the long run. It is likely that such effects cancel each other out in most cases and thus there is little in the way of wider economic benefits which can be added.

There may be some exceptions to this where new links are created which have such a significant impact on transport costs (which are already very high) that significant market restructuring takes place introducing competition to previously protected local monopolies. This is the 'unlocking' argument advanced by SACTRA

(1999) and reaffirmed in its latest guidance by the UK Department for Transport (2005). These are likely to be rare in most developed market economies.

4.2. Agglomeration effects

Much more significant than the market competition effects are the agglomeration benefits which may result from the change in transport costs. The argument here is that the rise in output which follows from the lower transport costs has cumulative effects through the way in which firms interact in a market. This involves both localisation economies, in which firms within the same industry benefit from proximity to each other through such factors as specialised labour pools or shared R&D, and urbanisation economies, in which firms obtain a form of public goods benefit from the existence of an urban infrastructure including knowledge, research and culture as well as the physical infrastructure. The larger the market the greater the likely net additional impact which arises because there is an additional impact on productivity.

There has been a long debate over the extent to which urban size and productivity are related, and the direction of causality, but there is an increasing consensus that there is a strong positive relationship which can have a significant additional impact on the benefits from transport improvements (Fujita and Thisse, 2002; Venables, 2007; Graham, 2005). This argues that although the lower transport costs may cause firms to increase the size of their market, that increased size provides an incentive for the firm to enjoy scale economies and to benefit from proximity to other more efficient firms. Typical productivity elasticities are in the range 0.01 to 0.1. Ciccone (2002), using data for EU regions, finds an elasticity with respect to employment density of 0.05. Graham (2005) finds for UK industries a weighted average elasticity of 0.04 for manufacturing, but significant variations between industries with some as high as 0.2, and an average of 0.12 for service industries. Graham also identifies some important variations between regions reflecting different degrees of localisation of industry groups.[5]

A further element of this output benefit under imperfect competition is that because productivity is increasing, the direct user benefits will also be greater than would be the case under an assumption of perfect competition. The largest direct user benefits from most projects are time savings, valued relative to the wage level assuming that wages reflect productivity. The increase in productivity implies that a higher value of time savings should be applied. But the increased productivity enables firms to increase output (or produce the same output with fewer workers) which implies an uplift needs to be applied to the time savings.

4.3. Labour market effects

The basic advantage which some regions obtain in an imperfectly competitive world derives from a larger market size which enables firms to increase both output (scale) and productivity. However, it is useful to break that larger market size effect up into a pure market size effect and the backward and forward linkages which are associated with agglomeration. One of the key backward linkages relates to the labour market. As transport costs are reduced labour markets become larger as commuting times are reduced and firms have access to a larger labour supply. This enables firms to benefit both from wage levels which might be lower than they might be as result of more competition in the larger market, but access to more skilled labour which will be more productive for the reasons discussed above.

Normally it would be expected that there would be a wage premium at the market centre reflecting its greater accessibility, scale and productivity effects, but also to reflect the wage necessary to attract labour to commute in from across the wider region. As transport is improved more workers find it attractive to work in the market centre, both in terms of there being a larger catchment area for which commuting is feasible and more people at each location find it worthwhile to seek work in the centre rather than elsewhere (or not at all), or if they are working in the centre to be prepared to work longer hours. Hence there is an output effect which

arises because of the increased size of the labour market. Where there is also a productivity effect due to agglomeration effects at the market centre the output effect from the expansion of employment is added to by the increased output of all existing workers (see Venables, 2007).

4.4. Implications for appraisal

Whilst this provides an interesting academic debate on the existence of agglomeration economies and the way they can be manifested in terms of wider economic benefits from transport investment, do these approaches provide us with an effective means of enhancing appraisal techniques of new infrastructure? Most applications to date have been in the context of major investment projects. We have noted above the application of LUTI and SCGE models to such projects as the TENs, Dutch high-speed links and Japanese highways; similar exercises have been carried out for a variety of other major projects across the world. The most detailed application of agglomeration-based modelling has been in the context of the Crossrail project for a major cross-London underground rail link (Department for Transport, 2005). Such exercises remain difficult and costly in terms of both data and modelling and frequently can only be justified where the scale of a project is large enough to cover the cost of such modelling. The goal is to have a simple and easily applicable appraisal model which can capture the same effects for any project, not least because much network development is actually the result of a series of independently taken link decisions.

Note that it is not the size of an infrastructure project which determines the scale of the wider economic benefits. Large projects are likely to have a wider impact in terms of greater direct user benefits, but the wider benefits are not simply proportional to the direct user benefits. Some relatively minor projects, the 'unlocking' projects, can have disproportionately large wider benefits, whereas some very large projects may have relatively little impact on the key scale, productivity and linkage effects. This is why there is no a priori reason for applying a simple wider benefits multiplier. It also demonstrates that seeking a simple output elasticity as in the macro analyses can be misleading. However, even at this level the empirical evidence (such as that presented by Graham) demonstrates the variability between sectors and regions of the likely impacts of given level of infrastructure investment. It is this which argues strongly for the addition of more micro level evidence of the impacts within firms and households.

In the UK the official guidance following the 1999 SACTRA Report was to consider wider economic benefits through an Economic Impact Report where there was a confirmed regeneration benefit or where the capital value of the project was greater than £5 million. The Eddington Report (2006) identified the importance of all these processes and particularly importantly wanted these to be identified at an early stage of project development – there is a clear problem that if the wider benefits are only ever considered for a fully developed project proposal many more effective options may have been rejected.

5. MICRO-LEVEL EVALUATION OF INFRASTRUCTURE

At the micro level there has been much less systematic work showing how infrastructure changes the behaviour of firms and individuals. Some work on the impact of high speed rail has shown that the impacts on the internal organisation of firms may be more significant than the overall redistribution of activity. The increasing interest on the impact on labour markets also demonstrates the need to make more of a connection between the different levels of analysis as the micro-behavioural decisions can be

linked to overall labour market operation and to productivity and agglomeration effects. To explore this further requires in-depth studies of changes which have occurred as the result of the introduction of new infrastructure.

5.1. Labour market effects

Gibbons and Machin (2005) provide some evidence of the impact of new infrastructure on individual behaviour by estimating the impact on house prices of the provision of a new Underground line in London. This looked at the effect of new stations on values at different distances from the station assuming that the new station increased accessibility to workplaces in Central London. The results showed that there was a clear positive link, average values rose by 9.3% more in areas affected by the new stations and values rose by about 1.5% for a 1 km reduction in access to the station. Such results are based on assumptions about where people might work and ignore job creation in the areas affected by the new stations, but do seem remarkably robust econometrically. Moreover they imply rather higher values of accessibility (as measured in house values) than do cross-section results comparing areas of different accessibility. This suggests that there is a strong positive response to the addition of new infrastructure which traditional model approaches based on assumptions of market equilibrium may fail to identify.

5.2. Business organisation effects

Turning to the impact on business, most studies have been carried out into the impact of the French TGV lines, particularly to examine the relative impacts on Paris and the provincial cities. Although such services led to a substantial growth of traffic the impact on the local economies of the cities served was much less certain. Generally such services cannot be shown to have had a major impact on the net redistribution of economic activity between Paris and the provincial cities, or on the overall rate of growth of these cities.

The evidence includes studies of the TGV Sud-Est, Paris-Lyon, opened in 1981 (Plassard and Cointet-Pinell, 1986), the TGV Atlantique, including a study of Nantes, opened in 1989 (Klein and Claisse, 1997; Dornbusch, 1997), and early studies of TGV Nord, including studies of Lille and Valenciennes, opened in 1993 (SES, 1998; Burmeister and Colletis-Wahl, 1996). All of these studies demonstrate a considerable growth in traffic between Paris and each of the provincial cities since the opening of TGV. The impact on business traffic is more mixed. In the case of TGV Sud-Est there was a substantial growth, in the case of TGV Atlantique as a whole there was a marginal reduction in business traffic, but the period immediately after opening coincided with a serious recession.

The Paris-Lyon study showed a major impact on the pattern of mobility, but with changes in both directions. Essentially many businesses in both locations modified their pattern of working leading to increases in travel in both directions. There was no overall net impact on the economies of either of the major cities, but a general tendency towards the concentration of economic activity towards these major cities from the regional hinterland, especially in the Bourgogne and Rhônes-Alpes regions. This centralising effect of high speed rail is now a well established impact.

In the case of TGV Atlantique, the development of business traffic showed interesting contrasts. Tours, at 240km (1h 10m) from Paris showed a significant reduction in business traffic of 24 % in total and 40% by rail between 1989 and 1993. Nantes, 380km (2h 05m) from Paris showed a total increase in business traffic between the two cities of 66% with a tripling of rail traffic. In 1989 some 73% of the traffic originated in Nantes, but there was a much larger increase in Paris originating traffic (+99%) compared with that originating in Nantes (+55%) with the coming of TGV. In Nantes there was considerable anticipation of the coming of the TGV in the light of some of the experiences of Lyon, but this was mainly felt in property development and relatively little impact on, for example inward movement of enterprises was identified. As in the case of Paris-Lyon there

was evidence of a degree of internal reorganisation within firms to take advantage of changing transport costs for business travel. For Toulouse, 700km from Paris (5h 06m), the increase in total business traffic after the TGV was introduced was 21%. In this case however, more of the increase in traffic was locally based (+35%) and Paris originating traffic actually fell by 5%. However, much of the driving force behind these changes was seen to be the business cycle rather than changes in the supply price of transport. The key factor here is seen to be the differential impacts on the cities around 2 hours from Paris, those closer and further away did not benefit to the same extent. This is consistent with other evidence that high speed rail has its major impact in the 2-3 hour journey time band.

For TGV Nord the distances are shorter than would be likely to make a major impact – Lille is just 1 hour from Paris. Nevertheless total traffic grew substantially over the first three years of operation, 5% in the first year, 6% in the second year and 11% in the third year. Except in year two the growth was stronger for traffic originating in Nord-Pas-de-Calais region. What is of interest is that rail showed much stronger growth in the latter market than for traffic originating in the Paris region.

The Lille study suggested that about one-third of all business travel was changed as a result of the introduction of TGV (both outward from regional based enterprises and inward by clients of such enterprises). However, 90% of enterprises identified no impact of TGV on their overall activity. As in the earlier studies there was evidence of some internal reorganisation, described in this study as a form of "spatial dualisation". Some considerable differences were noted between Lille and Valenciennes, just as in the Paris-Lyon study there was some evidence of centralisation of activity towards Lille, the major regional centre, at the expense of the weaker one, Valenciennes.

The French studies demonstrate the critical importance of time thresholds in the impact which TGV services will have on the relationships between major centres. Thus the headline time of two hours between Paris and Lyon was very significant. This is particularly true of the diversion of trips from air to rail, but it has also affected the potential for generation of new journeys reflecting new activity possibilities. A further issue is that although much of the success of the TGV in generating new traffic has been by providing through services from locations off the new infrastructure the economic spin-off for these centres has not been as great as the for the locations on the main lines.

Thus there does seem to be clear potential for further work on the direct impact of new infrastructure on the behaviour of individual, households and firms which may produce rather different implications than the traditional market-based or macro-based models.

6. CONCLUSIONS AND IMPLICATIONS

The main theme of this paper is that it is necessary to be clear as to the objective of any study of the impact of transport infrastructure on economic activity as the nature of the answer required will affect the appropriateness of different methodologies and different methodologies may give very different answers. These differences do not necessarily reflect inconsistency in results but rather incompatibility in method.

Much more development has been carried out of macro studies of the overall impact of transport on the economy. These have their place as part of our understanding of the basic relationship, but are not necessarily compatible with methods for the planning or appraisal of new infrastructure. The endogeneity question

remains central to the problems which such methods pose, though recent work has been able to produce more stable results, especially where the infrastructure itself is disaggregated and made more homogeneous.

The recent (renewed) growth of interest in measuring agglomeration effects is central to our understanding of the more market based approaches. These allow for specific variations in the degree of imperfection of competition, both in product markets and in labour markets. This makes them more suitable as inputs to the appraisal process, although there is a question as to how far the data requirements of such procedures can be met in the case of other than abnormally large projects. What is clear is that there is little evidence of there being standard transferable multipliers region to region or project to project.

Where there is still a considerable need for further work is in genuinely micro studies of the response to specific changes in order to understand something of the process of decision making in response to changed transport provision by both individuals and households, and firms. The evidence from both labour market studies and firm studies of the impact of new rail links is that these responses may be more significant than otherwise assumed.

But full appraisal will continue to require inputs from all three types of study to be able to understand the overall economic impact of new transport infrastructure. Each has its role to play according to the policy priority and the initial situation, such that where the lack of transport infrastructure is a constraint on economic growth the best understanding will still arise from traditional macro studies. Where questions of regional competitiveness are paramount, market based studies of agglomeration will be central to any appraisal. Where it is about improving efficiency and maximising social benefit then more detailed micro studies will be essential. There remains much still to do.

NOTES

1. This paper draws on a number of recent papers by the author, especially Vickerman (2007a, b).

2. It is not intended to provide a complete review of the development of this literature as several comprehensive reviews exist already, see for example Gramlich (1994); SACTRA (1999) and Vickerman (2000, 2002).

3. This is not to ignore a huge volume of previous work which had sought to identify the 'social' value of transport, that above its direct value to users, which can be found in the works of such diverse authors as Dupuit (1844); Pigou (1920); Knight (1924); Fogel (1964) and Fishlow(1965).

4. For a valuable discussion of this literature see Jiwattanakulpaisarn (2007).

5. See also further discussion in Graham (2006, 2007).

BIBLIOGRAPHY

Aschauer, D.A. (1989), 'Is public expenditure productive?', *Journal of Monetary Economics*, 23, 177–200.

Boarnet, M.G. (1998), 'Spillovers and the locational effects of pubic infrastructure', *Journal of Regional Science*, 38, 381-400.

Bröcker, J. (2000), 'Trans-European effects of trans-European networks' in F. Bolle and M. Carlberg (eds), *Advances in Behavioural Economics*, Heidelberg: Physica.

Bröcker, J. (2004), 'Computable general equilibrium analysis in transportation economics' in Hensher, D.A., K.J. Button, K. Haynes and P. Stopher (eds) *Handbook of Transport Geography and Spatial Systems: Handbooks in Transport Volume 5*, Oxford: Elsevier.

Bröcker, J., R. Capello, L. Lundquist, T. Pütz, J. Rouwendal, N. Schneekloth, A. Spairani, M. Spangenberg, K. Spiekermann, R. Vickerman, and M. Wegener (2004), *Territorial Impact of EU Transport and TEN Policies*, Final Report of Action 2.1.1. of the European Spatial Planning Observation Network ESPON 2006, Kiel, Institut für Regionalforschung, Christian-Albrechts-Universität.

Burmeister, A. and K. Colletis-Wahl (1996), 'TGV et fonctions tertiaires: grand vitesse et entreprises de service à Lille et Valenciennes', *Transports Urbains,* 93.

Ciccone, A. (2002), 'Agglomeration effects in Europe', *European Economic Review*, *46*: 213–227.

Department for Transport (2005) *Transport, Wider Economic Benefits, and Impacts on GDP*, Technical Paper.

Dodgson, J.S. (1973), 'External effects in road investment', *Journal of Transport Economics and Policy*, 7, 169–185.

Dornbusch, J. (1997), 'Nantes, sept ans après l'arrivée du TGV Atlantique', *Notes de Synthese du SES,* Mai-Juin

Dupuit, J.A. (1844), 'De la mesure de l'utilité des travaux publics', *Annales des Ponts et Chaussées*, 8.

Eddington, R. (2006), *Transport's Role iIn Sustaining the UK's Productivity and Competitiveness.* London: HMSO.

Elhorst, J.P., J. Oosterhaven, and A.E. Romp (2004), 'Integral Cost-Benefit Analysis of Maglev Technology Under Market Imperfections' SOM Research Report, University of Groningen.

Fishlow, A. (1965), *American Railroads and the Transformation of the Ante-Bellum Economy*. Harvard Economic Studies; Vol. 127, Cambridge, MA: Harvard University Press.

Fogel, R.M. (1964), *Railroads and American Economic Growth: Essays in Economic History*, Baltimore, MD: Johns Hopkins Press.

Flyvbjerg, B., N. Bruzelius and W. Rothengatter (2003) *Megaprojects and Risk: An Anatomy of Ambition*, Cambridge: Cambridge University Press.

Fujita, M., P. Krugman and A.J. Venables (1999) *The Spatial Economy: Cities, Regions and International Trade*, Cambridge, MA: MIT Press.

Fujita, M. and J.F. Thisse (2002), *Economics of Agglomeration*, Cambridge: Cambridge University Press.

Gibbons, S. and S. Machin (2005), 'Valuing Rail Access Using Transport Innovations', *Journal of Urban Economics*, 57, 148-69.

Graham, D. (2005), *Wider economic benefits of transport improvements: link between agglomeration and productivity. Stage 1 Report*, London: Department for Transport.

Graham, D. (2006), *Wider economic benefits of transport improvements: link between agglomeration and productivity, Stage 2 Report*. London: Department for Transport.

Graham, D. (2007), 'Agglomeration, productivity and transport investment' *Journal of Transport Economics and Policy*, 41, 1-27.

Gramlich, E. (1994), 'Infrastructure investment: a review essay', *Journal of Economic Literature*, 32, 1176-1196.

Gunn, H. (2004), *SCGE Models: Relevance and Accessibility for Use in the UK, with emphasis on Implications for Evaluation of Transport Investments*, Final Report to Department of Transport, London., Cambridge RAND Europe.

Hay, A., K. Meredith, and R. Vickerman (2004), *The impact of the Channel Tunnel on Kent and Relationships with Nord-Pas de Calais*, Final Report to Eurotunnel and Kent County Council, Canterbury: University of Kent Centre for European Regional and Transport Economics.

Holtz-Eakin, D. (1994), 'Public-sector capital and the productivity puzzle', *The Review of Economics and Statistics*, 76, 12-21.

Holtz-Eakin, D. and A.E. Schwartz (1995), 'Spatial productivity spillovers from public infrastructure: Evidence from state highways', *International Tax and Public Finance*, 2, 459-68.

Jara-Diaz, S.R. (1986), 'On the relationships between users' benefits and the economic effects of transportation activities', *Journal of Regional Science*, 26, 379-391.

Jiwattanakulpaisarn, P. (2007), 'Granger Causality and Spatial Spillover Effects of Highway Infrastructure on Regional Economic Development: Evidence from an Application of Spatial Filtering in a Panel Vector Autoregressive Framework' paper to European Regional Science Meeting, Paris, 2007.

Klein, O. and G. Claisse (1997), *Le TGV-Atlantique: entre recession et concurrence*, LET, Lyon.

Knight, F.H. (1924), 'Some fallacies in the interpretation of social costs', *Quarterly Journal of Economics*, 38, 582-606.

Miyagi, T. (1998), 'A Spatial Computable General Equilibrium Approach for Measuring Multiregional Impacts of Large Scale Transportation Projects', *Network Infrastructure and the Urban Environment*, Heidelberg: Springer.

Miyagi, T. (2001), 'Economic Appraisal for Multi-regional Impacts by a Large Scale Expressway Project', Tinbergen Institute Discussion Paper TI 2001-066/3, Amsterdam: Tinbergen Institute.

Oosterhaven, J., J.P. Elhorst (2003), 'Indirect economic benefits of transport infrastructure investments', in W. Dullaert, B. Jourquin, J.B. Polak (eds), *Across the Border: Building on a Quarter Century of Transport Research in the Benelux*, Antwerp: De Boeck.

Pereira, A.M. and J.M. Andraz (2004), 'Public highway spending and state spillovers in the USA', *Applied Economics Letters*, 11, 785-8.

Pigou, A.C. (1920), *The Economics of Welfare*, London: Macmillan.

Plassard, F. and O. Cointet-Pinell (1986), *Les effets socio-économique du TGV en Bourgogne et Rhônes Alpes*, DATAR, INRETS, OEST, SNCF, 1986.

SACTRA (Standing Advisory Committee on Trunk Road Assessment) (1999), *Transport and the Economy*, London: Stationery Office.

SES (1998), *Evaluation de l'impact du TGV Nord-Européen sur la mobilité*, Les Etudes du SES.

Simmonds, D. (David Simmonds Consultancy in collaboration with Marcial Echenique and Partners) (1999), *Review of Land Use/Transport Interaction Models* Report to Standing Advisory Committee on Trunk Road Assessment, London: DETR.

Venables, A.J. (1999), 'Road transport improvements and network congestion', *Journal of Transport Economics and Policy*, 33, 319-328.

Venables, A.J. (2007), 'Evaluating urban transport improvements: cost-benefit analysis in the presence of agglomeration and income taxation'. *Journal of Transport Economics and Policy* 41, 173-188.

Venables, A. and M. Gasiorek (1999), *The Welfare Implications of Transport Improvements in the Presence of Market Failure Part 1*, Report to Standing Advisory Committee on Trunk Road Assessment, London: DETR.

Vickerman, R.W. (2000), 'Economic growth effects of transport infrastructure', *Jahrbuch für Regionalwissenschaft*, *20*: 99-115.

Vickerman, R.W. (2002), 'Transport and Economic Development', in *Transport and Economic Development*, Round Table 119, Economic Research Centre, European Conference of Ministers of Transport, OECD, Paris: 139-177.

Vickerman, R.W. (2007a), 'Cost-benefit analysis and large-scale infrastructure projects: state of the art and challenges', *Environment and Planning B*, 34, 598-610.

Vickerman, R.W. (2007b), 'Cost Benefit Analysis and the Wider Economic Benefits from Mega-Projects' in *Decision-Making on Mega-Projects: Cost-benefit Analysis, Planning and Innovation'* ed. H. Priemus, B. van Wee and B. Flyvbjerg, Cheltenham: Edward Elgar.

Wegener, M. and D. Bökemann (1998), *The SASI Model: Model Structure*, SASI Deliverable 8. Berichte aus dem Institut für Raumplanung 40, Dortmund: Institut für Raumplanung, Universität Dortmund.

THE WIDER ECONOMIC BENEFITS OF TRANSPORTATION

T.R. LAKSHMANAN
Boston University
Boston
United States

SUMMARY

1. INTRODUCTION AND OVERVIEW 55

2. MACROECONOMIC MODELING OF ECONOMIC IMPACTS OF TRANSPORT INFRASTRUCTURE 55

3. LESSONS FROM ECONOMIC HISTORY 60

4. THE WIDER ECONOMIC BENEFITS OF TRANSPORT: AN OVERVIEW 62

5. CONCLUDING COMMENTS 64

NOTES 66

BIBLIOGRAPHY 67

Boston, December 2007

1. INTRODUCTION AND OVERVIEW

Economic contributions of investments of transport infrastructure are typically assessed from a microeconomic perspective, which tries to identify the link between specific transport infrastructure improvements and the productivity of specific production units. The traditional economic tool of the microeconomic perspective is cost benefit analysis (CBA), an *ex ante* tool which tries to capture the benefits of time and cost savings—as well as further gains from logistical improvements and facilities consolidation made possible by transport improvements—and the associated costs including external costs. The objective of this Round Table sponsored by OECD / ECMT and Boston University is to identify and move towards methods which incorporate the *wider economic benefits* of transport infrastructure, not typically captured in the CBA estimates of benefits and costs.

The aim of this brief paper is to offer an overview of such wider economic benefits ensuing from transport infrastructure investments. Section II offers a brief review of the recent literature on macroeconomic models which argue that there are externalities to investments in infrastructure which are not captured in microeconomic CBA studies. The economy-wide cost reductions and output expansion due to transport infrastructure are identified in these macroeconomic models. While the overall inference of a positive and modest economic contribution of transport infrastructure is offered, the utility of such a result is open to question in view of two serious drawbacks of this macroeconomic modeling stream: *first*, the sharp differences and conflicts on *the magnitudes and direction* of economic impacts of infrastructure, *second*, the macroeconomic models offer little clue to the mechanisms linking transport improvements and the broader economy. Section III attempts to identify the wider economic benefits of transport capital and the economic processes involved in the generation of these wider economic benefits as gleaned from the Economic History literature on studies of economic transformation attendant on large investments in railroads and waterways around the world. Section IV provides a discussion of our contemporary understanding how transport infrastructure improvements open up markets, achieve gains from trade, promote interregional integration, and enhance performance of factor markets. Further, there are two other mechanisms, in activity clusters made possible by transport improvements, one dealing with spatial agglomeration benefits, and the other with innovation and commercialization of new knowledge. This section discusses these mechanisms in the context of recent theoretical research respectively in the 'New Economic Geography' and Innovation research associated with the 'Economies of Variety'. Section V concludes the paper.

2. MACROECONOMIC MODELING OF ECONOMIC IMPACTS OF TRANSPORT INFRASTRUCTURE

Macroeconomic models offer *ex post* econometric analyses of the contributions that transport infrastructure investments offer an economy in terms of cost reductions and output expansion – such effects typically captured by cost functions and production functions. The idea of the macroeconomic models is that there are externalities to investments in transport infrastructure which are not captured in microeconomic

CBA studies. The incorporation of these externalities allows the macroeconomic models potentially to identify social rates of return to transport infrastructure. However, models which represent aggregate output by GDP, can capture the value of time savings from infrastructure *only to the extent* that time saved is applied to production – missing time savings devoted to leisure (which can be picked up by CBA). Further, analyses focusing on aggregate output may ignore relative price effects of transport facility construction, which can yield a sizeable welfare effect (Haughwout, 1998).

Such macroeconomic analyses of productivity of transport infrastructure have been carried out over the last three decades in Japan, U.S., Sweden, U.K., France, Germany, India, Mexico, and elsewhere. These different studies vary along many dimensions:

- in the functional specification of those models, (Cobb-Douglas, CES, or flexible functional forms);
- in the types of measures they apply to different model variables such as output (e.g. GDP, personal income, Gross State Product, etc.), or public capital (Value of capital stock or other measures of physical infrastructure);
- in the level of disaggregation of economic sectors (e.g. from aggregate output in the Aschauer (1989) model to outputs by 35 sectors in the Nadiri-Mamaneus (1996) model)
- in the size of the geographic areas used (nation, region, state, metro area, or county), and
- in the temporal level of analysis (time-series, cross section, or pooled)

Agreements and Sharp Disagreements in the Literature

The major agreement that can be gleaned from these macroeconomic analyses of transport-economy linkages is the broad support for the view that transport infrastructure contributes to economic growth and productivity. However, this contribution is modest and variable over time. This inference about the economic impact of infrastructure is robust, as it reflects a great many studies which use various specifications of production and cost functions over different time periods, in different countries, and with slightly different representations of several variables (See Table 1).

However, this inference of a modest positive economic contribution of infrastructure investments masks some sharp differences and conflicts in the results of recent studies. If one compares the different measures of economic contribution of infrastructure (e.g. output elasticities, cost elasticities or rates of return of transport infrastructure), there appear to be sharply different results among the recent studies:

- for the same country overall, and at different periods of time,
- for different countries at comparable stages of development,
- for countries at different stages of development and,
- where threshold effects and accelerated growth are evident.

This large variety of conflicting results can not be attributed to methodological deficiencies as many of them are associated with recent studies employing sophisticated functional forms and statistical methods.

Differential Results for the Same Country or Countries at Similar Stages of Development

Table 2 illustrates one aspect of this dissonance among the studies about the impact of public capital. Pereira (2001) and Demetriades and Mamuneas (D-M, 2000) apply sophisticated production functions to

Table 1. **Summary of output and cost elasticities of highway and other public capital in various countries**

Country	Sample	Infrastructure Measure	Elasticity Range
United States	aggregate (ts) states (xs) states (ts/xs) regions, trucking industry (ts/xs)	public capital public capital highway capital highway capital	output: 0.05 to 0.39 output: 0.19 to 0.26 output: 0.04 to 0.15 cost: −0.044 to −0.07
Japan	regions (ts/xs)	transportation & communication infrastructure	Output: 0.35 to 0.42
United Kingdom	aggregate (ts)	public capital	cost: negative, statistically significant
France	regions (xs)	public capital	output: positive, statistically significant
Germany	industry (ts/xs)	public capital, highway capital	cost: negative, statistically significant
India	aggregate (ts), states (xs)	economic infrastructure: roads, rail, electric capacity	cost: −0.01 to −0.47
Mexico	national, 26 industries	transportation, communication & electricity, public capital	returns to public capital: 5.4% - 7.3%

Note: ts = time-series; xs = cross-section.

analyze the relationships between public capital and output in 12 OECD countries for approximately the same period, using respectively a Vector Auto Regressive / Error Correction Mechanism (VAR / ECM) framework and a flexible functional form for the profit function.

First, the D-M (2000) study estimates output elasticities of public capital for the U.S. (and for Sweden and Germany) *four times* as large as the Pereira (2001) study does. For U.K. and Japan, the estimates are *twice* as large. Further, the five OECD countries in Table 2 are affluent industrialized countries with comparable levels of technological evolution, industrial composition and income and consumption. As the various *transport-using firms* respond to transport infrastructure and service improvements in an economy, the many market mechanisms and structural processes interact and generate the economic effects rippling through the economy and culminating in the growth in GDP. Such effects in these five economies can be expected to be of comparable magnitude. Yet, D-M (2000) study's estimates of the output elasticities, however, range from 1.03 (U.S.) to O.358 (U.K.); Pereira's estimates range from 0.2573 (U.S.) to 0.143 (U.K.). Such sharp differences in parameters for the same country and for countries in comparable levels of development need an explanation.

Figure 1 traces the variation of infrastructure productivity over time in the U.S. The Nadiri-Mamuneas (1996) identify net rates of return of Highway capital (which makes up a major part of public capital):

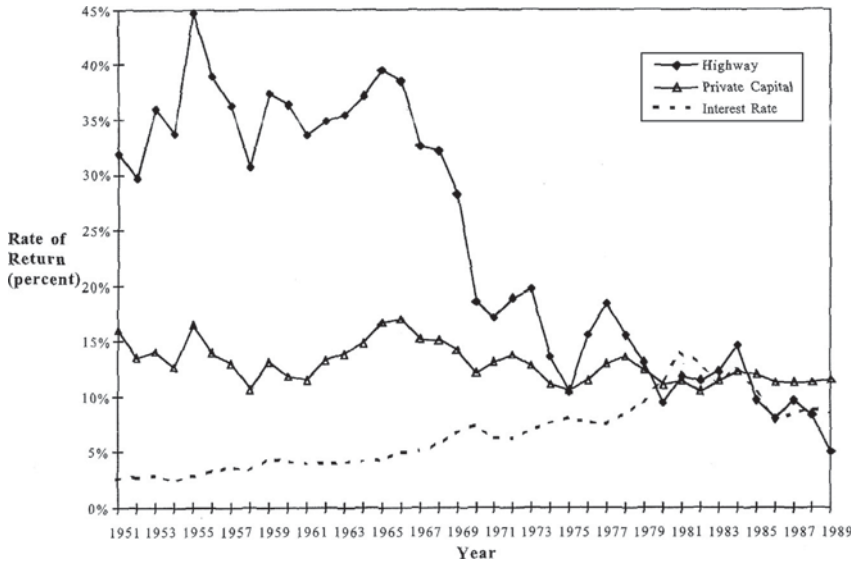
- Between 30% to 45% for years 1951-67,
- from 15% to 30% for years 1968-78 and,
- Below 15% for 1979 to 1987.

The net rate of return of public capital was higher than that of private capital from 1951 to 1978. In subsequent years, private capital had higher rates of return than highway capital.

Table 2. **Productivity effects of public capital: sharply dissonant results**

	U.S.	Japan	U.K.	Sweden	Germany
Pereira (2001) Vector Auto Regressive / Error Correction. Mechanism-data early 1960s to later 1980s	L.R. (10 yr) 0.2573	0.2525	0.1430	0.2270	0.1905
Demetriades and Mamuneas (2000). Flexible functional form for profit function (data for 1972-1991)	1.03	0.499	0.358	1.217	0.768

Figure 1. **Net rate of return of highway capital, private capital, and private interest rate (1951–1989) from Nadiri and Mamuneas (1996)**



Fernald's (1999) analysis of public capital's contribution to U.S. industry productivity between 1953 and 1989 suggests a similar time pattern of effects. He suggests that the massive road-building of the 1950s and 1960s (the interstate system) offered 'a one-time' increase[1] in the level of productivity (in the pre-1973 period).

Demetriades and Mamuneas (2000), on the other hand, arrive at a time pattern of productivity effects in the U.S., different from that espoused by Nadiri- Mamuneas and Fernald. They identify net rates of return of public capital, which exceed consistently private net rates of return of private capital in the U.S. all the way from 1972 to 1991. Indeed, the estimated long-run net rates of return to public capital in the U.S. (and Canada, Japan, Germany, France, Italy and U.K.) remained above those of private capital. In other words, an extra dollar of investment in the early 1990s (according to Demetriades and Mamuneas) would have been socially more productive in the long-run if it were invested as public capital.

Thus, for the period of the mid 1970s to early 1990s, two different patterns of productivity performance of public capital are offered by the Nadiri - Mamuneas and Fernald studies on the one hand and by Demetriades and Mamuneas on the other.

Countries at Different Stages of Development and Threshold Effects

Figure 2 presents the estimates of the elasticities of output with respect to public capital for a panel of countries in different stages of development (Canning and Bennathan, 2000). There is an inverted U shape, with higher elasticities in middle income countries and somewhat lower in the low and the high ends of the income distribution.

The rates of return to paved roads displayed in Figure 2 and categorized in Table 3 are obtained from a translog production function (Canning and Bennathan, 2000) in a set of countries which span the world income distribution. High rates of return to paved roads are evident in some middle income developing countries (Chile, Columbia, South Korea, and the Philippines). By contrast, low rates of return accrue to paved roads in affluent developed countries and in some developing countries.[2]

Figure 2. **Elasticity of output with respect to paved roads**

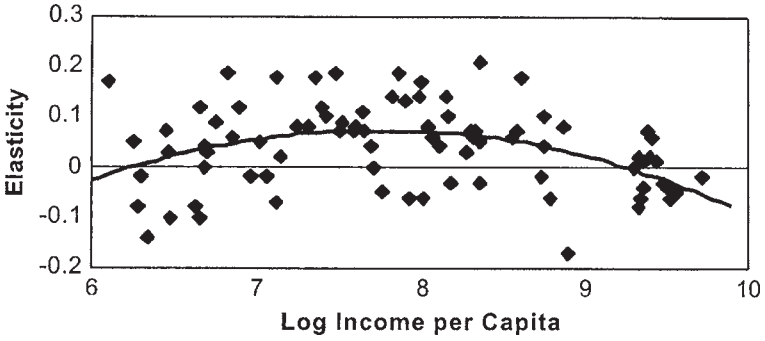


Table 3. **Transport infrastructure productivity in countries at different stages of development**

	Countries in Lower Quartile of Incomes	Countries in Middle Quartile of Incomes	Countries in Upper Quartile of Incomes
Output Elasticity of Paved Roads	0.05	0.09	0.04

Source: Canning and Besanthan, 2000.

Mechanisms Linking Transport Improvements and the Economy

While the macroeconomic models help determine whether and to what degree transport infrastructure lowers production costs, increases the level of economic output, and enhances the productivity of private capital, its analytical apparatus is a 'black box' variety. We have little inkling about the causal mechanisms and processes which translate infrastructure improvements into output and productivity enhancements. Such mechanisms are activated by the monetary and time savings induced by transport infrastructure improvements, and experienced at the regional and interregional levels by economic agents in different types of markets. The lowered costs and greater accessibility for transport-using production sectors and firms shipping goods from firms to retail outlets, and for households engaged in shopping and in commuting are likely to lead to the types and sequences of consequences such as: expansion of markets, higher efficiencies through scale economies,

economic restructuring through entry and exit firms exposed to new competition, spatial agglomeration economies and innovation benefits in spatial clusters made possible by transport infrastructure, etc.

This brief review of the drawbacks of macroeconomic models—the uncertainities in the magnitudes and direction of economic effects of transport infrastructure and the little light they shed on the mechanisms and processes underlying transport-economy linkages—suggests that we look elsewhere for guidance on delineating and estimating the broader economic benefits of transportation. Indeed, the many economic economic mechanisms and processes which translate transport improvements into a wide range of (and often transformational) consequences in the broader economy have been analyzed and reported in the case of railroads and waterways in a many countries in the Economic History literature. We turn briefly to this literature to highlight the broad range of wider consequences of transport infrastructure.

3. LESSONS FROM ECONOMIC HISTORY

Economic historians have attempted to measure in many countries the impact of the diffusion of railroad networks on economic growth and development.[3] In the process, they have shown how the time and cost savings induced by railroad expansion course through the countries' economies linking product and factor markets, promoting interregional trade, specialization and, increasing returns to scale, and reallocating economic activities.

A frequently used measure of the importance of railroads to a country's economy is *Social Savings,* computed as the *costs* of coping without railroads for one year. A counterfactual situation is envisaged where the producers, in the context of closing down of the railroad network for a year, transport the same volume of freight to the same destinations using alternative modes[4]. Table 4 provides estimates of social savings for railroads (which have been in full operation) in 10 countries.

The closure of a fully operational rail network has a considerable penalty in terms of GNP loss, especially for Spain, Mexico, and Argentina. In continental economies such as U.S. and Russia railroads did not provide a much cheaper service than waterways per ton-mile of freight over long and similar routes, with the result that social savings are lower.

However, Fogel's social savings measure is viewed currently as static and ignoring the 'forward linkages' from railways to the economy (Williamson, 1974), and a variety of indirect and induced effects of railways as gleaned from many studies of long run impacts of railways and case studies (Foreman-Peck, 1991).

Table 5. lists some of these wider effects of railroad infrastructure from 7 case studies. In 19th century India railroads lowered transport costs 80% per mile, thereby initiating grain bulk shipments, creating an India-wide market for foodgrains, and promoting a convergence of prices across India (Hurd, 1975).[5] In a separate study of factor markets in India, Collins (1999) showed that falling transport costs in Indian railroads facilitated regional wage convergence by facilitating both labor mobility and interregional commodity trade, especially in the areas surrounding the premier cities of Calcutta and Bombay. In late 19th and early 20th centuries European Russia, rail networks promoted market integration, based on the realization of gains from trade (Metzer, 1974). The narrowing of commodity price differentials increased regional specialization of production thereby improving resource allocation. In this regard, Metzer (1982) and O'Brien (1991) argue that the benefits from market integration are *additional* to those embodied in Fogel's Social savings, and these

Table 4. **Estimates of social savings on freight transported by railways, 1865–1913**

Country	Date	S.S. Expressed as a Share of G.N.P.
England and Wales	1865	4.1%
England and Wales	1890	11.0%
USA	1859	3.7%
USA	1890	8.9%
Russia	1907	4.6%
France	1872	5.8%
Germany	1890s	5.0%
Spain	1878	11.8%
Spain	1912	18.5%
Belgium	1865	2.5%
Belgium	1912	4.5%
Mexico	1910	25%–39%
Argentina	1913	21–26%

Source: Patrick O'Brien (1983).

integration benefits lead to internal and external economies that promote efficiency and enhance production (as compared to the pre-railroad situation).

Railroad investments in Brazil represented a purchase of specialization and enhanced productivity (Summerhill, 2005). This impact was large for overland movements given the absence of an affordable alternative to railroads, which further attracted large inflows of labor and capital which was used in other activities that raised national income. In the case of Argentina, the benefits from railroads built with British capital went largely to Argentine producers and consumers, enhanced aggregate productivity gains, and the transformation of the Argentine pampas (Summerhill 2001). The TFP gains deriving from the Spanish railroads were substantial, both through the shift from alternative modes of transport and through productivity improvements within the railroad networks.

A far more comprehensive analysis of the wider economic impacts of railroads has been carried out for U.S. railroad investments in the 19th century (Fishlow, 1965, Chandler 1965). Only a selective listing of the cascade of successive economic effects that ensued from the cost and time savings due to railroad expansion in 19th century from the Northeast U.S. to Midwest first and later to the rest of the country is possible here. As the lower costs and increased accessibility due to railroads coursed through markets and were experienced by different market actors (producers, consumers, laborers) a successive series of economic impacts ensued. This cascade of economic consequences include: expansion of settlements and agriculture; market expansion and integration; regional specialization in agriculture and industry; promotion of volume production and the realization of scale economies; enablement of lower inventories and the rise of a logistical revolution and the rise of wholesaling; the need to tap idle savings and channel them into railroad investment inducing development of financial institutions and raising the savings rate; the extension of mass production techniques (e.g. volume production of goods with interchangeable parts developed in New England) to mass produce a whole range of goods; the promotion of the complementary communication service (postal service); and eventually the integration of the Northeast to the Midwest to form the "Manufacturing Belt" (Chandler, 1965; Lakshmanan and Anderson, 2007; Kim and Margo, 2003).

Table 5. **The wider effects of railroad investments (1850-1914)**

Author	Country	Broader Effects of Transportation Infrastructure
Hurd (EEH, 1975)	India (1861-1921)	Prices across India began to converge and India-wide market in grains developed.
Collins (EEH, 1999)	India (1873-106)	Wage dispersion narrowed, Real wages in initially low wage areas grew faster
J. Metzer (JEH, 1974)	Czarist Russia (1870-1910)	Evolution of a national grain market. Improved interregional terms of trade. Narrower prices ⇒ regional specialization ⇒ Better resource allocation
Summerhill (JEH, 2005)	Brazil (1898-1913)	A purchase of specialization that boosted productivity
Summerhill (Mimeo, 2001)	Argentina (1857-1913)	Social savings 12-26% of GDP, Most gains went to Argentina producers and consumers
Heronz-Lancon (JEH, 2006)	Spain (1850-1913)	Growth accounting studies. TFP gains of Spanish RR. By 1914 11% of income per capita growth (cf. 14% in UK) Case against Fogel
Fishlow (1965)	U.S. Midwest (1848-1890)	Agricultural and industrial expansion of Great Lakes States and Integration into U.S. and World Economies

4. THE WIDER ECONOMIC BENEFITS OF TRANSPORT: AN OVERVIEW

Figure 3 offers one view of the mechanisms and processes underlying the wider economic benefits of transport infrastructure investments. It is a contemporary version of what Williamson (1974) and O'Brien (1983) call "forward linkages" of transport infrastructure. The lower costs and increased accessibility due to transport improvements modify the marginal costs of transport producers, the households' mobility and demand for goods and services. Such changes ripple through the market mechanisms endogenizing employment, output, and income in the short run. Over time *dynamic development effects* derive from the mechanisms set in motion when transport service improvements activate a variety of interconnected economy-wide processes and yield a range of sectoral, spatial, and regional effects, that augment overall productivity.

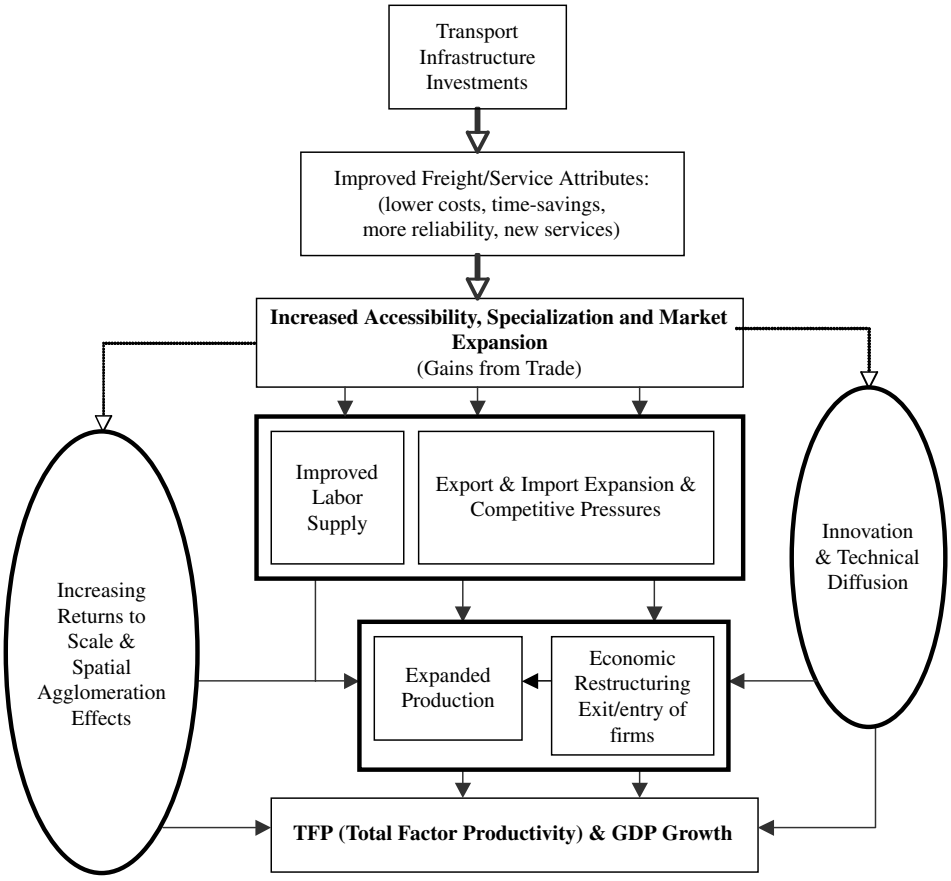
The lower costs and enhanced accessibility due to transport infrastructure and service improvements expand markets for individual transport-using firms. As such market expansion links the economies of different localities and regions, there is a major consequence in terms of shifting from local and regional autarky to increasing specialization and trade and the resultant upsurge in productivity. Thus, the U.S. Interstate Highway System, the Trans-European Network Programme and super-efficient ocean ports all contribute to "Smithian" growth—growth arising from specialization and trade.

Opportunities for exporting and importing goods are enhanced, in turn opening up several channels of economic effects, both in product markets and in factor markets – in a manner analogous to the results from tariff reduction and trade area expansion.

First, export expansion will lead to higher levels of output, which allow higher sales to cover fixed costs of operation, yielding efficiencies; Second, increasing imports put competitive pressures on local prices. Such pressures lead not only to the removal of monopoly rents but also to improved efficiency. Schumpeterian dynamics come into play—firm entry, exit, expansion, and contraction. As firms promote leaner production processes, which lower costs of production and raise productivity, further restructuring of the economy occurs. Third, lower transport costs and increased accessibility enlarge the markets for labor and other factor inputs. Firms will likely draw labor from a broader area and with a greater range of attributes improving labor supply and with lower costs. Similar effects in land and other factor markets are likely as transport improvements open up new land for economic activities.[6]

Finally, Figure 3 suggests that the two mechanisms in the oval boxes, one dealing with *innovation* and the other with *spatial arrangements* in the economy. These two mechanisms create, *in the context of transport infrastructure improvements*, conditions (in activity clusters) which enhance economic performance, and promote total factor productivity and endogenous growth. Our understanding of these two mechanisms of

Figure 3. **Transport infrastructure and economy-wide benefits**



innovation and spatial arrangements derive from recent research on "Innovation-Friendly Locales" and the 'New Economic Geography'.

Transport improvements can have an endogenous growth effect to the degree they impact the rate of growth of the economy through the creation and commercialization of new knowledge – thereby promoting Total Factor Productivity (TFP) growth, and the *rate of growth* of the economy. In the contemporary knowledge economy, firms are concerned with the reduction of a new class of costs—*adaptive costs*—incurred by the firm as it monitors the environment for changes in technology and products, identifies competitive strategies, and implements such strategies quickly enough to retain or improve market share (Hage and Alter, 1997; Lakshmanan and Button, 2008) The key notion in this case of spatial proximity is that innovation derives from the *Jacobsian Economies* (1969) or *the Economies of Variety* (Quigley, 1998) and the firms minimize their adaptive costs by participating in economic networks in the activity cluster or agglomeration—made possible by transport infrastructure improvements.

Research on imperfect competition and the increasing returns to scale extends to locational analysis and emphasizes the importance of the interactions between transport costs on the one hand and market size and economies of scale on the other.[7] With dropping transport costs and economies of scale, a firm in a location gains a larger market area and dominance, which in turn promotes the concentration of other firms in the same location. This idea of a location with good access to markets and suppliers for one firm improves market and supply access for other producers there, and the process of *cumulative causation* (where a location becomes more attractive to successive firms as more firms locate) derives from earlier ideas in Economic Geography. The central feature of this theory of agglomeration (as has been noted for a long time in economic geography and regional science) is the presence of external economies of scale in the Marshallian sense. Different firms clustered in a location experience positive externalities in the form of agglomeration economies, industrial complexes and social networks engaged in untraded interdependencies. In short order, regional specialization develops. Indeed, without increasing returns to scale in the context of transport improvements, it is impossible to account for the observed spatial concentration of firms and regional specialization in regional and national economies.

In contemporary spatial agglomerations of economic activity—where there are frequent transactions between suppliers and customers and where high-end business services often accompany goods delivery – the cost of transactions are likely to be lower inside such centers than outside them. Further, some interregional links gain advantages from the existence of increasing returns to transportation and transactions, which may help form transportation and transaction hubs as noted by Krugman (1999). The notion of *density* (of economic activities, social opportunities and transaction options) and *economic milieu* in such locations as leading to self-reinforcing and cumulative causation effects have been used by Johansson (1998) and Ciccone and Hall (1996).

5. CONCLUDING COMMENTS

The purpose of our discussion is to show how transport infrastructure and transport improvements open up markets and create conditions, in the context of spatial agglomerations and technical change and diffusion, which influence economic structure and performance. A broad variety of interactions take place within firms and between firms, within sectors and between sectors and more broadly within and between households and organizations. Hence the first inference we draw is *the importance of general equilibrium analysis of transport-economy linkages*. The implication is that the impacts of transport improvements must

be examined in a general equilibrium fashion, dealing with linkages between sectors and within sectors, where sectors exhibit different transport requirements, varying competitive strengths, and diverse spatial markets. These effects are realized through the operation of product markets and factor (labor, land, etc.) markets and technological and structural changes. Since these interactions are not only numerous and multiple and complex but may also operate to enhance or dampen the initial economic impacts of transport improvements, a more disaggregate analysis than is currently the case is called for in future analyses of transport-economy linkages.

NOTES

1. Fernald (1999) argues that the aggregate correlation between productivity and public capital primarily reflects causation from public capital to productivity, and the slowdown in productivity growth after 1973 may reflect the public investment patterns in that period.

2. It is generally observed that private returns to capital are quite low in the poorer developing countries, and that diminishing returns to capital set in slowly in affluent industrialized countries—because they can keep up their marginal productivity up by accumulating large amounts of human capital (Canning and Bennathan, 2000). The higher returns to private capital are also understandable in the middle income newly industrializing countries (NICs), which have received in recent years considerable flows of foreign direct investment (and associated technologies) from developed countries and participate in the global production system. If one assumes that NICs have invested in transport infrastructure to facilitate participation in global production and trade, a legitimate question arises: whether the high rates of return to paved roads observed in such countries reflect an expansion of transport networks to a critical density at which interregional economic integration occurs, thereby promoting regional specialization and accelerated growth in those economies.

3. There was a rapid expansion of rail networks across Europe—growing from 3000 kms of track in 1840 to 362,000 kms by 1913 (O'Brien, 1982). U.S. and many countries in Latin America and India witnessed rapid growth in their railroads in a comparable period.

4. Extra costs are incurred since freight will now move along longer and circuitous routes, at lower speeds, and at higher tariffs. First formulated by Fogel (1964) for the U.S., social savings have been computed for many countries. There can be some problems with the data quality and assumptions on prices in these estimates.

5. The prices of grain in some districts in 1860s were 8 to 10 times higher than prices in others (Hurd, 1975).

6. However, in an integrated market, there are likely some feedback effects associated with expanded production, which may dampen the initial strong positive impacts of transport improvements noted above. Since production expansion deriving from market expansion will raise the demand for labor and land, wages and rents will go up offsetting part of the initial lowering of costs and gains in competitiveness. The wage rises, if persistent, will have migration consequences. Finally, higher production may induce congestion in the networks and a rise in transport costs. The point to be made here is that transport improvements initiate a sequence of economic effects and feedback effects in a number of interacting markets.

7. The core idea of the 'new economic geography' is the notion of increasing returns, an idea that has earlier transformed both trade theory and growth theory (Fujita, Krugman, and Venables, 1999). Taking advantage of Dixit and Stiglitz's (1977) formalization of monopolistic competition, tractable models of competition in the presence of increasing returns have been developed in the fields of industrial organization, international trade, economic growth and location theory.

BIBLIOGRAPHY

Aschauer, D.A. 1989. "Is public expenditure productive?" *Journal of Monetary Economics*, 23: 177-200.

Canning, David and Esra Bennathan 2001. *The Social Rate of Return on Infrastructure Investments,* World Bank Research Project on "Infrastructure and Growth: A Multicountry Panel Study," 48 pages.

Chandler, Alfred D. 1965. *The Railroads, the Nation's First Business,* Harcourt, Brace & World, Inc., New York.

Collins, William J. 1999. "Labor mobility, market integration, and wage convergence in late 19th century India," *Explorations in Economic History,* 36: 246-277.

Ciccone, A. and R.E. Hall 1996. "Productivity and density of economic activity," *American Economic Review,* 86: 54-70.

Demetriades, Panicos and T.F. Mameneus 2000. "Intertemporal output and employment effects of public infrastructure capital: evidence from 12 OECD countries," *The Economic Journal,* 110: 687-712.

Fernald, John G. 1999. "Roads to prosperity? Assessing the link between public capital and productivity," *The American Economic Review,* 89: 3, 619-638.

Fishlow, Albert 1965. *American Railroads and the Transformation of the Ante-bellum Economy, Cambridge,* MA: Harvard University Press.

Fogel, Robert W. 1964. *Railroads and American Economic Growth: Essays in Econometric History,* Baltimore: The Johns Hopkins University Press.

Foreman-Peck, James 1991. "Railways and Late Victorian Economic Growth" in *New Perspectives in the Late Victorian Economy,1860-1914,* (ed.) James Foreman-Peck, Cambridge University Press, 73-95.

Fujita, M., Paul Krugman and A.J. Venables 1999. *The Spatial Economy,* The M.I.T. Press. Cambridge, MA.

Hage, J. and C. Alter 1997. "A Typology of Interorganizational Relationships and Networks" in *Contemporary Capitalism,* (eds.) J.R. Hollingsworth and R. Boyer, New York; Cambridge University Press. 94-126.

Haughwout, A.F. 1998. "Aggregate production functions, interregional equilibrium, and the measurement of infrastructure productivity," *Journal of Urban Economics,* 44: 216-227.

Herraz-Loncan, Alfonso 2006. "Railroad impact in backward economies: Spain, 1850-1913," *The Journal of Economic History,* 66: 853-881.

Hurd II, John 1975. "Railways and the expansion of markets in India, 1861-1921," *Explorations in Economic History,* 12: 263-288.

Jacobs, Jane 1969. *The Economy of Cities.* New York: Random House.

Kim, S. and R.A. Margo 2003. "Historical perspectives in U.S. economic geography" in *Handbook of Regional and Urban Economics* (eds.) V. Henderson and J-F. Thisse, Vol. 4. North-Holland, New York.

Lakshmanan, T.R. and K.J. Button 2008. (forthcoming) "Institutions and Regional Economic Development" in *Advances in Regional Economics* (eds.) R. Cappello and P. Nijkamp.

Lakshmanan, T.R., and William P. Anderson 2002. *Transport Infrastructure, Freight Services Sector and Economic Growth:* A White Paper prepared for the U.S. Department of Transportation, January. 127 pages.

_____ 2007. "Transport's Role in Regional Integration Processes" in *Market Access, Trade in Transport Services and Trade Facilitation,* Round Table 134. Paris: OECD-ECMT, 45-71.

Metzer, Jacob 1974. "Railroad development and market integration: The case of tsarist Russia," *The Journal of Economic History,* 34: 529-550.

_____ 1984. "Railroads and the efficiency of internal markets: Some conceptual and practical considerations," *Economic Development and Cultural Change,* 33: 61-70.

Nadiri, Ishaq M. and T. P. Mamuneas 1996. *Constitution of Highway Capital to Industry and National Productivity Groups*. Report prepared for FHWA. Office of Policy Development.

O'Brien, Patrick 1983. "Transport and Economic Development in Europe, 1789-1914" in R*ailways and the Economic Growth of Western Europe,* (ed.) Patrick O'Brien, 1-27, London: Macmillan.

Quigley, John M. 1998. "Urban diversity and economic growth," *The Journal of Economic Perspectives*. 12: 2, 127-138.

Summerhill, William R. 2005. "Big social savings in a small laggard economy: Railroad-led growth in Brazil," *The Journal of Economic History*, 65: 72-102.

_____ "Profit and Productivity on Argentine Railroads, 1857-1913", Los Angeles: Department of History UCLA (Mimeo).

Williamson, Jeffrey G. 1974. *Late Nineteenth-Century American Development: A General Equilibrium History.* London: Cambridge University Press.

WIDER ECONOMIC BENEFITS OF INVESTMENTS IN TRANSPORT INFRASTRUCTURE

Jeffrey P. COHEN
Barney School of Business
University of Hartford
West Hartford, CT
United States

THE WIDER ECONOMIC BENEFITS OF TRANSPORT—ISBN 978-92-821-0160-5 - © OECD/ITF, 2008

SUMMARY

1. INTRODUCTION 74

2. MOTIVATION 74

3. GENERAL BACKGROUND 76

4. SPATIAL ECONOMETRICS 79

4.1. Spatial autocorrelation 79
4.2. Spatial lag 82

5. APPLICATIONS 83

6. CONCLUSIONS AND FUTURE WORK 88

BIBLIOGRAPHY 91

West Hartford, August 2007

ABSTRACT

This paper begins by motivating the need for including "wider economic effects" when conducting transport infrastructure appraisal, followed by a discussion of various techniques to do so. The major focus is on studies from the cost function perspective that incorporate spillover benefits from public infrastructure capital, with a presentation of applications on highways, airports, and ports infrastructure stocks. The substantial differences between approaches focusing on "narrow" and "wider" impacts is evaluated, along with discussion of how application of the tools of spatial econometrics has facilitated estimation of models that capture wider economic benefits.

1. INTRODUCTION

There are many studies since the 1980's that attempt to quantify the effects of public infrastructure on the U.S. economy. There are a broad range of findings in these studies, including large positive, small positive, as well as negative effects. In recent years, research on the impacts of public infrastructure capital has started to incorporate assessments of the spillover benefits and costs across geographic boundaries. This revolution in the field comes at approximately the same time as growth in the area of spatial econometrics, which has facilitated the development of this strand in the infrastructure literature.

Despite these recent advances, there is still more that could be done, some of which depends on data availability. Namely, applying the approaches of recent cost function studies to other industries besides the manufacturing sector would require detailed data on input prices at the industry level. Another aspect that is worthy of additional attention is modeling cross-boundary spillovers in a general equilibrium framework that accounts for both consumers and firms.

In this paper, first I begin by introducing and motivating the need for incorporating measures of "wider" benefits of transport infrastructure in studies of the impacts of public infrastructure capital. In the context of this paper, "wider" benefits refer to the benefits beyond the geographic region in which the investment is undertaken. This motivation is followed by a description of several techniques used in the literature for measuring the "wider" (or spillover) benefits and how these measurement techniques differ from those for local benefits, for a variety of types of transportation infrastructure in general. These techniques include spatial spillovers (or lags) and spatial autocorrelation, both of which can be addressed through the empirical tools of spatial econometrics. Next I describe results of a variety of studies in the literature on highways, airports, ports, and various combinations of more than one type of transportation infrastructure. Finally, I elaborate on possible extensions and future work in this area, including research in progress and data sources that could be useful for addressing these issues.

2. MOTIVATION

An economic principles approach (supply and demand analysis) is instructive to motivate the problem of transportation infrastructure spillovers. Consider an average manufacturing firm in New York. The equilibrium amount of goods produced by this firm is given by the intersection of its supply and demand curves. What causes a shift in these curves? For the supply curve, holding all else constant, a decrease in the cost of "inputs" (such as wages, or the cost of private capital machinery or equipment) is one possibility. Another potential cause of a shift in supply is an improvement in technology. Finally, a "spillover" benefit (or a positive spillover) can shift the supply curve to the right.

A positive spillover occurs when other agents' actions confer benefits on an individual while the individual does not provide any compensation for these benefits. For example, if Connecticut improves its roads, the employees that travel to work from Connecticut to New York may have shorter commuting times,

which would be expected to enhance productivity of workers in New York. Similarly, the cost of shipping goods out of New York can be expected to go down when Connecticut improves its roads, so this would be another way in which better roads in Connecticut would confer spillover benefits on New York firms. The key difference between the roads in Connecticut and those in New York are that the Connecticut roads may not be financed by the firms in New York. While a portion of highway infrastructure is paid for by the federal government, a major portion of road financing in a neighboring state is paid for (indirectly) by residents and firms in that neighboring state, opposed to individuals in other states who pass through on a regular basis.

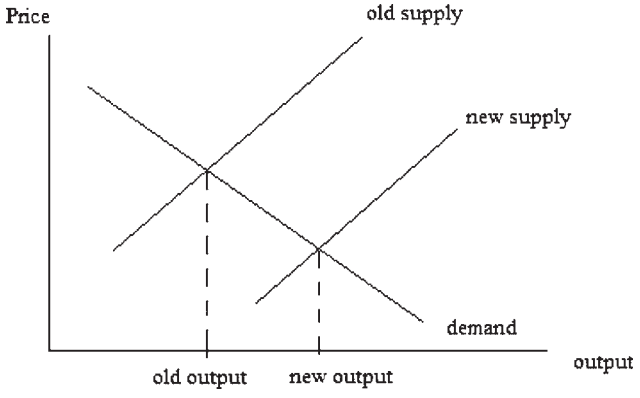
So when Connecticut expands its stock of public infrastructure, it causes the supply curve for firms in New York to shift to the right (see Figure 1). The new equilibrium level of production in New York is now higher than previously. In our analysis, the number of workers employed in New York is not changed, so output per worker, or productivity, now increases.

Researchers implicitly use similar reasoning to explain the impacts of public infrastructure within a particular geographic region while ignoring the impacts of spillovers across boundaries. Accordingly, much of the empirical literature on public infrastructure is concerned with the question of: by how much is productivity enhanced when the stock of public infrastructure increases? In other words, by how much does the supply curve shift, and how large is the associated output change, when public infrastructure increases?

The early empirical literature focused on national-level data using a production function approach of Aschauer (1989), and found a tremendous effect of infrastructure on productivity. Subsequent studies, such as Munnel (1990) assessed state-level data (Munnel), followed by studies that focused on the cost impacts of infrastructure (Morrison and Schwartz, 1996; Nadiri and Mameaunus, 1994). These subsequent studies found a range of infrastructure elasticities that were more reasonable than the initial Aschauer findings. Although the cost function study results are not directly comparable with the earlier production function studies, it is expected that they should be roughly in line with the production function results.

But most of these studies ignore an important aspect of public infrastructure. The network structure of many types of public infrastructure might imply that there are benefits to individuals beyond the state or locality where the infrastructure is located. On the other hand, better infrastructure in one location could assist firms in neighboring locations with drawing away the most productive resources, which could be detrimental to firms in the locality with the enhanced infrastructure. These network effects (both positive and negative) could have significant ramifications for the infrastructure elasticities worth examining in studies of state or county level infrastructure. A major focus of this paper is on the research, most of which developed in the late 1990's and 2000's, of the spatial spillover effects of public infrastructure capital.

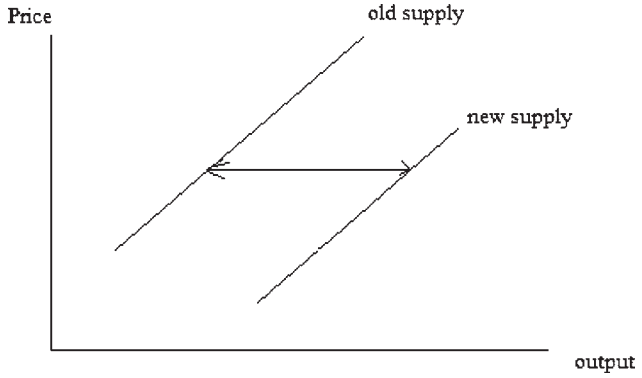
Figure 1. **Change in equilibrium output from an increase in public infrastructure stock in a neighboring locality**



At this point, it is also worth noting that most infrastructure productivity studies are done in a partial equilibrium context. Haughwout (2002) is an exception. He estimates a general equilibrium model of production and consumption, with public infrastructure as a local public good for several large U.S. cities. He finds that public infrastructure is beneficial to firms and consumers, but a significant expansion of infrastructure capital would leave producers and consumers worse off. However, Haughwout's model does not incorporate spatial spillovers across different cities due to public infrastructure, and estimating the net benefits of such a spillover model in a general equilibrium framework is worthy of attention.

Unlike Haughwout's study, most of the partial equilibrium studies in the literature ignore the impact of the demand curve on the equilibrium change in production from public infrastructure. In other words, the researchers really are concerned with the magnitude of the rightward shift of the supply curve from improvements in public infrastructure (Figure 2), opposed to the change in the equilibrium level of output resulting from the supply curve shift (Figure 1). This implies that the researchers assume a flat demand curve. Thus, there may be an overstatement of the impacts of public infrastructure in partial equilibrium studies, assuming the "true" private demand curve slopes downward. Another aspect deserving of greater attention in the infrastructure literature is the wider benefits to other sectors, such as the approach of Lakshmanan *et al*. (2007). Studies that ignore these benefits may underestimate the impacts of public infrastructure investment. Overall, the net effect is unknown, but it would need to be determined empirically. Although describing the models behind such a general equilibrium approach are beyond the scope of the present paper, they are worthy of attention, and the reader is encouraged to see Lakshmanan et. al. (2007) for additional details.

Figure 2. **Change in supply from an increase in public infrastructure stock in a neighboring locality**



3. GENERAL BACKGROUND

There are at least a couple of ways researchers attempt to quantify the changes in productivity from greater infrastructure investments in neighboring jurisdictions. One of these approaches is the production function approach, which incorporates the stock of infrastructure in neighboring jurisdictions as a "shift" factor in the production function. The production function approach requires panel (cross-section and time series) data on the amount of output (Y), labor (L), other "variable" factors such as materials (M), the stock of fixed factors such as private capital stocks (K), and measures for public capital stocks for neighboring (G) and within-locality (I).

The production function from the early studies on infrastructure that ignore inter-jurisdictional spillovers could be written (in vector notation) as the product of two functions, as follows:

$$Y = h(I)\, f(K, L, M) + u, \tag{1}$$

where u is a stochastic error term, which in general is implicitly assumed to have the desirable properties of zero mean, constant variance, and zero correlation across observations. Possible violation of the last of these assumptions can lead to inefficient estimates for the parameters, in which case their statistical significance may be understated. This potential problem is described in the spatial autocorrelation section below. The production function in (1) allows for infrastructure to shift the production function.

The more recent production function studies that incorporate spatial spillovers across jurisdictions (such as Boarnet, 1998) use a more general production function, such as the following:

$$Y = h(I, G)\, f(K, L, M) + u \tag{2}$$

In this specification, infrastructure in the own-jurisdiction, as well as in neighboring jurisdictions, can cause a shift in the production function.

Another approach, often referred to as a cost function approach, relies on duality theory. Duality theory (Varian, 1992) tells us that if we assume firms minimize costs, then cost minimization is essentially the same problem as profit maximization (which is based on the production function). The cost function approach is appealing because it incorporates optimizing behavior by firms, and it estimates an implied reduced-form cost function. This approach requires information on factor prices (such as P_{LP}, the wages of production workers; P_{LN}, the wages of non-production workers; and P_M, the price of materials inputs); the stock of fixed factors (such as private capital, K) and their associated prices (P_K); output (Y); as well as separate measures of infrastructure stocks for within-jurisdiction (I) and in other jurisdictions (G). Specifically, the total cost (TC) function model that ignores inter-jurisdictional infrastructure spillovers (similar to Morrison and Schwartz, 1996) can be written as follows:

$$TC = VC(Y, P_{LP}, P_{LN}, P_M, K, I, t) + P_K K + u, \tag{3}$$

where VC(·) is the variable cost function, and t is a "time" counter representing the passage of time.

Incorporating neighboring jurisdictions' infrastructure (G), such as in Cohen and Morrison Paul (2004), yields

$$TC = VC(Y, P_{LP}, P_{LN}, P_M, K, I, G, t) + P_K K + u \tag{4}$$

A useful rule (called Shepard's Lemma) that is a special case of the envelope theorem (see Varian, 1992) states that the derivative of VC with respect to any of the input prices yields a demand function for that particular input. So as an example, for production labor (L_P),

$$L_P = \partial VC(\cdot)/\partial P_{LP} \tag{5}$$

With both the cost function and production function approaches, regression analysis is used to estimate parameters necessary to obtain elasticities of the infrastructure variables. For the cost function approach, an input demand function similar to (5) is derived for each of the variable factors, and a stochastic error term is appended to each of these equations. These input demands are estimated together with the variable cost function, using Seemingly Unrelated Regression (SUR) techniques.

In terms of assessing spillover benefits, with the production function approach the goal is to obtain estimates of the elasticity of output with respect to neighboring jurisdictions' infrastructure:

$$\varepsilon_{Y,G} = [\partial Y/\partial G][G/Y] \tag{6}$$

For the cost function analysis, in assessing the wider benefits of infrastructure, one objective is to estimate the elasticity of variable costs with respect to neighbors' infrastructure:

$$\varepsilon_{VC,G} = [\partial VC/\partial G][G/VC] \tag{7}$$

When researchers compare results from production function studies with cost function studies, they tend to compare elasticities (6) and (7), respectively. However, the comparison is not completely valid since (6) shows the impact of neighbors' infrastructure on output, while (7) shows the effect of neighbors' infrastructure on variable costs.

A similar way of writing (7) is as the "shadow" value of neighboring localities' public infrastructure stocks (Z_G), as it reveals how additional infrastructure in neighboring localities affects a particular locality's variable costs:

$$ZG = [\partial VC/\partial G] \tag{8}$$

For $Z_G < 0$, neighboring jurisdictions' public infrastructure can be thought of creating "value" for firms in a particular jurisdiction, since variable costs fall as the size of the public infrastructure stock in neighboring jurisdictions increases.

The cost function approach also enables an examination of other revealing elasticities that provide insight into the wider benefits of public infrastructure. For instance, the elasticity of labor demand with respect to neighboring jurisdictions' infrastructure, which for production labor (LP) is (building on the result from equation (5), which is based on Shepard's Lemma):

$$\varepsilon_{LP,G} = \partial L_P/\partial G = \partial(\partial VC(\cdot))/\partial P_{LP}\partial G \tag{9}$$

Also, the elasticity of the "shadow" value of the neighbors' infrastructure with respect to the own-jurisdiction infrastructure is written as:

$$\varepsilon_{G,I} = [\partial Z_G/\partial I][I/Z_G] \tag{10}$$

This shadow value elasticity (10) is useful in determining whether infrastructure in neighboring jurisdictions is a substitute for or complement to an individual jurisdiction's infrastructure stock. Namely, if greater infrastructure in a particular jurisdiction increases the value of neighboring jurisdictions' infrastructure, then the two are complements. On the other hand, if greater infrastructure in a jurisdiction decreases the value of neighboring jurisdictions' infrastructure, the two are substitutes. The outcome for this elasticity can have important implications for regional infrastructure coordination policies.

Since it is clear that estimating these elasticities is an objective of the analysis, now a major question is how to construct the "neighbor" infrastructure stocks, test for and possibly adapt the model for spatial autocorrelation, and estimate the resulting equations. This is the focus of the next section on spatial econometrics.

4. SPATIAL ECONOMETRICS

Spatial Econometrics (Cliff and Ord, 1981, Anselin, 1981) has grown in popularity over the past 25 years, and only recently has been applied in the area of infrastructure studies. There are two aspects of spatial econometrics, commonly referred to as spatial autocorrelation and spatial lags (Kelejian and Prucha, 1999).

4.1. Spatial autocorrelation

Spatial autocorrelation occurs when one locality's error term in the regression depends on "neighboring" localities' shocks or innovations, instead of merely being normally distributed with zero mean, constant variance, and zero covariances over time and space. Spatial autocorrelation implies interdependencies among different localities, and in general researchers can accommodate for spatial autocorrelation after conducting a procedure that generates an estimate of the magnitude of the autocorrelation. The word "neighboring" is in quotations because the word does not necessarily imply that the neighbor is at a contiguous location. That is, it could imply that localities are similar (or dissimilar) in other ways, such as average incomes of residents, volume of trade between individual locations, or other demographic characteristics.

Mathematically, spatial autocorrelation is represented in the following form:

$$u_i = \lambda \Sigma_j w_{i,j} u_j + \gamma_i \tag{11}$$

or, in vector notation,

$$u = \lambda Wu + \gamma \tag{11'}$$

In equation (11), u_i is the error term for locality i, λ is the spatial autocorrelation coefficient, $w_{i,j}$ is the weight that locality j's error term has on locality i's error term (described as W in matrix notation), and γ_i is locality i's error term with the "desirable" properties (described below). Depending on the estimation technique for λ, researchers impose different assumptions on the distribution of γ_i. Namely, the Generalized Moments (GM) approach of Kelejian and Prucha (1999) assumes that γ_i is independently, identically distributed with zero mean, constant variance, and zero covariances across observations. The other commonly used approach, known as maximum likelihood (ML) estimation (Anselin, 1981), assumes normality of the γ_i, along with the same assumptions of zero mean, constant variance, and zero covariances.

Before the estimation can be implemented, researchers must choose the specification for the spatial weights, $w_{i,j}$. One common approach is contiguity weights, where all jurisdictions that are contiguous geographic neighbors to a particular jurisdiction are weighted equally. In other words,

$$w_{i,j} = 1/c \text{ if j is a contiguous neighbor to i} \tag{12}$$

$$= 0 \text{ otherwise,}$$

where c is the total number of i's contiguous geographic neighbors.

Other approaches, such as that of Boarnet (1998), specify more complicated spatial weight structures. One common example is the following:

$$w_{i,j} = [1/|D_i - D_j|] / [1/ \Sigma_j |D_i - D_j|] \quad (13)$$

In this weights specification, D_i and D_j can represent demographic variables, such as population, per capita income, or others (Boarnet, 1998). Intuitively, this gives greater weight to jurisdictions that are "similar" to each other, and less weight to jurisdictions that are "dissimilar". Since two jurisdictions (i and j) that are similar based on some demographic information will have D_i and D_j relatively close together, the inverse of the absolute value of their difference will be a large number, so jurisdiction j will have greater weight on jurisdiction i. The term involving the summation in the denominator is a normalization to ensure that $\sum_j w_{i,j} = 1$.

The next step after specification of the spatial weights is the estimation. Often, researchers estimate the production or cost function (along with the associated input demand equations), and perform a test for spatial autocorrelation (such as the Moran I test). Assuming the null hypothesis of no spatial autocorrelation is rejected, the next step is to determine the appropriate estimation technique for λ. One approach is to test whether the fitted residuals are normally distributed, using a test for normality (such as the Jarque-Bera test). If normality is rejected, the GM approach is followed to appropriately estimate λ, otherwise the ML estimation approach is used. Finally, once an estimate of λ is obtained, researchers use it to perform a spatial Cochrane-Orcutt transformation (analogous to a time-series Cochrane-Orcutt transformation) before re-estimating the transformed system.

Namely, to demonstrate this process consider the production function $Y = h(I,G)f(K,L)$, which we rewrite as:

$$Y = X\beta + u, \quad (14)$$

where X represents a matrix of the explanatory variables (I,G,K,L), β is a vector of parameters to be estimated (and subsequently used to obtain the infrastructure elasticities), and u is as represented in the spatial autocorrelation error process described in (11') above. Substituting equation (11') into equation (14) yields:

$$Y = X\beta + \lambda Wu + \gamma \quad (15)$$

Also, since we can rewrite the production function equation as:

$$u = Y - X\beta, \quad (14')$$

then multiplying through both sides by W yields

$$\lambda Wu = \lambda WY - \lambda WX\beta, \quad (14'')$$

and substituting this result into the equation (15) above,

$$Y = X\beta + \lambda WY - \lambda WX\beta + \gamma \quad (15')$$

and rewriting:

$$Y - \lambda WY = X\beta - \lambda WX\beta + \gamma \quad (15'')$$

or

$$Y^* = (X^*)\beta + \gamma \quad (16)$$

where $Y^* \equiv Y\,[I_N - \lambda W]$,

$$X^* \equiv X\,[I_N - \lambda W]$$

and I_N is an N by N identity matrix (where N is the number of observations in the sample).

After obtaining parameter estimates for λ and substituting them into equation (16) above, the resulting estimation equation has an error term (γ) that does not exhibit spatial autocorrelation, and thus yields efficient parameter estimates for the elasticities of output or costs with respect to infrastructure (either in the own or neighboring jurisdictions).

There are a number or potential reasons why a model might be expected to exhibit spatial autocorrelation. These include possible omitted variables that vary spatially; decisions in one location that are made for entities in other locations; and/or common shocks that spill over across geographic boundaries. An example of the latter is the weather and its impact on firms' costs or production process. A weather "shock" (for instance, either a storm or a heat wave) hitting some states and impacting production or costs can spill over to an adjacent state, and thus there can be some degree of persistence over geographic space that may lead to spatial autocorrelation.

Ignoring spatial autocorrelation can lead to parameter estimates with higher standard errors than if spatial autocorrelation had not been present. These higher standard errors can translate into t-statistics that are smaller than they should be. In other words, ignoring significant spatial autocorrelation can impact hypothesis testing, as researchers might fail to reject a null hypothesis that is actually a true hypothesis. In the context of infrastructure, ignoring spatial autocorrelation can lead a researcher to erroneously accept a null hypothesis that the infrastructure elasticity is equal to zero.

One of the first known infrastructure studies that addressed spatial autocorrelation is Kelejian and Robinson (1997). They estimate a Cobb-Douglas production function and incorporate a spatial autocorrelation adjustment, and they are careful to try many other specifications as well. They find that there can be a wide range of estimates on the infrastructure elasticities, depending on the econometric specification employed by the researchers.

Two subsequent studies find less convincing evidence of spatial autocorrelation. Holtz-Eakin and Schwartz (1995) test for spatial autocorrelation but find no evidence of its presence in their model. Boarnet (1998) finds no evidence that accommodation of spatial autocorrelation affects the sign and significance of the infrastructure elasticity estimates in his model.

The form of spatial autocorrelation in equation (11) is analogous to a first-order time series autoregressive process. Just as there are much more complicated time series processes in the econometrics literature, there are now some more complicated spatial processes addressed in the infrastructure literature to allow for more general forms of spatial autocorrelation. Cohen and Morrison Paul (2007) address the problem of higher order spatial autocorrelation in the context of assessing the impacts of transportation infrastructure on manufacturing costs. Namely, they consider more general forms for the spatial process, such as:

$$u_i = \Sigma_m \lambda_m \Sigma_j w_{m,i,j}\, u_j + \gamma_i \qquad (17)$$

where m represents the "order" of the neighbor. Equation (17) is similar to but more general than equation (11), since here $w_{m,i,j}$ stands for the weight that state j has on state i in neighbor band m. Also, λ_m is the spatial autocorrelation parameter for the impact of the weighted average of errors in neighbor band m on state i's error term. For instance, at the state level and using contiguity weight matrices, New York, Connecticut, Rhode Island, New Hampshire, and Vermont would be first-order neighbors (m=1) to Massachusetts; New Jersey, Maine, and Pennsylvania would be second-order neighbors (m=2) to Massachusetts, etc. Such an error structure allows for more complex interactions among error terms for states (or other geographic units),

so that in the previous example, shocks hitting New Jersey, Maine, and Pennsylvania might spill over to Massachusetts, whereas they would not with the first order contiguity neighbor matrix. Also, since each order neighbor has a separate spatial autocorrelation coefficient, it is possible in models of higher order spatial autocorrelation that the shocks hitting Massachusetts' second order neighbors have different impacts on the state than the shocks that hit its first-order neighbors. This error structure can be preferable to an approach where all other units are neighbors in varying degrees but with the same spatial autocorrelation coefficient. With higher order spatial autocorrelation, one can test whether the autocorrelation impact dissipates (or even dies out) beyond a certain range, instead of merely imposing a cutoff distance for neighbors to be included in the weighted average.

In determining the appropriate number of neighbors (m), Cohen and Morrison Paul (2007) apply a variation of the Kelejian and Robinson (1992) test for spatial autocorrelation as follows. First, Cohen and Morrison Paul test for first order spatial autocorrelation. When they find evidence of first order spatial autocorrelation, they proceed to test for second order, otherwise they stop. If they find second order spatial autocorrelation, they proceed to test for third order, otherwise they stop. They perform these tests on each of the estimation equations (the variable cost and the 3 input demands) separately. They find evidence of first order spatial autocorrelation in the non-production labor demand equation; second order spatial autocorrelation in the materials demand and variable cost equations; and third order spatial autocorrelation in the production labor demand equation. Accordingly, they estimate the spatial autocorrelation coefficients for each equation using the Kelejian and Prucha (2004) Generalized Moments techniques for systems of equations, then use these estimates to perform a spatial Cochrane-Orcutt transformation on each equation, before estimating the transformed system to obtain consistent parameter estimates.

Cohen and Morrison Paul (2007) find that the magnitude of the spatial autocorrelation coefficients for each equation decreases as the order of the neighbors increases. In other words, the impact of a "band" of neighbors' error terms on a particular state's error term is higher for states that are closer neighbors to a particular state, and it dissipates for bands of states that are more distant neighbors.

4.2. Spatial lag

The other form of spatial spillovers that can be assessed with spatial econometrics is known as a spatial lag. A spatial lag (or spatial dependence) occurs when the "neighbors" of a particular geographic unit's variable(s) are included as explanatory variables in a regression. These spatially lagged variables can be of the dependent variable, as in Boarnet (1998), who includes a spatial lag of output. Such a spatial lag is interpreted as the weighted average of other jurisdictions' dependent variable. It is also common for researchers to include a spatial lag of some variable(s) other than the dependent variable. Examples of such spatial lags described below in more detail include Cohen and Morrison Paul (2003a, 2004), who include the weighted average of other states' airports, and highways, respectively.

A production function regression equation with a spatial lag can be written as follows:

$$Y = \rho WY + X\beta + u, \tag{18}$$

where ρ and β are parameters to be estimated. In this equation, WY is the spatial lag, and it represents the weighted average of other jurisdictions' endogenous variable (which is output in the case of the production function). In Boarnet (1998), the endogenous variable is output. Since we know that Y is correlated with the error term u, it follows that WY is also correlated with u. Thus, WY is also an endogenous variable. In this case, ordinary least squares (OLS) is not the appropriate estimation technique. Instead, two-stage least squares (2SLS), or instrumental variables (IV) should be used to estimate equation (18). It can be shown (Kelejian and Prucha, 1998) that X is the appropriate instrument for itself, and WX is an instrument for WY. It is also possible, but not necessary, to include additional instruments for WY, such as WWX, WWWX, etc.

In situations where there is a spatially lagged dependent variable and spatial autocorrelation in the same model (that is, when equation (18) has the error structure described in equation (11')), the procedure for estimating λ described above is somewhat different. The first step is to estimate equation (18) by 2SLS, using X and WX as instruments. The second step is to retrieve the fitted values of the error terms u, and use them in either the GM or ML procedure described above to generate an estimate for λ. The final steps are to transform (18) with a spatial Cochrane-Orcutt transformation, plug in the estimate for λ, and estimate the transformed equation(s) by 2SLS, using X and WX as instruments for X and WY, respectively. This process yields efficient parameter estimates for β and ρ, and in turn, estimates for the infrastructure elasticities.

It is also possible to model spatial dependence by including spatial lags of other exogenous variables in the model. One example is the weighted average of other jurisdictions' public infrastructure stocks. In such a situation, the production function is written as:

$$Y = X\beta + WZ\delta + u, \tag{19}$$

where Z is some subset of the variables included in X (such as the stock of public infrastructure), and β and δ are parameters to be estimated. It is also possible, but not necessary, to add a spatially lagged dependent variable in the model. Once the estimates of β and δ are obtained, either through OLS, the spatial autocorrelation adjusted OLS procedure, or 2SLS (if there is a spatially lagged dependent variable), it is possible to generate insights on the wider benefits of infrastructure. By calculating the elasticity of output with respect to neighboring jurisdictions' infrastructure ($\varepsilon_{Y,G}$), or the elasticity of variable costs with respect to neighbors' infrastructure ($\varepsilon_{VC,G}$), it is possible to assess these wider benefits. Also, if spatial autocorrelation is found to be present in the earlier estimation stages, that can provide additional information about wider benefits by shedding light on the innovations that spill over among "neighboring" jurisdictions.

5. APPLICATIONS

Recent applications of spatial lags and spatial autocorrelation in U.S. public infrastructure capital studies (both production function and cost function) have been done at the state and county levels, and have focused on airports, ports, highways and roads. Boarnet (1998) includes a spatial lag of the public infrastructure variables (roads and highways). He conducts an analysis of California counties with a Cobb-Douglas production function, allowing the infrastructure and neighboring county infrastructure stocks to be "free" variables that would shift the production function. Boarnet also tries a variety of different spatial weights matrices, and he finds significantly negative spatial lags with the weights for counties with more similar population densities ($\varepsilon_{Y,G} = -.307$), as well as those with similar levels of per-capital income ($\varepsilon_{Y,G} = -.806$). The magnitudes of these effects seem quite large, as the impacts of own-state infrastructure $\varepsilon_{Y,I}$ are 0.268 and 0.300 for the population and income weights, respectively.

Boarnet's results represent evidence of leeching behavior. Namely, improved infrastructure in neighboring counties would enable firms in those neighboring counties to draw away productive resources from a nearby county, leaving less productive workers in the nearby county. Thus, he finds some evidence showing that improvements in infrastructure in neighboring counties lead to a decrease in output in a particular county, assuming that workers are mobile.

Other subsequent state-level infrastructure studies by Cohen and Morrison Paul (2003a, 2004) find evidence of positive spillovers across states. The former paper focuses on airports, while the latter on

highways. These studies incorporate spatial autocorrelation adjustments as well. They estimate cost functions and input demand equations for the U.S. manufacturing sector, so any benefits that they find accrue only to this particular sector.

Cohen and Morrison Paul (2003a) is motivated by the hub and spoke structure of the U.S. air transportation network. This system consists of airlines transporting passengers and freight from spoke airports to hub airports, followed by the passengers and cargo deplaning at the hubs and boarding other flights that transport them to their final destinations. With such a system, a delay at any particular node in the network can have system-wide effects, since passengers and cargo waiting to be transported by connecting flights at other nodes can be delayed as well. Improving infrastructure at a particular airport may reduce congestion throughout the entire system, leading to a decrease in travel time for business travelers and for cargo throughout the country. This lower travel time can translate into a decrease in firms' costs and enhance worker productivity.

A distinctive characteristic of the Cohen and Paul (2003a) analysis is that external benefits are different for airports than for highways or roads. In order for an airport to generate any benefits at all, there must be another node somewhere in the system for departing planes to land. Highways or roads infrastructure, on the other hand, can provide benefits with as little as several miles length within one city. Thus, one might expect the out-of-state-benefits for airports to be relatively large compared with those of highways, since better infrastructure at congested airports in other states should have a similar impact on travel savings (and in turn, costs) as if the improvements had been made at a congested departure airport in the firm's state.

Cohen and Morrison Paul (2003a) estimate a state-level variable cost function (which is the VC(·) expression in equation (4) above) and input demand equations similar to equation (5), where I represents within-state airport infrastructure stocks, and G represents a weighted average of airport infrastructure stocks in other states. They use Seemingly Unrelated Regressions (SUR) to estimate the system of equations, and they also find that applying a spatial autocorrelation adjustment to this system based on parameter estimates from the Kelejian and Prucha (2004) Generalized Moments approach does not substantively affect their results. They obtain the data for I by applying the perpetual inventory method to state-level capital spending data on air transportation, for the years 1982–1996. They obtain an estimate of the average service life of airports of 25 years, which they multiply by the average air transportation capital spending from 1977 to 1981, to obtain a base-year airports capital stock. Their depreciation rate is obtained by the inverse of the estimated average service life, and their investment deflator is from the *2000 Economic Report of the President*.

Their G variable is based on the extent of the interaction between a particular state and other states. This interaction is measured by the number of person-trips by air between individual states, from data in the 1995 American Travel Survey (Bureau of Transportation Statistics). So, as an example, a destination-state (j) with fewer person trips ($a_{i,j}$) between it and an origin state (i) has a lower weight on the origin state than another destination state with a larger number of person trips between it and the origin state. They define the weight that a particular destination state has on an individual state i as:

$$w_{i,j} = a_{i,j}/\Sigma_j(a_{i,j}) \tag{20}$$

with the term in the denominator ensuring that the $w_{i,j}$ sum to 1 (and $w_{i,j}$ represents the (i,j) element of the spatial weight matrix, W). Equation (20) represents the spatial weights that they use to perform a spatial autocorrelation adjustment in the variable cost and each of the input demand equations.

They also construct R_j, the ratios of Gross State Product (GSP) in state i to GSP in state j in a given year. Then, they define the average "neighbors'" airport infrastructure, G_i, in any given year as

$$G_i \equiv \Sigma_j \, w_{i,j} \, I_j \cdot R_j, \tag{21}$$

where I_j is the airport infrastructure stock in state j in a given year. The intuition behind R_j is that one might expect a disproportionately large number of flights in larger states (such as Texas) to enter into G for smaller states (such as Rhode Island). Multiplying each "neighbor" state's infrastructure stock (I_j) by the inverse of its GSP times state i's GSP essentially eliminates the size effect arising due to the large neighbor states.

Cohen and Morrison Paul note that many large hub airports in the U.S. are facing more congestion during the period of their sample than airports that are not large hubs. Thus, it might be expected that the cost elasticities with respect to own-state and other state airports are not the same for states with at least one hub airports opposed to states with no hub airports. So they present two sets of elasticity results, for states with hub airports and for states with no hub airports.

First, for states with large hubs, the $\varepsilon_{VC,I}$ and $\varepsilon_{VC,G}$ elasticities are very similar and significant, with values of -0.113 and -0.116, respectively. This implies that better airports in other hub states are just as effective as airports in the origin state at reducing costs for manufacturing firms in that particular state. As discussed above, this supports the notion that unlike highways, airport improvements at origin and destination points should provide approximately the same level of cost-reduction benefits. In other words, for states with large hubs, an out-of-state airport can be just as important as the origin airport because two points are necessary to complete a trip.

For the input demand elasticities with respect to G for states with large hub airports, both production and non-production labor demand are negative and significant. These imply that increased airport infrastructure stocks in other states leads to lower demand for both types of labor in an individual state with large hub airports. With these lower numbers of workers, increased marginal product of labor is implied as a result of the higher levels of G. The results are similar in direction for materials inputs, while the magnitude of the effect of G on materials demand is smaller than the impacts for both types of labor.

The results are somewhat different for states with no major hubs. Namely, while $\varepsilon_{VC,G}$ and $\varepsilon_{VC,I}$ are negative and significant, $\varepsilon_{VC,G}$ is much larger in magnitude. The authors explain this difference by the fact that G includes states with large hubs, many of which are congested, while I represents airport infrastructure stocks for non-hub origin states, which in general are not as congested. Thus, the cost savings from expanding airports in other states is much larger in magnitude than the cost savings from larger airports in the origin states. Furthermore, the negative and significant shadow value elasticities $\varepsilon_{I,G}$ and $\varepsilon_{G,I}$ imply that G and I are substitutes, as increases in I imply lower Z_G (and vice-versa for G and Z_I).

Cohen and Morrison Paul (2004) focus on highway interdependencies across state borders. They note that the magnitudes and directions of such network effects have been elusive in previous infrastructure studies. The highways problem is motivated by the possibility of travel time savings for firms' workers in a particular state who travel through neighboring states on their way to and from work. Also, firms generate cost-saving benefits from shipment of materials through neighboring states with improved infrastructure stocks.

The authors estimate a variable cost function for the U.S. manufacturing industry similar to that of Cohen and Morrison Paul (2003a), except here I represents within-state highways infrastructure (obtained from Paul et. al., 2001, who apply the perpetual inventory method to state-level investment data); and G is the weighted average of neighbors' highway infrastructure. They calculate the spatial weights $w_{i,j}$ as in equation (20) above, where here $a_{i,j}$ is the average value of goods shipped from state i to state j, and j consists of states that are contiguous neighbors of state i. After the $w_{i,j}$ are determined, G is calculated as in equation (21).

Another element of the Cohen and Morrison Paul (2004) estimation system is that they allow for first order spatial autocorrelation in the cost function and input demand equations, by appending an error structure to each estimation equation similar to (11). They estimate a Generalized Leontief variable cost function, as well as input demand functions based on (5) for production labor, non-production labor, and materials inputs. Their annual data are for the manufacturing industry at the state level, covering the period 1982–1996.

They find that the parameters for the terms involving G are jointly significant, which justifies their inclusion of spatial spillover effects in the variable cost function model. They also reject the hypothesis that the I and G parameters together are jointly zero. They find that the mean of the elasticity $\varepsilon_{VC,I} = -0.230$ and is statistically significant, while the mean of $\varepsilon_{VC,G} = -0.011$ and is statistically insignificant. The inconsistency between the joint significance of the terms involving G in the regressions and the insignificance of the mean $\varepsilon_{VC,G}$ elasticity may be explained by the difference in how the standard errors are calculated for $\varepsilon_{VC,G}$. Namely, the latter are based on the mean of the data over the entire sample.

The authors also find that when spatial effects (both G and spatial autocorrelation) are not recognized, the $\varepsilon_{VC,I}$ is only about -0.15, so they conclude that incorporating G and spatial autocorrelation increases the absolute value of the magnitude of the own-state infrastructure elasticity. Furthermore, the combined effect of G and I is approximately -0.24, which is about 50% larger in magnitude than when both G and spatial autocorrelation are ignored. The upshot is that accounting for these spatial effects appears to have a substantial effect on estimates of the cost-saving impacts of public infrastructure.

Another finding is that several of the inputs (namely, private capital, materials, and non-production labor) are substitutes with I, while production labor is a complement with I. The finding that private capital and I are substitutes is consistent with other findings in the public infrastructure literature.

There are somewhat different relationships between G and the inputs. Namely, capital, non-production labor and production labor are substitutes with G, while materials and G are complements. Cohen and Morrison Paul (2004) note that the substitutability between G and both types of labor is similar to the Boarnet (1998) findings.

Interestingly, Cohen and Morrison Paul (2004) note differences in the regional elasticities involving G. They find that $\varepsilon_{VC,G}$ is slightly positive for the Pacific states, implying that within-state infrastructure is more important than inter-state infrastructure improvements for those states. This may be partly because California, a relatively large state, is included in the Pacific region. On the other hand, $\varepsilon_{VC,G}$ is largest for states in the Mountain and West North Central regions. The authors note that since these states have relatively small populations, interstate highways may be more important for manufacturing firms in those states.

Cohen and Monaco (2007) examine the impacts of ports on manufacturing costs at the state level. They look at the within-state port effects (through I) and the inter-state port effects (through G) based on estimating a Generalized Leontief variable cost function, with I and G as shift factors. They construct ports capital stocks using the perpetual inventory method on state-level ports investment data. The authors also incorporate highway infrastructure variables to test for complementarity or substitutability between ports and highways. They test for and allow for spatial autocorrelation in their analysis as well. The spatial autocorrelation parameters are positive and significant, implying that a shock to states neighboring a particular state spill over to the particular state.

Regarding their elasticity estimates, Cohen and Monaco find that increases in ports infrastructure within a particular state decrease variable costs, with a variable cost elasticity of about -0.04 and statistically significant. The results for the variable cost elasticity with respect to neighboring states' ports infrastructure are quite different. Namely, greater levels of ports infrastructure in neighboring states leads to a rise in variable costs in a particular state. The variable cost elasticity with respect to neighboring states' ports is 0.129. The authors argue that these inter-state findings are consistent with Boarnet (1998), and imply that improved ports in nearby states may draw away the most productive workers from a particular state, leading to higher manufacturing costs in that particular state. In other words, the positive and significant infrastructure elasticity is evidence of external diseconomies of scale. From the perspective of manufacturing firms in a particular state, neighboring states may have too much ports infrastructure during the sample period, and lower ports infrastructure in neighboring states may be expected to lower manufacturing costs in a particular state.

Cohen and Monaco (2007) also find that the elasticity of the shadow value of neighbors' ports with respect to the own state's ports infrastructure is negative and significant. This result implies that states with decreasing ports infrastructure face larger external diseconomies of scale resulting from changes in ports infrastructure in neighboring states. On the other hand, they find that the elasticity of the own state's ports shadow value with respect to the stock of ports in neighboring states is insignificant, implying that additional ports infrastructure in neighboring states has no significant impact on the shadow value of ports infrastructure in a particular state.

Based on the elasticities relating ports and highways, the authors find no significant relationship between the shadow value of ports (highways) and additional highways (ports). The ports shadow value elasticities with respect to both types of labor (production and non-production labor) are positive. In other words, the cost-reduction potential (or shadow value) of ports increases with more workers, so there appear to be some complementarities between workers and ports. Finally, the shadow value of ports increases over time, after controlling for all factor prices and other shift factors, as is seen by the sign and significance of the elasticity of the ports shadow value with respect to the time counter (t).

The functional forms for the cost function studies discussed so far all are Generalized Leontief. Also, the focus of most previous spatial cost function studies is on the impacts of various types of infrastructure on the U.S. manufacturing sector. Another recent study by Moreno et. al. (2004) assesses spillovers for 15 Spanish regions over the years 1980 to 1991, for 12 manufacturing industries. They estimate a translog variable cost function, for two separate classes of models. They classify the first type of model as the "sectoral" case, where the weighted average of other industries' and/or geographic regions' output are included as external inputs. Their sectoral case is similar in spirit to the approach of Morrison and Siegel (1999), who incorporate external shift variables in the cost function for other industries' output. Moreno et. al.'s other group of models is the "regional" case, where they add measures of public capital for neighboring regions. They include measures of public capital (I) within a particular region for each industry, by apportioning the aggregate infrastructure stock to the individual industries based on the output share of each manufacturing industry in total manufacturing output. For the regional case, the authors have a somewhat different specification of G than the spatial lag approach of the other cost function studies described above. Namely, they denote G as W times ln(I), where ln(I) represents the natural logarithm of I, and W is a contiguity matrix based on geographic neighboring Spanish regions. Then, total public capital (which here will be called "T") is assumed to be a geometric mean of the own-region public capital (I) and the neighbors' public capital (G):

$$T \equiv I^{\theta} G^{1-\theta}, \tag{22}$$

where θ is a parameter between 0 and 1 to be estimated empirically together with the rest of the cost function. They argue that one advantage of such a specification for public capital is that it allows for complementarities between I and G. This specification also averts the need to add several additional interaction terms for both I and G, while instead interaction terms for only one infrastructure variable (T) needs to be added to the basic cost function. They argue that inclusion of minimal interaction terms mitigates potential multicolinearity problems. A disadvantage of this approach, however, is that now with the addition of T the model must be estimated with nonlinear regression techniques.

For their regional case, Moreno et. al. build up their model by starting with a translog variable cost function model containing input prices for labor and intermediate materials, an output measure, and a fixed factor for capital. They also perform 3 tests for spatial autocorrelation, and find significant evidence of spatial autocorrelation in this basic model with one of the 3 tests. Next, they add public capital (I), and find that all of the parameters that are involved with terms for I are jointly significant. Once again, they find evidence of spatial autocorrelation with one of their 3 tests for this specification. They find that on average over all Spanish regions, $\varepsilon_{VC,I} = -0.034$. Their estimates for input demand elasticities imply that labor and infrastructure are complementary, while infrastructure and intermediate materials are substitutes. Finally,

when they add cross-region externalities in the form of G and the weighted average of neighboring regions' output, they find no evidence of spatial autocorrelation, but they find that $\theta = 0.58$. This value for θ (and the associated value for $(1-\theta)$) implies that both G and I are important determinants of variable costs, and supports the notion that transportation networks are present. But the elasticity of variable costs with respect to the composite infrastructure measure is now positive, which leads to a conclusion that these Spanish regions may have too much infrastructure during the 1980's. Also, the elasticity of labor with respect to the composite infrastructure measure T is now negative, implying that workers and infrastructure are now substitutes. Furthermore, the elasticity of intermediate materials with respect to infrastructure also switches signs, with an interpretation that these two inputs are now complements. The authors also note, however, that the spatial weight matrix specification may be driving their results with this particular estimation approach, but they do not report results of testing with alternative weight matrices.

In the sectoral case, they assume that $\theta = 1$, so that they do not incorporate public capital spillovers across regions. First, they find that $\varepsilon_{VC,I} = 0.305$, implying once again that there is an excess of public infrastructure capital during the 1980's in Spain. They also find strong evidence of spatial autocorrelation across sectors (which they call "sectoral autocorrelation") based on all 3 tests. Finally, in a separate estimation procedure they add the weighted average of neighboring regions' output as a fixed factor. This additional fixed factor, together with the inclusion of public capital (I), completely eliminates the evidence of significant "sectoral autocorrelation". They also find that the average $\varepsilon_{VC,I} = -0.341$, implying that public infrastructure capital in Spain confers cost-saving benefits on manufacturing firms in that country. In both of the estimation procedures that incorporate public capital for the sectoral case, they find that labor and public infrastructure capital are complements, while intermediate materials and public capital are substitutes.

6. CONCLUSIONS AND FUTURE WORK

Recent advances in spatial econometrics have facilitated analysis of the wider benefits of public infrastructure. In particular, researchers over the past decade have assessed both the impacts of spatial autocorrelation and spatial lags on estimates of the benefits of public infrastructure capital. Various modes of transportation infrastructure have been studied, including highways, air, and ports. Coverage has focused on U.S. counties, states, as well as regions of Spain. Studies have been conducted using both production function and cost function approaches, and have led to a broad range of results. Namely, some studies have found that additional infrastructure capital leads to greater output or lower costs, while others have found the opposite. Despite this lack of consensus on infrastructure's impacts, it is clear that incorporating measures of "wider benefits" has enhanced the precision of the effects of infrastructure relative to the state of the art in the early 1990's. Thus, the innovations in the tool set of spatial econometrics have contributed to understanding in this field. However, there is still more that can be done in future research to improve the accuracy of impact measures for public infrastructure.

One area of potential further work would be to utilize firm-level manufacturing data to estimate the elasticity of variable costs with respect to public infrastructure. Such a disaggregate analysis would allow for greater heterogeneity among the individual agents, which may generate different results for the infrastructure elasticities. Such data are housed at the U.S. Census Bureau Research Data Centers (RDC's). There are a number of obstacles to overcome before obtaining these data, but the potential richness of the data may be worth the effort needed to gain access to the RDC's resources. One potential benefit of the firm-level analysis is that once elasticities are estimated, one could impute for each firm a dollar value of the estimated cost-reduction resulting from additional infrastructure. Such an approach could lead to innovative alternative

approaches for financing infrastructure improvements by charging firms based on their expected (or realized) benefits from infrastructure improvements.

Along with the advances in the area of spatial econometrics over the last 15 years, Geographic Information Systems software has grown in popularity and usage in the economics profession. While its usage in other areas within the economics profession has become common, such as in hedonic housing price studies, there is much that could be done with GIS software in infrastructure studies. For example, researchers could utilize GIS more heavily so as to generate more sophisticated spatial weights in assessing the spillover benefits from "neighboring" jurisdictions. Constructing a greater variety of spatial weights and estimating either the cost function or production function for several different weights specifications can provide a robustness check for the spatial modeling.

Related to the notion of checking robustness of using different spatial weights matrices is incorporating alternative variations of the measure of other localities' infrastructure stocks. Namely, many studies calculate G for a particular locality as the weighted average of other localities' infrastructure, and G enters as a separate shift factor in the analysis. One exception is Moreno *et al.* (2004), who instead use I and G to derive a net infrastructure measure, which we call T in equation (22) above. As noted by Moreno *et al.*, using T instead of separate terms for both I and G reduces the number of interaction terms (and in turn, the number of parameters to estimate with the more sophisticated functional forms), although it introduces nonlinearities that preclude classical linear estimation techniques. But it would be a worthwhile exercise to compute such a composite infrastructure measure and check the robustness of results. One potential drawback, however, is that such a structure imposes additional interdependencies between G and I instead of testing for such interrelationships empirically.

While there have been studies of public infrastructure impacts on manufacturing costs involving multi-modal transportation, such as Cohen and Morrison Paul (2007) for airports and highways, and Cohen and Monaco (2007) for ports and highways, a large scale intermodal study would generate new insights on the complementarity and/or substitutability between different types of infrastructure. A more detailed analysis of spillovers from intermodal transportation at the disaggregate (county) level, incorporating ports, rail, air, and highways would integrate the more complex structure of transportation networks into the current literature.

Another possible extension would be to examine the impacts of infrastructure on other sectors besides manufacturing. Cohen and Monaco have work in progress that explores the impacts of ports on the textiles and wholesale goods sectors, at the California county level. Studies for additional industries and locations that examine other types of infrastructure as well could be insightful.

In addition to looking at the benefits across sectors, another possibility would be to examine the general equilibrium impacts of G along the lines of Haughwout (2002). Namely, this would consist of a model with consumers making consumption choices while minimizing their total expenditures, with infrastructure as an exogenous shift factor. Additionally, the model would have a production side, with firms choosing inputs to minimize production costs, and infrastructure would also enter the cost function. Here, "infrastructure" could consist of both I and G, so one might assess the general equilibrium impacts on welfare from infrastructure spillovers, both across jurisdictions as well as within a particular jurisdiction.

Another more macro approach would be to look at benefits across countries, such as individual European countries that are highly interdependent, along with benefits across regions that are within countries. Cohen and Morrison Paul (2003b) assess production-related spillovers across EU countries, but they do not incorporate infrastructure. Yet another aspect would be having different layers of G that start at micro level, and then aggregate up. This approach would avoid missing spillovers that accrue within individual countries when doing a cross-country spillover analysis. While the spillover public capital stocks (G) would likely be larger here, this does not necessarily imply that the benefits would be greater as well. The sign of the net

benefits would depend on the sign of the elasticities with respect to infrastructure based on the econometric estimation of the model.

Finally, rolling up many of these ideas and examining them together would be a complex exercise. But it would also be an excellent springboard for introducing CGE models, as presented by Lakshmanan, *et al.* (2007). Needless to say, there is much more work that still can be done in assessing the wider benefits of public infrastructure capital.

BIBLIOGRAPHY

Anselin, L. 1981. "Small sample properties of estimators for the linear model with a spatial autoregressive structure in the disturbance," *Modeling and Simulation* 12: 899-904.

Aschauer, D.A. 1989. "Is Public Expenditure Productive?" *Journal of Monetary Economics*, 23(2): 177-200.

Boarnet, M.G. 1998. "Spillovers and the Locational Effects of Public Infrastructure" *Journal of Regional Science*, 38(3): 381-400.

Cliff, A. and Ord, J. 1981. *Spatial Processes, Models and Applications*. London: Pion.

Cohen, J.P. and K. Monaco. 2007. "Ports and Highways Infrastructure: An Analysis of Intra- and Inter-state Spillovers," manuscript.

Cohen, J.P. and C.J. Morrison Paul. 2007. "The Impacts of Transportation Infrastructure on Property Values: A Higher-Order Spatial Econometrics Approach" *Journal of Regional Science*, 47(3): 457-478.

Cohen, J.P, and C.J. Morrison Paul. 2004. "Public Infrastructure Investment, Interstate Spatial Spillovers, and Manufacturing Costs" *Review of Economics and Statistics* 86: 551-560.

Cohen, J.P. and C.J. Morrison Paul. 2003a. "Airport Infrastructure Spillovers in a Network System" *Journal of Urban Economics* 54(3): 459-473.

Cohen, Jeffrey P. and Catherine Morrison Paul. 2003b. "Production Externalities, Integration and Growth: The Case of the European Union 'Single Market'", *Growth and Development in the Global Economy*, (Harry Bloch, ed.), Edward Elgar Press, chapter 4, pages 53-66.

Haughwout, A. 2002. "Public Infrastructure Investments, Productivity and Welfare in Fixed Geographic Areas" *Journal of Public Economics,* 83: 405-425.

Holtz-Eakin, D. and A.E. Schwartz. 1995. "Spatial Productivity Spillovers from Public Infrastructure: Evidence from State Highways" *International Tax and Public Finance* 2: 459-468.

Kelejian, H.H. and I.R. Prucha. 2004. "Estimation of Simultaneous Systems of Spatially Interrelated Cross Sectional Equations" *Journal of Econometrics* 118: 27-50.

Kelejian, H.H. and I.R. Prucha. 1999. "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model." *International Economic Review* 40, 509-533.

Kelejian, H.H. and I.R. Prucha. 1998. "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances," *Journal of Real Estate Finance Economics*, 17, 99-121.

Kelejian, H.H. and D. Robinson. 1992. "Spatial Autocorrelation: A New Computationally Simple Test With an Application to Per Capital County Police Expenditures" *Regional Science and Urban Economics* 22: 317-331.

Kelejian, H.H. and D. Robinson. 1997. "Infrastructure productivity estimation and its underlying econometric specifications: a sensitivity analysis." *Papers in Regional Science* 76: 115-131.

Lakshmanan, T.R., W. Anderson, and I. Sue Wing. 2007. Supply and demand side meso effects of infrastructure investments. Manuscript.

Moreno, R., E. Lopez-Bazo, E. Vaya, and M. Artis. 2004. "External Effects and Costs of Production," Chapter 14 in *Advances in Spatial Econometrics: Methodology, Tools, and Applications* (L. Anselin, 1981, R.J.G.M. Florax, and S.J. Rey, eds.), Berlin: Springer.

Morrison, C.J. and A.E. Schwartz. 1996. "State Infrastructure and Productive Performance" *American Economic Review* 86: 1095-1111.

Morrison, C.J. and D. Siegel, 1999. "Scale Economies and Industry Agglomeration Externalities: A Dynamic Cost Function Approach," *American Economic Review* 89: 272-290.

Munnell, A.H. 1990. "How Does Public Infrastructure Affect Regional Economic Performance?" *New England Economic Review*, September/October, 11-32.

Nadiri, M.I. and T.P. Mameaunus. 1994. "The Effects of Public Infrastructure and R&D Capital on the Cost Structure and Performance of U.S. Manufacturing Industries," *Review of Economics and Statistics* 76: 22-37.

Paul, C.J. Morrison, V.E. Ball, R.G. Felthoven, and R. Nehring, 2001. "Public Infrastructure Impacts on U.S. Agricultural Production: A State-Level Panel Analysis of Costs and Netput Composition," Public Finance and Management 1, http://www.spaef.com.

Varian, H.R. 1992. *Microeconomic Analysis*, third edition. New York: W.W. Norton.

AGGLOMERATION ECONOMIES AND TRANSPORT INVESTMENT

Daniel J. GRAHAM[1]
Imperial College
University of London
London
United Kingdom

SUMMARY

1. INTRODUCTION 98

2. AGGLOMERATION ECONOMIES AND TRANSPORT INVESTMENT 98

2.1. Agglomeration and productivity 98
2.2. Transport investment and agglomeration 101

3. ESTIMATING AGGLOMERATION ECONOMIES 103

3.1. Firm data 104
3.2. Measuring agglomeration 104
3.3. Estimating the link between agglomeration and productivity 104

4. RESULTS 105

4.1. Production function estimates 105
4.2. Applying the agglomeration elasticities in transport appraisal 106
4.3. Limitations of the approach and future research directions 108

5. CONCLUSIONS 108

NOTES 110

BIBLIOGRAPHY 111

APPENDIX 1: THE TRANSLOG PRODUCTION INVERSE INPUT DEMAND MODEL 113

London, August 2007

ABSTRACT

This paper is concerned with the links between agglomeration, productivity and transport investment. If improvements in transport systems give rise to changes in the mass of economic activity accessible to firms, for instance by reducing travel times or the costs of travel, then they can induce positive benefits via agglomeration economies. The paper presents empirical results from an econometric analysis of the relationship between productivity and accessibility to economic activity for different sectors of the UK economy. The results show that agglomeration economies do exist and that they can be substantial, particularly for services. Furthermore, the effect of agglomeration externalities is not trivial when considered in the context of transport appraisal. Initial calculations typically indicate additions to conventional user benefits of 10%-20% arising from increasing returns to economic mass.

1. INTRODUCTION

A recent paper by Venables (2007) develops a theoretical model to demonstrate some important links between transport provision and agglomeration. He shows that if, as urban economic theory suggests, there are increasing returns to agglomeration, then transport investments may induce positive productivity benefits by effectively raising accessibility to economic mass. Any such agglomeration externalities can be classed as 'wider benefits' of transport investment in the sense that they are typically not captured in a standard cost-benefit appraisal.

To understand the magnitude of the potential 'wider benefits' of transport investment we first need quantitative estimates of returns to agglomeration. In other words we require some empirical verification of the existence and magnitude of the relationship between productivity and accessibility to economic mass. Preferably, we want to examine this relationship separately for different sectors of the economy because the benefits derived from agglomeration are unlikely to be uniform across industries.

This paper describes the results of new empirical research on the relationship between agglomeration and productivity for different sectors of the UK economy. It also considers the implications of agglomeration economies for the evaluation of transport investment. The results show that agglomeration economies do exist and that they can be substantial, particularly for services. If transport investments change the densities available to firms, for instance through a reduction in travel times or in the cost of travel, then there are likely to be positive gains from agglomeration.

Furthermore, the effect of agglomeration externalities is not trivial when considered within the framework of transport appraisal. Initial calculations typically indicate additions to conventional user benefits of 10%–20% arising from increasing returns to economic mass.

The paper is structured as follows. Section 2 reviews the literature on agglomeration and productivity and discusses the relationship between transport investment and agglomeration. Section 3 describes the methodology used to estimate agglomeration economies. Empirical results are presented in section 4, including a review of some recent applications of agglomeration benefits within transport appraisal. Conclusions are then drawn in the final section.

2. AGGLOMERATION ECONOMIES AND TRANSPORT INVESTMENT

2.1. Agglomeration and productivity

The tendency towards concentration or agglomeration is perhaps the most widely observed feature of the spatial organisation of economic activity. It can be discerned across the Globe at a variety of different geographical levels. Agglomeration is evident, for instance, in the existence and growth of cities,

in the formation of industrial regions and districts, and in the clustering of like activities within the same neighbourhood of a town or city.

Attempts to explain the microfoundations of agglomeration generally start from the premise that cities and industrial concentrations would not form if there were not some tangible benefits that accrue to firms. The advantages derived through the spatial concentration of economic activities are referred to generically as *agglomeration economies*.

Typically, a distinction is made between agglomeration effects that arise from the scale or density of activity within a particular industry and from those due to urban scale or city size. Economies of industry concentration, termed localization economies, are external to the firm but internal to the industry and are principally thought to be sourced through labour market pooling, the sharing of intermediate inputs, and knowledge sharing or 'technological spillovers'. Economies of urban concentration, or urbanization economies, are external to the firm and the industry but internal to the city with benefits arising from the existence of local public goods, the scale of markets, the proximity of input-output sharing, and other kinds of inter-industry interaction.

The theoretical foundations for the existence of agglomeration economies are now well established (see for example Fujita and Thisse 2002; Duranton and Puga 2005). There is also a body of empirical work that has sought to identify these externalities and to quantify their effects on productivity. There are a number of excellent up-to-date surveys of the empirical literature on agglomeration (see in particular Rosehthal and Strange 2004; Eberts and McMillen 1999). This literature has concentrated largely on manufacturing with, until very recently, few published results on the link between agglomeration and service sector productivity. This is almost certainly due the poor quality of service sector data for most countries compared to the manufacturing statistics. Nevertheless, since services now comprise such a large share of many national and urban economies, the emphasis on manufacturing represents a real limitation.

To identify agglomeration economies empirical work typically proceeds by constructing variables that measure the extent of industry and urban concentration, and uses these within a production or cost function framework to estimate effects on productivity. Urbanization is often represented by the total population or total employment of an urban area. Localization is proxied using some measure of local industry scale such as employment. Table 1 provides a summary of some prominent studies of the effects of agglomeration on productivity. It summarises those studies that have produced an actual elasticity estimate of the effects of agglomeration, rather than those that have detected agglomeration effects through the use of dummy variables or other limited variable methods.

With the exception of studies 17 and 18, which are concerned with effects on total economic productivity, the estimates shown in table 1 are for manufacturing industries. Elasticities describing the strength of localization economies are given in 13, 14, 15, and 16; the remaining estimates show the effect of urbanization economies on productivity.

The estimates of urbanization economies for manufacturing industries shown in table 1 range from 0.01 to 0.20, but the majority of values are under 0.10. This indicates that a doubling of city size is typically associated with an increase in productivity of somewhere between 1% and 10%. The estimates given in the table are all positive although Henderson (1986) and Henderson (2003) do report difficulties in identifying urbanization effects on productivity.

Table 1 shows four estimates of localization economies. Nakamura (1985) estimates the effect of localization economies on the productivity of 20 manufacturing industries. He quotes an unweighted average elasticity of productivity with respect to industry size of 0.05. Henderson (1986) using industry level data for US MSAs and Brazilian cities also find positive localization economies. His estimates for Brazil vary by industry, with a maximum elasticity estimate of 0.20 and a minimum of 0.03, the mean value over 10 industries is 0.11. For US MSAs estimated localization elasticities range from 0.09 to 0.45 with a mean value of 0.19. Henderson (2003) estimates a mean localization elasticity of 0.03.

Table 1. **Estimates of agglomeration economies from production function analyses**

Author	Unit of Analysis	Independent Variable	Elasticity
1 Aaaberg (1973)	Swedish cities	city size (population)	0.02
2 Shefer (1973)	US MSAs	RTS at MSA aggregation	0.2
3 Sveikauskas (1975)	US MSAs	city size (population)	0.06
4 Kawashima (1975)	US MSAs	city size (population)	0.2
5 Fogarty and Garofalo (1978)	US MSAs	city size (population)	0.1
6 Moomaw (1981)	US MSAs	city size (population)	0.03
7 Moomaw (1983)	US MSAs	city size (population)	0.05
8 Moomaw (1985)	US MSAs	city size (population)	0.07
9 Nakamura (1985)	Japanese Cities	city size (population)	0.03*a*
10 Tabuchi (1986)	Japanese Cities	city size (population)	0.04
11 Louri (1988)	Greek Regions	city size (population)	0.05
12 Sveikauskas *et al*. (1988)	US MSAs	city size (population)	0.01*b*
13 Nakamura (1985)	Japanese Cities	industry size (employment)	0.05
14 Henderson (1986)	Brazilian Cities	industry size (employment)	0.11*c*
15 Henderson (1986)	US MSAs	industry size (employment)	0.19*d*
16 Henderson (2003)	US MSAs	industry size (no. of plants)	0.03*e*
17 Ciccone and Hall (1996)	US States	employment density	0.06
18 Ciccone (2002)	EU regions	employment density	0.05
19 Rice et al. (2006)	GB NUTS 3	proximity / travel time	0.04

Notes: MSA - Metropolitan Statistical Area,
a - mean value for 14 manufacturing industries,
b - mean value from 5 model specifications,
c - mean value for ten industries,
d - mean value for 9 industries,
e - mean value for 4 model specifications.

In addition to studies using MSA population and employment to represent city and industry size there other studies that have incorporated some measures of distance or density into the specification of agglomeration effects. Two papers are particularly interesting in this respect. Ciccone and Hall (1996) derive an equation to estimate the effects of county-level employment density on aggregate state productivity for the United States. They find that over 50% of the variance in aggregate labour productivity across states can be explained by variance in the density of employment and that a doubling of employment density is associated with a 6% increase in average labour productivity (i.e. an elasticity of 0.06). Ciccone (2002) extends the analysis to European data and estimates an elasticity of labour productivity with respect to employment density of 0.045.

Thus, from the empirical literature, we find evidence to support the theory of increasing returns to urban density and of returns to industry size.

2.2. Transport investment and agglomeration

It seems intuitively reasonable to suppose a further link between transport provision and the benefits that arise from the spatial concentration of economic activity. Transportation costs are crucial in determining the mass of economic activity (including population) that firms can access. New investments in transport can render a larger scale of activity more accessible by reducing travel times or the costs of travel, giving rise to positive agglomeration benefits. Conversely, where transports systems work inefficiently, or where there are constraints on accessibility, these may inhibit the generation and distribution of agglomeration externalities.

A crucial issue here is that agglomeration economies are *externalities*, that is, they arise as a side effect of the activities of firms which have consequences for the wider economy. This is very important from the point of view of transport appraisal because traditional methods of appraisal based on valuation of travel times do not recognise these types of externalities. For this reason agglomeration effects of transport investment can be classes as *wider economic benefits* because they represent market imperfections that are not accounted for in a standard cost-benefit appraisal.

Venables (2007) formalises this argument and shows that estimates of the elasticity of productivity with respect to agglomeration can be used to shed light on the magnitude of the external benefits of transport improvements. He develops a theoretical model of an urban economy that links productivity to transport investment via effects on city size. His objective is to distinguish real income changes that result from transport investment due to a productivity-city size (agglomeration) effect, from those economic benefits that are captured in standard transport appraisals and which arise from resources saved in commuting and from an increase in urban output.

Venables' paper provides a clear demonstration of the key relevant arguments that link transport and agglomeration. A diagraming representation of the model taken from Venables' paper is given in figure 1.

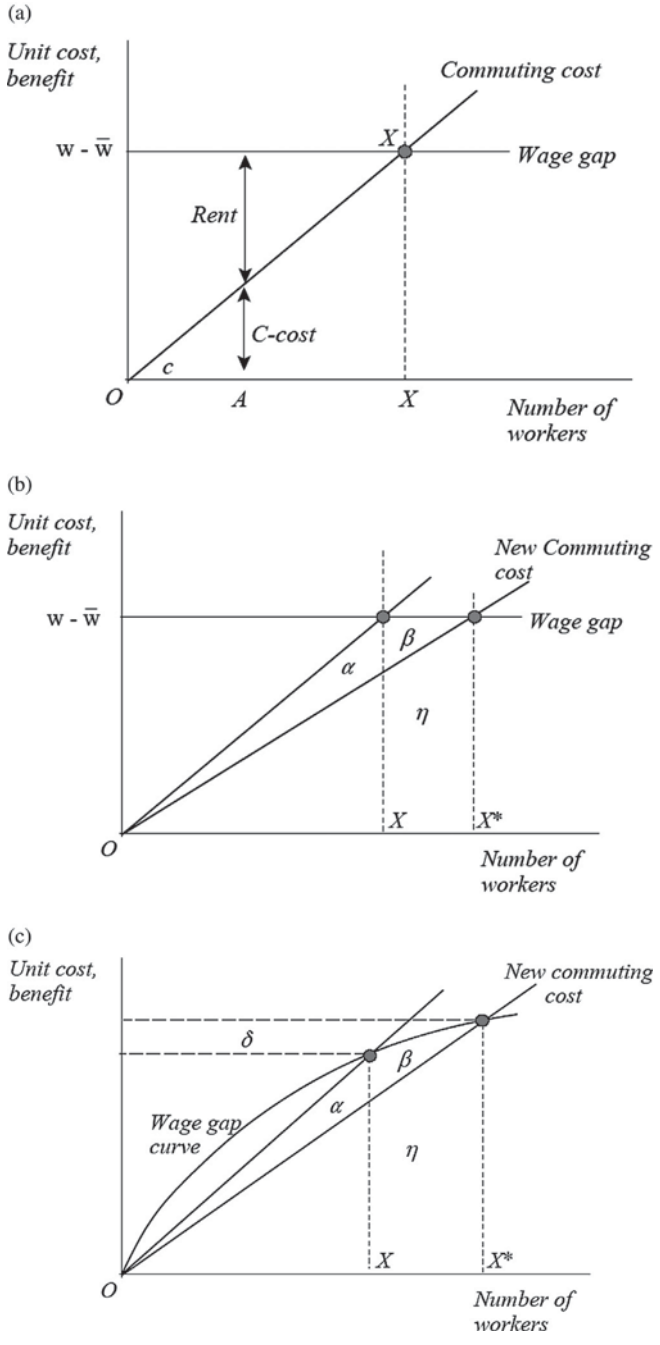
Figure 1a shows an urban equilibrium in which the size of the city is determine at point *X*, where the wage gap between city workers and non-city workers is taken up in the travel costs of the city worker who is most distant from the CBD.

Figure 1b shows that when a transport improvement is made commuting costs are shifted downwards and consequently the city expands to point X^*. The total change in the resources used in commuting is $\eta - \alpha$, which combined with the change in output ($\beta + \eta$), yields a net benefit from the transport improvement of $\alpha + \beta$.

In Figure 1c, Venables shows the implications of the existence of a city size-productivity gradient. If larger cities have higher productivity due to agglomeration externalities then the wage gap can be expressed, not as a constant gap, but as a concave curve that increases with city size. Equilibrium is found at the intersection of the commuting cost and wage gap curves. The fact that productivity is non-constant with respect to city size means that the real income gain from a transport improvement is $\alpha + \beta + \delta$; where δ measures the increase in productivity experienced by city workers and is akin to a measure of the elasticity of productivity with respect to city size.

In this way Venables demonstrates that there are external benefits from transport investment related to agglomeration and that these can be quantified using elasticities of productivity with respect to some measure of agglomeration.

Figure 1a, **Urban equilibrium;** b, **Net gains from transport improvement;** c, **Net gains form transport improvement with endogenous productivity**



3. ESTIMATING AGGLOMERATION ECONOMIES

The previous literature has indicated that agglomeration externalities do exist for manufacturing, but the sectoral coverage of existing work is incomplete and the analysis of agglomeration is typically based on data for relatively aggregated industries and spatial areas.

The purpose of the research described in this paper is to estimate a set of agglomeration elasticities that are compatible with the objective of assessing the wider economic benefits of transport investment. We do this for a comprehensive range of industries to find out whether agglomeration externalities really matter across sectors and whether they might be important in assessing the benefits of transport investment[2].

The empirical analysis uses firm level data to represent spatial variance in productivity along with data for small areas to construct measures of the agglomeration 'experienced' by firms. The analysis proceeds in four steps. First, we gather data on the production characteristics of firms across a range of sectors. We then use Geographical Information System (GIS) software to identify the location of each firm in an electronic map. In the third step, we overlay on top of our map of firms a framework of small spatial units and use these to construct measures of the agglomeration experienced by each firm in each location. Finally, we use the firm data and the measures of agglomeration within a production function framework to estimate the effect of agglomeration on firm productivity.

There are several advantages to the micro firm-level approach we adopt here rather than the conventional method which uses aggregate spatial areas as the units of observation:

(i) **Consistency with theory** - the assumptions we use to analyse production behaviour presuppose firms as the basic decision making units, not aggregate spatial areas. Thus, modelling at the firm level provides consistency with the theory we draw upon to analyse productivity.

(ii) **Compatible measures of agglomeration** - by locating each firm geographically we can capture a high degree of spatial detail in our measures of agglomeration and avoid using data based on large pre-defined geographic units such as administrative areas or metropolitan definitions. Furthermore, using a distance based approach we can include an implicit transport dimension in the measure of agglomeration by considering not just the scale of economic activity within some concentration, but how accessible (proximate) this scale is to each firm.

(iii) **Flexible representation of production technology** - analysis using production data aggregated over firms require us to assume homogenous technology across those firms and constant returns to scale, restrictions which can give rise to aggregation bias. Firm level data permit the use of more flexible functional forms to represent technology.

(iv) **Econometric estimation** - in estimating productivity the use of extensive firm level data can help to reduce multicollinearity and provide more identifying variance (e.g. Griliches and Mairesse 1995).

3.1. Firm data

The firm data we use for estimation describe the production and cost characteristics of registered UK companies in 2 digit sectors. Under UK legislation each registered company is required to provide accounting and other information about their operations to an executive agency of the Department of Trade and Industry know as Companies House. These data are made available in a commercial software package called Financial Analysis Made Easy (FAME), which is produced jointly by Jordans and Bureau Van Dijk (BVD 2003). The production data are for companies not plants. It is, however, possible to identify and remove multi-plant firms from the sample because they report more than one trading address.

The FAME data record extensive financial information for each firm and are available for a number of years, although the time-series reporting for individual firms is irregular. The basic input data we have on each firm include a measure of capital stock and the number of employees. Capital stock is the value of assets possessed by the firm and includes 'fixed assets' such as the depreciated value of buildings, plant, machinery and equipment; 'current assets' such as stocks and various debts owed to the company; and 'current liabilities' or the amount owed by the company as a result of normal trading. Sales are used as a proxy for output. We also have data on wages and on the total costs of each firm, which includes all direct elements of the cost of the ordinary activities used to produce the firm's output.

3.2. Measuring agglomeration

The measure of agglomeration we use is calculated using a *ward* framework because there are extensive economic data available for these areas3, and because they allow for a high level of spatial disaggregation dividing Britain (230,700 square kilometres) into approximately 10,760 units.

Using the ward data we represent agglomeration with an 'effective density' measure. This is essentially an accessibility based measure of agglomeration for very small areas. The total effective density (U) of employment that is accessible to any firm located in ward *i* is

$$U_i = \frac{E_i}{\sqrt{A_i/\pi}} + \sum_{j}^{i \neq j} \left(\frac{E_j}{d_{ij}} \right) \tag{1}$$

where E_i is total employment in ward i, A_i is the area of ward i, E_j is total employment in ward j, and d_{ij} is the distance between i and j. Note that the density effect that arises within the ward in which the firm is actually located (i.e. the first term on the right hand side of equation (1)) is measured by total ward employment divided by a proxy for average ward radius that is calculated assuming that the wards are roughly circular.

It is worth stressing here the implicit transport dimension of (1). Our effective density measure captures the scale and proximity of economic activity that is available in particular locations. We assume that investment in transport will change effective densities because it will alter the relative proximities of activity. Note, that we could also use travel times, or a measure of the generalised cost of travel, as the denominator in equation (1) (e.g. Graham 2007c).

3.3. Estimating the link between agglomeration and productivity

As externalities, agglomeration economies are treated as a kind of technology component that serves to shift the firm's production or cost function. For instance, at the firm level a typical specification of the production function would be

$$Y = g(U)\, f(X) \tag{2}$$

where Y is the output level of the firm, X is a vector of factor inputs, and $g(U)$ is a vector of influences on production that arise from agglomeration economies.

We use firm level data to provide an empirical representation of the production function, allowing us to estimate the effect of agglomeration on firm productivity. Specifically, we a variant of the translog production function which includes a primary production function along with a set of inverse input demand equations, which introduce additional information on costs and factor prices. Using this particular approach we can sketch out a reasonably complete specification of the production technology of firms and analyse some distinct effects of agglomeration on productivity. A description of the translog model used is given in appendix 1. A full demonstration of the model for the estimation of agglomeration economies is provided by Graham and Kim (2007).

4. RESULTS

In this section we present estimates of the relationship between agglomeration and productivity for UK industries derived using the production function methodology outlined above. We then review some recent attempts to use these results within appraisal methodology to assess the agglomeration benefits of transport investments. Finally we note some limitations of the approach and suggest some future directions for research.

4.1. Production function estimates

The results presented in this sub-section are taken from Graham (2005) and Graham (2006), and the intention here is to provide only an overview of the empirical finding of this previous work. For a full description of methodology, data sources, or other technical aspects of the research the reader should refer to these more detailed reports.

The results are presented for eight industry groups which comprise the following SIC codes:

(i) Manufacturing (MAN) (SIC 15-40)

(ii) Construction (CON) (SIC 45)

(iii) Distribution, Hotels & Catering (DHC) (SIC 50-55)

(iv) Transport, Storage & Communications (TSC) (SIC 60-64)

(v) Real Estate (RE) (SIC70)

(vi) Information Technology (SIC 72)

(vii) Banking, Finance & Insurance (BFI) (SIC 65-67)

(viii) Business Services (BUS) (SIC741 -745)

Separate estimates of agglomeration economies from the production function analyses are obtained for each group4. These are expressed as elasticities showing the proportional change in productivity associated with a proportional change in the level of agglomeration. Table 2 below shows estimates of the elasticities of productivity with respect to agglomeration.

Table 2. **Estimated elasticities of productivity with respect to agglomeration**

Industry	Elasticity
Manufacturing	0.077
Construction	0.072
Distribution, hotels & catering	0.153
Transport, storage & communications	0.223
Real estate	0.192
IT	0.082
Banking, finance & insurance	0.237
Business services	0.224
Whole economy	0.119

We estimate positive agglomeration externalities for manufacturing, construction and for each of our six service industries. The lowest agglomeration elasticity shown in the table is for manufacturing (0.077). The largest agglomeration elasticities are for transport storage & communications (0.223)5, banking finance & insurance (0.237), business services (0.224), and real estate (0.192).

The weighted average elasticity for the service sector as a whole, where the weights are based on industry group employment shares, is 0.186. This indicates that a doubling of accessibility to economic mass is associated with an increase in productivity of just under 20%. The service sector elasticity is over twice as large in magnitude than the manufacturing estimate of 8%. So it seems, on the basis of the results given in table 2 that services enjoy higher returns from agglomeration than manufacturing, and particularly the types of activities that we expect to find in CBD locations such as banking finance & insurance, business services, and real estate. Calculating a weighted average elasticity over all industries, gives an estimated elasticity of productivity with respect to agglomeration for the whole economy of 0.119 (12%).

4.2. Applying the agglomeration elasticities in transport appraisal

The results given above support the theory of increasing returns to agglomeration across a range of different industries. Proximity to economic mass appears to matter, and for this reason, we may suppose that an increase in effective densities induced through transport investment could have associated productivity benefits via agglomeration. However, we may still wonder about the actual magnitude of these effects in the context of transport appraisal. Would they appear insignificant relative to conventional travel time savings, or could they actually make a real difference to the benefit-cost calculations?

The answers to these questions will ultimately depend on the characteristics of any particular scheme. However, by way of illustration we can draw upon some recent examples of ex ante evaluation that have calculated the agglomeration benefits of certain transport investments for the UK.

The first such evaluation was carried out by the UK Department for Transport (DfT 2005). Using agglomeration elasticities given in Graham (2005), and employing a methodology similar to that suggested by Venables (2007), the DfT have reappraised a proposed London rail scheme called Crossrail to see how these externalities would affect the projected benefits of investment. Table 3 below shows the results of this exercise6. The table shows that inclusion of the urban economic effects, the so called agglomeration benefits, increase the total benefits of the Crossrail project by 25%.

Table 3. **Applying the new appraisal to CrossRail (DfT calculations)**

Benefits	Welfare ($ million)
Business time savings	4 847
Commuting time savings	4 152
Leisure time savings	3 833
Total user benefits (conventional)	**12 832**
Agglomeration benefits	3 094
Total benefits (new approach)	**15 926**

The second recent evaluation of agglomeration benefits has been undertaken by Steer Davies Gleave (SDG) consultants, again using agglomeration estimates given in Graham (2005). They have undertaken a full economic appraisal of various proposed schemes for the Yorkshire & Humberside regionof England. The results relating to estimated agglomeration benefits are shown in table 4. SDGs calculations typically indicate additions to conventional user benefits of somewhere between 10% – 20% arising from increasing returns to agglomeration.

The calculations shown in tables 3 and 4 indicate that the inclusion of agglomeration effects could substantially increase the estimated benefits of transport projects. If these are as large as these recent applications show, there are some important implications for those charged with making decisions on transport investment:

Table 4. **Appraisal of agglomeration benefits from transport investments**

Mode	Scheme	Agglomeration
Road	Leeds to Bradford Improved Highway	21%
Road	Leeds Urban Area Improved Highway	22%
PT	Leeds to Bradford PT Improvements	15%
Bus	Intra Leeds bus subsidy	11%
Road	Leeds to Sheffield Improved Highway	19%
Road	M6 shoulder	12%
Bus	West Yorkshire County bus subsidy	9%
PT	Leeds Urban Area Major PT Investment	9%
Bus	South & West Yorkshire Bus subsidy	7%
Bus	South Yorkshire bus subsidy.	3%

- the inclusion of additional information on agglomeration benefits could help inform the prioritisation of schemes for funding allocation.
- the estimation of higher returns to transport could release more public funds for investment.
- identifying impacts on GDP and on welfare could help to assess the trade-offs between scheme objectives.
- the quantification of GDP effects could help support calls for private contributions to infrastructure investment.

4.3. Limitations of the approach and future research directions

The research on agglomeration and transport investment has only very recently emerged and there are a number of limitations of the existing approach that require further attention. The empirical identification of agglomeration economies is fraught with difficulties. The actual processes that give rise to these externalities are generally not observed; instead, we use variables reflecting urban or industrial densities to measure the aggregate efficiency gains that we believe are offered by cities and industrial clusters. The measurement and analysis of productive efficiency itself also poses a number of problems, as do the classifications available to describe industrial and functional heterogeneity. In this sub-section, we emphasize some priorities for future research that could address limitations of the existing treatment of agglomeration in transport appraisal.

The first obvious limitation of the existing approach, and of the empirical work presented in this paper, is that it does not actually tell us much about where the productivity benefits of agglomeration come from. The theoretical literature does propose a number of sources of agglomeration benefits, (i.e. labour market benefits, knowledge interactions, and input sharing), but the empirical literature has not yet uncovered the relative magnitude of productivity effects arising from each source. In the context of transport appraisal, this means that we do not know how the sources of agglomeration might relate to transport movements.

This may in fact prove to be an important gap in our knowledge. When transport investments are made they usually affect different types of journey in different ways. Some transport investments will have their greatest impact on business trips, others on commuting, and others perhaps on freight trips. The extent of the overall agglomeration benefit from a scheme, therefore, will depends on the degree to which agglomeration externalities are currently constrained by transport provision, but also by the extent to which agglomeration is driven by different journey purposes made by different modes.

A second important research theme that requires further attention relates to the geographic scope of agglomeration economies and how we represent this in appraisal. Essentially, the issue here concerns our understanding of the spatial distribution of the agglomeration benefits that might arise from transport spending. For instance, if we make a transport investment in Central London are the agglomeration benefits of this investment available only in the immediate locality of the project, or are they distributed further perhaps through the whole of Central, Inner, or Outer London, or even beyond? This is clearly a very important issue. Whether the productivity benefits of investment via agglomeration affect only a relatively small number of firms or a very large number of firms will radically alter estimates of the magnitude of those benefits. We therefore need to know more about how agglomeration economies diminish with distance from source. Rice et al. (2006) address this theme and it certainly requires more attention in the empirical work on agglomeration and productivity.

Finally, there are a range of other limitations of much of the existing work on agglomeration and productivity that are the subject of ongoing research. These include problems of identification arising from endogeneity and measurement error and the issue of urban functional specialization and the ‘quality’ of inputs7.

5. CONCLUSIONS

This paper has considered the link between agglomeration and productivity for sectors of the UK economy. It has developed an 'effective density' measure of accessibility to economic mass for small spatial areas which incorporates an implicit transport dimension. The analysis presented above tests the association between productivity and effective density in a firm level translog production function analysis.

The motivation for the study is to identify if there might be any external benefits that arise from the provision of transport infrastructure that are not included in standard transport appraisals. The results show that agglomeration economies do matter and that they can be substantial, particularly for services. We calculate a weighted average agglomeration elasticity of 0.119 for the economy as whole, 0.186 for the service sector and 0.077 for manufacturing. We also find considerable variation across industries in the magnitude of the elasticities.

If transport investment changes the densities available to firms, for instance through a reduction in travel times or in the cost of travel, then there are likely to be positive gains from agglomeration. Having reliable estimates of the economic mass-productivity relationship allows us to quantify these `wider' economic benefits. Some recent applications find that the effect of agglomeration externalities is not trivial when considered in the context of transport appraisal. Initial calculations typically indicate additions to conventional user benefits of 10%-20% arising from increasing returns to economic mass.

NOTES

1. Centre for Transport Studies, Imperial College London, London, SW7 2AZ, UK, Tel: +4420-7594-6088, Fax: +4420-7594-6107, Email: d.j.graham@imperial.ac.uk

2. In this analysis we concern ourselves with the agglomeration of all economic activity and do not distinguish between urbanization and localization economies. Graham (2007b) uses a similar approach to estimate externalities arising from both sources.

3. We use ward level employment data taken from the Annual Business Inquiry (ABI), the official census of employment in Great Britain.

4. Estimation using a finer industrial disaggregation at the 2 digit level can be found in Graham (2007a).

5. It is interesting that such a high elasticity is estimated for transport services. This result may be indicative of the increasing returns to density which tend to affect transport operators such that unit costs fall as the density of traffic increases. (e.g. Berechman 1993, Graham *et al*. 2003).

6. It is important to emphasise that these calculations have been made by the DfT. The full methodology and a background to Crossrail can be found in DfT 2005.

7. There is evidence to show that functional specialization may vary systematically across the urban hierarchy with larger cities tending to have a higher proportion of firms engaged in specialized functions involving skilled occupations (for instance Duranton and Puga 2005, Rice *et al*. 2006, Combes *et al*. 2007).

BIBLIOGRAPHY

Aaaberg, Y. (1973). Regional productivity differences in Swedish manufacturing. *Regional and Urban Economics 3*, 131-156.

Berechman, J. (1993). *Public transport economics and deregulation policy*. Amsterdam: North Holland.

BVD (2003). *FAME: UK and Irish company information in an instant*. London: Bureau van Dijk.

Ciccone, A. (2002). Agglomeration effects in Europe. *European Economic Review 46*, 213-227.

Ciccone, A. and R. Hall (1996). Productivity and the density of economic activity. *American Economic Review 86*, 54-70.

Combes, P., G. Duraton, and L. Gobillon (2007). Spatial wage disparities: Sorting matters! *Journal of Urban Economics (in press)*.

DfT (2005). *TRansport, wider economic benefits and impacts on GDP*. London: HMSO.

Duranton, G. and D. Puga (2005). From sectoral to functional urban specialisation. *Journal of Urban Economics 57*, 343-370.

Eberts, R. and D. McMillen (1999). *Agglomeration economies and urban public infrastructure*, Chapter in HP Cheshire and E S Mills (eds) Handbook of regional and urban economics, Volume III. New York: North Holland.

Fogarty, M. and G. Garofalo (1978). Environmental quality income trade-off functions with policy applications. *paper presented at the Southern Regional Science Association Meeting*, .

Fujita, M. and J. Thisse (2002). *The economics of agglomeration: Cities, industrial location and regional growth*. Cambridge: Cambridge University Press.

Graham, D.J. (2005). *Wider econonmic benefits of transport improvements: link between agglomeration and productivity, Stage 1 Report*. London: DfT.

Graham, D.J. (2006). *Wider econonmic benefits of transport improvements: link between agglomeration and productivity, Stage 2 Report*. London: DfT.

Graham, D.J. (2007a). Agglomeration, productivity and transport investment. *Journal of Transport Economics and Policy 41*, 1-27.

Graham, D.J. (2007b). Identifying urbanization and localization externalities in manufacturing and service industries. *Papers in Regional Science (in press)*.

Graham, D.J. (2007c). Variable returns to agglomeration and the effect of road traffic congestion. *Journal of Urban Economics 62*, 103-120.

Graham, D.J., A. Couto, W. Adeney, and S. Glasiter (2003). Economies of scale and density in urban rail transport: effects on productivity. *Transportation Research E 39*, 443-458.

Graham, D.J. and H.Y. Kim (2007). An empirical analytical framework for agglomeration economies. *Annals of Regional Science (in press)*.

Griliches, Z. and J. Mairesse (1995). Production functions: the search for identification. *NBER 5067*.

Henderson, J. (1986). Efficiency of resource usage and city size. *Journal of Urban Economics 19*, 47-70.

Henderson, J.V. (2003). Marshall's scale economies. *Journal of Urban Economics 53*, 1-28.

Kawashima, T. (1975). Urban agglomeration economies in manufacturing industries. *Papers of the Regional Science Association 34*, 157-175.

Kim, H. (1992). The translog production function and variable returns to scale. *Review of Economics and Statistics 74*, 546-552.

Louri, H. (1988). Urban growth and productivity: the case of Greece. *Urban Studies 25*, 433-438.

Moomaw, R.L. (1981). Productivity and city size: a review of the evidence. *Quarterly Journal of Economics 96*, 675-688.

Moomaw, R.L. (1983). Spatial productivity variations in manufacturing: a critical survey of cross sectional analyses. *International Regional Science Review 8*, 1-22.

Moomaw, R.L. (1985). Firm location and city size: reduced productivity advantages as a factor in the decline of manufacturing in urban areas. *Journal of Urban Economics 17*, 73-89.

Nakamura, R. (1985). Agglomeration economies in urban manufacturing industries: a case of Japanese cities. *Journal of Urban Economics 17*, 108-124.

Rice, P., A. Venables, and E. Patacchini (2006). Spatial determinants of productivity: analysis for the regions of Great Britain. *Regional Science and Urban Economics 36*, 727-752.

Rosehthal, S. and W. Strange (2004). *Evidence on the nature and sources of agglomeration economies*, Chapter in Henderson JV and Thisse JF (eds) Handbook of Regional and Urban Economics, Volume 4. Amsterdam: Elsevier.

Shefer, D. (1973). Localization economies in SMSA's: a production function analysis. *Journal of Regional Science 13*, 55-64.

Sveikauskas, L. (1975). The productivity of cities. *Quarterly Journal of Economics 89*, 392-413.

Sveikauskas, L., J. Gowdy, and M. Funk (1988). Urban productivity: city size or industry size. *Journal of Regional Science 28*, 185-202.

Tabuchi, T. (1986). Urban agglomeration, capital augmenting technology, and labour market equilibrium. *Journal of Urban Economics 20*, 211-228.

Venables, A. J. (2007). Evaluating urban transport improvements: cost-benefit analysis in the presence of agglomeration and income taxation. *Journal of Transport Economics and Policy 41* (2), 173-188.

APPENDIX 1: THE TRANSLOG PRODUCTION INVERSE INPUT DEMAND MODEL

The generalised translog production function system based on an inverse input demand framework was first proposed by Kim (1992) and has been applied in full for the estimation of agglomeration economies by Graham and Kim (2007).

Let

$$Y = f(X,U) \tag{3}$$

be the production function for the firm where Y is the output level of the firm, X is a vector of factor inputs with elements X_i $(i = 1,\ldots,n)$, and U represents influences on production that arise from agglomeration economies.

If inputs are rented in competitive markets the first-order conditions for output maximisation subject to an expenditure constraint are

$$\frac{\partial Y}{\partial X_i} = \lambda W_i, \tag{4}$$

where W_i is the price of the ith input, and λ is a Lagrange multiplier which is the reciprocal of marginal cost $\partial C / \partial Y$. The expenditure constraint is given by,

$$\sum_i W_i X_i = C, \tag{5}$$

where C is total cost.

From (4) and (5)

$$\lambda = \frac{\sum_i \left(\partial Y / \partial X_i\right) X_i}{C} \tag{6}$$

and substituting (6) back into (4) after rearrangement yields the inverse input demand equations

$$\frac{W_i}{C} = \frac{\partial Y / \partial X_i}{\sum_i \left(\partial Y / \partial X_i\right) X_i} \equiv g_i\left(X,U\right) \tag{7}$$

Note that these inverse input demand functions determine prices as functions of quantities as opposed to ordinary demand functions which determine quantities in terms of prices. Equation (7) can be written in cost share form s_i^c as

$$S_i^C = \frac{W_i X_i}{C} = \frac{\partial \log Y / \partial \log X_i}{\sum_i \partial \log Y / \partial \log X_i} \tag{8}$$

Taking a translog approximation to equation (3) we have

$$\log Y = \alpha_0 + \sum_{i=1}^{i} \alpha_i \log X_i + \gamma_U \log U + \frac{1}{2}\sum_{i=1}^{i}\sum_{j=1}^{i} \gamma_{ij} \log X_i \log X_j + \sum_{i} \gamma_{iU} \log X_i \log U + \frac{1}{2}\gamma_{UU} (\log U)^2 \quad (9)$$

where $\gamma_{ij} = \gamma_{ji}$ $(i \neq j)$, and given (8) appropriate differentiation of (9) yields the cost share equations

$$S_i^C = \frac{\alpha_i + \sum_j \gamma_{ij} \log X_j + \gamma_{iU} \log U}{\sum_i \alpha_i + \sum_i \sum_j \gamma_{ij} \log X_j + \sum_i \gamma_{iU} \log U} \quad (10)$$

The translog parameters can be efficiently estimated by simultaneously estimating (9) and (10) as a nonlinear multivariate regression system. Since the factor share equations sum to unity, however, estimation of the full system cannot be undertaken because the disturbance covariance matrix is singular and non-diagonal. The singularity problem is addressed by simultaneously estimating the primary translog production function and $n-1$ share equations.

Equations (9) and (10) are estimated as a non-linear SURE system using a two stage procedure which first estimate the error covariance matrix using non-linear least squares and then estimates the parameters that minimizes the generalized sum of the squares for the system as a whole. The random errors in each equation are assumed to be distributed independently of the regressors, have expected values of zero and constant variance. The SURE procedure also allows for cross-equation correlation between error terms and the error in each equation can have different variance.

The translog production-inverse demand system provides a generalisation over previous specifications allowing for a non-homothetic production technology in which returns to scale (RTS) and the elasticity of substitution vary with the level of production and with factor proportions. Homotheticity, homogeneity and linear homogeneity each represent restricted versions of the non-homothetic function. From equation (9) RTS are measured by

$$\sum_i \frac{\partial \log Y}{\partial \log X_i} = \sum_i \alpha_i + \sum_i \sum_j \gamma_{ij} \log X_j + \sum_i \gamma_{iU} \log U \quad (11)$$

Flexibility in RTS across the sample is particularly important for our purposes because we wish to distinguish between scale economies and the effects of agglomeration. To do this effectively we need to ensure as far as possible that the agglomeration term is really distinct; that it does not capture some residual RTS effect arising due to the inadequate fit of a restrictive production function.

An analytical framework for agglomeration economies

The system we outline above offers a comprehensive analytical framework for the analysis of agglomeration. The very general specification does not require particularly onerous assumptions about the impact of agglomeration (such as Hick's neutrality) and allows for the possibility of non-linear effects. We can distinguish three distinct but highly inter-related dimensions of agglomeration that can be identified using our model. First, there are the external returns to agglomeration that influence TFP and the productivity of individual factors and which we will term *productivity effects*. Second, there are *effects on factor prices* which arise as a result of the increase in the productivity of factors induced by agglomeration. Third, there are *effects on factor demands* which follow from the influence that agglomeration has on factor prices. Below we consider each dimension in turn.

Productivity effects

The aggregate productivity effect of agglomeration economies is captured by the elasticity of output with respect to agglomeration. Differentiating equation (9) with respect to U we have

$$\frac{\partial \log Y}{\partial \log U} = \gamma_U + \gamma_{UU} \log U + \sum_i \gamma_{iU} \log X_i \tag{12}$$

Equation (12) measures the total shift in output that arises from agglomeration and thus the effect of agglomeration on TFP. We can decompose this aggregate productivity effect in two parts. First, we have what we will call a *direct* agglomeration effect, which is independent of factor levels and which varies depending on the level of agglomeration ($\gamma_U + \gamma_{UU} \log U$). Note that the quadratic specification allows for non-linear agglomeration effects and thus for the kind of diminishing returns that might be predicted by theory. Second, we have a component that arises through the effect that agglomeration has on the productivity of factor inputs given the volume of factor inputs employed ($\sum_i \gamma_{iU} \log X_i$).

To determine the effect of agglomeration on the productivity of individual factors we use the output elasticities because these are the logarithmic marginal products of each input

$$\frac{\partial \log Y}{\partial \log X_i} = \alpha_i + \gamma_{ij} \log X_j + \sum_i \gamma_{iU} \log U. \tag{13}$$

If $\gamma_{iU} > 0$ then agglomeration is positively associated with the productivity of factor X_i, if $\gamma_{iU} < 0$ then agglomeration is negatively associated with the productivity of factor X_i. Agglomeration economies are Hick's neutral only if $\gamma_{iU} = 0$.

So in terms of productivity, our framework allows us to identify three types of impact from agglomeration externalities: an aggregate effect on TFP, a 'direct' effect that is independent of factor inputs, and the particular effects on the efficiency of each factor. We can therefore provide some distinction of total productivity effects. This is important because we might expect agglomeration to affect efficiency in different ways. The possibility of accommodating non-linear agglomeration effects, for instance, due to diminishing returns, is also advantageous.

Price effects

If there are productivity improvements from agglomeration then these should be capitalized in the prices of factor inputs and these price effects can also be identified within our framework. The inverse input demand equations (7) measure shadow prices or the marginal willingness to pay for inputs by firms at a pre-determined level of expenditure. In equilibrium the marginal willingness to pay for an input should be equal to price. Denoting W_i/C as $\hat{W}i$ and rewriting equation (10) in inverse input demand form we have

$$\hat{w}_i = \frac{\alpha_i + \sum_j \gamma_{ij} \log X_j + \gamma_{iU} \log U}{\left(\sum_i \alpha_i + \sum_i \sum_j \gamma_{ij} \log X_j + \sum_i \gamma_{iU} \log U \right) X_i} \tag{14}$$

Logarithmically differentiating (14) with respect to U gives

$$\frac{\partial \log \hat{w}_i}{\partial \log U} = \frac{\gamma_{iU}}{\left(\partial \log Y / \partial \log X_i \right)} - \frac{\sum_i \gamma_{iU}}{\sum_i \left(\partial \log Y / \partial \log X_i \right)} \tag{15}$$

Equation (15) identifies the effect of agglomeration on the willingness to pay for input *i*, or its price. This is expressed in terms of the impact of agglomeration on the productivity of factor *i* (γ_{iU}) given the contribution of this factor to total output ($\partial \log Y / \partial \log X_i$), less the impact of agglomeration on the productivity of all factors given the contribution of all factors to output change.

Effects on factor demands

If the price of factor inputs varies systematically with the level of agglomeration then we might expect to find an effect on factor demands. From (14) the inverse price elasticities of each factor are

$$\frac{\partial \log \hat{w}_i}{\partial \log X_i} = \frac{\gamma_{ii}}{\left(\partial \log Y / \partial \log X_i\right)} - \frac{\sum_j \gamma_{ij}}{\sum_i \left(\partial \log Y / \partial \log X_i\right)} - 1, \tag{16}$$

and so using (15) we can determine the effect of agglomeration on factor demands as follows

$$\frac{\partial \log X_i}{\partial \log U} = \left(\frac{\partial \log \hat{w}_i}{\partial \log X_i}\right)^{-1} \cdot \left(\frac{\partial \log \hat{w}_i}{\partial \log U}\right) \tag{17}$$

Equations (12) to (17) provide a useful comprehensive empirical framework to analyse agglomeration economies. They allow us to derive the effect of agglomeration on TFP, on the efficiency of each individual factor, on factor prices, and on factor demands.

TRANSPORT INFRASTRUCTURE INSIDE AND ACROSS URBAN REGIONS: MODELS AND ASSESSMENT METHODS

Börje JOHANSSON
Jönköping International Business School (JIBS)
Jönköping
and Centre of Excellence for Science and Innovation Studies
Royal Institute of Technology (KTH)
Stockholm
Sweden

SUMMARY

1. NETWORKS AND THE SPATIAL ORGANISATION OF ECONOMIES 122

1.1. Infrastructure networks and location patterns 122
1.2. Identifying infrastructure properties 123
1.3. Identifying infrastructure impacts on the economy and welfare 124
1.4. Outline of the presentation 124

2. TRANSPORT NETWORKS AND AGGLOMERATION ECONOMIES 125

2.1. Local and distant markets 125
2.2. Classifying distance sensitivity 125

3. TRANSPORT INFRASTRUCTURE AND NEW GROWTH THEORY 127

3.1. Endogenous growth and growth accounting 127
3.2. Assessing dissonant results 128
3.3. Productivity impacts of infrastructure measured by physical attributes 130

4. NETWORKS AND ACCESSIBILITY 132

4.1. Spatial organisation and accessibility 132
4.2. Job accessibility, random choice and commuting 133
4.3. Different ways to make use of accessibility measures 135

5. EMPIRICAL RESULTS FROM ACCESSIBILITY-BASED STUDIES 137

5.1. Commuting and the spatial organisation of an FUR 137
5.2. Sector development in cities and regions 139
5.3. FUR growth and interdependencies in the spatial organisation 140
5.4. Estimation of growth with a simultaneous equation system 142

6. CONCLUSIONS AND REMARKS 144

6.1. The issue of spatial organisational and geographical scale 144
6.2. Discussion of models in section 5 144

BIBLIOGRAPHY 146

Jönköping, September 2007

ABSTRACT

Infrastructure investment represents large capital values, whereas the benefits and other consequences are extended into the future. This makes methods to assess investment plans an important issue. This paper develops a framework in which infrastructure networks are interpreted as determinants of the spatial organisation of an economy, while the very same organisation is assumed to influence the growth of functional urban regions (FUR) and thereby the entire economy. The suggested framework is formulated so as to facilitate the modeling of agglomeration economies, and hence to separate intra-regional and interregional transport flows. A basic argument is that transport networks should preferably be described by their (physical) attributes, and several accessibility measures are presented as tools in this effort. This type of accessibility measures combine information about time distances between nodes in an FUR and the corresponding location pattern.

The attempts to estimate aggregate production functions and associated dual forms is assessed in view of the so-called new growth theory are discussed, and it is concluded that this approach has been more successful when cross-regional data are employed in combination with infrastructure measures that reflect attributes.

The discussion of macro approaches is followed by a detailed presentation of how accessibility measures can depict the spatial organisation of FURs and the urban areas inside an FUR. Such measures are candidates as explanatory variables in macro models, although the presentation concentrates on applications in commuting models, and sector growth models. In particular, the paper presents a model in which an individual urban area's accessibility to labour supply interact with the same area's accessibility to jobs, in the context of an FUR. Empirical results from Sweden are used to illustrate how the spatial organisation and its change is influenced by the inter-urban networks of urban areas in an FUR. It is also argued that the model is capable of depicting essential aspects of recent contributions to the economics of agglomeration.

1. NETWORKS AND THE SPATIAL ORGANISATION OF ECONOMIES

1.1. Infrastructure networks and location patterns

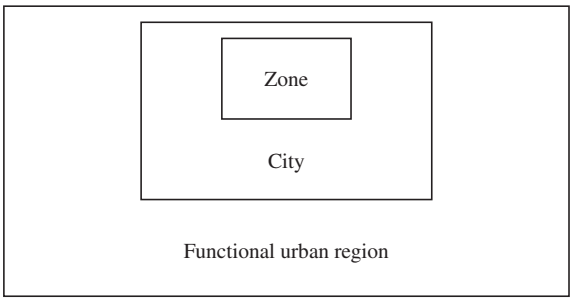
In the subsequent presentation transport services are divided into intra-regional (local) and extra-regional (inter-regional) flows, which result in displacements of goods, persons, and information (messages). Infrastructure networks enable and facilitate these movements. This statement implies that infrastructure consequences should reflect transport service opportunities, and our understanding of such opportunities depends on how we describe and measure the properties of infrastructure networks.

The purpose of this paper is threefold. The first task is to elucidate how transport infrastructure influences the spatial organisation of the economy, both at the regional level and the multi-regional, country-wide level. The second task is to examine – with the help of recent theory development – how the spatial organisation of an economy impacts the efficiency and growth of the economy. The third task is to suggest approaches to assess existing infrastructure and infrastructure changes on the basis of its impact on the economy.

In order to provide a scheme for analyzing and discussing spatial organisation, the study introduces concepts that recognize that urban areas are basic in an urbanized economic geography. The basic entity in the scheme is the functional urban region (FUR) or, with an alternative terminology, city region. The prefix "functional" indicates that all locations in an FUR share the same labour market as well as market for local supply of producer or business services. Typically, an FUR is composed of several cities and smaller urban-like settlements. When the region has one largest city, the region may be classified as a monocentric or rather one-polar region. Each city is finally decomposed into zones, which means that the spatial "entities" are ordered as illustrated in Figure 1.

With the above scheme, the economy-wide organisation of space is described by a system of FURs, often labelled city system (Henderson, 1982; Fujita and Thisse, 2002). The empirical observation that such a multi-regional system is hierarchical in identified in Christaller, 1933; Lösch, 1940; Tinbergen, 1967). In all essence, a system of cities extends beyond country borders, although each border between two countries represents a trade barrier that influences cross-boarder interaction and transport flows (Ottaviano, Tbuchi and Thisse, 2002).

Figure 1. **Spatial concepts for a Functional Urban Region**



The concepts introduced above and illustrated in Figure 1 can now be applied to formulate a consistent principle for studying spatial organisation. At the lowest level of spatial resolution we can observe the time distance and the associated transport cost between each pair of zones in a city, between each pair of cities in an FUR, and between each pair of FURs. These are all "link values" for nodes at the local, regional and interregional level. These link values are basic components of the decision information used by firms and households when they chose where to locate, and thus they will influence location patterns (spatial organisation). Moreover changes in a spatial transport system will affect the link values and thereby over time change the spatial organisation (Johansson and Klaesson, 2007).

1.2. Identifying infrastructure properties

The previous subsection identifies time distances between nodes or, more generally, link values reflecting generalized transport costs as a basic infrastructure property. This type of information has also been the most basic input to the established cost-benefit analysis (CBA) of infrastructure investments, which is focused on efficiency improvements. This approach has remained static in nature and focuses on marginal or piecemeal changes in transport opportunities. In Starret (1988) it is convincingly argued that CBA methods were designed to accurately solve this type of assessment problems.

In particular, welfare assessment of CBA type have especially been applied in the evaluation of investments in specific links, although there are interesting examples of approaches were changes occur in a network context (e.g. Mattsson, 1984). In a true network-based analysis the interaction flows are so-called activity based, and when this is the case, the infrastructure properties are identified and described in a way that also has an interface with emerging theories of spatial economics such as new economic geography (Krugman, 1991), agglomeration economics (Fujita and Thisse, 2002), knowledge and innovation spillover economics (Karlsson and Manduchi, 2001), new growth theory (Roemer, 1990), and new trade theory (Helpman, 1984). All these emerging strands include elements of imperfect competition, scale economies and externalities. In most cases they also imply that changes in transport costs and other geographic transaction costs matter (Johansson and Karlsson, 2001), and thus spatial organisation matters for productivity and growth – for regions and for summations across regions.

Given the above discussion, let us tentatively accept the idea that infrastructure properties impact the spatial organisation, which in turn is assumed to affect productivity as well as productivity growth. How can we then identify infrastructure properties? With reference to Lakhsmanan and Andersson (2007a, 2007b), the following alternatives should be contemplated:

(i) The capital value of infrastructure objects and the sum of such values, where the capital values are included as production factors in models that apply production, cost and profit functions to determine the infrastructure impact on the economy.

(ii) Physical or tangible properties of infrastructure objects and of infrastructure networks. Such measures include a specification of time distances, capacity, comfort and transport costs. Capacity aspects are, e.g. road length and flow capacity.

(iii) Compound measures of physical and value properties of a network, such as connectivity and accessibility of nodes to other nodes. Accessibility measures, in particular, combine link properties and features of the nodes in the network, and this provides a way to describe interaction opportunities with a vector of accessibility measures. This approach is theoretically linked to activity-based transport flow models.

1.3. Identifying infrastructure impacts on the economy and welfare

Consider that the existing transport infrastructure influences the spatial organisation and economic growth of the economy in cities and FURs. This implies that the infrastructure impact on economic development may focus on different spatial scales such as

- Consequences in individual FURs
- Consequences in macro regions such as federal states in Germany and the USA
- Country-wide consequences

In a standard CBA approach the consequences emphasized are (i) time gains of different categories of users of the transport system, (i) reduced accident risks, reduced vehicle costs, other cost effects, including monetary value of environmental effects. A correct CBA should be based on the development network users over time, which implies that it should consider the impact of a changing spatial organisation associated with traffic system changes.

How can the effects of a changing spatial organisation of the economy be categorized? Aggregate approaches that apply production functions and dual forms such as profit and cost functions consider changes in output, productivity, and cost level. Production functions may be specified for the entire economy or for separate sectors, and they may refer to FURs, macro regions and an entire country. The idea is that an aggregate function is able to summarise micro-level effects.

In contradistinction to the production-function approach, the messages from recent developments in agglomeration economics, innovation economics and new economic geography imply that the analysis has consider the spatial organisation in a more direct way, may it by at the level of city zones, cities or FURs. The idea then is that infrastructure properties affect phenomena such as firms' (i) labour markets, (ii) intermediary input markets, (iii) customer markets, and (iv) interaction with other firms and knowledge providers in their development activities, including R&D. These phenomena may be reflected by firms' accessibility to labour supply, to input suppliers, to customers, and to knowledge providers. As the accessibility to input suppliers grows, increased diversity is assumed to cause augmented productivity, and as accessibility to customers improve, firms can better exploit scale economies. Changing perspective, there is also households' accessibility to job opportunities, to supply of household services etc. The log sum of such accessibility measures may be used as welfare indicators (e.g. Mattsson, 1984).

1.4 Outline of the presentation

Section 2 outlines a framework for understanding intra-regional and extra-regional transport networks by distinguishing between local and distant markets and by classifying time distances. This forms a reference to agglomeration economies. Section 3 utilizes the framework to assess macro models that focus on the productivity impact of transport infrastructure. Section 4 presents a method to depict a region's spatial organisation by means of infrastructure measures. This method is shown to be compatible with random choice models in trip-making models and similar transport models. Section 5 presents a set of econometric exercises with Swedish data to model and predict (i) commuter flows inside and between urban areas, (ii) growth of jobs and industries in urban areas and FURs, and (iii) interdependent evolution of labour supply and jobs in urban areas as well as for entire FURs. Section 6 concludes and suggests new directions of future research.

2. TRANSPORT NETWORKS AND AGGLOMERATION ECONOMIES

2.1. Local and distant markets

Still in the 1970s analyses of regional economic growth relied on the so-called export-base model, according to which a region's economy is stimulated to expand as demand from the rest of the world increases (Armstrong and Taylor, 1978). The model than predicts that service production grows in response to augmented income in the region. Already in the 1950s this perspective was modified by inter-regional input-output analyses, in models that combine intra-regional and inter-regional deliveries of goods and services (e.g. Isard, 1960).

From the beginning of the 1980s the perspective on economic growth changes in many fields of economics. New macroeconomic growth models are developed to emphasize other factors than labour and capital, and to model the growth as an endogenous process (e.g. Romer, 1986; Barro and Sala-i-Martin, 1995). These and related efforts form a background to models where public capital and infrastructure capital are included as explanatory factors in aggregate (macro) production functions. The increased focus on such phenomena also influenced the development of regional growth modeling and empirical studies.

A prime novelty in this avenue of research was the clear ambition to model economies of scale in theory-consistent way. In this atmosphere, the New Economic Geography (NEG) is developed, with models that make a clear distinction between local deliveries to customers inside a region and customers outside a region (e.g. Krugman, 1990, 1991). Other contributions emphasized agglomeration economies as a productivity and growth enhancing aspect of urban economic life (e.g. Hendersson, 1981; Fujita, 1986; Fujita and Thisse, 2002). Still another route of research focused on the innovativeness of regions, referring to the so-called Jacobs hypothesis about the role of urban diversity (Jacobs, 1969, 1984; Feldman and Audretsch, 1999). In essence, these various contributions clarify that urban economic life is distinctly different from inter-urban exchange processes, and they stress that size of urban regions matter.

Some of the conclusions drawn from the described theory development are summarized in Table 1, which attempts to shed light on the separation of intra-regional and extra-regional interaction and transaction. In the intra-regional context, distance-sensitive exchange and deliveries are a key feature, and require intra-urban contact networks. In contradistinction, interregional interaction and transaction is a matter for goods and service-like deliveries that have a low distance sensitivity and which may be packed and distributed in large bundles. The corresponding infrastructure networks have other features and efficiency conditions than intra-regional face-to-face (FTF) oriented interaction.

2.2. Classifying distance sensitivity

In the subsequent presentation we consider a geography with the following structure. The basic unit is a functional region, with few exceptions a functional urban region, i.e. an FUR, which usually encompasses several cities of different size. In this sense an FUR is multicentric. However, with few exceptions one city is the largest, and the FUR is thus a one-polar region. For each city we will consider a set of zones and a set of links which make the city as well as the region as a whole a network of transport links and activity nodes, hosting residential buildings and firm premises.

Table 1. **The role of local and distant markets in economic development**

Intra-regional market phenomena	Extra-regional market phenomena
Self-supporting production	Production for extra-regional demand
Local markets which allow frequent FTF-contacts between buyers and sellers	Distant markets with mediated contacts between buyers and sellers and schedules delivery systems
Local competition	Global competition
Infrastructure is designed to create local accessibility	Infrastructure is designed to establish accessibility in global networks
Low intra-regional transaction costs stimulate development	Low extra-regional transaction costs stimulate development
Economic growth is driven by population growth and regional enlargement	Economic growth drives population growth
Endogenous, self-generated economic growth	Exogenous demand and self-generated productivity improvements stimulate economic growth
Diversity and welfare depend on the size of the region	Diversity can stimulate productivity growth and export expansion

Consider two zones (nodes in urban areas), labelled k and l, and let the time distance on the link (k, l) be t_{kl}. Such link distances may be associated with several alternative transport modes, and then we could specify mode-specific time distances for each link. For the moment we shall only consider one time distance value for each link. Before proceeding, it should be stressed that the importance to a city of a link (k, l) depends on characteristics of node k and node l, such as the number of node inhabitants, the number of jobs, the size and diversity of service supply for household and for firms.

Referring to Swedish data, which according to the literature seem fairly representative, time distances can be divided into local (intra-city), regional (intra-regional) and interregional (extra-regional) as specified in Table 2

Table 2. **Classification of time distances between zones**

	Time interval in minutes	Average travel time in minutes
Between zones in the same city (local)	0–15	8–12
From a zone in a city to zones in other part of the FUR (regional)	15–50	25–35
From a city in a FUR to a city in another FUR (inter-regional)	More than 60	More than 60

From Table 2 we can make several observations. The first has to do with sparsely populated land between cities and hence also such land between FURs. If a country's area is divided into exhaustive and mutually exclusive FUR areas, some parts of the geography will not match the time specifications in the table. However, from a transport point of view flows on links to such places are so thin (or infrequent) that they statistically will have close to measure zero, and hence can be disregarded for all practical purposes.

The second observation is that the separation between intra-regional and extra-regional links has an empty interval, from 50 to 60 minutes. Again, that reflects that FURs or city regions normally are sufficiently far away from each other to be divided by "empty land", just as mentioned above.

As a third observation we note that Table 2 provides an implicit definition of an FUR. It is a functional area, for which the time distance between any (or most) pairs of zones is shorter than 50 minutes. From this point of view an FUR allows firms and households to have frequent contacts with suppliers of household and producer or business services. In this way the city region is also an arena for knowledge interaction and diffusion. Moreover, the FUR can be an integrated labour market area. In addition, each city itself is an arena for very frequent face-to-face interaction, although it is only the largest cities in region that host sufficiently many actors to offer frequent interaction opportunities.

There is a final aspect of Table 2 that should be discussed. The model suggestions in sections 4 and 5 are twofold. First, as time distances are reduced an increasing share of all deliveries are not planned or scheduled in advance, but can take place on short notice. For long time distances the opposite holds and they will therefore generally be associated with more logistic-systems arrangements, like large shipments, multipurpose trips, supply-chain optimisation and the like. Second, theoretical development of the economics of agglomeration tells us that activities with frequent interaction have an incentive to cluster in the neighbourhood of each other.

We may also remark that a measure of time distances incorporate both economies and diseconomies of density. When an urban area becomes too dense of activities and interaction, congestion phenomena emerge and time distances will rise. New infrastructure networks may again remedy this type of development.

3. TRANSPORT INFRASTRUCTURE AND NEW GROWTH THEORY

3.1. Endogenous growth and growth accounting

Transport infrastructure affects options to interact inside and between regions, and in this way it influences economic efficiency. We may then ask: does improved efficiency imply anything about regional economic growth? In a strict neoclassical setting, there is no direct link between efficiency and growth. A step towards a link between infrastructure and economic growth is present in Mera (1973), where public infrastructure influences productivity. During the 1980s we can identify a sequence of studies applying national and regional production and cost functions, where infrastructure is a factor of production (e.g. Wigren, 1984; Elhance and Lakshmanan, 1988; Deno, 1988). The discussion of the productivity impact of infrastructure was strongly intensified by several papers by Aschauer (1989, 2000).

The attempts to model and estimate the role of transport infrastructure may be classified into two avenues. Along the first, transport infrastructure is represented by capital value, as one form of public capital, and thus relates to the general question: Is public capital productive? Two typical studies of this kind can be found in Aschauer (1989) with an aggregate production-function model of the US economy, and in Aschauer (2000) with an aggregate model specified for a set of macro regions.

The second avenue is to measure transport infrastructure in terms of its "physical" attributes, an approach that primarily is applied to regional cross-sectional or panel data from at set of regions. With this approach transport infrastructure may be represented by a variable like highway density or degree of agglomeration (e.g. Moomaw and Williams, 1992; Carlino and Voith, 1992).

The two approaches to assess the productivity and growth effects of infrastructure capital differ in a fundamental way. Infrastructure capital is a one-dimensional measure, and such a measure should be expected to fail when applied to different investments or different regions. A kilometre highway that solves exactly the same way in two different regions should have the same effect in both regions. However, if it is much more

expensive to construct the road in the first region, the capital value would be higher in this region and, as a consequence; the output elasticity of a kilometre highway would differ between the regions. With a physical measure this problem disappears.

A similar issue is the option to describe transport-infrastructure capital with a vector instead of a single value, where each component refers to a specific type of transport capital, such as road, rail, air terminals, etc.

If capital values are used for large regions or for an entire country, the above problems could be expected to disappear with the help of the law of large numbers. We may also observe the following pattern:

- Time series econometrics for countries tend to use an aggregate capital value of transport infrastructure
- Cross-regional and panel-data econometrics tend to use physical attributes of transport infrastructure

What is then the theoretical framework of the aggregate country-wide analyses of the role of transport infrastructure in economic growth? From one point of view they adhere to the idea behind endogenous growth. However, the studies contain very little of explicit references to endogenous growth, although the approach most likely would benefit from examining such model formulations. For one thing, infrastructure capital is to some extent public in the same way as knowledge in the core model of endogenous growth. In spite of this the studies referred primarily have the form of growth accounting.

Another issue in these studies is the choice of estimating a production function or a cost function for the economy or for a set of different sectors. Examples of studies using a cost-function approach are Seitz (1993) and Nadiri and Mamanueas (1991, 1996). What are then the advantages of a cost-function (or profit function) approach? In brief, a cost function estimation has direct support from microeconomic theory, because

- The estimation is based on optimization assumptions
- Duality conditions such as Shephard's lemma allows for controlled conclusions
- The approach makes it possible to distinguish between variable and fixed costs
- The approach makes it possible to consider scale economies
- The estimation considers how both supply and demand adjustments influence productivity growth
- The approach comprise not only capital and labour inputs, but also intermediary inputs

3.2. Assessing dissonant results

In Lakshmanan and Anderson (2007) it is observed that the whole range of studies that examine the productivity of infrastructure have generated quite dissonant estimates of output and cost elasticities. These results differ sharply for the same country, for countries at comparable stages of development and for countries at different stages of development. In view of this they pose the question: Is macroeconomic modeling of transport infrastructure unable to incorporate key transport-economy linkages? In this context they point at several problems such as (i) the network character of roads and other transport modes, (ii) threshold phenomena in transport development, (iii) the state of the pre-existing transport network, (iv) the state of development in regions undergoing transport improvements, (v) the structure of markets in regions, (vi) the presence of spatial agglomeration economies, and (vii) the potential for innovation economies.

Lakshmanan and Andersson (2007) discuss what they call the traditional view that transport infrastructure contributes to economic growth and productivity. In this discussion they emphasize that a set of recent methodologically sophisticated studies produce markedly dissonant estimates of the productivity of transport infrastructure, where the return to transport capital varies in a disturbing way. In the subsequent presentation

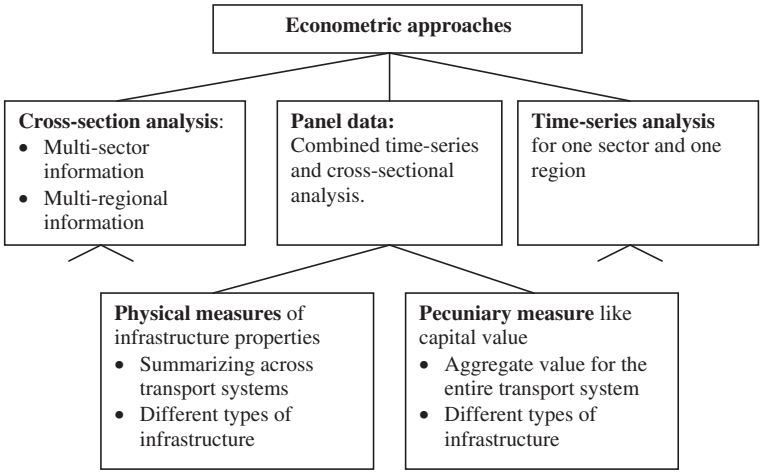
it claimed that one reason for the lack of consistency between those empirical efforts is related to how transport infrastructure is identified and measured.

The measurement and definition of transport infrastructure involves a set of partly interrelated choices such as

- A compound measure of transport infrastructure versus a vector specifying different types of infrastructure
- National versus regional specifications of available transport infrastructure
- The capital value of transport infrastructure versus physical and systems properties of the infrastructure

The options above can be included in alternative econometric approaches. For example, some studies apply cross-section analysis, whereas others employ time-series analyses. Moreover, the cross-section choice comprises the option to distinguish between industries (sectors of the economy), as well as multi-regional information. Combining these different observations the options of econometric approaches can be specified as illustrated in Figure 2.

Figure 2. **Overview of approaches to estimate infrastructure productivity**



Time-series, cross-section and panel data analysis all allow a choice between measuring (i) physical attributes and (ii) pecuniary values of infrastructure. The initial studies of infrastructure productivity were applying time-series analysis with an aggregate capital value and using GDP as the dependent variable. Naturally, the result from such studies can only be useful decision support for macroeconomic problems, such as the typical Aschauer questions: Is public expenditure productive or Do states optimize? Estimated elasticities are not useful for individual investment decisions for the following reasons:

- Two different highway projects that generate the same "amount of transport services" may differ in cost by factor 2 or 3. Thus, capital values are not correlated with the amount of service. This argument is weakened in the aggregate due to the law of large numbers.
- If the value of different types of infrastructure like roads, railway and air terminal are aggregated together results will be ambiguous. Instead a vector with capital value components referring to different types of infrastructure may reveal system composition or substitution effects. Again, the acquired result will be relevant only for an "average investment project".

The second alternative is labeled "physical measure". Obviously, such measures must be collected at a disaggregate level, with information from regions, in particular FUR-level data. However, first we have to clarify what is meant by "physical attributes". For roads, one may use variables such as (i) kilometer highway per regional area, (ii) flow capacity per regional area, (iii) time distances to neighbouring metropolitan regions, (iv) time distances to terminals for international freight. Measures of this kind can be applied in regional system models as demonstrated by followers of Mera, such as Sasaki, Kunihisa and Sugiyama (1995), and Kobayashi and Okumura (1997). They estimate relations between production, transport deliveries for regions and apply these estimates in multi-regional models with consistency constraints for each region and for the multi-regional system as a whole. This type of model is then used as a means to predict effects on the system of changes in the infrastructure in one or several regions. This approach has a clear interface with so-called activity-based models for transport forecast, and it captures the fact that regional context matters.

Another way to reflect physical attributes of a transport infrastructure is to calculate how it affects time distances between nodes in a transport network. Improved road and railroad infrastructure quality can reduce such time distances. This measure will also indirectly reflect capacity, since insufficient capacity will cause time delays (congestion) and thereby reduce speed, which implies that time distances increase. In Section 4 we will demonstrate how time distance information about transport infrastructure can be combined with information about activities in the nodes of the infrastructure network to yield purposeful characterization of infrastructure networks. Information about time distances and activity location is combined into accessibility measures.

3.3. Productivity impacts of infrastructure measured by physical attributes

Spatially aggregated models are not designed to reflect how and why transport infrastructure can have different effects on productivity in different regions. The impact of additional infrastructure may be weaker in a region which is already infrastructure affluent than in other regions with less developed infrastructure. The effects could also be greater in dense metropolitan regions than elsewhere. However, we also know that when a smaller region gets shorter time distances to a larger region, then the income may increase for the smaller region. In addition, such regional integration implies that the larger region increases its market potential, which should imply higher productivity in view of models of agglomeration economies and new economic geography (NEG).

A major conclusion is that aggregate models provide information that is macro relevant, by estimating effects which reflect consequences attributed to "an average bundle of infrastructure objects or to an average infrastructure investment project. The meaningfulness of estimates has to rely on the law of large numbers. The same conclusion applies with regard to using GDP as dependent variable contra using sector-specific output values.

The way to avoid the problems addressed is two-fold. First and foremost, when the infrastructure is recorded in terms its attributes instead of capital values, then econometric exercises will reflect effects that have an interface with effects that are included in orthodox CBA evaluations. Second, cross-regional observations give rise to enough variation for more reliable results that are also open for more insightful interpretations. A third possibility is to employ panel data.

With the suggested approach one may consider three major issues:

- How does infrastructure attributes stimulate structural changes in the economy, with exit and entry of activities?
- Will a region's output rise or fall? How fast is the change of GRP?
- What happens with a regions productivity in terms of GRP or income per capita?

In Table 3 a set of regression results are presented. They are all based on cross-regional information. In addition, all studies – except the Merriman study – use information about infrastructure attributes. As a consequence we can observe that productivity impacts vary considerable between regions.

Table 4 contains studies that employ panel data, with regional specification for a sequence of dates or just for a start year and a final year. Three of the studies use infrastructure attributes as explanatory variables, and these may be considered as panel data variants of the studies in Table 3. All studies in the table report that productivity effects and rate of return to investments vary strongly between regions.

Table 4 presents examples of estimations that (i) use panel-data information and (ii) information about infrastructure attributes. An overall observation is that these estimations tend be robust with regard to variation in parameter values. Together with ordinary cross-regional studies they produce lower parameter values than aggregate production function specifications. This is the background to the conclusion that they are more reliable. Does this mean that they can replace CBA approaches? The conclusion that we will arrive at later is that they are rather complements than substitutes.

Table 3. **Regional productivity impacts from physical attributes of transport infrastructure in regional cross-section analysis**

Researcher	**Estimation results**
Andersson *et al*. (1990)	Large productivity effects which vary considerably between regions
Anderstig (1991)	The rate of return to an investment varies with regard to in which region the investment takes place, generating examples with both high and low returns
Wigren (1984, 1985)	Considerable productivity effects which vary in size between regions
Sasaki *et al.* (1995)	Considerable productivity effects that vary markedly between regions
Bergman (1996)	Productivity effects vary strongly between regions of different size. Considers both intra-regional and inter-regional infrastructure networks
Merriman (1990)*	Considerable effects

* The study by Merriman does not employ physical measures of infrastructure attributes.

Table 4. **Regional productivity impacts from transport infrastructure in panel data estimations, with physical infrastructure attributes in three cases**

Researcher	**Estimation results**
Carlino and Voith (1992)	Large productivity effects of (i) highway density and (ii) agglomeration level
Johansson (1993)	Rate of return to an investment varies across regions and hence attains both high and low values. Effects of both intra-regional and inter-regional infrastructure networks
Mera (1973a, 1973b)	Productivity effects differ considerably with regard to region of investment
Seitz (1995)*	Rate of return to an investment varies across urban regions and hence attains both high and low values.
McGuire (1995)	Clear productivity effects that vary between regions

4. NETWORKS AND ACCESSIBILITY

4.1. Spatial organisation and accessibility

The previous section observes that the production function (or cost function) approach capture urban and other density and collocation externalities in a crude and indirect way. At the same time we clarified that modern spatial economics modelling promotes such externalities as important and use them as necessary to explain the very existence of cities and city regions.

Section 2 introduces time distances between nodes (zones) in regions as an important aspect of a region's spatial organisation. In the present subsection we start with these distances and add information about activities in each node to get full picture of the spatial organisation. Referring back to Figure 1.1 we can observe that a one-polar FUR consists of a major city (central city) together with other neighbouring cities and urban areas, where "city" is included in the notion "urban area". Each city and urban area consists of zones, and the region's transport system is reflected by the time distances between al zones. Reducing the dimensionality of such a time-distance matrix, we can focus on the following "aggregate" time distances:

(i) *Intra-urban*: The average time distance between all nodes in a city (urban area), denoted by t_{kk} for urban area k.

(ii) *Intra-regional*: The average time distance between urban area k and l inside region R, denoted by t_{kl}, for $l \in R(k)$, where $R(k)$ is the set of urban areas that belong to the same FUR as k, except k itself.

(iii) *Extra-regional*: The average time distance between urban area k in region R and urban area l outside region R, denoted by t_{kl}, for $k \neq l$ and $l \in E(k)$, where $E(k)$ is the set of urban areas that do not belong to the same FUR as k.

Next, consider that we can collect information about the number of jobs in each urban area k, denoted by J_k. Then we can select another urban area s and make the following calculations for a household with residence in s:

$$\text{The accessibility to jobs in } s \text{ equals } T_{ss}^{J} = \exp\left\{-\lambda(t_{ss})t_{ss}J_s\right\} \tag{4.1a}$$

$$\text{The accessibility to jobs in } R(s) \text{ equals } T_{R(s)}^{J} = \sum\nolimits_{k\in R(s)} \exp\left\{-\lambda(t_{sk})t_{sk}J_s\right\} \tag{4.1b}$$

$$\text{The accessibility to jobs in } E(s) \text{ equals } T_{R(s)}^{J} = \sum\nolimits_{k\in R(s)} \exp\left\{-\lambda(t_{sk})t_{sk}J_s\right\} \tag{4.1c}$$

Two properties of the formulas in (4.1) need comments. The first is that the time-sensitivity parameter λ is modelled as a function of the actual time distance. The reason for this is that empirical studies with Swedish data strongly suggest that the time sensitivity for short, intermediate and long distances are different (Johansson, Klaesson and Olsson, 2002, 2003). The second observation is that the three accessibility measures are determined only by time distances and job location. One might argue that the value $\lambda = \lambda_{sk} = \lambda(t_{sk})$ should reflect generalized transport costs. As shown in the following subsection, an

estimation of λ will reflect time costs and other trip costs accurately if these two components both are proportional to time distance.

The basic message now is that the vector $\left[T_{ss}^{J}, T_{R(s)}^{J}, T_{E(s)}^{J}\right]$ provides us with one description of the spatial organisation of a region from the perspective of an urban area (city) *s* in region *R*. As we shall see, this is just one out of several such descriptions that will be suggested. Before any further step is taken, two basic changes in the spatial organisation will be illustrated. As the first type of change, consider that the number of jobs in urban area *k* increases from J_k to $J_k+\Delta J_k$. The resulting change in the accessibility to jobs on link (*s*, *k*) is calculated in (4.2a)

$$\Delta T_{sk}^{J} = \exp\left\{-\lambda_{sk} t_{sk}\right\} \Delta J_k \tag{4.2a}$$

The second type of change is generated by a change in the time distance t_{sk}. Suppose that the distance increases by Δt_{sk}. This will result in the following reduction of job accessibility on link (*s*, *k*):

$$\left[\exp\left\{-\lambda_{sk} t_{sk}\right\}(1-\exp\left\{-\lambda_{sk}\Delta t_{sk}\right\})\right] J_k \tag{4.2b}$$

Returning to formula (4.1), it should be observed that we can shift from the location variable J_s to a variable showing the labour supply from households in *s*, denoted by L_s, to a variable referring to the supply of business services, denotes by F_s, or to a variable informing about the supply of household services, denoted by H_s. Applying the technique in formula (4.1), this would allow us to characterize an urban area *s* in the following complementing ways:

A household's accessibility to jobs, depicted by the vector $T_s^J = \left[T_{ss}^{J}, T_{R(s)}^{J}, T_{E(s)}^{J}\right]$, and to household services, given by the vector

$$T_S^H = \left[T_{ss}^{H}, T_{R(s)}^{H}, T_{E(s)}^{H}\right]. \tag{4.3a}$$

A firm's accessibility to labour supply, depicted by the vector $T_s^L = \left[T_{ss}^{L}, T_{R(s)}^{L}, T_{E(s)}^{L}\right]$, and to business services, given by the vector

$$T_s^F = \left[T_{ss}^{F}, T_{R(s)}^{F}, T_{E(s)}^{F}\right]. \tag{4.3b}$$

The accessibility measures calculated in this way evidently reflect interaction and contact opportunities of households and firms, respectively. Classical references would be Lakshmanan and Hansen (1965, and Weibull (1976)). If we introduce a variable that can represent customer budgets in different locations, it is also possible to calculate sales and delivery opportunities.

The second requirement for the accessibility measures is that they should be compatible or consistent with models designed to predict trip making and transport flows. This issue is illustrated for labour-market commuting in the next subsection.

4.2. Job accessibility, random choice and commuting

Consider now a set of urban areas (cities and towns) belonging to the same FUR $k \in R$. For urban area *k*, L_k is the potential labour supply and $M_k \leq L_k$ is the realized labour supply at any point in time. This means that supply is recognized as all persons in place *k* who live there and have a job in the same place or somewhere else. For the same group of urban areas we can also identify the number of available jobs in each municipality *k*, denoted by J_k. Commuting from area *k* to area *l* is denoted by m_{kl} such that

$$\sum_l m_{kl} = M_k, \text{ and } \sum_k m_{kl} = J_l \tag{4.4}$$

In formula (4.4) we observe that intra-urban commuting is denoted by m_{kk}. For urban areas in the same FUR, it is expected that either m_{kl}/M_k or m_{kl}/J_l is large. In transport models the commuting on the link (k, l) may be explained by two factors. The first is the benefit an individual in k obtains by commuting to a certain location l, and this may be related to (i) a higher wage level and (ii) better job opportunities in l. The second factor is the generalized commuting costs on links between municipalities. Let us assume that individuals' commuting incentives can be described by a random utility function. For an individual living in k, the utility of working in l may be expressed as follows:

$$U_{kl} = a_l + b(w_l - w_k) - \gamma c_{kl} - \mu_{kl} t_{kl} + \varepsilon_{kl} \tag{4.5}$$

where a_l refers to attributes in l, $(w_l - w_k)$ is the difference between the wage in urban area k and l for those jobs that match the individual's qualifications, c_{kl} denotes the pecuniary commuting costs, whereas the parameters b and γ translate the pecuniary values to a common preference base. Moreover, t_{kl} denotes the time distance between k and l, μ_{kl} is a time-value parameter and ε_{kl} denotes the random influence of not observed factors. This formulation allows us to differentiate between categories of jobs and between types of labour supply. Moreover, we can consider that the time sensitivity may be different for different labour categories.

Suppose now that individuals maximize their preference functions as specified in (4.5). Suppose also that wage differentials are small and that the direct commuting costs, c_{kl}, are approximately proportional to time distances so that $c_{kl} = \mu_c t_{kl}$

Consider now that a_l in formula (4.5) represents an attraction factor of municipality l and that ε_{kl} is an extreme value distributed error term. Moreover, let $V_{kl} = U_{kl} - \varepsilon_{kl}$. If the error term in (4.5) is extreme-value distributed, we can derive the following probability of choosing the commuting link (k, l):

$$P_{kl} = \exp\{V_{kl}\} / \sum_s \exp\{V_{ks}\} \tag{4.6}$$

Thus, the probability of choosing a specific link is described by a logit model. Next, let us define the attraction factor a_l as $a_l = \ln J_l$, where J_l signifies the number of jobs in urban area l. The numerator in (4.6) represents the preference value of the labour market in municipality, and the denominator is the sum of such values. Hence, the probability of commuting on the link (k, l) is the normalized preference value. In this way one may view P_{kl} as a ratio between the potential utility on link (k, l) and the sum of such utility values, given by $\sum_s \exp\{V_{ks}\}$.

Let us now assume that $c_{kl} = \mu_c t_{kl}$ and that $(w_k - w_l) = 0$, which yields

$$T_{ks}^J = \exp\{-\gamma\mu_c t_{ks} - \mu_{kl} t_{kl}\} A_s = \exp\{-\lambda_{kl} t_{kl}\} A_s, \text{ pour } \lambda_{kl} = \gamma\mu_c + \mu_{kl} \tag{4.7}$$

which is the standard measure of job accessibility on a link (k, l) introduced in the preceding subsection. It provides an exact measure only if the assumption about equal wages is valid. We should also observe that the new time-sensitivity parameter $\lambda_{kl} = (\gamma\mu_c + \mu_{kl})$.

Given the exercises above, how do the accessibility measures relate to predictions of transport (commuter) flows? To see this, consider the expression in (4.6). From this we can predict the number of commuter trips between k and l as $m_{kl} = M_k \exp\{V_{kl}\} / \sum_s \exp\{V_{ks}\}$, where $\exp\{V_{kl}\} = T_{kl}^J$, and where the denominator is a normalizing factor, based on the sum of all link accessibilities originating from urban area k. Moreover, it is also possible to include other attractiveness factor in the specification of V_{ks}, that may distinguish between intra-urban, intra-regional and extra-regional flows as shown in Johansson, Klaesson and Olsson (2003). This approach provides empirical model results which reveal that the time sensitivity parameter (variable) $\lambda_{kl} = \lambda(t_{kl})$ is a non-linear function of t_{kl}, represented by three different values such that $\lambda_{kk} = \lambda_0$, $\lambda_{ks} = \lambda_1$ for $s \in R(s)$ and $\lambda_{ks} = \lambda_2$ for $s \in E(s)$, as presented in Table 5.

Table 5. **Nonlinear commuter response to time distance**

	Intra-urban commuting	Inter-regional commuting	Extra-regional
Time distance t_{kl}	0–15 minutes	15–50 minutes	More than 60 minutes
Time sensitivity λ_{kl}	λ_0 is very low	$\lambda_1 \approx 3.8\lambda_1$	$\lambda_2 \approx 2.1\lambda_0$
Additional destination preference	Strong preference for local commuting	Medium preference for regional commuting	No preference

Source: Johansson, Klaesson and Olsson (2003).

The properties presented in Table 5 refer to factors that can be included in an accessibility measure, and hence add to the possibility to reflect the spatial organisation of an FUR.

4.3. Different ways to make use of accessibility measures

The previous presentation attempts to illuminate how a region's spatial organisation can be revealed by means of accessibility measures. The presentation aims at making precise how these measures change (i) as firms (jobs) and households (labour supply) migrate into or out from urban areas in a region, and (ii) as time distances change inside each urban area and between different areas. However, it remains to discuss how accessibility measures can be employed in the assessment of transport infrastructure policies.

First, let us consider the set of accessibility measures presented in Table 6. The table presents an overview of alternative measures and the processes and consequences associated with each measure.

With information of the type illustrated in Table 6, it is possible to consider at least four areas where the accessibility measures can be applied. These areas will be treated under the following labels:

- *Prediction of flows*. Accessibility to household services could for example be used to predict shopping trips but also migration flows. In section 5 empirical results are provided for commuting to work. Obviously, such predictions should be an important subtask in CBA calculations.
- *Prediction of location patterns*. Accessibility to labour supply can be used predict changes in the number of jobs in different parts of an FUR, and thus in the entire FUR. In an analogous way, the size of labour supply may be predicted. Observe that if jobs and labour supply increases in a region, this should potentially generate additional agglomeration effects, with productivity implications. In addition, changes in job location can indeed be carried out for specific groups of sectors.
- *Prediction of economic change*. With similar approaches as for prediction of location patterns, the growth and decline in employment and output (value added) can be predicted for an FUR. Such predictions can also focus on groups of sectors like private services, business services, etc.
- *Prediction based on output elasticities*. In this case FUR-specific accessibility measures are used as indicators of the services provided by transport infrastructure. This opportunity would, for example, fit a cross-regional version of the cost-function and total factor-productivity analyses by Nadiri and Manueas (1996).

The fundamental idea behind the four suggestions above is that accessibility measures are assumed to reflect the potential services that a region's internal and external transport infrastructure networks afford. If they offer such services, as claimed in this presentation, then it should be possible to estimate relations for both level and growth specifications, based on cross-region and panel data specifications, respectively.

Moreover, accessibility measure should emerge as significant inputs in production function formulations. This latter opportunity may also take the form cost-function formulations of the type employed by Nadiri and Mamuneas (1996).

Table 6. **Overview of optional accessibility measures**

Types of Accessibility	Associated Processes and Consequences
Households' Accessibility to	
Jobs	In terms of dynamics, households are attracted to locate in places with high accessibility to jobs. In terms of efficiency, high accessibility implies better labour-market matching
Household services	Households are attracted to locate in towns and cities with high accessibility to services, as well as diversity of services
Wage sum in firms located in different areas	Households are attracted to locate in towns and cities with high accessibility to economic activities, that may reflect job diversity, higher than average wages and productivity.
Firms' Accessibility to	
Labour supply	In terms of dynamics, firms are attracted to places with high accessibility to labour supply. In terms of efficiency, high accessibility implies better labour-market matching
Knowledge intensive labour supply	Growing economic sectors are oriented towards knowledge-intensive advanced services. Accessibility to a matching labour supply attracts firms belonging to growth sectors
Wage sum of households residing in different areas	Reflects the size of market demand for firms supplying household services; with an expanding local market scale economies can be exploited and diversity can increase
Wage sum in firms located in different areas	Reflects the size of market demand for firms supplying business (producer) services; with an expanding local market scale economies can be exploited and diversity can increase

Let us first consider the fourth option, labeled output elasticity estimation. This may take the form of estimating production functions with cross-regional information with FURs as observation units. Since some FURs have limited size, this also implies a restrictive sector specification. It also implies that accessibility measures have to be calculated as averages for each FUR. Each FUR, signified by R, could then be represented by the following three-component:

$$\begin{aligned} T_R^I &= \sum_{s\in R} g_s T_{ss}, \sum_{s\in R} g_s = 1 \\ T_R^{II} &= \sum_{s\in R} h_s T_{R(s)}, \sum_{s\in R} h_s = 1 \\ T_R^{III} &= \sum_{s\in R} q_s T_{E(s)}, \sum_{s\in R} q_s = 1 \end{aligned} \tag{4.8}$$

where g_s, h_s and q_s are weighting factors. The T-variables in (4.8) as separate observations of a region's transport system, and each T-value could represent local, regional and extra-regional accessibility GRP (gross regional product or wage sum) of each FUR. Other options are accessibility to port capacity (Johansson, 1993), airport capacity Andersson, Anderstig and Hårsman, 1990) or knowledge resources (Andersson and Karlsson, 2005).

In order to predict location patterns one might consider cross-section estimations generating information about location pattern and accessibility structure in terms of levels. However, it may be more rewarding to consider estimations of change processes, such that the change of, for example, jobs in urban areas is regressed against the accessibility pattern for each urban area at the start year. This approach comes close to the so-called Carlino-Mills model (Mills and Carlino, 1989).

This paper argues that the spatial organisation of an FUR can be described by an FUR-vector $T_R = \left[T_R^I, T_R^{II}, T_R^{III} \right]$, and by vectors $\left[T_{ss}, T_{R(s)}, T_{E(s)}^I \right]$ for each urban area *s*. Are there any other structural aspect that adds important information about the spatial organisation? The empirical results in Section 5 indicate convincingly that additional insights can be gained by incorporating the Christaller conjecture about a hierarchical pattern, known as the central-place system (CPS) model. In view of this, the following arrangement of urban areas is suggested, with three groups of urban areas, labeled C1, C2, and C3, where

C1 = The central and largest city in each region

C2 = Other urban areas in large FURs (more than 100 000 inhabitants) (4.9)

C3 = Other urban areas in small FURs (less than 100 000 inhabitants)

It is possible to make use of the CPS idea by estimating model parameters in a separate regression equation (or equation system) for each category of urban areas, while still using measures of accessibility to all types of urban areas.

Consider now that there is a prediction of job changes for each urban area in an FUR. Then the total effect for the FUR is assumed to be the sum of the change in each of the individual areas. Suppose now that the total number of jobs has increased. Is this growth an addition to the entire economy, across different FURs. The suggestion here is that it is an addition, in line with models of agglomeration economies. Thus, the number of jobs is not governed by zero-sum game restrictions.

There are empirical observations supporting the conclusion above. First, when using accessibility measure to explain (or predict) growth in FURs, one can observe that not all FURs have a positive growth. Second, empirical observations for the last two decades in Sweden tell us the following:

- The labour productivity as reflected by the wage level is positively correlated with an urban area's accessibility to jobs, to labour supply and to the wage sum.
- The labour market participation rate is positively correlated with an urban area's accessibility to jobs, to labour supply and to the wage sum.

These two observations support the assumption that accessibility properties have productivity effects.

5. EMPIRICAL RESULTS FROM ACCESSIBILITY-BASED STUDIES

5.1. Commuting and the spatial organisation of an FUR

Commuting can be viewed as a by-product of the spatial organisation of an FUR, and reflecting forces which are striving to equilibrate the supply of labour and the demand for labour. Viewing the labour market in

this way makes it natural to relate commuting to spatially separate locations where supply in one location meets demand in another. In this context, the aim of this subsection is to illustrate how well measures of labour-market accessibility can depict the spatial organisation and the corresponding commuter flows. This accomplished by means of two equations with the following structure (Johansson, Klaesson and Olssson, 2002):

- Commuting into an urban area is positively affected by (i) the intra-urban accessibility to jobs, and by (ii) the area's accessibility to labour supply in neighbouring (surrounding) urban areas.
- Commuting out of an urban area is positively affected by (i) intra-urban accessibility to residents in the area who have a job somewhere, and by (ii) the area's accessibility to jobs in neighbouring urban areas.

Following the structure above we make use of two accessibility measures for in-commuting to area k, denoted by I_k. The first measure is the job accessibility in k, T^J_{kk}, and the second is $T^L_{RE} = T^L_{R(k)} + T^L_{E(k)}$, which summarizes the entire accessibility to labour supply outside the urban area. This yields the regression equation

$$I_k = \alpha + \beta T^J_{kk} + \gamma T^L_{RE} + \varepsilon_k \tag{5.1}$$

The results of the regressions for 1990 and 1998 are shown in Table 7. All slope coefficients are positive and highly significant. In particular, we observe that most of the variation is captured by the two-variable equation.

It should be observed that equation (5.1) provides a measure of each inter-urban flow, because γT^L_{RE} is the sum of link-specific elements like $\exp\{-\lambda_{kl} t_{kl}\} L_k$.

The regression equation for out-commuters is described in (5.2). The two explanatory variables are T^L_{kk} and T^J_{RE}, denoting intra-urban accessibility to residents in the area who have a job somewhere and the area's accessibility to jobs in neighbouring urban areas, respectively. The dependent variable, O_k, denotes the total out-commuting from urban area k.

$$O_k = \alpha + \beta T^L_{kk} + \gamma T^J_{RE} \tag{5.2}$$

where O_k denotes out-commuting from municipality k, T^L_{kk} denotes the internal accessibility to realized labour supply, and T^J_{RE} denotes the external accessibility to jobs outside the urban area k. The results of estimating equation (5.2) for 1990 and 1998 are displayed in Table 8. All slope coefficients are positive and highly significant, and a large portion of the variation is explained by the two independent accessibility variables.

Tables 7 and 8, taken together, illustrate the strong correspondence between the spatial organisation of an FUR and the commuter transport flows. It may also be remarked that the time distances employed in the regression of (5.1) and (5.2) refer to commuting by car, which is the overwhelmingly dominating mode in all FURs except in the Stockholm region, for which automobile commuting still dominates but with a considerable share of public-transport commuting.

Table 7. **Commuting into municipalities 1990 and 1998**

	1990	**1998**
Intercept, α	−3374.2 (−6.3)	−2780.6 (−6.3)
Internal accessibility to jobs, β	0.37 (41.6)	0.39 (49.1)
External accessibility to realized labour supply, γ	0.05 (5.4)	0.06 (6.5)
R^2 adj	0.94	0.95

Remark: t-values are shown within parentheses. Number of observations is 288.

Table 8. **Commuting out of municipalities 1990 and 1998**

	1990	1998
Intercept, α	−291.1 (−6.3)	−237.9 (−0.9)
Internal accessibility to realized labour supply, β	0.15 (41.6)	0.18 (28.2)
External accessibility to jobs, γ	0.07 (15.0)	0.07 (16.0)
R^2 adj	0.93	0.95

Remark: *t*-values are shown within parentheses. Number of observations is 288.

5.2. Sector development in cities and regions

The economics of agglomeration is a field of models that focus on activities that may cluster in space to allow for distance-sensitive interaction. As presented in Fujita and Thisse (2002) these types of interaction include information and knowledge exchange between firms, and distance-sensitive contacts between supplier and customers. In dynamic context, the economies of scale may also be related to the so-called home-market effect in models classified as new economic geography.

This subsection presents models which reflect agglomeration economies with each urban area's accessibility to the wage sum in the area itself and in other parts of the FUR, to which the area belongs. The wage sum corresponds to a large share of each city's and town's total value added

For each urban area *s*, the accessibility to wage sum is expressed by a vector $T_s^W = \left[T_{ss}^W, T_{R(s)}^W \right]$, which only consists of intra-urban and intra-regional accessibility. The reason for excluding the extra-regional accessibility has a statistically insignificant and minimal influence on the change processes examined here. The very straightforward approach is formulate a simple growth equation for jobs, supply of household services, and supply of business services as described in (5.3). The two service-supply variables are reflected by the number of jobs in the pertinent industries. Having said this, one could observe that job growth during the 1990s (and later) is primarily a growth of jobs in private services.

$$\begin{aligned} \Delta J_s &= \alpha_0 + \alpha_1 T_{ss}^W + \alpha_2 T_{R(s)}^W + \varepsilon_s \\ \Delta H_s &= \alpha_0 + \alpha_1 T_{ss}^W + \alpha_2 T_{R(s)}^W + \varepsilon_s \\ \Delta F_s &= \alpha_0 + \alpha_1 T_{ss}^W + \alpha_2 T_{R(s)}^W + \varepsilon_s \end{aligned} \tag{5.3}$$

where the first equation refers to change of all jobs, the second to the change persons employed in household service industries, and the third to the change of persons employed in business service industries.

Table 9 presents results for the central (largest) city in each FUR. Similar regressions for non-central cities and towns shows that a similar pattern but with somewhat lower R^2-values for C2-areas, and much lower

Table 9. **Growth 1993–2000 in central cities induced by the accessibility to wage sum 1993**

Change Process	α_0	α_1	α_3	R^2
ΔJ_s Growth of jobs	−684 (−4.4)	0.58 (8.7)	0.75 (2.1)	0.97
ΔH_s Growth of household service supply	−425 (−3.4)	0.42 (7.8)	0.64 (2.2)	0.97
ΔF_s Growth of business service supply	−939 (4.3)	0.95 (10.2)	1.31 (2.7)	0.98

for C3-areas. For the latter group, the regression results imply that the estimated equation has a low prediction accuracy. In all essence, this reflect that service supply in Sweden is concentrated in the central city of each FUR and that this also implies that overall employment growth is strongly associated with the central cities, because overall growth is driven by service sector growth. For the three metropolitan FURs, this observation is less accentuated, which means that metropolitan growth is shared between the C1-area and the C2-areas.

It should be stressed that the models in (5.3), the accessibility measures are given for the start year, and growth is the result of a given spatial organisation at the start year. This means that different FURs can be compared and assessed with regard to their inherent change qualities. How is then assessment of planned infrastructure investments and other network changes carried out? The strategy for this is simple. First the growth path (change process) with the given accessibility pattern is calculated. In a second step the growth path associated with a new accessibility structure is calculated. The consequence of infrastructure changes is then the differences between the first and the second growth path.

5.3. FUR growth and interdependencies in the spatial organisation

In this and the following subsection the focus is on the labour market, with reference to a model presented in Johansson and Klaesson (2007). This model assumes (i) that labour supply (households) is attracted to an urban area in response to the area's accessibility to jobs, and (ii) that jobs (firms) are attracted to an urban area in response to the area's accessibility to labour supply. Such change processes operate also when the infrastructure properties remain unchanged, but will change as accessibility patterns are altered. It should be emphasized that the two change processes introduced constitute "coupled dynamics", and the initial task is to clarify these dynamics.

The dynamics on the labour market will be specified for each urban area, m. The supply of labour is depicted by the accessibility vector $[T^L_{mm}, T^L_{R(m)}, T^L_{E(m)}]$ as specified in (4.3), and labeled urban area m's accessibility to labour supply. The demand for labour is reflected by the accessibility vector $[T^J_{mm}, T^J_{R(m)}, T^J_{E(m)}]$ as specified in (4.3), and labeled area m's accessibility to jobs. The two vectors are dynamically related through the two equations in (5.4a) and (5.4b), specified as follows for each of the three groups of municipalities, C1, C2 and C3, which are introduced earlier in (4.9):

$$\Delta J_m = f(T^L_{mm}, T^L_{R(m)}, T^L_{E(m)}) \tag{5.4a}$$

$$\Delta A_m = f(T^J_{mm}, T^J_{R(m)}, T^J_{E(m)}) \tag{5.4b}$$

Consider first the variable ΔJ_m in equation (5.4a) above, which shows how the number of jobs in urban area m are expected to change from time $t = 0$ to time $t = \tau$. Next, let J_m denote the number of jobs at date $t = 0$. Then we can define $J^*_m = J_m + \Delta J_m$. Once the J^*_m-value is given, it will affect the values in all vectors like $[T^J_{kk}, T^J_{R(k)}, T^J_{E(k)}]$. These new future values (at time τ) are denoted by $[T^{J^*}_{kk}, T^{J^*}_{R(k)}, T^{J^*}_{E(k)}]$ for each k. This shows that we could consider this latter vector as an attractor for the change of labour supply in municipality k between time $t = 0$ to time $t = \tau$, described by ΔL_k

Having reached this point we recognize that we can introduce L_k to denote the labour supply at time $t = 0$ and define $L^*_k = L_k + \Delta L_k$ for urban area k. The value L_k affects in principle all T^L-values for each urban area m at time τ, denoted by $[T^{L^*}_{mm}, T^{L^*}_{R(m)}, T^{L^*}_{E(m)}]$ for each m.

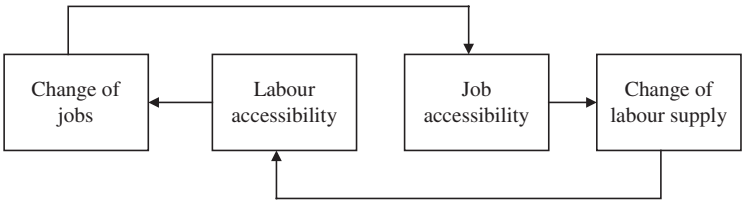
As described above the two equations (5.5a) and (5.5b) are now coupled such that we can write

$$\Delta J^*_m = f(T^{L^*}_{mm}, T^{L^*}_{R(m)}, T^{L^*}_{E(m)}) \tag{5.5a}$$

$$\Delta L^*_m = f(T^{J^*}_{mm}, T^{J^*}_{R(m)}, T^{J^*}_{E(m)}) \tag{5.5b}$$

where equation (5.5a) implies that the change of jobs in urban area m is influenced by the T^L-values for the same municipality at time $t = \tau$, i.e., the T^L-values that are a consequence of the change process during the time interval $(0, \tau)$. In a similar way equation (5.5b) describes how the future T^J-values for urban area m influence the change of labour supply, ΔL_m^*, during the time interval $(0, \tau)$. The coupled change processes in (5.5a)–(5.5b) are illustrated in Figure 3.

Figure 3. **Simultaneous change of labour supply and jobs**



The interpretation of the equation system in (5.5) is that the gradual change from $[T_{mm}^L, T_{R(m)}^L, T_{E(m)}^L]$ towards $[T_{mm}^{L^*}, T_{R(m)}^{L^*}, T_{E(m)}^{L^*}]$ is consistent with the simultaneous gradual change of labour supply towards $L_1^*,\ldots,L_N^*$ in a set of N urban areas. Moreover, the gradual change towards $[T_{kk}^{J^*}, T_{R(k)}^{J^*}, T_{E(k)}^{J^*}]$ is consistent with the simultaneous gradual change of jobs towards $J_1^*,\ldots,J_N^*$. For both processes in (5.5), the time distances in the transport network is assumed to be invariant and given at time $t = 0$.

Econometrically, the coupled change processes are represented by the two equations in (5.6). The two equations are determined in a simultaneous estimation procedure, where the implicit future accessibility values predict the future number of jobs and the future amount of labour supply:

$$\Delta L_m^* = \alpha_0 + \alpha_1 T_{mm}^{J^*} + \alpha_2 T_{R(m)} J^* + \alpha_3 T_{E(m)}^{J^*} + \varepsilon_m \tag{5.6a}$$

$$\Delta J_m^* = \beta_0 + \beta_1 T_{mm}^{L^*} + \beta_2 T_{R(m)}^{L^*} + \beta_3 T_{E(m)}^{L^*} + \varepsilon_m \tag{5.6b}$$

When the simultaneous system in (5.6) is regressed the variables on the right-hand side of the two equations may have either positive or negative parameter values. A positive parameter means that the associated variable is a positive attractor in the implicit change process. A negative value corresponds to a "negative attractor", i.e., a force of repulsion. Before the estimation results are presented, this matter is discussed in terms of a CPS-model.

The organisation of a multiregional system is analyzed in Christaller (1933) and Lösch (1940). Their contributions are known under the label central place system (CPS), a system description that is further developed in Beckmann (1958), Bos (1965) and Tinbergen (1967). In these models the geography is considered to be well-structured when it is organized in a hierarchical way, such that a large city is surrounded by smaller cities/settlements. In this way larger cities are separated from each other.

Applying the CPS model to the Swedish city system, we recognize that a functional region normally consists of a central city, which is embedded in a set of smaller towns. Moreover, the regions themselves can be grouped into (i) 58 small regions with less than 100 000 inhabitants, and (ii) 23 large urban regions of which three may be classified as metropolitan. These regions and their constituent urban areas are not always organized in a perfect CPS-structure. Instead, in some cases the central city (C1) in one region may be located fairly close to the central city of another region, and then the two regions will compete in a marked way for the same labour force. It is conjectured that this manifests itself as a negative effect from extra-regional accessibility for central cities.

For non-central urban areas, the situation may be different. In many cases a C2-area as well as a C3-area has small time distances to urban areas in other regions. In cases like this the individual urban area has an advantage of accessibility from both its own region and an "external region". These considerations motivate the decomposition of urban areas into category C1, C2 and C3.

5.4. Estimation of growth with a simultaneous equation system

Given the considerations above, we are ready to discuss the regression results. First the results associated with equation (5.5a) are presented in Table 7. The table presents a regression where the dependent variable is the change in labour supply in a sequence of five consecutive 8-year periods, with 1990–1998 as the first and 1994–2002 the last period. Each period is recognized by a time dummy.

By using a sequence of 8-years periods the number of observations gets larger, and the influence from short-term fluctuations is moderated when the regression refers to observations across a business cycle. This latter aspect is likely to be more important for changes in jobs than for changes in labour supply. The total number of jobs varies with the business cycle, whereas labour supply represents the number of persons in the age interval 20–64 years, and this number is affected by short-term fluctuations in a more modest way.

Consider first equation (1) where all urban areas are treated as one group. In this case α_1 and α_2 are positive and have large t-values. The extra-regional parameter α_3 is not significantly different from zero. Thus, overall there is no extra-regional influence. Turning to C1 areas, i.e., central cities in equation (2), the parameters α_1 and α_2 are positive and significant, whereas α_3 is negative and significant. This result is compatible with the idea of competition between central cities.

Equation (3) shows that C2-municipalities have a similar structure as the central municipalities, though with smaller parameter values and with the exception that α_3 is positive and significant. Thus, the C2-municipalities are positively affected by extra-regional accessibility, which is in line with our earlier remarks about Christaller-like patterns. Finally, equation (4) tells us that C3-areas are influenced primarily by intra-regional accessibility, and by a negative intercept, which is relatively large for these smaller and peripheral areas.

Table 10. **Change in labour supply in response to the accessibility to jobs. Regression parameters based on equation (5.5a)**

Municipality type		α_0	α_1	α_2	α_3	R^2
All	(1)	**−948.7**	**0.1307**	**0.002099**	−0.00145	0.91
		(−11.15)	(116.87)	(7.02)	(−0.46)	
C1	(2)	**−1170.7**	**0.1338**	**0.009504**	**−0.04311**	0.95
		(−5.79)	(64.48)	(3.08)	(−4.97)	
C2	(3)	**−662.9**	**0.0507**	**0.0039**	**0.01958**	0.42
		(−6.16)	(9.03)	(12.13)	(5.39)	
C.3	(4)	**−380.5**	**0.00575**	**0.00515**	−2.60E-05	0.11
		(−8.34)	(0.58)	(5.90)	(−0.01)	

Remark: Estimated for the change 1990–1998, 1991–1999, 1992–2000, 1993–2001, 1994–2002. Significant parameters in bold and t-values in parenthesis.

A general remark about the results in Table 7 is that the size of coefficients is larger in equation (2) than in equations (3) and (4), which indicates that the accessibility to jobs tends to affect the change of labour supply with greater force in central cities than in other urban areas.

The second part of the simultaneous equation system in (5.2) shows the relation between accessibility to labour supply and change of jobs in municipalities. The pertinent regression results are presented in Table 8, again based on data from a sequence of 8-year periods. We observe that the number of jobs in 1990 was extremely high and then fell dramatically for three years. In 1998 it had still not returned to the level from 1990 in most urban areas. Thus, one can argue that the period 1990–1998 is affected very strongly by business-cycle fluctuations, and this would provide arguments to employ the chosen approach with a series of 8-years periods.

Equations (1) and (2) in Table 8 provide us with a similar pattern, where the parameters β_1 and β_2 are positive and significant, whereas β_3 is negative, indicating competition with other regions. The competition effect is significant for the C1-cities. For C2-areas, the parameters β_2 and β_3 are positive and significant, indicating that growth of jobs in these municipalities is positively influenced by intra-regional and extra-regional accessibility to labour supply, whereas the local supply of labour has no clear effect. Finally, equation (4) shows that C3-areas have only one significant factor, namely intra-regional accessibility to labour supply.

A general remark about the findings in Table 8 is that the prediction power seems to be about the same for all three categories of urban areas. In other words, the development of jobs across municipalities follows a more volatile pattern than does the development of labour supply. In particular we note:

- It is only the central cities that benefit from their intra-urban labour supply.
- All urban areas benefit from the intra-regional labour supply. The central cities have a larger coefficient, but the rest of the region is always smaller for the central (largest) city than for its neighbours.

The estimated model presented above has been applied in assessment of combined road and railway investment programs. The method has then been to make a forecast without any infrastructure changes, followed by a new prediction with new accessibility conditions reflecting the investment program. To calculate the consequence of the investment program, the two predictions have been compared for all urban areas, and by adding up results for functional regions. The growth impact may be characterized as modest, but still high enough to match the investment costs.

Table 11. **Change in jobs in response to the accessibility to labour supply. Regression parameters based on equation (5.2b)**

Municipality type		β_0	β_1	β_2	β_3	R^2
All	(1)	101.1	**0.0733**	**0.00292**	−0.00463	0.34
		(0.41)	(22.23)	(3.70)	(−0.66)	
C1	(2)	1135.8	**0.0721**	**0.0223**	**−0.0585**	0.43
		(1.50)	(8.72)	(2.63)	(−2.24)	
C2	(3)	135.4	0.00320	**0.00437**	**0.01357**	0.36
		(0.90)	(0.46)	(9.87)	(3.49)	
C3	(4)	122.0	−0.01474	**0.00274**	−.00309	0.39
		(1.79)	(−1.22)	(2.66)	(−1.39)	

Remark: Estimated for the change 1990–1998, 1991–1999, 1992–2000, 1993–2001, 1994–2002. Significant parameters in bold.

6. CONCLUSIONS AND REMARKS

6.1. The issue of spatial organisational and geographical scale

Through the five sections of this paper there is a message, which may sound as a "mantra": the spatial organisation of an economy can be depicted by properly selected accessibility measures, and the spatial organisation of functional urban regions affects the productivity and the change of an economy. Since accessibility measures are based on time distances and other components of "generalized" transport or interaction costs, the approach suggested here presupposes detailed information of networks and the corresponding time distances between nodes of the network for each relevant transport mode. To keep such a data base updated is a demanding task. However, public investment in transport infrastructure has to be motivated by the accessibility it creates. There are two aspects of spatial organisation: the intra-FUR and the inter-FUR networks. In this presentation a large share of the attention has been directed towards the first of these two aspects. Such a focus can be supported by contributions to the economics of agglomeration, which indicates that large urban regions are key drivers of a country's economy and of the global economy. Of course this also implies that inter-FUR networks have an important role to play.

If one turns to the so-called new economic geography, the pertinent class of models focuses on interregional trade and transport. Thereby it also extends to the so-called new trade theory. In the contributions of Krugman (1990, 1992) there is a two-fold message. The first is that the cost of transport between regions is essential for the specialization and growth of regional economies. However, the second message is that the spatial organisation is governed by the difference between intra-regional and inter-regional transport costs.

Some transport links (and network associated links) have direct implications both for intra-regional and interregional accessibility. This is often the case for highways that are motivated by inter-regional interaction, but which at the same time strongly affect intra-regional accessibility. On the other hand, a high-speed train link may have much less intra-regional effects, although for some pairs of urban areas the time distance may shrink far below 45 minutes – thereby transforming an interregional link to a link with intra-regional properties.

Having reached this point, a question of tractability arises. Is it necessary to analyze intra-regional consequences of an infrastructure network separately from the inter-regional consequences? Furthermore, if the answer points in the direction of a yes, how can the two separate results be combined into an overall conclusion?

6.2. Discussion of models in section 5

The change processes modelled in sections 5.2–5.4 all employ a linear structure. The regression results in all cases include a significant and negative intercept. That should imply that the processes are non-linear, approximated by a linear equation. This conclusion has an important temporal implication: the approximation could only be considered valid for a limited time period. In the examples provided the time period is limited

to 8 years, and other experiments seem to indicate that the time period should not be extended much beyond 8–10 years. The conclusion drawn here are as follows:

- Less than 8 years is too short to reasonably depict very slow adjustment processes. In particular part of the mechanism improving accessibility is a spatial relocation process.
- On the other hand, if the time period is made much longer, the linear approximation becomes questionable. In practice this means that the intercept may stop to be valid for longer time spans.

The studies reported in the preceding sections and other supporting Swedish research efforts have been constrained by not having any complete time-distance matrices for periods before 1990. Thus, examination of non-linear, long-period models remains a task in the future.

In the paper a set of different accessibility measures have been presented. All these measures are – for a given urban area – related to the same networks and the same location pattern. As a consequence, the different measures are highly collinear. This means, for example, that an area with high accessibility to jobs also tends to have a high accessibility to labour supply. It also means that high accessibility household services imply high accessibility to jobs, since households services are executes by people who work in place where the services are supplied. This is not to say that the measures are identical, but it means that they have to be combined in a thoughtful way. One such example is provided in the two-equation model in sections 5.3 and 5.4.

Multi-colinearity extends even further. If one introduces accessibility to airport capacity to reflect inter-regional transport opportunities, such measures are also strongly correlated with intra-regional accessibility measures. Still, a major ambition should be to find ways to combine information about an urban areas intra-regional accessibility and its inter-regional accessibility (e.g. Hugosson, 2001).

BIBLIOGRAPHY

Andersson Å.E., Batten D.F. and Kobayashi K. (1993), Logistical Dynamics, Creativity and Infrastructure, in Å.E. Andersson, D.F. Batten, K. Kobayashi and K. Yoshikawa (eds). *The Cosmo-Creative Society – Logistical Networks in a Dynamic economy*, Berlin: Springer-Verlag.

Armstrong H. and Taylor, J. (1978) *Regional Economic Policy and Its Analysis*. Philip Allen, Oxford.

Aschauer D.A. (1989) Is Public Expenditure Productive? *Journal of Monetary Economics*. 23:177–200.

Aschauer D.A. (2000) Do States Optimize? Public Capital and Economic Growth, *Annals of Regional Science*, 34:343–364.

Barro R.J. and Sala-i-Martin X. (1995), *Economic Growth*, McGraw-Hill, New York.

Beckmann M.J. (1958). City Hierarchies and the Distribution pf City Sizes. *Economic Development and Cultural Change*, IV 3:343–348

Bos H. (1965). *Spatial Dispersion of Economic Activity*. Rotterdam: Rotterdam University Press.

Cheshire P.C. and Gordon I.R. (1998). Territorial Competition: Some Lessons for Policy. *Annals of Regional Science* 32:321–346.

Christaller W. (1933) *Die Zentralen Orte in Süddeutschland*, Gustav Fischer, Jena.

Feldman M.P. and Audretsch D.B. (1999), Innovations in Cities:Science-Based Diversity, Specilization and Localized Competition, *European Economic Review*, 43:409–429.

Forslund M. and Johansson B. (1995), Assessing Road Investments: Accessibility Changes, Cost Benefit and Production Effects, *Annals of Regional Science* 29:155–172

Fujita M. Krugman P. and Venables A.J. (1999). *The Spatial Economy – Cities, Regions and International Trade*, Cambridge Mass.: The MIT Press.

Fujita M. and Thisse J-F. (2002). *Economics of Agglomeration: Cities, Industrial Location and Regional Growth*. Cambridge: Cambridge University Press.

Helpman E. (1984) Increasing Returns, Imperfect Markets and Trade Theory. In R. Jones and P. Kenan (eds), *Handbook of International Economics*, vol. 1. North_Holland, Amsterdam, pp 325–365.

Henderson, J.V. (1982), Systems of Cities in Closed and Open Economies, *Regional Science and Urban Economics*, 12:325–350.

Hugosson P. (2001). Interregional Business Travel and the Economics of Business Interaction. *JIBS Dissertation Series No. 009*, Jönköping International Business School.

Isard, W. (1960), *Methods of Regional Analysis*, MIT Press, Cambridge MA.

Jacobs J. (1969), *The Economy of Cities*, Random House, New York.

Jacobs J. (1984), *Cities and the Wealth of Nations*, Random House, New York.

Johansson B (1993), Infrastructure, Accessibility and Economic Growth. *International Journal of Transport Economics* XX, 2:131–156.

Johansson B., Klaesson J. and Olsson M. (2002) Time Distances and Labour Market Integration. *Papers of Regional Science* 81:305–327

Johansson B., Klaesson J. and Olsson M. (2003). Commuters' Non-Linear Response to Time Distances. *Journal of Geographical Systems* 5:315–329.

Johansson B. and Klaesson J. (2007), Infrastructure, Labour Market Accessibility and Economic Development. In C. Karlsson, W. Anderson, B. Johansson and K. Kobayashi (eds), *The Management and Measurement of Infrastructure – Performance, efficiency and innovation*, Edward Elgar, Cheltenham.

Karlsson C. and Manduchi A. (2001), Knowledge Spillovers in a Spatial Context – A Critical Review and Assessment. In M.M. Fischer and J. Fröhlich (eds) *Knowledge, Complexity and Innovation Systems*. Springer, Berlin, pp 101–123.

Kobayashi K. and Okumura M. (1997), The Growth of City Systems with High-Speed Railway Systems, *Annals of Regional Science*, 31:39–56.

Kopp A. (2007). Aggregate Productivity Effects of Road Investment: A Reassessment for Western Europe. In C. Karlsson, W. Anderson, B. Johansson and K. Kobayashi (eds), *The Management and Measurement of Infrastructure – Performance, efficiency and innovation*, Edward Elgar, Cheltenham.

Krugman P. (1990), *Rethinking International Trade*, MIT Press, Cambridge, Mass.

Krugman P. (1991), *Geography and Trade*. Leuven University Press, Leuven, Belgium.

Lakshmanan T.R. and Hansen W.G. (1965). A retail market potential model. *Journal of the American Institute of Planners* 31:134-143. [An early empirical application of accessibility representation of spatial attractors]

Lakshmanan T.R. and Anderson W.P. (2007) Contextual Determinants of Transport Infrastructure Productivity: The Case for a New Modelling Strategy. In C. Karlsson, W. Anderson, B. Johansson and K. Kobayashi (eds), *The Management and Measurement of Infrastructure – Performance, efficiency and innovation*, Edward Elgar, Cheltenham.

Lakshmanan T.R. and Anderson W.P. (2007). Infrastructure Productivity: What are the Underlying Mechanisms? In C. Karlsson, W. Anderson, B. Johansson and K. Kobayashi (eds), *The Management and Measurement of Infrastructure – Performance, efficiency and innovation*, Edward Elgar, Cheltenham.

Lösch A. (1940) *Die Räumliche Ordnung der Wirtschaft*. Gustav Fischer, Jena.

Mattsson L-G (1984), Equivalence Between Welfare and entropy Approaches to Residential Location. Regional Science and Urban Economics 14:147–173.

Mera K. (1973) Regional Production Functions and Social Overhead Capital: An Analysis of the Japanese Case. *Regional and Urban Economics* 3:157–185.

Moomaw R. and Williams M. (1991) Total Factor Productivity in Manufacturing. Further Evidence from the States. *Journal of Regional Science* 31:17–34.

Nadiri M.I. and Mamanueas T.P. (1991) The Effects of Public Infrastructure and R&D Capital on the Coststructure and Performance of US Manufacturing Industries. *Working Paper No 3887*, NBER, Cambridge Mass.

Nadiri M.I. and Mamanueas T.P. (1996) Contribution of Highway Capital to Industry and National Productivity Growth. Federal Highway Authority Administration, Washington DC.

Romer P.M. (1986), Increasing Returns and Long-Run Growth, Journal of Political economy, 94:1002–1037

Romer P.M. (1990), Endogenous Technological Change. *Journal of Political Economy* 98:71–102.

Sasaki K., Kunihisa S. and Sugiyama. (1995), Evaluation of Road Capital and its Spatial Allocation, *Annals of Regional Science*, 29:143–154

Tinbergen J. (1967), The Hierarchy Model of the Size Distribution of Centres, *Papers of the Regional Science Association*, 20:65–80.

Tsukai M., Ejiri R., Kobayashi K. and Okukura M. (2007). Productivity of Infrastructure with Spillover Effects: A Study of Japan. In C. Karlsson, W. Anderson, B. Johansson and K. Kobayashi (eds), *The Management and Measurement of Infrastructure – Performance, efficiency and innovation*, Edward Elgar, Cheltenham.

Weibull J.W. (1976) An Axiomatic Approach to the Measurement of Accessibility. *Regional Science and Urban Economics* 6:357–379. [Examines the theoretical underpinnings for exponential and other distance decay assumptions.]

THE BROADER BENEFITS OF TRANSPORTATION INFRASTUCTURE

Ian SUE WING, William P. ANDERSON and T.R. LAKSHMANAN
Boston University
Center for Transportation Studies
Boston
United States

SUMMARY

1. INTRODUCTION 154

2. CONTEXT: THE BROADER ECONOMIC IMPACTS OF INFRASTRUCTURE INVESTMENT 155

3. CONVENTIONAL METHODS OF IMPACT ASSESSMENT 157

4. A REVIEW OF GENERAL EQUILIBRIUM ANALYSES OF CONGESTION 158

5. A HYBRID MESO-MACRO APPROACH 162

5.1. Algebraic summary of the CGE model 165
5.2. Data and calibration 169

6. DISCUSSION AND SUMMARY 169

7. APPENDIX: IMPLEMENTATIONAL DETAILS 171

7.1. Zero profit conditions and associated demand functions 171
7.2. Market clearance conditions 175
7.3. Income balance conditions and auxiliary variables 177
7.4. General equilibrium in complementarity format 177

NOTES 178

BIBLIOGRAPHY 179

Boston, September 2007

ABSTRACT

Assessments of the economic benefits of transportation infrastructure investments are critical to good policy decisions. At present, most such assessments are based of two types of studies: micro-scale studies in the form of cost-benefit analysis (CBA) and macro-scale studies in the form of national or regional econometric analysis. While the former type takes a partial equilibrium perspective and may therefore miss broader economic benefits, the latter type is too widely focused to provide much guidance concerning specific infrastructure projects or programs. Intermediate (meso-scale) analytical frameworks, which are both specific with respect to the infrastructure improvement in question and comprehensive in terms of the range of economic impacts they represent, are needed. This paper contributes to the development of meso-scale analysis via the specification of a computable general equilibrium (CGE) model that can assess the broad economic impact of improvements in transportation infrastructure networks. The model builds on recent CGE formulations that seek to capture the productivity penalty on firms and the utility penalty on households imposed by congestion (Meyers and Proost, 1997; Conrad, 1997) and others that model congestion via the device of explicit household time budgets (Parry and Bento, 2001, 2002). The centerpiece of our approach is a representation of the process through which markets for non-transport commodities and labor create derived demands for freight, shopping and commuting trips. Congestion, which arises due to a mismatch between the derived demand for trips and infrastructure capacity, is modeled as increased travel time along individual network links. Increased travel time impinges on the time budgets of households and reduces the ability of transportation service firms to provide trips using given levels of inputs. These effects translate into changes in productivity, labor supply, prices and income. A complete algebraic specification of the model is provided, along with details of implementation and a discussion of data resources needed for model calibration and application in policy analysis.

1. INTRODUCTION

Most contemporary assessments of the economic effects of transportation infrastructure investments fall into two major categories, one at the micro-scale and the other at the macro-scale. Micro-scale assessments follow the procedures of cost-benefit analysis (CBA). They use information on the likely outcomes of a proposed project – its effect on travel times, traffic flows, emissions, accidents, etc. – to estimate a pecuniary value of its lifetime benefit. That benefit estimate is then contrasted with lifetime project costs to determine whether it is economically productive. Such ex ante analyses are often required as a justification for devoting public funds to a proposed project. (For a review see Mackie and Nellthorp, 2001.)

Macro-scale studies include econometric analyses that relate the aggregate investment in (or stock of) transportation infrastructure to economy-wide measures of economic performance. For the most part, they specify production or cost functions in which public infrastructure is regarded as an input to production by private firms in a region or nation. The estimated production and cost functions provide evidence of the contribution that infrastructure investment makes towards augmenting the productivity of private firms and, in some cases, make it possible to calculate a rate of return on aggregate infrastructure investment. (For a review see Lakshmanan and Anderson, 2002.)

The two approaches are complementary. Micro-scale analyses have the advantage of being able to measure the impacts of adding or improving a specific infrastructure element, but the scope of their economic assessment is limited to effects on users of the element in question or closely related elements and to firms and individuals in its immediate locale. The macro-scale analyses capture a broader range of economic impacts, but they treat infrastructure investment as a homogenous good (measured in dollars or network miles.) and are therefore of little use for assessing the worth of specific investments. Further, the macro-scale approach sheds little, if any, light on the mechanisms that drive the observed economic impacts.

To provide a more complete picture of the economic impacts of infrastruture, an intermediate level of analysis is needed. For convenience, we refer to this level as "meso-scale," although models in this category might be applied at a variety of geographical scales. We define three requisites for models in this class.

1. Unlike macro-scale analyses they should incorporate information about specific additions or improvements to transportation infrastructure networks (although not necessarily at the level of detail found in micro-scale analyses.)
2. They should trace the economic processes that are triggered by infrastructure improvements. (As we will explain below, these may take the form of static general equilibrium effects or dynamic developmental effects.)
3. Finally, in order to assess the relative magnitude of different economic mechanisms and to inform policy, they should be amenable to empirical implementation using data that are either available or obtainable at reasonable cost.

As a contribution toward the development of meso-scale analyses, we introduce a computable general equilibrium (CGE) model that incorporates a number of novel mechanisms for tracing the effects of additions to the capacity of a transport network through the broader economy. Infrastructure investments are modeled as reducing travel times over links in a network. The key novelty is to incorporate travel time explicitly in the

utility and profit maximization problems of households and firms. For households, travel time for commuting and consumption activities enters a time budget that also includes time devoted to work and leisure. For firms providing transportation services, travel time affects the number of trips that can be provided by a given stock of vehicles, which in turn affects the prices of intermediate and final goods.

The remainder of the paper is organized as follows. Sections 2 and 3 discuss the broad economic impacts of transportation infrastructure and the state of the art in assessing those impacts. Section 4 reviews the relatively brief literature on assessing economic impacts of infrastructure investment with a CGE framework. Section 5 constitutes the meat of the paper, presenting an overview of our model, a complete algebraic specification, details of implementation and a discussion of data needs. Section 6 provides a discussion and summary.

2. CONTEXT: THE BROADER ECONOMIC IMPACTS OF INFRASTRUCTURE INVESTMENT

The role of transportation infrastructure in the economy is multifaceted and plays out over a long period of time. It is unlikely that any modeling framework can capture all possible mechanisms. For our purpose, it is useful to make a distinction between two classes of economic impacts, which we call static general equilibrium impacts and dynamic developmental impacts. Static general equilibrium impacts comprise a broad range of effects coursing through the economy consequent on the time and monetary savings induced by the infrastructure improvements. Such temporal and monetary savings alter, in turn, the marginal costs of transport producers, individuals' mobility and the demand for goods and services in the context of lowered congestion. As these changes ripple through the market mechanisms, endogenous changes occur in employment, output, and incomes. Dynamic developmental impacts ensue from the mechanisms set in motion when transport infrastructure improvements activate a variety of interacting processes that yield over time many sectoral, spatial, and regional effects which augment productivity. They produce transformations in the structure and pattern of the economy – such as changes in the spatial pattern of production; creation of new industries and inter-industry linkages; changes in the lifestyles and preferences of households; and the evolution of institutions and markets. While static general equilibrium impacts arise from the actions of a well-defined set of economic agents through the medium of markets, dynamic developmental impacts involve complex interactions of economic, social, cultural and institutional factors and are more idiosyncratic in nature. We therefore attempt to capture only the former category of impacts in the CGE model.

General equilibrium effects occur within a system of market relationships that is stable and relatively well understood. Most economic activities require some movement of goods and people. Production requires the movement of intermediate inputs to the production site, the movement of workers back and forth between their homes and places of employment (commuting) and the movement of finished goods to market. Consumption activities also require movement as in the case of household trips for shopping and recreation. To the extent that improvements to transportation infrastructure reduce the cost of movement of goods and people, they affect the levels of economic activity in all parts of the economy.

A number of general equilibrium mechanisms are described in detail below. But for the purpose of illustration, consider the effect of infrastructure on employment. Most transportation analyses start with the explicit or implicit assumption that the number of people who commute to work over a given network is fixed. In an economy like the US, however, labor supply is by no means perfectly inelastic because significant

segments of the population, such as mothers with children and individuals past the normal retirement age, face decisions as to whether to enter or remain in the labor force. Labor supply is normally associated with the wage, but since commuting represents a significant cost of labor force participation, infrastructure improvements could entice more people to work. Of course, in a general equilibrium framework this would represent a shift in the labor supply function, which in turn would affect the equilibrium wage and employment level.

A peculiar aspect of transportation infrastructure investments is that the cost reductions they generate are often realized in time savings rather than monetary savings. Returning to the commuting example, a road expansion that relieves congestion might have a minor effect on a commuter's out of pocket cost (e.g. lower fuel costs due to efficiency improvements that stem from changes in the driving cycle) but a major effect on commuting time. Time is a scarce resource for any potential worker, so less time spent commuting means more time is available for work, leisure, consumption activities, childcare, etc. Thus, in assessing the general equilibrium impacts of transportation infrastructure, the household time budget is as important as the household expenditure budget. Depending on the magnitude of the wage relative to the marginal utility of leisure, the impact of the decrease in commuters' time costs on the labor supply may be larger than that of their pecuniary savings.

A general equilibrium perspective on transportation infrastructure recognizes that reductions in the pecuniary and time costs of transportation can lead to increases in the levels of various economic activities and thereby to increased derived demand for transportation services. Thus, induced traffic flows are a natural outcome of market mechanisms. To many transportation analyses, such flows are seen as negating benefits from transportation infrastructure. A project whose congestion reduction effect disappears due to increased traffic within a few years of its implementation is seen as a failure. This point of view may be appropriate from an environmental perspective, where the goal of the project is to reduce emissions via improved traffic flow, but there are conceptual difficulties from a broad economic perspective. Induced trips are derived from increases in economic activities (labor supply, production, consumption, recreation) that lead to increased welfare, so as long as there are more trips there is presumably a benefit. This has an important implication: from a broader economic perspective, the benefits of a transportation infrastructure project cannot be assessed solely in terms of resultant travel time savings. This is especially true over the medium to long run, when the additional economic activity made possible by the expansion of infrastructure capital stock increases the derived demand for transportation to the point where it once again approaches the transportation network's capacity.

The fact that we do not try to capture dynamic developmental impacts in the CGE model is not meant to detract from their importance. Impacts of this type are most pronounced in low-income countries, where infrastructure improvements often represent significant and non-marginal enhancements of infrastructure capacity, which (along with the transport services they make possible) can facilitate interregional trade and integration. As infrastructure and service improvements lower money and time costs and increase accessibility to various market actors—input suppliers, workers and customers—market expansion, increased interregional integration and sustaining growth occurs over time. The underlying mechanisms include gains from trade, technology shifts, and gains from agglomeration supported by transport. A well-studied example of such developmental transformation is the experience of the U.S. Midwest consequent on a 400% expansion of the rail network between 1848 and 1860 – essentially linking the Midwest to Northeastern U.S. and the world economy. There is considerable evidence that the development of railroads accelerated the settlement, agricultural expansion, and growth and diversification of manufacturing, and initiated dynamic sequences that integrated the New England and Mid-Atlantic regions with the Midwest (Fogel, 1964, Fishlow 1965, Lakshmanan and Anderson 2007). A more recent example of such developmental effects of major road investments is discernable in Sri Lanka (Gunasekara, Anderson and Lakshmanan 2007 forthcoming). The broader literature on transport and economic development suggests that transport infrastructure facilitates the transformation of low-income economies from subsistence to commercial agriculture, the development of basic, transport-intensive industries and the growth of cites (Haynes and Button, 2001).

It would be a mistake to think that developmental impacts occur only at an early stage of economic development. Even in a mature economy, transportation infrastructure improvements might promote structural changes such as increased decentralization or agglomeration of economic activity; changes in the way business enterprises conduct operations such as inventory management, logistics and other practices; enhanced opportunities for face-to-face interaction; and a range of new recreational opportunities (Anderson and Lakshmanan, 2007.) These impacts, which affect the long-term evolution of the economy, are difficult to measure and even more difficult to predict. Nevertheless they are important, and a better understanding of developmental effects should lead to better decision-making on transportation infrastructure.

3. CONVENTIONAL METHODS OF IMPACT ASSESSMENT

As we have stated earlier, current methods of impact assessment include the micro-scale CBA and macro-scale econometric studies. CBA is nearly universal as a means of assessing the desirability of specific projects. Conceptually, economic benefits are assessed as the consumer surplus, defined in relation to the demand curve for the infrastructure facility in question. The effect of the infrastructure improvement is represented as a rightward shift in the infrastructure supply curve, which results in a fall in the price of using the facility—usually defined in units of time as opposed to money—for any given level of demand. The associated economic benefit thus has two components: one based on the cost savings enjoyed by the number of travelers who used the facility prior to the improvement, and a second representing the benefits to new travelers who now choose to use the facility because of its lower price.

Since the benefit is calculated in terms of time savings, it is necessary to apply a value of time to recast the total benefit in monetary terms so that it can be compared against the project's cost. Benefits may also be adjusted for the value of environmental externalities and traffic accidents. Since benefits accrue annually over the lifetime of the facility and most costs are incurred at the beginning of its lifetime, present values of the flows of benefits and costs are calculated to make them comparable.

In practice, the result of CBA can be highly sensitive to the assumed value of time and discount rates. If these values are accurate, however, the beauty of CBA lies in the theoretical argument that consumer surplus, which is a measure of travelers willingness-to-pay, captures the full range of economic benefits.[1] For example, other measurable benefits, such as property appreciation near the improved facility, are chiefly outcomes of reduced travel time so including them in benefit calculation constitutes double-counting (Forkenbrock and Foster, 1990).

Even proponents of CBA concede that there are broader economic impacts that are not captured, but argue that the magnitude of these impacts for any particular project is probably small (Mackie and Nellthorp, 2001). But such impacts summed across a number projects may be substantial, which suggests that CBA is more appropriate for assessing individual projects than for assessing a program of infrastructure spending. As an indication that certain broader impacts are excluded from CBA results, notice that economic benefits are measured almost exclusively in terms of time savings. As we noted earlier, general equilibrium benefits can accrue even in the absence of time savings.

To the extent that an infrastructure spending program significantly influences relative prices, its effects are likely to be felt in markets that are removed from those under the narrow consideration of micro-level CBA. (e.g., consider the impacts on West-coast commodity markets of a significant infrastructure investment

at the Port of Long Beach.) In such cases, analysis which (i) ignores changes in prices by treating the latter as strictly exogenous and (ii) considers only those impacts which are spatially or temporally proximate—or confined to transportation or related sectors—may well fail to fully account for the benefits of the investment in question. In traditional CBA the issue boils down to the conditions under which the value of time is a theoretically valid measure of for the monetary impacts of these myriad inter-market adjustments, and the extent to which these conditions are likely to be satisfied in practice.

Macro-scale assessments of the economic impact of productivity analysis generally take the form of production and cost functions in which transportation infrastructure is included as an argument on the right-hand-side. (For a review see Lakshmanan and Anderson, 2002.) Despite their rigorous grounding in economic theory, there is a "black-box" quality about them because public capital does not function like private capital in the production technology. For example, no firm has exclusive use of a highway, and for any firm one might consider, there are large segments of the highway network that it does not use at all. Still, a firm might benefit from a highway that it does not use directly via the indirect means of reduced input costs. Clearly the mechanisms by which private productivity is enhanced by transportation infrastructure are varied and complex. Thus, a positive output elasticity tells us that some economic benefit is occurring, but sheds little light on the underlying mechanisms (Anderson and Lakshmanan, 2007). In particular, it is often very difficult to discern how much of the observed impact is due to developmental as opposed to general equilibrium influences.

A further limitation of macro-level studies is that they treat transportation infrastructure as a homogeneous good that can be measured in dollar terms. Such a measurement has some validity in the case of private capital, because it is not unreasonable to assume that the value of a capital good reflects its competitively determined marginal revenue product. In the case of transportation infrastructure, which is allocated via mechanisms that are likely to emphasize distributional goals or political expediency over economic efficiency, such an assumption is questionable. It is highly likely that investments of some types and in some locations are more productive than others.

In short, the results of macro-studies point to an important relationship between public capital and private productivity, but provide little in the way of either explanation or policy guidance.

4. A REVIEW OF GENERAL EQUILIBRIUM ANALYSES OF CONGESTION

We focus our attention on two sets of simulation studies, which develop models of the interplay between infrastructure and congestion at the level of the aggregate economy. The first, by Mayeres and Proost (1997), Conrad, 1997 and Conrad and Heng (2000), define an explicit index of congestion (Z), modeled a function of the level of utilization of aggregate transportation infrastructure or capacity, where the latter is expressed in terms of either aggregate transport activity or the size of the vehicle capital stock. Congestion incurs a productivity penalty on firms and a utility penalty on consumers. The former manifests itself through the reduced speed with which firms are able to ship their goods to market, while the latter does do via the diminished quality of transport services consumed by households. The second set of studies (Parry and Bento, 2001; 2002) model congestion through the device of an explicit household time budget. Increases in travel times with the expansion of transport activity cause a reduction in labor supply and the consumption of both leisure and services of transport producers.

Mayeres and Proost (1997) construct a stylized applied general equilibrium model which captures the essence of the congestion problem without simulating the process by which infrastructure spending affects the value of time. They consider a simple economy made up of a utility-maximizing representative household and two representative firms, summarized algebraically as follows:

$$\max_{C,\Lambda,q_P,R} U(C,\Lambda;Z^{-\xi}q_P) \tag{MP1}$$

subject to:

$$C + q_P + R \le Z^{-\varpi} f_1(h_1, q_F) \tag{MP2}$$

$$q_F = f_2(h_2) \tag{MP3}$$

$$h_1 + h_2 + \Lambda \le H \tag{MP4}$$

$$Z = 1/[1-(q_P+q_F)/(CAP+R)]^{1.5} \tag{MP5}$$

In eq. (MP1) the household derives utility (U) from consumption of a final good (C), passenger transport (q_p) and leisure (Λ). Eq. (MP2) says that firm 1 combines inputs of labor (h_1) and freight transportation (q_F) according to the production function f_1 to produce the final good, which may be consumed directly, allocated to passenger transport services, or used to create new transport infrastructure (R). Firm 2 produces freight transportation services from labor (h_2) according to the production function f_2 in (MP3), and the household's labor endowment (H) constrains labor-leisure choice in (MP4): Eq. (MP5) specifies how the imbalance between aggregate transport activity and infrastructure capacity (CAP) gives rise to congestion according to a capacity restraint function based on Evans (1992). In turn, Z adversely influences both the productivity of the final goods producer and the quality of passenger transport enjoyed by the household, according to the elasticities ξ and ϖ, respectively. Infrastructure investment alleviates congestion by expanding transit capacity, though at the cost of reduced consumption.

Conrad (1997) and Conrad and Heng (2002) apply these ideas in the context of a large-scale recursive-dynamic CGE model (GEM-E3). They elaborate the mechanisms underlying eq. (MP5) by developing an explicit model of the influence of aggregate infrastructure on the utilization of vehicle capital stocks. Their economy is made up of a representative utility-maximizing household and $j = \{1, \ldots, J, Tr\}$ firms, where firm Tr is a producer of transportation services. Firms' capital stocks are partitioned into intersectorally mobile "jelly" capital (k_j) and transportation capital (kt_j), which represents vehicles and is a fixed factor. In the simplest version of their model the aggregate quantity of transportation infrastructure (KI) is constant, and its divergence from the socially optimal level (KI^*) is responsible for congestion which reduces the productivity of kt:

$$\max_{C_j} U(C_1,\ldots,C_J;Z^{-\xi}C_{Tr}) \tag{CH1}$$

subject to

$$\sum_v X_{v,j} + C_j \le f(X_{1,j},\ldots,X_{I,j};h_j,k_j,kt_j^e) \tag{CH2}$$

$$kt_j^e = kt_j Z^{-\varpi} \tag{CH3}$$

$$kt_j = kt_j^0 \exp(-a/KI) \tag{CH4}$$

$$Z = \exp\left(-\sum_j a\omega_j\left(\frac{1}{KI} - \frac{1}{KI^*}\right)\right), \sum_j \omega_j = 1 \tag{CH5}$$

$$H = \sum_j h_j, \tag{CH6}$$

$$K = \sum_j k_j. \tag{CH7}$$

$$KI \text{ fixed} \tag{CH8}$$

$$KI^* \approx \kappa \sum_j \pi_j^{KT} kt_j / P^{KI} \tag{CH9}$$

In (CH1) the household derives utility from consumption of C_j units of each good, with congestion diminishing the quality of consumed transportation services. The jth firm produces a unique good which is both consumed and used as an intermediate input ($X_{i,j}$) to the i other firms (CH2). Production is described by a nested CES function, f_j, which combines intermediate inputs with labor (h_j), jelly capital (k_j) and effective units of transportation capital (kt_j^e). The latter consists of a benchmark quantity of fixed capital (kt_j^0) whose productivity is exponentially augmented by infrastructure in (CH4) and attenuated by congestion in (CH3). These influences are modulated by the coefficient a and the elasticity ϖ, respectively, and the factor $\exp(-a/KI) < 1$ can be interpreted as a capacity utilization measure. Equilibrium between the demands for labor, capital and infrastructure and the endowments of these factors (*H, K* and *KI*) is given by eqs. (CH6)-(CH8), and eq. (CH5) defines congestion in terms of the weighted average utilization rate of transportation capital relative to the optimal utilization level, with industry weights ω_j. Conrad (1997) derives the condition for the optimum (CH9) under the assumption that there exists an exogenous government-cum-social planner whose objective is to minimize the economy's total expenditure on transportation. The resulting supply function for *KI** is denominated over the quantities of firms' transportation capital stocks, their shadow prices (π_j^{KT}), the marginal social cost of infrastructure provision (P^{KI}), and the elasticity of transport capital with respect to infrastructure (κ).

This approach has the advantage of being straightforward to numerically parameterize.[2] However, its main limitation is that it does not explicitly relate congestion to investment in infrastructure and the value of time (e.g., the Lagrange multiplier on eq. (MP4)), whose role in CBA is to indicate when the marginal benefits of alleviating the former exceed the marginal costs of the latter. The relevant mechanism is captured by the second set of studies, which model the production of travel as requiring inputs of time, which explicitly enter into a household time budget constraint.

Parry and Bento (2001) emphasize the impact of substitution among differentially congested modes of travel on time expenditures. Theirs is a stylized model of commuting—production is modeled in the simplest possible way and freight transport is not considered. The economy is made up of a utility-maximizing household, a final goods producer and three transport firms (indicated by the subscript m), each of which corresponds to a particular mode: congested roads (R), public transit (P) and non-congested roads (F).

$$\max_{C,\Lambda,q_R,q_P,q_F} U(C,\Lambda) \tag{PB1}$$

subject to:

$$C + \sum_m X_m \leq \min(H, Q) \tag{PB2}$$

$$Q = f(q_R, q_P, q_F) \tag{PB3}$$

$$q_m = \begin{cases} \min(X_m / \nu_m, T_m / \tau_m) & m = P, F \\ \min[X_m / \nu_m, D(T_m, d_1 - d_2 q_m) / \tau_m] & m = R \end{cases} \tag{PB4}$$

$$H + \Lambda + \sum_m T_m \leq \bar{T} \tag{PB5}$$

The household derives utility from consumption of the final good and leisure, (PB1). Eq. (PB2) says that the output of the final goods firm is produced from labor and aggregate transportation services (Q) according to a fixed-coefficients technology, and can either be consumed or allocated to intermediate uses by the transport firms (X_m). Transport services are defined in (PB3) as a composite of the trips on the different modes (q_m), with f used to indicate a constant elasticity of substitution (CES) aggregator function. In turn, the production of trips in eq. (PB4) necessitates use of the intermediate commodities and travel time (T_m). For public transit and uncongested roads, trip generation is modeled using a Leontief transformation function, whose coefficients (ν_m and τ_m) indicate the per-trip expenditures of money and time. The implication is that for these modes the level of congestion is exogenous, with constant marginal time expenditures τ_m. By contrast, on congested roads the level of congestion is endogenous. The modeling device used to represent this is the CES aggregator function D, which defines the degree of substitutability between travel time and "available road capacity", given by the linear function $d_1 - d_2 q_R$ (where d_1 and d_2 are constants). Finally, the time budget constraint (PB5) requires that the sum of labor supply, leisure and total commuting time exhaust the household's time endowment ($\bar{T}$).

The simple logic of the model is that production creates a derived demand for transport. As trips via congested modes (in this case roads, R) rise, so does congestion, which in turn reduces available capacity and time spent traveling by those modes, inducing substitution of trips to less congested alternatives. The critical parameters governing this process are the elasticities of substitution among transit modes in f and between travel time and unused mode capacity in D, and the coefficients of the road availability function.

Parry and Bento's (2002) extension enumerates trips on congested freeways (RF) and alternate back roads (RB) as additional modes of travel, includes negative externalities such as accidents and air pollution (which we indicate using the function E), and represents congestion in terms of travel time using a more traditional approach.

$$\max_{C, \Lambda, q_{RF}, q_{RB}, q_P, q_F} \quad U(C, Q, \Lambda) - E(q_{RF}, q_{RB}, q_P, q_F) \tag{PB1′}$$

subject to (PB3), (PB5) and:

$$C + \sum_m X_m \leq H \tag{PB2′}$$

$$X_m = V_m q_m \tag{PB6}$$

$$T_m = \tau_m q_m \tag{PB7}$$

$$\tau_m = \tau_m^0 [1 + 0.15(q_m / CAP_m)^4] \,, \; m = RF, RB \tag{PB9}$$

Aggregate transportation services are now included as an additional argument in the household's utility function, along with non-congestion externalities (PB1′). As before, the sole factor of production is labor, whose supply is traded off against leisure and travel according to the time budget constraint (PB5). Here, however, (PB2′) assumes that each unit of labor produces one unit of the final good, which can be consumed or used to pay for trips, whose transformation into aggregate transport services follows (PB3). As in (PB4), trips incur fixed marginal pecuniary costs (PB6), but marginal expenditures of time (PB7) which increase with congestion. Eq. (PB10) defines the latter relationship using the classic Bureau of Public Roads (BPR) capacity restraint formula.

In both Parry-Bento models, the Lagrange multiplier on eq. (PB5) represents the "true" marginal utility of time, which takes into account the general equilibrium interactions among the labor supply, the consumption of the final good and leisure, and the supply-demand balance for trips by different modes. Nevertheless, the value of time which emerges from this analysis still does not completely account for the channels through which congestion's effects are felt. In particular, the simple representation of production fails to capture the way in which travel delays impact firms or may themselves be exacerbated by the shipment of finished goods to retail markets or households' retail purchasing behavior.

Likewise, the specification of substitution possibilities in transportation is simplistic. The Parry-Bento models are "maquettes" which resolve only a few, very aggregate modes of travel and can afford to rely on synthetic benchmark distributions of trips.[3] In section 5.2.5 we caution that moving to the use of real-world data to numerically calibrate the aggregator functions for transportation services may be quite involved. The remainder of the paper addresses these issues and examines their implications for constructing large-scale transport-focused CGE models.

5. A HYBRID MESO-MACRO APPROACH

Our proposed approach is a hybrid one which seeks to capture meso-level details of infrastructure, congestion and transport within the traditional framework of a macro-level CGE model. We consider a static economy with N representative profit-maximizing firms, each of which produces a single, distinct commodity. Firms and commodities come in two varieties, I non-transport producers and their associated goods and services, which we index using the subscript $i = \{1, \ldots, I\}$, and M transport or logistics firms and their associated services, which we index with the subscript $m = \{1, \ldots, M\}$. To distinguish between firms and the goods which they produce, we introduce the subscript $j = \{1, \ldots, I\}$ to enumerate non-transport producers. Furthermore, we define the set of transport producers in such a way that each firm corresponds to a single disaggregate mode of transit, e.g., rail as well as truck freight shippers, air passenger and freight travel, own-supplied passenger road travel using private vehicles, purchased local/interurban passenger transit by road and rail, etc. Non-transport firms supply goods and services to satisfy the intermediate demands of other firms as well as the final demands of households. Transportation firms provide freight transport services to the non-transport firms and passenger transport services to households. Households in the economy are modeled as a representative utility-maximizing agent who owns the factors of production (hours of labor, H, and capital, K) and rents them out to the firms in exchange for factor payments which finance the consumption of commodities.

The centerpiece of our approach is a representation of the process though which the operation of markets for non-transport commodities and labor creates derived demands for transportation. In particular, we assume that:

(a) Each unit of non-transport commodity requires freight trips to be shipped to sources of intermediate or final demand.

(b) The representative agent's final purchases of these goods and services require retail shopping trips in order for them to be converted into utility, and

(c) The agent's rental of labor to firms requires commuting trips.

The three kinds of mobility are distinguished using the superscripts *TF*, *TC* and *TH*, respectively.

We assume that households face two budget constraints, a pecuniary constraint that commodity purchases not exceed factor income, and a temporal constraint that the duration of travel for shopping and commuting, hours of work and leisure not exceed an aggregate endowment of time. The latter sets up a tension between travel time expenditures for the purposes of consumption and income generation. For households to increase their consumption they need more income, which in the short run can only be obtained by renting additional hours of labor to firms, with the possible side-effect of more and/or longer journeys to work. However, their ability to earn is constrained by the fact that consuming more non-transport goods requires additional retail trips, and concomitant expenditures of time.

Firms do not have explicit time budgets, nevertheless we assume that time constraints influence production in an implicit fashion. We treat non-transport firms as "mills", whose products are manufactured using labor, capital and intermediate inputs from other non-transport firms. In order for these products to be consumed they must be shipped to markets, which creates a demand for trips supplied by the transportation firms. A key feature of the model is that the latter firms do not produce trips directly. Rather, their outputs take the form of generalized transportation services such as vehicles operated on the road, trains on tracks, planes in the air, etc.—whose capacity is determined by the firms' stocks of transportation capital (i.e., vehicles). This allows us to model the mechanism by which congestion imposes a productivity penalty on firms: *speed*, which along any given segment of the transportation network is equivalent to the inverse of the travel time, is necessary to transform these services into trips. Thus, given a certain capacity to produce transport services, increases in travel time translate into fewer trips. Other things equal, the main consequence is a fall in the productivity of inputs to transportation and in rise in the average cost of trips, a decline in movements of passengers and non-transport goods, and a reduction in production and consumption.

These devices allow us to model the impacts of congestion in a novel way. Congestion is the increase in travel time arising out of the imbalance between the aggregate derived demand for mobility and the capacity of the stock of transportation infrastructure to support the desired flux of trips. Our way of representing travel admits three channels through which congestion may exert a drag on activity, corresponding to (a)–(c), above:

- An increase in the duration of households' retail trips per unit of consumption, which attenuates consumption through the time budget constraint. Depending on the relevant elasticities, the time spent on work or leisure may rise or fall as well, but the standard result is a decline in utility.

- A reduction in average productivity of transport-producers. Because congestion increases the duration of freight trips of a given distance, it reduces the number of trips which transport producers can supply with a given fleet of vehicles. The consequent dissipation of operating time—and thus revenue—drives a wedge between the marginal cost of each transportation firm's output and the unit value of its trips consumed by firms and households, much like a tax.

- Dissipation of time in commuting, acting through the time budget constraint to reduce the economy's aggregate labor supply. This effect is similar to a tax on labor.

To effectively inform traditional microeconomic cost benefit analysis, any attempt to capture these influences using top-down economy-wide models must address a number of issues. First, considering the impacts of congestion *in the aggregate* will yield limited insights, as typically only a fraction of the links in an economy's transport network will be congested. These are the ones which are candidates for infrastructure projects. Following from this observation, a second consideration is that trips should be thought of as differentiated commodities, whose equilibrium allocation among network links will be a function of transport producers' marginal costs, firms and households' demands for goods and passenger mobility, and the distribution of travel times/congestion. In general, infrastructure investments that are sufficiently large will give rise to *simultaneous* non-marginal changes in *all* of these variables, the character of which will depend on both magnitude of the investments and their location on the network. Finally, in order to properly capture the dissipative effect of congestion, a distinction needs to be made between the production of transport services and the consumption of the trips which they make possible. Logistics and passenger transport firms will allocate their outputs to the network segments which yield the highest marginal revenue, while households and non-transport firms will allocate their demand for trips to links with the lowest marginal cost. The key challenge is therefore to develop a computationally tractable way of modeling the equilibrium between the supply-side transformation of transport services to trips and the demand-side aggregation of trips into passenger and freight movements so as to resolve the substitution of trips from congested links to uncongested alternatives.

The simplest way of doing this is to keep the spatial details of the network structure to a minimum. Our strategy is to assume the existence of a generic transport network with $l = \{1, \ldots, L\}$ links, amongst which trips generated by the m transport producers are allocated in a competitive fashion. This choice allows us to specify a top-down model of intermediate complexity which is able to capture the macroeconomic feedbacks which affect—and are affected by—Wardropian equilibria, while serving as a bridge to more disaggregate network equilibrium models (e.g., Ferris et al, 1999). We make the key simplifying assumption that trips across different mode-link alternatives are imperfect substitutes, with differing marginal costs to transport consumers and differing marginal revenues to transport producers. Thus, when non-transport firms ship their product to market, or households supply labor or consume a particular commodity, each of these actors simultaneously chooses travel distances/routes and modes by allocating trips over l and m so as to minimize total transportation expenditure. Symmetrically, each logistics or passenger transport firm simultaneously chooses travel distances/routes and payloads by allocating the transportation services it produces to trips by l and j in the case of freight, l and i in the case of retail shopping, and just l in the case of commuting, so as to maximize revenue.

We operationalize these ideas by specifying transportation demands as constant elasticity of substitution (CES) functions of trips by mode. Thus, freight transport demand by the jth firm, Q_j^{TF}, is modeled as a CES aggregate of the trips, $q_{j,l,m}^{TF}$, made by transport mode m on network link l to ship j's product to intermediate and final consumers. Similarly, we model the aggregate household demand for transportation to consume the ith commodity, Q_i^{TC}, as a CES function of the retail trips, $q_{i,l,m}^{TC}$, across all combinations of transit modes and links, and the aggregate demand for transportation to supply labor, Q^{TH}, as a CES function of the various mode- and link-specific commuting trips, $q_{l,m}^{TH}$. We use the same device on the supply side, specifying the trips undertaken by each transport producer as a constant elasticity of transformation (CET) expansion of that firm's output. Thus, firm m's trips $q_{j,l,m}^{TF}$, $q_{i,l,m}^{TC}$ and $q_{l,m}^{TH}$ are modeled as a CET function of the aggregate supply of transportation services by mode, Y_m. Parry and Bento (2001) note that the ability to substitute between transport modes mitigates the cost of reducing congestion. Our assumption that households and firms substitute among both transport modes and network segments means that the elasticities in the aforementioned CES and CET functions will likely be a key influence on the marginal benefit of investments to increase the capacity of congested links.

The attractive feature of this formulation is that it automatically generates different levels of congestion for each transport mode and network link. Congestion is a function of the total flux of trips generated by each mode across a given link,

$$\vartheta_{l,m} = \sum_i \left(q_{i,l,m}^{TC} + q_{i,l,m}^{TF} \right) + q_{l,m}^{TH}, \tag{1}$$

and the design capacity of the particular segment, $CAP_{l,m}$, such that as $\vartheta_{l,m}$ exceeds $CAP_{l,m}$, travel time on that link, $\tau_{l,m}$, increases rapidly. A convenient representation of this phenomenon is the BPR formula (PB9):

$$\tau_{l,m} = \tau^0_{l,m}\left(1 + 0.15\left(\vartheta_{l,m}/CAP_{l,m}\right)^4\right) \tag{2}$$

in which $\tau^0_{l,m}$ is the mode- and link-specific free-flow travel time.

The dependence of capacity on infrastructure investment is an exogenous input to the model that must be developed from the existing characteristics of the transit network, the links which are candidates for improvement, and the projected changes in traffic flows resulting from the proposed project. Eqs. (1) and (2) establish the crucial connection between household trips to consume goods or commute back and forth to work, and the trips undertaken by logistics firms to deliver commodities. Growth in any one type of transportation adds to the total flux of trips across the transport network, raising trip times, and inducing economic actors to re-allocate trips to less congested mode-link alternatives, as well as cut back on overall travel. As mentioned above, the consequence of this is a decline in both the quantity of labor supplied to producers and the goods and services consumed by households. In this way infrastructure capacity acts as a fundamental brake on the expansion of economic activity.

5.1. Algebraic Summary of the CGE model

We begin with a description of the households in our simulated economy. We assume the existence of a representative agent whose utility is represented by the nested CES function shown in Figure 1A. At the top level of the nesting hierarchy, the agent obtains utility (U) from non-transport consumer commodities ($\hat{C}_i$) and leisure (Λ), with elasticity of substitution σ^U and technical coefficients α_i and α_Λ:

$$U = \left(\sum_i \alpha_i \left(\hat{C}_i\right)^{(\sigma^U-1)/\sigma^U} + \alpha_\Lambda \Lambda^{(\sigma^U-1)/\sigma^U}\right)^{\sigma^U/(\sigma^U-1)} \tag{3}$$

The second level of the utility hierarchy describes how the demands for commodities create derived demands for personal transportation. We specify each unit of delivered commodity as a CES composite of transportation services, Q_i^{TC} (i.e., trips for the purpose of retail purchases), and the relevant commodity ($\overline{C}_i$) whose consumption necessitates transport expenditures, with elasticity of substitution $\sigma_i^C < 1$ and technical coefficients β_i^{TC} and β_i^C:

$$\hat{C}_i = \left(\beta_i^{TC}\left(Q_i^{TC}\right)^{(\sigma_i^C-1)/\sigma_i^C} + \beta_i^C\left(\overline{C}_i\right)^{(\sigma_i^C-1)/\sigma_i^C}\right)^{\sigma_i^C/(\sigma_i^C-1)}. \tag{4}$$

Each retail commodity is itself a composite of the goods and services actually being purchased by consumers and freight transportation services, which we discuss in more detail below. Note that eq. (4) implicitly captures Lakshmanan and Hua's (1983) distinction between discretionary and non-discretionary transportation: while travel is a necessary input to consumption ($\sigma_i^C < 1$) the actual quantity of travel undertaken by households is discretionary. Consequently, our formulation captures the ability of consumers to substitute passenger travel (Q_i^{TC}) for the freight transportation component of $\overline{C}_i$, e.g., by opting to travel to retail outlets to purchase goods versus having them delivered directly to the consumer's place of residence. At the lowest level of the hierarchy the transportation services necessary to consume good i are a CES composite of the shopping trips $\left(q^{TC}_{i,l,m}\right)$ which occur on each link of the transport network and mode of transit:

$$Q_i^{TC} = \left(\sum_l \sum_m \gamma^{TC}_{i,m,l}\left(q^{TC}_{i,m,l}\right)^{(\sigma_i^{TC}-1)/\sigma_i^{TC}}\right)^{\sigma_i^{TC}/(\sigma_i^{TC}-1)}. \tag{5}$$

Here σ_i^{TC} is the substitution elasticity and $\gamma_{i,l,m}^{TC}$ are technical coefficients which indicate how the retail trips associated with each commodity are distributed across mode-link alternatives in the benchmark data used to calibrate the model.

Households' rental of their factor endowments is modeled in a similar fashion. We assume that transport services are not necessary to supply capital to firms. However, to supply H units of labor the agent must utilize transportation services, Q^{TH} (i.e., trips for the purpose of commuting). Accordingly, the aggregate supply of labor ($\overline{H}$) is modeled as a CES composite, with elasticity σ^H and coefficients β^{TH} and β^H:

$$\overline{H} = \left(\beta^{TH} \left(Q^{TH} \right)^{(\sigma^H - 1)/\sigma^H} + \beta^H H^{(\sigma^H - 1)/\sigma^H} \right)^{\sigma^H/(\sigma^H - 1)}. \tag{6}$$

As with retail trips, we model Q^{TH} as a CES composite of the commuting trips $\left(q_{l,m}^{TH} \right)$ by each network link and transit mode, with elasticity σ^{TH} and mode-link coefficients $\gamma_{l,m}^{TH}$:

$$Q^{TH} = \left(\sum_l \sum_m \gamma_{l,m}^{TH} \left(q_{l,m}^{TH} \right)^{(\sigma^{TH} - 1)/\sigma^{TH}} \right)^{\sigma^{TH}/(\sigma^{TH} - 1)}. \tag{7}$$

The representative agent's budget constraint mandates that the value of consumption at retail goods prices $\overline{P}_i$ exhaust the income from factor rentals:

$$\sum_i \overline{P}_i \overline{C}_i \leq \theta H + rK, \tag{8}$$

where θ denotes the marginal utility of time—i.e., the wage net of the marginal cost of commuting, and r is the capital rental rate. The agent's time constraint mandates that the total expenditure of time on trips for retail consumption and commuting (summed over all network links and modes), plus labor and leisure time, exhaust the agent's endowment of time, given by $\overline{T}$:

$$\sum_l \sum_m \tau_{l,m} \left(\sum_i q_{i,l,m}^{TC} + q_{l,m}^{TL} \right) + H + \Phi \leq \overline{T}. \tag{9}$$

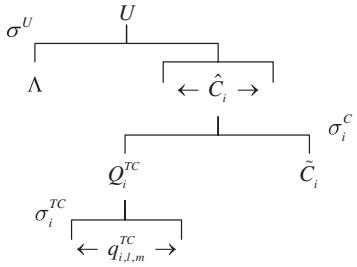
The key variable in this expression is $\tau_{l,m}$, the average trip time on each network segment, which by (1) and (2) reflects the tension between the total flux of trips on that segment and its capacity.

The organization of production is shown in panels B and C Figure 1. In panel B, the output of each of the j non-transport firms (Y_j) is modeled using a CES production function denominated over inputs of intermediate commodities $(\overline{X}_{i,j})$, labor $\overline{h}_j$ and capital (k_j):

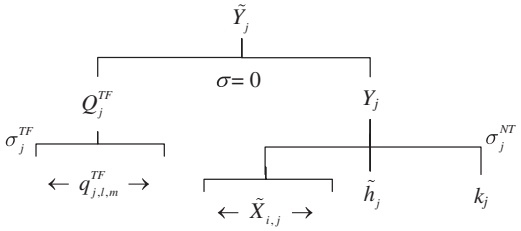
$$Y_j = \left(\sum_i \delta_{i,j}^{NT} \overline{X}_{i,j}^{(\sigma_j^{NT} - 1)/\sigma_j^{NT}} + \delta_{H,j}^{NT} \overline{h}_j^{(\sigma_j^{NT} - 1)/\sigma_j^{NT}} + \delta_{K,j}^{NT} k_j^{(\sigma_j^{NT} - 1)/\sigma_j^{NT}} \right)^{\sigma_j^{NT}/(\sigma_j^{NT} - 1)} \tag{10}$$

with substitution elasticity σ_j^{NT} and distribution parameters δ^{NT}.

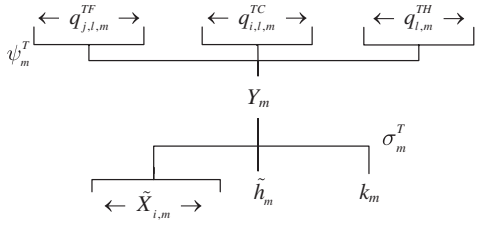
Figure 1. **Nested CES utility and production functions used in the model**



A. Utility



B. Production: Non-Transport Firms($i \in I$)



C. Production: Transportation Firms ($m \in M$)

Notes: U=utility; Λ=leisure; $\hat{C}_i$=consumption goods-retail mobility composite; $\tilde{C}_i$=consumption goods-freight mobility composite; Q_i^{TC} =retail mobility; σ^U, σ_i^C=goods-leisure and transport-goods, substitution elasticities; $\tilde{Y}_j$=delivered non-transport goods; Y_j = non-transport goods production; Y_m=transport services production; Q_j^{TF} =freight mobility; $\tilde{X}_{i,j}$, $\tilde{X}_{i,m}$ = intermediate inputs; $\tilde{h}_j$, $\tilde{h}_m$=labor inputs; k_j, k_m=capital inputs; σ_j^{NT}, σ_m^T=firm input substitution elasticities; $q_{j,l,m}^{TF}$, $q_{i,l,m}^{TC}$, $q_{l,m}^{TH}$=freight, retail and commuting trips; ψ_m^T = elasticity of transformation of transport services into trips; σ_j^{TF}, σ_i^{TC}, σ^{TH}=freight, retail and commuting trip mode-link substitution elasticities.

Similar to the households, the supply of commodities creates a derived demand for transportation services. We assume that the delivery of χ_i^{TF} units of commodity *i* to intermediate and final users requires a unit of freight transportation services (Q_i^{TF}). As a result, the supply of non-transport commodities are a Leontief composite of produced commodities and transportation:

$$\tilde{Y}_i = \min\left(\chi_i^{TF} Q_i^{TF}, Y_i\right). \tag{11}$$

In turn, freight transport is a CES composite of delivery trips ($q_{j,l,m}^{TF}$) by network link and transit mode:

$$Q_j^{TF} = \left(\sum_l \sum_m \gamma_{j,l,m}^{TF} \left(q_{j,l,m}^{TF}\right)^{(\sigma_j^{TF}-1)/\sigma_j^{TF}} \right)^{\sigma_j^{TF}/(\sigma_j^{TF}-1)}, \tag{12}$$

with elasticity σ_j^{TF} and mode-link coefficients $\gamma_{j,l,m}^{TF}$.

The production function for the *m* transportation services is illustrated in panel C, and is essentially the same as that for non-transport commodities:

$$Y_m = \left(\sum_i \delta_{i,m}^T \overset{\square}{X}_{i,m}^{(\sigma_m^T - 1)/\sigma_m^T} + \delta_{H,m}^T \overset{\square}{h}_m^{(\sigma_m^T - 1)/\sigma_m^T} + \delta_{K,m}^T k_m^{(\sigma_m^T - 1)/\sigma_m^T} \right)^{\sigma_m^T/(\sigma_m^T - 1)}, \tag{13}$$

with substitution elasticity σ_m^T and distribution parameters δ^T. However, to translate between transport producers' outputs and the trips necessary to deliver passengers and freight along each link, we use the following CET formulation with transformation elasticity ψ_m^T and distribution parameters μ:

$$Y_m = \left(\sum_l Z_{l,m}^{-1} \cdot \left(\sum_j \mu_{j,l,m}^{TF} \left(q_{j,l,m}^{TF} \right)^{(\psi_m^T - 1)/\psi_m^T} + \sum_i \mu_{i,l,m}^{TC} \left(q_{i,l,m}^{TC} \right)^{(\psi_m^T - 1)/\psi_m^T} + \mu_{l,m}^{TH} \left(q_{l,m}^{TH} \right)^{(\psi_m^T - 1)/\psi_m^T} \right) \right)^{\psi_m^T/(\psi_m^T - 1)} \tag{14}$$

The variable $Z_{l,m}$ is particularly important. It is a link-specific productivity penalty which is a function of each link's average travel time in (2), and is intended to capture the adverse impact of congestion on the ability of transport firms to translate service outputs into movements of goods and passengers. Thus, the more congested a given link, the more units of Y_m necessary to generate an additional trip on that link, reducing the average and marginal productivity of the inputs to (13).

The model is closed by specifying market clearance conditions for the supply of composite non-transport commodities:

$$\overset{\square}{Y}_i = \sum_j \overset{\square}{X}_{i,j} + \overset{\square}{C}_i, \tag{15}$$

and the representative agent's primary factor endowments:

$$H = \sum_j \overset{\square}{h}_j + \sum_m \overset{\square}{h}_m, \tag{16}$$

$$K = \sum_j k_j + \sum_m k_m. \tag{17}$$

In an appendix to the paper we provide a detailed mathematical elaboration of how the foregoing elements may be used to construct an operational CGE model. The assumption of Walrasian competitive equilibrium allows us to express the production and aggregation relations in eqs. (3)–(7) and (10)–(14) in terms of their dual cost functions and associated conditional input demand functions. The latter can then be substituted into the supply-demand balance constraints (6), (8), (9) and (15)–(17) to yield a system of nonlinear equations (Ξ) in commodity and factor prices, the activity levels of firms and the income level of the representative agent. The result is the Walrasian excess demand correspondence of the economy.

The advantage of our approach is two-fold. First, on the consumer side, the price variable consisting of the Lagrange multiplier on eq. (9) is the endogenous value of time, which takes full account of not only the channels through which congestion's impacts manifest themselves, but also the general equilibrium interactions among these effects. Second, in both eq. (9) and the cost functions for transport producers (i.e., the dual of eq. (14)) congestion plays the role of a vector of differentiated, link-specific taxes on trips. This is very useful, as it enables us to simulate congestion as an endogenous, nonlinear tax whose level is determined by trip volumes according to eqs. (1)–(2).

5.2. Data and calibration

Numerical solution of the model requires that values be specified for the parameters of the excess demand correspondence, Ξ. This procedure, known as calibration, is likely to be especially challenging given that it requires the integration of economic and transportation data which are often incommensurate.

Calibrating the purely macroeconomic, non-transport related components of the CGE model is a fairly simple task, involving the selection of values for the elasticities σ^U, σ_j^{NT} and σ_m^T based on empirical estimates, and the computation of values for the coefficients, α_i, α_Λ, δ^{NT} and δ^T using a national- or regional-level social accounting matrix (SAM). (For details see, e.g., Sue Wing, 2004.) We anticipate that it will be somewhat more difficult to find econometric estimates for, or calibrate using related empirical studies, values for the substitution elasticities σ_i^C, or to infer values for the coefficients χ_j^{TF}, β_i^{TC} and β_i^C from data on freight transport margins. And since published estimates for the elasticities σ_i^{TC}, σ_j^{TF}, σ_i^{TC}, ψ_m^T do not exist, developing the data and econometric procedures to estimate these parameters will likely involve a large amount of effort. A rough-and-ready way to proceed is to set up the model using the values in the range assumed by Parry and Bento (2001, 2002), and perform sensitivity analysis.

Calibrating the coefficients $\gamma_{j,l,m}^{TF}$, $\gamma_{i,l,m}^{TC}$, $\gamma_{l,m}^{TH}$, $\mu_{j,l,m}^{TF}$, $\mu_{i,l,m}^{TC}$, and $\mu_{l,m}^{TH}$ in the trip aggregation and transformation functions is likely to be a significant undertaking. The key difficulty lies in defining the topology of the transport network and associated traffic flows at a geographic scale for which there are published inter-industry economic accounts. Although survey data on commuting and freight traffic flows are readily available at the level of metropolitan statistical areas (MSAs), input-output data are rarely tabulated at such a fine spatial scale. At the opposite extreme, while it is straightforward to construct an aggregate CGE model using a SAM constructed from the transportation satellite account make and use tables (Fang et al, 2000), for modes of surface travel such as commuting or retail shopping which are important contributors to congestion, it is not obvious how to represent major congested network links at such a highly aggregate scale.

Thus, apart from the normative question of what is the most appropriate geographic scale at which our model should be specified, data constraints dictate the practical necessity of structuring a reduced-form representation of the transportation network so that it is both sufficiently simple to calibrate the γ and μ parameters and able to be matched to a regional SAM. At the current stage of this research, the most promising source of economic data would seem to be county-level SAMs developed by IMPLAN which are coterminous with MSAs that straddle major transportation corridors in the eastern U.S.

6. DISCUSSION AND SUMMARY

Provision of transportation infrastructure is one of the most visible, vital and costly ways in which the public sector contributes to the private economy. Yet decisions about the levels and allocations of transportation infrastructure investments must currently be made with incomplete information about their economic impacts. Analytical tools are limited to micro-scale analyses, which may not capture the full range of economic benefits induced by a project or program, and macro-scale analyses, which are too broadly defined to provide guidance on the relative benefits of specific projects and programs. The situation calls for analytical tools defined as a "meso-level" that can provide impacts assessments that are both comprehensive and capable of representing specific expansions of infrastructure capacity, following the three criteria we

define in the introduction. This paper contributes to the development of such tools by specifying a CGE model that is specifically designed to assess the broader economic impacts of transportation infrastructure investments.

The model specified above draws on the limited CGE literature on transportation infrastructure – especially the household time budgets of Parry and Bento (2001, 2002). It goes beyond existing models both in terms of its technical specification and its overall scope. It specifies a set of derived demands for transportation services that arise from production, consumption and labor supply activities. It represents transportation infrastructure as a set of capacitated mode-link combinations on which flows are assigned and congestion is modeled as increases in travel time. By embedding travel time explicitly in the determination of household utilities and the prices and quantities of commodities produced in the economy, it incorporates in the model the phenomenon of congestion that is based on a comprehensive, endogenous definition of the value of time.

Existing CGE models designed for the analysis of transportation infrastructure can, for the most part, be classified s maquettes – that is, simplified or scaled down models designed to make rough estimates of relative magnitudes or as steps toward the creation of more comprehensive models. By contrast, the model specified above is intended as a practical tool for policy analysis. There are, however, significant hurdles to overcome before it can be made operational, including defining an appropriate geographical scale for its application, an appropriate level of detail for the set of mode-link combinations and appropriate data and parameters for calibration.

The question that naturally arises is whether it is worth the effort. To some extent this comes down to the empirical question of whether the broader economic benefits captured in the CGE model are of significant magnitude relative to the more direct effects captured in CBA. But looking beyond the "bottom line" of aggregate benefits, the CGE model generates a range of information that cannot be obtained from existing models such as whether the benefits of a capacity expansion accrue mostly to firms or households, whether household benefits are mostly in consumption activities or commuting and whether some industries benefit more than others. Such information may be useful in assessing whether specific objectives that policy makers attach to a project are likely to be met. Also, the CGE specification is especially well-suited to assessing the impact of infrastructure programs because the implementation of two or more capacity expansions can be modeled simultaneously. This will be useful in identifying complementarities among projects by seeing, for example, whether the benefits of projects A and B implemented simultaneously exceed the sum of the benefits of A and B implemented independently.

Ultimately, the value of a model such as the one we have specified lies in its laying bare a plausible set of economy-wide interactions that are triggered by an improvement in transportation infrastructure. In other words, it is an attempt to move beyond "black box" and "bottom line" approaches to policy models to an approach that explains rather than just captures economic impacts. Naturally, laying the underlying mechanisms bare opens the door to criticism based on underlying assumptions – especially as regards market imperfections. Furthermore, we recognize that there are a range of dynamic, developmental impacts that the model does not include. Still, we believe that specifying and calibrating our model is a useful step toward a better understanding of the economy-wide consequences of transportation infrastructure in the economy.

7. APPENDIX: IMPLEMENTATIONAL DETAILS

Walrasian general equilibrium prevails when the price of commodities equals their marginal cost of production with firms earning zero profits, there is zero excess demand for commodities and factors, and consumer' income equals their expenditure. These conditions form the basis for CGE models in a complementarity format, which specify the economy as a vector of zero profit, market clearance, income balance, and auxiliary equations. Each equation is paired with an associated dual variable with respect to which it exhibits complementary slackness (see, e.g., Rutherford, 1995; Sue Wing, 2004):

1. *Zero profit conditions for firms and households*. These specify the equilibrium between commodity prices and firms' unit cost functions, and between the marginal utility of income and the aggregate expenditure function. They complementary to the activity levels of firms and the utility level of the representative agent.
2. *Market clearance conditions for commodities and factors*. These specify the equilibrium between the aggregate demands for commodities and factors—which are functions of prices and activity levels, and their aggregate supplies—typically indicated by firms' activity levels and households' factor endowments. They are complementary to commodity and factor prices.
3. *Income balance conditions*. These specify the equilibrium between the value of households' expenditures and the value of their income, and are complementary to the income levels of the households.
4. *Auxiliary equations*. These typically represent some sort of constraint on the economy that is a function of both an auxiliary variable and other endogenous variables, which requires them to be solved for along with the remaining variables. They are complementary to the auxiliary variable.

Henceforth we use the shorthand symbol "⊥" to represent these complementary relationships.

7.1. Zero profit conditions and associated demand functions

As before, we begin with the households in the economy. Recasting the representative agent's utility maximization problem as a dual expenditure minimization permits us to solve for the unit expenditure function, ε, dual to (3):

$$\varepsilon = \left(\sum_i \alpha_i^{\sigma^U} \hat{P}_i^{1-\sigma^U} + \alpha_\Phi^{\sigma^U} \theta^{1-\sigma^U} \right)^{1/(1-\sigma^U)}. \tag{18}$$

where $\hat{P}_i$ is the price of the ith consumption good-transport services aggregate, and θ is the value of time. This expression can be thought of as a zero-profit condition for the "production" of a utility good, to which aggregate utility is the complementary activity variable. By Shepard's Lemma, the derivatives of the zero-profit condition with respect to the prices of the inputs yields the conditional input demands. Accordingly, final demands for commodities and leisure are given by:

$$\hat{C}_i = \alpha_i^{\sigma^U} \hat{P}_i^{-\sigma^U} \varepsilon^{\sigma^U} U, \tag{19}$$

$$\Phi = \alpha_{\Phi}^{\sigma^U} \theta^{-\sigma^U} \varepsilon^{\sigma^U} U, \tag{20}$$

Cost minimization in the aggregation of transport services and physical goods in eq. (4) gives rise to the following zero profit condition, which is the unit cost function for $\hat{C}_i$:

$$\hat{P}_i = \left(\left(\beta_i^{TC}\right)^{\sigma_i^C} \left(P_i^{TC}\right)^{1-\sigma_i^C} + \left(\beta_i^C\right)^{\sigma_i^C} \square P_i^{1-\sigma_i^C} \right)^{1/(1-\sigma_i^C)}, \tag{21}$$

where P_i^{TC} and $\square P_i$ are the consumer prices of retail mobility and final commodity sales associated with good *i*. The conditional demands for these inputs are given by:

$$Q_i^{TC} = \left(\beta_i^{TC}\right)^{\sigma^C} \left(P_i^{TC}\right)^{-\sigma^C} \hat{P}_i^{\sigma^C} \hat{C}_i, \tag{22}$$

$$\square C_i = \left(\beta_i^C\right)^{\sigma^C} \left(\square P_i\right)^{-\sigma^C} \hat{P}_i^{\sigma^C} \hat{C}_i . \tag{23}$$

Similarly, the unit cost function arising from cost-minimizing aggregation of labor hours and commuting to produce supplied labor in (6) is:

$$\square w = \left(\left(\beta^{TH}\right)^{\sigma^H} \left(P^{TH}\right)^{1-\sigma^H} + \left(\beta^H\right)^{\sigma^H} \theta^{1-\sigma^H} \right)^{1/(1-\sigma^H)}, \qquad \perp \square H \tag{24}$$

where $\square w$ is the wage and P^{TH} is the marginal cost of commuting trips. Then, the conditional demands for trips and aggregate labor are given by:

$$Q^{TH} = \left(\beta^{TH}\right)^{\sigma^H} \left(P^{TH}\right)^{-\sigma^H} \square w^{\sigma^H} \square H \tag{25}$$

$$H = \left(\beta^H\right)^{\sigma^H} \theta^{-\sigma^H} \square w^{\sigma^H} \square H \tag{26}$$

The zero profit conditions corresponding to the cost-minimizing allocation of trips by mode and link in eqs. (5) and (7) are

$$P_i^{TC} = \left(\sum_l \sum_m \left(\gamma_{i,l,m}^{TC}\right)^{\sigma_i^{TC}} \left(p_{i,l,m}^{TC}\right)^{1-\sigma_i^{TC}} \right)^{1/(1-\sigma_i^{TC})}, \qquad \perp Q_i^{TC} \tag{27}$$

$$P^{TH} = \left(\sum_l \sum_m \left(\gamma_{l,m}^{TH}\right)^{\sigma^{TH}} \left(p_{l,m}^{TH}\right)^{1-\sigma^{TH}} \right)^{1/(1-\sigma^{TH})}, \qquad \perp Q^{TH} \tag{28}$$

where $p_{i,l,m}^{TC}$ and $p_{l,m}^{TH}$ are the marginal costs of trips on a given mode-link alternative incurred by the representative agent in order to consume good *i* and journey to work, respectively. The associated conditional demands for trips by link, mode and commodity are:

$$q_{i,l,m}^{TC} = \left(\gamma_{i,l,m}^{TC}\right)^{\sigma_i^{TC}} \left(p_{i,l,m}^{TC}\right)^{-\sigma_i^{TC}} \left(P_i^{TC}\right)^{\sigma_i^{TC}} Q_i^{TC}, \tag{29}$$

$$q_{l,m}^{TH} = \left(\gamma_{l,m}^{TH}\right)^{\sigma^{TH}} \left(p_{l,m}^{TH}\right)^{-\sigma^{TH}} \left(P^{TH}\right)^{\sigma^{TH}} Q^{TH}. \tag{30}$$

Turning now to the firms in the economy, cost minimization by producers of non-transportation goods and services in eq. (10) results in the following zero-profit condition:

$$P_j = \left(\sum_i \left(\delta_{i,j}^{NT}\right)^{\sigma_j^{NT}} \left(\overline{P}_i\right)^{1-\sigma_j^{NT}} + \left(\delta_{H,j}^{NT}\right)^{\sigma_j^{NT}} \left(\overline{w}\right)^{1-\sigma_j^{NT}} + \left(\delta_{K,j}^{NT}\right)^{\sigma_j^{NT}} r^{1-\sigma_j^{NT}} \right)^{1/(1-\sigma_j^{NT})}, \qquad \perp Y_j \tag{31}$$

where P_j is the producer price of each non-transport commodity and r is the capital rental rate. The zero-profit condition for logistics firms in eq. (13) takes a somewhat different form, owing to the CET specification of production. In particular, transportation services are not traded, and so do not have an explicit price within the model. Producers therefore equate the marginal revenue from revenue-maximizing allocation of trips in (14) to the marginal cost from cost-minimizing transport service production in (13):

$$\sum_l \left(\sum_j \left(\mu_{j,l,m}^{TF}\right)^{\psi_m^T} \left(Z_{l,m} p_{j,l,m}^{TF}\right)^{1-\psi_m^T} + \sum_i \left(\mu_{i,l,m}^{TC}\right)^{\psi_m^T} \left(Z_{l,m} p_{i,l,m}^{TC}\right)^{1-\psi_m^T} + \left(\mu_{l,m}^{TH}\right)^{\psi_m^T} \left(Z_{l,m} p_{l,m}^{TH}\right)^{1-\psi_m^T} \right)^{1/(1-\psi_m^T)}$$

$$= \left(\sum_i \left(\delta_{i,m}^{T}\right)^{\sigma_m^T} \left(\overline{P}_i\right)^{1-\sigma_m^T} + \left(\delta_{H,m}^{T}\right)^{\sigma_m^T} \left(\overline{w}\right)^{1-\sigma_m^T} + \left(\delta_{K,\boldsymbol{m}}^{T}\right)^{\sigma_m^T} r^{1-\sigma_m^T} \right)^{1/(1-\sigma_m^T)} \qquad \perp Y_m \tag{32}$$

The left-hand side of the foregoing expression clearly demonstrates that the impact of congestion is identical to a tax on trips that is differentiated by link. This result turns out to be very useful, because it enables the level of congestion, $Z_{l,m}$, to be modeled as an endogenous, nonlinear tax. We elaborate on this point below.

The associated conditional demands for inputs of intermediate commodities, labor and capital are found by applying Shepard's lemma to the right-hand sides of (31) and (32):

$$\overline{X}_{i,j} = \left(\delta_{i,j}^{NT}\right)^{\sigma_j^{NT}} \left(\overline{P}_i\right)^{-\sigma_j^{NT}} \left(P_j\right)^{\sigma_j^{NT}} Y_j, \tag{33}$$

$$\overline{X}_{i,m} = \left(\delta_{i,m}^{T}\right)^{\sigma_m^T} \left(\overline{P}_i\right)^{-\sigma_m^T} \left(\sum_i \left(\delta_{i,m}^{T}\right)^{\sigma_m^T} \left(\overline{P}_i\right)^{1-\sigma_m^T} + \left(\delta_{H,m}^{T}\right)^{\sigma_m^T} \left(\overline{w}\right)^{1-\sigma_m^T} + \left(\delta_{K,m}^{T}\right)^{\sigma_m^T} r^{1-\sigma_m^T} \right)^{\sigma_m^T/(1-\sigma_m^T)} Y_m, \tag{34}$$

$$\overline{h}_j = \left(\delta_{H,j}^{NT}\right)^{\sigma_j^{NT}} \left(\overline{w}\right)^{-\sigma_j^{NT}} \left(P_j\right)^{\sigma_j^{NT}} Y_j, \tag{35}$$

$$\overline{h}_m = \left(\delta_{H,m}^{T}\right)^{\sigma_m^T} \left(\overline{w}\right)^{-\sigma_m^T} \left(\sum_i \left(\delta_{i,m}^{T}\right)^{\sigma_m^T} \left(\overline{P}_i\right)^{1-\sigma_m^T} + \left(\delta_{H,m}^{T}\right)^{\sigma_m^T} \left(\overline{w}\right)^{1-\sigma_m^T} + \left(\delta_{K,m}^{T}\right)^{\sigma_m^T} r^{1-\sigma_m^T} \right)^{\sigma_m^T/(1-\sigma_m^T)} Y_m, \tag{36}$$

$$k_j = \left(\delta_{K,j}^{NT}\right)^{\sigma_j^{NT}} r^{-\sigma_j^{NT}} \left(P_j\right)^{\sigma_j^{NT}} Y_j, \tag{37}$$

$$k_m = \left(\delta_{K,m}^T\right)^{\sigma_m^T} r^{-\sigma_m^T} \left(\sum_i \left(\delta_{i,m}^T\right)^{\sigma_m^T} \overline{P}_i^{1-\sigma_m^T} + \left(\delta_{H,m}^T\right)^{\sigma_m^T} \overline{w}^{1-\sigma_m^T} + \left(\delta_{K,m}^T\right)^{\sigma_m^T} r^{1-\sigma_m^T} \right)^{\sigma_m^T/(1-\sigma_m^T)} Y_m. \quad (38)$$

As well, the associated conditional supplies for trips are found by invoking Shepard's lemma on the left-hand side of (32):

$$q_{j,l,m}^{TF} = Z_{l,m}^{1-\psi_m^T} \left(\mu_{j,l,m}^{TF}\right)^{\psi_m^T} \left(p_{j,l,m}^{TF}\right)^{-\psi_m^T} \left(\sum_i \left(\delta_{i,m}^T\right)^{\sigma_m^T} \overline{P}_i^{1-\sigma_m^T} + \left(\delta_{H,m}^T\right)^{\sigma_m^T} \overline{w}^{1-\sigma_m^T} + \left(\delta_{K,m}^T\right)^{\sigma_m^T} r^{1-\sigma_m^T} \right)^{\psi_m^T/(1-\sigma_m^T)} Y_m, \quad (39)$$

$$q_{i,l,m}^{TC} = Z_{l,m}^{1-\psi_m^T} \left(\mu_{i,l,m}^{TC}\right)^{\psi_m^T} \left(p_{i,l,m}^{TC}\right)^{-\psi_m^T} \left(\sum_i \left(\delta_{i,m}^T\right)^{\sigma_m^T} \overline{P}_i^{1-\sigma_m^T} + \left(\delta_{H,m}^T\right)^{\sigma_m^T} \overline{w}^{1-\sigma_m^T} + \left(\delta_{K,m}^T\right)^{\sigma_m^T} r^{1-\sigma_m^T} \right)^{\psi_m^T/(1-\sigma_m^T)} Y_m, \quad (40)$$

$$q_{l,m}^{TH} = Z_{l,m}^{1-\psi_m^T} \left(\mu_{l,m}^{TH}\right)^{\psi_m^T} \left(p_{l,m}^{TH}\right)^{-\psi_m^T} \left(\sum_i \left(\delta_{i,m}^T\right)^{\sigma_m^T} \overline{P}_i^{1-\sigma_m^T} + \left(\delta_{H,m}^T\right)^{\sigma_m^T} \overline{w}^{1-\sigma_m^T} + \left(\delta_{K,m}^T\right)^{\sigma_m^T} r^{1-\sigma_m^T} \right)^{\psi_m^T/(1-\sigma_m^T)} Y_m. \quad (41)$$

The zero profit condition corresponding to the cost-minimizing allocation of freight trips by mode and link in eq (12) is:

$$P_j^{TF} = \left(\sum_l \sum_m \left(\gamma_{j,l,m}^{TF}\right)^{\sigma_j^{TF}} \left(p_{j,l,m}^{TF}\right)^{1-\sigma_j^{TF}} \right)^{1/(1-\sigma_j^{TF})}, \qquad \perp Q_j^{TF} \quad (42)$$

The associated conditional demands for freight trips by link, mode and commodity are:

$$q_{j,l,m}^{TF} = \left(\gamma_{j,l,m}^{TF}\right)^{\sigma_j^{TF}} \left(p_{j,l,m}^{TF}\right)^{-\sigma_j^{TF}} \left(P_j^{TF}\right)^{\sigma_j^{TF}} Q_j^{TF}. \quad (43)$$

Finally, using eq. (11), the consumer price of non-transport commodities, $\overline{P}_i$, is given by the following zero-profit condition:

$$\overline{P}_i = P_i^{TF} / \chi_i^{TF} + P_i, \qquad \perp Y_i \quad (44)$$

whose first term indicates the transportation margin. The associated demands are

$$Q_i^{TF} = \overline{Y}_i / \chi_i^{TF}, \quad (45)$$

$$Y_i = \overline{Y}_i. \quad (46)$$

7.2. Market clearance conditions

Substituting eqs. (35), (36) and (46) into (16) and (37), (38) and (46) into (17) yields the supply-demand balances for labor and capital

$$\overline{H} = \sum_j \left(\delta_{H,j}^{NT}\right)^{\sigma_j^{NT}} \overline{w}^{-\sigma_j^{NT}} \left(P_j\right)^{\sigma_j^{NT}} Y_j + \sum_m \left(\delta_{H,m}^{T}\right)^{\sigma_m^{T}} \overline{w}^{-\sigma_m^{T}} \left(\sum_i \left(\delta_{i,m}^{T}\right)^{\sigma_m^{T}} \overline{P}_i^{1-\sigma_m^{T}} + \left(\delta_{H,m}^{T}\right)^{\sigma_m^{T}} \overline{w}^{1-\sigma_m^{T}} + \left(\delta_{K,m}^{T}\right)^{\sigma_m^{T}} r^{1-\sigma_m^{T}} \right)^{\sigma_m^{T}/(1-\sigma_m^{T})} Y_m. \perp \overline{w} \quad (47)$$

$$K = \sum_j \left(\delta_{K,j}^{NT}\right)^{\sigma_j^{NT}} r^{-\sigma_j^{NT}} \left(P_j\right)^{\sigma_j^{NT}} Y_j + \sum_m \left(\delta_{K,m}^{T}\right)^{\sigma_m^{T}} r^{-\sigma_m^{T}} \left(\sum_i \left(\delta_{i,m}^{T}\right)^{\sigma_m^{T}} \overline{P}_i^{1-\sigma_m^{T}} + \left(\delta_{H,m}^{T}\right)^{\sigma_m^{T}} \overline{w}^{1-\sigma_m^{T}} + \left(\delta_{K,m}^{T}\right)^{\sigma_m^{T}} r^{1-\sigma_m^{T}} \right)^{\sigma_m^{T}/(1-\sigma_m^{T})} Y_m. \quad \perp r \quad (48)$$

Substituting eqs. (33), (34) and (46) into (15) yields the market clearance condition for delivered non-transport commodities:

$$\overline{Y}_i = \sum_j \left(\delta_{i,j}^{NT}\right)^{\sigma_j^{NT}} \left(\overline{P}_i\right)^{-\sigma_j^{NT}} \left(P_j\right)^{\sigma_j^{NT}} Y_j + \sum_m \left(\delta_{i,m}^{T}\right)^{\sigma_m^{T}} \left(\overline{P}_i\right)^{-\sigma_m^{T}} \left(P_m\right)^{\sigma_m^{T}} Y_m + \left(\beta_i^{C}\right)^{\sigma^{U}} \left(\overline{P}_i\right)^{-\sigma^{U}} \hat{P}_i^{\sigma^{U}} \hat{C}_i, \quad \perp \overline{P}_i \quad (49)$$

while the corresponding equation for non-transport firms' outputs is given by (46):

$$Y_i = \overline{Y}_i. \quad \perp P_i \quad (46')$$

We note that a similar condition for the services produced by transportation firms (Y_m) does not exist, as we assume that there are only markets for trips.

The balance between supply and demand for the final use of the commodity-retail transport aggregate is given by (19), and is complementary to the composite final commodity price:

$$\hat{C}_i = \alpha_i^{\sigma^{U}} \hat{P}_i^{-\sigma^{U}} \varepsilon^{\sigma^{U}} U, \quad \perp \hat{P}_i \quad (19')$$

which enables us to specify analogous conditions for the retail, commuting and freight mobility aggregates, given by (22), (25) and (45):

$$Q_i^{TC} = \left(\beta_i^{TC}\right)^{\sigma^{C}} \left(P_i^{TC}\right)^{-\sigma^{C}} \hat{P}_i^{\sigma^{C}} \hat{C}_i, \quad \perp P_i^{TC} \quad (22')$$

$$Q^{TH} = \left(\beta^{TH}\right)^{\sigma^H} \left(P^{TH}\right)^{-\sigma^H} \overline{w}^{\sigma^H} \overline{H}, \qquad \perp P^{TH} \tag{25'}$$

$$Q_j^{TF} = \overline{Y}_j / \chi_j^{TF}. \qquad \perp P_j^{TF} \tag{45'}$$

Supply-demand equilibria for trips, which are complementary to mode- and link-specific marginal trip costs, are found by equating (27) and (39), (28) and (40), and (43) and (41):

$$Z_{l,m}^{1-\psi_m^T} \left(\mu_{j,l,m}^{TF}\right)^{\psi_m^T} \left(p_{j,l,m}^{TF}\right)^{-\psi_m^T} \left(\sum_i \left(\delta_{i,m}^T\right)^{\sigma_m^T} \overline{P}_i^{1-\sigma_m^T} + \left(\delta_{H,m}^T\right)^{\sigma_m^T} \overline{w}^{1-\sigma_m^T} + \left(\delta_{K,m}^T\right)^{\sigma_m^T} r^{1-\sigma_m^T} \right)^{\psi_m^T/(1-\sigma_m^T)} Y_m$$
$$= \left(\gamma_{j,l,m}^{TF}\right)^{\sigma_j^{TF}} \left(p_{j,l,m}^{TF}\right)^{-\sigma_j^{TF}} \left(P_j^{TF}\right)^{\sigma_j^{TF}} Q_j^{TF}, \qquad \perp p_{j,l,m}^{TF} \tag{50}$$

$$Z_{l,m}^{1-\psi_m^T} \left(\mu_{i,l,m}^{TC}\right)^{\psi_m^T} \left(p_{i,l,m}^{TC}\right)^{-\psi_m^T} \left(\sum_i \left(\delta_{i,m}^T\right)^{\sigma_m^T} \overline{P}_i^{1-\sigma_m^T} + \left(\delta_{H,m}^T\right)^{\sigma_m^T} \overline{w}^{1-\sigma_m^T} + \left(\delta_{K,m}^T\right)^{\sigma_m^T} r^{1-\sigma_m^T} \right)^{\psi_m^T/(1-\sigma_m^T)} Y_m$$
$$= \left(\gamma_{i,l,m}^{TC}\right)^{\sigma_i^{TC}} \left(p_{i,l,m}^{TC}\right)^{-\sigma_i^{TC}} \left(P_i^{TC}\right)^{\sigma_i^{TC}} Q_i^{TC}, \qquad \perp p_{i,l,m}^{TC} \tag{51}$$

$$Z_{l,m}^{1-\psi_m^T} \left(\mu_{l,m}^{TH}\right)^{\psi_m^T} \left(p_{l,m}^{TH}\right)^{-\psi_m^T} \left(\sum_i \left(\delta_{i,m}^T\right)^{\sigma_m^T} \overline{P}_i^{1-\sigma_m^T} + \left(\delta_{H,m}^T\right)^{\sigma_m^T} \overline{w}^{1-\sigma_m^T} + \left(\delta_{K,m}^T\right)^{\sigma_m^T} r^{1-\sigma_m^T} \right)^{\psi_m^T/(1-\sigma_m^T)} Y_m$$
$$= \left(\gamma_{l,m}^{TH}\right)^{\sigma^{TH}} \left(p_{l,m}^{TH}\right)^{-\sigma^{TH}} \left(P^{TH}\right)^{\sigma^{TH}} Q^{TH}. \qquad \perp p_{l,m}^{TH} \tag{52}$$

A particularly attractive feature of the model is the fact that the value of time exhibits complementary slackness with respect to the representative agent's time budget constraint. The associated market clearance condition is derived by substituting eqs. (19), (29) and (30) into the representative agent's time budget constraint, (9):

$$\sum_l \sum_m \tau_{l,m} \left(\sum_i \left(\gamma_{i,l,m}^{TC}\right)^{\sigma_i^{TC}} \left(p_{i,l,m}^{TC}\right)^{-\sigma_i^{TC}} \left(P_i^{TC}\right)^{\sigma_i^{TC}} Q_i^{TC} + \left(\gamma_{l,m}^{TH}\right)^{\sigma^{TH}} \left(p_{l,m}^{TH}\right)^{-\sigma^{TH}} \left(P^{TH}\right)^{\sigma^{TH}} Q^{TH} \right)$$
$$+ \left(\beta^H\right)^{\sigma^H} \theta^{-\sigma^H} \overline{w}^{\sigma^H} \overline{H} + \alpha_\Phi^{\sigma^U} \theta^{-\sigma^U} \varepsilon^{\sigma^U} U \le \overline{T}. \qquad \perp \theta \tag{53}$$

Because θ is the Lagrange multiplier on a constraint which takes into account the fully endogenous price, supply and demand responses across the entire spectrum of markets in the economy (as opposed to just transportation), it represents the true general equilibrium value of time. It is also useful to observe that in this expression the mode- and link-specific time costs play the role of differentiated taxes, whose values are endogenous determined by trip volumes according to the capacity restraint formula (1)-(2).

The final market clearance condition is a placeholder equation that specifies the quantity of "utility goods" as the ratio of the representative agent's aggregate income, Ω, to the unit expenditure index. This expression is complementary to unit expenditure:

$$U = \Omega / \varepsilon. \qquad \perp \varepsilon \tag{54}$$

7.3. Income balance conditions and auxiliary variables

Income-expenditure balance is defined by the representative agent's money budget constraint, (8), which is complementary to aggregate income:

$$\sum_i \overline{P}_i \overline{C}_i \leq \theta H + rK. \qquad \perp \Omega \tag{8′}$$

The auxiliary variables in the model are the average trip times by mode and link ($\tau_{l,m}$) in eq. (53) and the congestion penalty parameter ($Z_{l,m}$) in eqs. (32) and (50)-(52). Assuming that $Z_{l,m}$ can be expressed as a parametric function of $\tau_{l,m}$ (e.g., as in Mayeres and Proost, 1997), we may specify two auxiliary equations that are complementary to these variables:

$$\tau_{l,m} = \tau^0_{l,m}\left(1+0.15\left(\frac{\sum_i\left(q^{TC}_{i,l,m}+q^{TF}_{i,l,m}\right)+q^{TH}_{l,m}}{\kappa_{l,m}}\right)^4\right), \qquad \perp \tau_{l,m} \tag{55}$$

$$Z_{l,m}(\tau_{l,m}). \qquad \perp Z_{l,m} \tag{56}$$

7.4. General equilibrium in complementarity format

Given the above, we can now specify the general equilibrium of the economy as follows:

- $3 + 5I + M$ zero profit equations (18), (21), (24), (27)-(28), (31)-(32), (42) and (44) in as many unknown activity variables ($U, \hat{C}_i, \overline{H}, Q^{TC}_i, Q^{TH}, Y_j, Y_m, \overline{Y}_i, Q^{TF}_j$).
- $5 + 5I + (1 + 2I)(L \times M)$ income balance equations (19′), (22′), (25′), (45′)-(46′), and (47)-(54), in as many unknown price variables ($\hat{P}_i, P^{TC}_i, P^{TH}, P^{TF}_j, P_i, \overline{w}, r, \overline{P}_i, p^{TF}_{j,l,r}, p^{TC}_{i,l,m}, p^{TH}_{l,m}, \theta, \varepsilon$)
- A single income balance condition (8′) in one unknown income level (Ω), and
- The $2(L \times M)$ auxiliary constraints (55) and (56) in as many unknown auxiliary variables ($\tau_{l,m}, Z_{l,m}$).

The CGE model consists of the paired, stacked vectors of $9 + 10I + M + (3 + 2I)(L \times M)$ variables, $\mathbf{b} = \text{vec}[U, \hat{C}_i, \overline{H}, Q^{TC}_i, Q^{TH}, Y_j, Y_m, \overline{Y}_i, Q^{TF}_j, \hat{P}_i, P^{TC}_i, P^{TH}, P^{TF}_j, P_i, \overline{w}, r, \overline{P}_i, p^{TF}_{j,l,m}, p^{TC}_{i,l,m}, p^{TH}_{l,m}, \theta, \varepsilon, \Omega, \tau_{l,m} Z_{l,m}]$ and $9 + 10I + M + (3 + 2I)(L \times M)$ equations (18), (21), (24), (27)-(28), (31)-(32), (42)-(44), (19′), (22′), (25′), (45′)-(46′), (47)-(54), (8′), (55)-(56), which we denote $\Xi(\mathbf{b})$. The latter is the excess demand correspondence of the economy. By setting up the model in this way, the economy can be cast as a square system of nonlinear inequalities known as a mixed complementarity problem (Ferris and Pang, 1997; Ferris and Kanzow, 2002):

$$\Xi(\mathbf{b}) \geq \mathbf{0}, \qquad \mathbf{b} \geq \mathbf{0}, \qquad \mathbf{b}'\,\Xi(\mathbf{b}) = \mathbf{0},$$

which is straightforward to express and solve using computational tools such as the MPSGE subsystem (Rutherford, 1999) for GAMS (Brooke et al, 1998) in conjunction with the PATH solver (Dirkse and Ferris, 1995).

NOTES

1. The theoretical justification rests on the assumption of perfect competition. Venables and Gasiorek (1999) develop a theoretical framework for assessing impacts under the assumption of monopolistic competition.

2. The main empirically-derived inputs employed by Conrad-Heng are benchmark estimates of the transportation and infrastructure capital stocks, the aggregate cost of congestion, and the elasticity of congestion with respect infrastructure spending, which, along with assumed industry weights, ω_j, permits the *a* parameter in the congestion function to be calibrated.

3. Parry and Bento (2001) distribute trips equally among modes, while in Parry and Bento (2001) trips are allocated 33 percent to each of peak-period freeway and public transit and 17 percent each to back roads and off-peak freeway travel.

BIBLIOGRAPHY

Anderson, William P. and T.R. Lakshmanan (2007) "Infrastructure and Productivity: What are the Underlying Mechanisms?" (eds) Charlie Karlsson, William P. Anderson, Börje Johansson, and Kiyoshi Kobayashi. in *The Management and Measurement of Infrastructure: Performance, Efficiency and Innovation*. Edward Elgar, UK.

Brooke, A., D. Kendrick, A. Meeraus and R. Raman (1998). GAMS: A User's Guide, Washington DC: GAMS Development Corp.

Dirkse, S.P. and M.C. Ferris (1995). The PATH Solver: A Non-Monotone Stabilization Scheme for Mixed Complementarity Problems, *Optimization Methods and Software* 5: 123-156.

Evans, A.W. (1992). Road congestion pricing: When is it a good policy? Journal of Transport Economics and Policy 26: 213-43.

Fang, B., X. Han, S. Okubo and A.M. Lawson (2000). U.S. Transportation Satellite Accounts for 1996, *Survey of Current Business* 80: 14-22.

Ferris, M.C. and C. Kanzow (2002). Complementarity and Related Problems, in P.M. Pardalos and M.G.C. Resende (eds.), *Handbook of Applied Optimization*, New York: Oxford University Press, 514-530.

Ferris, M.C., A. Meeraus and T.F. Rutherford (1999). Computing Wardropian Equilibria in a Complementarity Framework, *Optimization Methods and Software* 10: 669-685.

Ferris, M.C. and J.S. Pang (1997). Engineering and Economic Applications of Complementarity Problems, *SIAM Review* 39(4): 669-713.

Fishlow, Albert (1965). American Railroads and the Transformation of the Antebellum Economy, Harvard University Press, Cambridge, MA.

Fogel, , R.W. (1964) Railroads and American Economic Growth: essays in econometric history, John Hopkins University Press, Baltimore.

Forkenbrock, D.J. and N.S. Foster (1990) Economic benefits of corridor investment projects, *Transportation Research*, 24A(3): 303-312.

Gunasekara, K., W.P. Anderson, and T. R. Lakshmanan (2007 forthcoming) "Highway Induced Development: Evidence from Sri Lanka", *World Development*.

Haynes, Kingsley and Kenneth J. Button (2001) Transportation systems and economic development, Chapter 16 in Kenneth J. Button and David A. Hensher (eds.) *Handbook of Transportation Systems and Traffic Control*, Amsterdam: Pergamon.

Lakshmanan, T.R. and W. Anderson (2002). A White Paper on "Transportation Infrastructure, Freight Services Sector, and Economic Growth", prepared for the U.S. Department of Transportation, Federal Highway Administration.

Lakshmanan, T.R. and W. Anderson (2007). "Transport's Role in Regional Integration Processes" *Market Access, Trade in Transport Services and Trade Facilitation*, Round Table 134. OECD-ECMT, Paris, pp. 45-71.

Lakshmanan, T.R. and C.-I. Hua (1983). A Temporal-Spatial Theory of Consumer Behavior, *Regional Science and Urban Economics* 13: 341-361.

Mackie, Peter and John Nellthorp (2001). Cost-benefit analysis in transport, Chapter 10 in Kenneth. J. Button and David A. Hensher (eds.) *Handbook of Transportation Systems and Traffic Control*, Amsterdam: Pergamon.

Mayeres, I. and S. Proost (1997). Optimal Tax and Public Investment Rules for Congestion Type of Externalities, *Scandinavian Journal of Economics* 99(2): 261-279.

Parry, I.W.H. and A.M. Bento (2001). Revenue Recycling and the Welfare Effects of Road Pricing, *Scandinavian Journal of Economics* 103: 645-671.

Parry, I.W.H. and A.M. Bento (2002). Estimating the Welfare Effect of Congestion Taxes: The Critical Importance of other Distortions within the Transport System, *Journal of Urban Economics* 51: 339-365.

Rutherford, T.F. (1995). Extensions of GAMS for Complementarity Problems Arising in Applied Economic Analysis, *Journal of Economic Dynamics and Control* 19(8): 1299-1324.

Rutherford, T.F. (1999). Applied General Equilibrium Modeling with MPSGE as a GAMS Subsystem: An Overview of the Modeling Framework and Syntax, *Computational Economics* 14: 1-46.

Sue Wing, I. (2004). Computable General Equilibrium Models and Their Use in Economy-Wide Policy Analysis, MIT Joint Program on the Science & Policy of Global Change Technical Note No. 6, Cambridge MA.

Venables, Anthony J. and Michael Gasiorek (1999) *Welfare Implication of Transport Improvement in the Presence of Market Failure*, Report to the Standing Committee on Trunk Road Assessment, London: Department of Environment Transportation and the Regions.

PROGRESS AND CHALLENGES IN THE APPLICATION OF ECONOMIC ANALYSIS FOR TRANSPORT POLICY AND DECISION MAKING:

Concluding Comments for the Research Roundtable on Infrastructure Planning and Assessment Tools

Glen E. WEISBROD and
Brian Baird ALSTADT
Economic Development Research Group, Inc.
Boston, United States

SUMMARY

1. INTRODUCTION: RESEARCH DIRECTIONS AND POLICY ASSESSMENT NEEDS 186

2. WHAT DO WE MEAN BY "WIDER" EFFECTS? 187

3. CLASSIFICATION OF PREDICTIVE TRANSPORT ECONOMIC MODELS 187

4. MODELING IMPLICATIONS OF RECENT RESEARCH 191

5. METHODOLOGICAL ENHANCEMENTS NEEDED FOR POLICY EVALUATION 193

NOTES 195

BIBLIOGRAPHY 196

Boston, September 2007

ABSTRACT

This concluding paper discusses key aspects of the five research papers presented at this Roundtable in terms of their policy applications. It notes problems concerning how policy makers make use of economic analysis findings, and then summarizes the breadth of macro-, meso- and micro-economic methods in terms of their predictive use for infrastructure assessment and planning. It then examines tradeoffs and limitations among all the methods that affect their policy application, and it identifies directions needed to enhance the applicability of future economic models for policy makers.

1. INTRODUCTION: RESEARCH DIRECTIONS AND POLICY ASSESSMENT NEEDS

Over time, policy-makers have seen research on transport-economic interactions evolve to become increasingly sophisticated in the breadth of interactions being recognized. Yet it is not the proliferation of complexity that policy makers seek, but rather, better coverage of applicable situations and more accuracy in findings and applicability for policy appraisal. In that respect, there are two important elements of this evolution of research that are highlighted by the OECD "Research Round Table on Macro-, Meso- and Micro-Infrastructure Planning and Assessment Tools."

- *Value of Different Spatial Perspectives* – One element of this research evolution is a more explicit recognition that the nature of transport problems and their interactions with the economy can appear different when viewed from alternative perspectives – the macro scale of nations, the meso scale of metropolitan areas or the micro scale of local communities. The effects of trade flows, agglomeration economies and spatial spillovers each tend to emerge as particularly important at a different level of spatial focus.

- *Importance of Recognizing Wider Effects* – A second element of this research evolution is the growing appreciation that the effects of transport on the economy can be significantly "wider" than has been recognized by traditional transport appraisal methods. The implication is that appraisal techniques need to be expanded to recognize broader interactions of transport systems and economic systems, such that they can enlarge, diminish or otherwise change our measurement of the economic benefits arising from our transport investments.

From the perspective of policy makers, these two elements of research progress are necessary and important, but they are still insufficient to enable better transport investment decisions. There are at least two additional needs. One is the need for models with adequate "policy levers." Whereas researchers often look for universal relationships that enable broad generalizations about the magnitude of economic effects, policy makers often seek differentiators that can help them distinguish among alternative policies or investments. So while researchers may bemoan a lack of consensus about whether economic spillover effects of highway investment are positive or negative, policy makers may see that both findings can apply in different situations and they may seek information to help make those differentiations. Similarly, while researchers may struggle to reconcile different findings on the importance of agglomeration economies, policy makers may seek to distinguish the conditions under which such effects actually become important. From the viewpoint of policy analysis, an unfortunate reality today is that many past research studies have not adequately differentiated the types of policies or situations in which they were meant to apply. The result, not surprisingly, is misinterpretation via overgeneralization of research results by both proponents and opponents of transport projects and policies.

The other need of policy makers is for economic models that can help improve the applicability of benefit-cost appraisal for decision making. The recognition of wider "external" benefits is critically important in accomplishing this objective. However, as we shift between macro and micro levels of spatial and economic perspective, we may also see shifts in our definitions of who is the "user" or "decision-maker" (e.g., vehicle drivers, travelers, commodity shippers and receivers, or larger industry units) and what constitutes so-called "wider effects." As a result, while we commonly refer to economic development or economic reorganization as externalities with regard to the effects of transport investment, they can actually be core motivations rather than just side effects of some projects or policy interventions.

Given these policy interests, the current research on wider economic effects of transport investment can indeed be quite relevant for decision makers. This paper reviews both the progress that is being made and the challenges that remain in applying economic research findings for transport policy and investment decisions. First, we review definitions of what constitute wider economic effects. Then we classify the different perspectives inherent in different types of economic modeling tools and methods of policy analysts, noting how they focus on different types of externalities and wider effects. Finally, we discuss limitations and challenges confronting the use of these economic modeling approaches for policy analysis.

2. WHAT DO WE MEAN BY "WIDER" EFFECTS?

The question, "what are the wider benefits of transport investment?" begs a follow-up: "wider than what?" Among the authors who have prepared papers for this Forum, these related questions are met with varying interpretations. Cohen (2007), for example, considers that "'wider' benefits refer to the "benefits beyond the geographic region in which the investment is undertaken." (p. 2) He then reviews empirical tools and results on wider "spatial spillover" effects. Others, such Graham (2007), discuss wider impacts as those that "are typically not captured in a standard cost-benefit appraisal" (p. 1). More specifically, he presents methods of expanding impact measures to include the productivity effects of agglomeration. Sue Wing, Anderson, and Lakshmanan (2007) interpret "wider" to mean the degree to which the mechanisms of economic adjustment are endogenized in the analytic process. Models such as they present in their paper, "provide a more complete [wider] picture of the economic impacts of infrastructure" (p. 2) For Johansson (2007), "wider" is interpreted simultaneously as the breadth of geographic scale and the inclusion of inter-urban network effects into modeling. He discusses ways in which access patterns can shift economic behavior and spatial organization *between and among* urban centers in functional urban regions.

Finally, Vickerman (2007) reinforces the ideas of the other authors by reviewing recent research with the goal of reconciling the "standard" benefit/cost approach with macroeconomic findings. He suggests that the standard analysis may be widened to include several phenomena, including spatial externalities, agglomeration, and firm-level effects (input substitution). More generally, he suggests that benefit/cost work can be expanded beyond the (unnecessarily narrow) market for transport to include the broader markets for activities that use transport.

For even this limited survey, the diversity of responses to the question of "wider impacts" is reassuring, and each paper helps to broaden our understanding of the relationship between transport and economic interactions. More importantly, these papers make more explicit the shortcomings of current appraisal techniques, and they identify ways to restructure future methods to incorporate this broader understanding.

3. CLASSIFICATION OF PREDICTIVE TRANSPORT ECONOMIC MODELS

Empirical analysis and statistical studies also provide a foundation for the development of *ex ante* models and other appraisal techniques that support policy and investment decision making. Indeed, existing predictive modeling methods represent a range of different macro-, meso- and micro-level perspectives that

reflect various elements of these "wider effects." Yet across that range of views, there are two consistent tradeoffs:

1. *Precision Tradeoff* – Models with greater precision along one dimension of effect (such as spatial or industrial detail) tend to have less precision along other dimensions, and
2. *Complexity Tradeoff* – Models with greater complexity and breadth of effects tend to require a greater amount of simplifying assumptions that also constrain their realism.

These types of tradeoffs tend to occur across all types of models. They do not necessarily undermine the usefulness of predictive models, but they do highlight the importance of continuing research to improve the accuracy and usefulness of such models for policy and investment decision making. To understand these relationships, it is useful to briefly review the breadth of *ex ante* appraisal techniques and models, the tradeoffs they embody, and how they have evolved over time. The review shows that every type of modeling approach and perspective has a different set of inherent advantages and inherent limitations.

Interaction of Transportation and Economic Models

Following the introduction of computers, the 1960s and 1970s saw the development of several useful tools. Among the most important of these were travel demand models and input-output models. Travel demand models greatly facilitated impact appraisal because they provided a method of simulating supply-demand relationships in the market for transport *at the level of the individual traveler*. These models were very conducive to benefit/cost calculations because they provided user-level metrics (travel time, travel cost) that could easily be converted into benefits on a project-specific basis.

Input-Output models were also extremely useful for policy analysis. They simulated the matrix of inter-industry interactions for one or more regions, and therefore provided a method for assessing the macroeconomic impacts *at the level of the specific industry*. Moreover, the macro-scale input-output framework was seen to complement the micro-scale travel demand model, because it predicted the economy-wide impacts of travel cost changes and project-related spending. Projects that used both could therefore predict a wide range of likely outcomes at a variety of scales.

Although these models represented great improvements in appraisal techniques, early generations were rather limited. Travel models, for example, relied on overly simple assumptions such as fixed trip matrices, straight-line growth in total demand, and simple assignment methods based primarily on travel times. Input-output models were limited as well, particularly because they were non-spatial and did not account for the effect of transport instrumentally, but only as commodity produced by a single sector. The result of these shortcomings was that benefit/cost appraisals were "agnostic" of wider macroeconomic interactions, just as region-wide economic impact assessments were naïve to changes in travel times and access.

These limitations were recognized and understood by many early researchers, and much progress has been made in addressing them. In particular, micro-level travel models and macro-level economic impact models are now frequently merged into larger "connected" modeling frameworks, or are otherwise mathematically integrated. These developments have blurred the once clear distinction between travel models and economic impact models. One consequence of this trend is that the concepts of *benefit* and *impact* have sometimes also been blurred.

Travel Demand Models

Travel demand models have evolved greatly since their early use. A general view of their evolution is one of relaxing restrictive assumptions and expanding the breadth and realism of the transport market being analyzed. In particular: fixed trip matrices can be replaced with dynamic ones; networks can be made more realistic with respect to traffic flow; traffic assignment techniques can be made using generalized cost

functions and can be stochastic rather than deterministic; models can incorporate multiple modes and trip purposes; and induced travel can be accommodated.

On the other hand, few planning processes today incorporate all of these features. Most economic impact models still use generalized costs that do not distinguish peak from off-peak effects. Most transportation models used in planning practice do not fully distinguish differences in the mix and time sensitivity value of freight moving through different corridors and regions. These shortcomings continue to frustrate business organizations, which believe that the result is a dilution of the apparent benefit of policies and actions that reduce congestion delays at peak times, or congestion at particularly critical locations such as airports, seaports, intermodal rail facilities and international borders.[1]

Land Use-Transport Interaction (LUTI) Models

LUTI models build on improvements to travel demand models by recognizing that over sufficiently long time periods, origin-destination patterns are endogenous to transportation demand. In effect, this improvement merely relaxes an assumption of "standard" travel models – that land use remains constant. LUTI models can vary widely in structure ("integrated" vs. "connected" models), as well as scope. In some applications, travel models interact with land use models only; in others, transport markets interact via social-accounting-matrices with land markets, labor markets, and commodity markets. In the latter case, the model can operate on several scales simultaneously: the input-output framework may operate for a small number of large areas, land use changes may operate on an intermediate scale, and travel demand may operate at the highest level of disaggregation.[2]

An advantage of these types of models are that they make it possible to assess the impacts on transportation projects on business market expansion and dispersion of residential and business locations at a highly detailed spatial level. However, one tradeoff that is commonly made to enable the greater spatial detail of land uses is reliance on less detail in the classification of industries and inter-industry flows associated with those regions. Another tradeoff is that they usually focus on just road system access and travel costs, and usually do not address rail, air or marine modes or specialized freight transportation requirements.

A notable modeling feature of many LUTI models is that the individual markets being simulated (transportation, land use, labor, commodity) are not solved simultaneously, but rather in a step-wise fashion. That aspect may not necessarily compromise their usefulness for planning purposes, but it may have implications for their use in benefit/cost analysis. Because the overall model is comprised of several sub-models that may be calibrated and solved separately (and not simultaneously), estimated benefits across all markets may not always capture or reflect all project-wide benefits. Notwithstanding, LUTI models have been successfully used to estimate economic impacts, with the majority of applications being for single metropolitan areas or states, where detailed spatial data is needed to calibrate dense travel demand networks.

General Equilibrium Models

As opposed to LUTI models, which endogenize broader market behavior by "connecting" separate market simulations via larger frameworks, general equilibrium models endogenize broader market behavior into a unified mathematical framework. These are frequently not solved analytically but rather computationally through iteration, and are therefore also referred to "computable" general equilibrium (CGE) models. As with LUTI models, CGE models vary considerably in their methods and scope, but most are based on a set of simultaneous equations representing supply, demand, equilibrium conditions, and interactions between the markets for transport, land, labor, and commodities. CGE models are typically based on a single- or multi-regional input-output framework, and are therefore not well suited to applications where great spatial detail is necessary. As such, the majority of applied CGE models have worked at the international, national, or inter-metropolitan area level.[3]

While CGE models operate at a coarser level of spatial detail than LUTI models, they can more easily provide multi-modal coverage of transportation conditions and be more detailed in terms of distinguishing industry-specific changes in inter-regional freight shipment costs. They also differ from LUTI models in that they are solved simultaneously. In theory, that allows them to obtain valid estimates of benefits across all markets at once, without double-counting (to the extent that market assumptions remain valid as well). However, one trade-off to the complexity and theoretical rigor of CGE models is the need for simplification of various cost measures and response mechanisms to enable simultaneous equations to be solved. That includes mathematical "tricks" such as iceberg costs that are typically used instead of solving for supply-demand equilibrium at the level of individual links and routes. It also includes reliance on production functions with constant elasticities, even as emerging empirical research is showing the existence of non-linearities and threshold effects in transport impacts relating to economies of scale, agglomeration, supply chain dispersion and spatial spillovers.

Economic Simulation Models

Economic simulation models are software tools available for general use in policy analysis. For transport appraisal, they are distinguished from general equilibrium models in the types of impacts they predict. Accordingly, Sue Wing et al (2007) note: "[I]t is useful to make a distinction between two classes of economic impacts, which we call *static general equilibrium* impacts and *dynamic developmental* impacts" (p. 4, italics added). The first of these reflects the short-term changes in travel, labor, and commodity markets, whereas the latter reflects longer-term endogenous induced impacts such as population and employment migration, input substitution, and changes in household preferences. Some economic simulation models also attempt to predict these additional dynamic impacts down to the county or sub-provincial level.[4]

Whereas CGE models most commonly focus on predicting economic growth, some economic simulation models also attempt to predict time paths of input substitution, housing and labor price shifts, migration shifts and changes in consumer purchasing patterns. Furthermore, this type of model is differentiated from the LUTI approach because it typically operates at a larger (regional or multi-regional) scale, and has a more naïve (less developed) treatment of land use and transportation interactions. That is the tradeoff: a greater detail of economic sectors at the expense of less detailed spatial zones.

For policy analysis, economic simulation models are commonly seen as an improvement over earlier "static" input-output models because they can forecast demographic and labor-force impacts and do so over a time-path. However, for transport appraisal, economic simulation models have limitations similar to CGE models – namely, that they incorporate simplifying assumptions about transport costs. In fact, their added complexity is achieved by adding yet more simplifying assumptions about the elasticities of import substitution, labor cost responses, migration responses and timing of impact adjustments. While there is a clear theoretical basis for including these additional effects, the empirical backing for their values (as model coefficients) is often thin, and simplifying assumptions of linear responses can also be suspect. The applicability of transferring large scale impact responses onto small scale study areas has also been questioned.

Access Models

Another type of model has emerged in policy research to predict economic growth following a transport investment. Access models are typically econometric models that draw from literatures on agglomeration, spatial spillovers, supply chain productivity, and new economic geography to predict the increase in local economic development likely to result from a particular transport investment. They are based on econometric studies showing that economic impacts on business location and attraction are subject to non-linear effects that are beyond traditional impacts of travel time costs and travel expenses, as demonstrated by Johansson (2007). These non-linear factors include economies associated with expanding labor market access, delivery market access and supply chain market access. Besides agglomeration economies of enlarged market access,

some access models also consider economies associated with greater supply chain connectivity to highway networks and intermodal rail, air and marine facilities (Weisbrod, 2007).

Such models tend to work independently of travel demand and macroeconomic adjustment models, and are in effect ad-hoc methods of capturing the economic impacts that each of these models miss in their "traditional" form. Graham (2007) makes this point explicit:

> "A crucial issue here is that agglomeration economies are *externalities*, that is, they arise as a side effect of the activities of firms which have consequences for the wider economy. This is very important from the point of view of transport appraisal because the traditional methods of appraisal based on valuation of travel times do not recognise these types of externalities. For this reason agglomeration effects of transport investment can be classed as *wider economic benefits* because they represent market imperfections that are not accounted for in a standard cost-benefit appraisal (p. 6, emphasis in original)."

Access models are very diverse in nature, and can be used to capture a wide variety of phenomena, but have frequently been used to estimate impacts relating to agglomeration. Johansson (2007) notes that infrastructure properties can be measured in three ways: (1) by the capital value of the investment, (2) by link properties, and (3) network or accessibility properties. The key feature of access models is that they focus on the third measure. Productivity gains or other benefits are thus predicted based on prior empirical work relating changes in these measures to past observed growth. Access models have the benefit of being flexible enough to work with traditional travel demand models but they are subject to a number of limitations. Graham (2007) notes several, including that fact that an access model "does not actually tell us much about where the productivity benefits of agglomeration come from" (p. 16). Similar comments can apply to models of the impact of airport, seaport and rail access improvements on economic growth, and also to some models of the spatial spillover impacts of transportation improvements. In each case, the predicted effects reflect a combination of net productivity gain and spatial transfer of activity (business location shift), but the models often do not distinguish the extent of each element.

4. MODELING IMPLICATIONS OF RECENT RESEARCH

Each of the papers presented at this forum (and the respective fields of research they represent) has implications for the different types of models discussed above.

At first glance, Cohen's (2007) review of production and cost function studies with spatial spillover adjustments might seem to have limited relevance to the predictive policy impact model for reasons identified by Vickerman (2007), who notes that a problem with such an approach is that "it takes no account of the way in which infrastructure is used by the activities within the economy in question" (p. 7). That is, it is blind to the mechanisms by which any measured impacts arise, and therefore has limited application to *ex ante* research. This does not, of course, diminish its importance in conditioning our overall *understanding* of transportation's affect on economic performance, particularly with respect to the existence of spatial spillover effects, but merely limits its applicability as a policy research tool.

However, we identify one very critical implication of this line of work: namely, the importance of addressing spatial autocorrelation in any empirical work. All of the modeling techniques discussed in the previous section must be calibrated to particular geographies in order to be valid for project appraisal. These

calibrations come in many forms, but frequently involve econometric analysis of spatial data. Travel demand models and input-output models, for example, both rely on "gravity models"; LUTI models may incorporate dozens of spatial regressions. In each case, residuals should be tested for spatial autocorrelation, but in practice rarely are. The critical point here, quoting Cohen (2007) is that "spatial autocorrelation implies interdependencies among different localities." (p. 7). However, in calibrating a spatial model, this is precisely what is trying to be captured in the *parameters* (and not among the residuals). Therefore, unidentified spatial autocorrelation is a form of bias in the model and amounts to misspecification. Unfortunately, a survey of applications of *ex ante* appraisal methods previously discussed is nearly void of any consideration of these phenomena.

Graham's (2007) research also has focused implications for certain types of policy analysis. As discussed above, his research outlines one approach to exogenously estimating economic impacts that are "external" to traditional benefit/cost and economic impact methods. He identifies several extensions of this line of research that may improve appraisal techniques, such as increasing the industrial resolution of results, accounting for differential impacts across space, and using generalized travel costs (on multimodal networks) to measure accessibility rather than distance-based measures.

More generally, we recognize that the work of estimating such "externalities" has the dilemma of remaining outside of broader modeling frameworks vs. being endogenized into LUTI, CGE, or economic simulation approaches. On the one hand, separately estimating these impacts is attractive because of the empirical difficulty of doing so, and because impacts can vary substantially from place to place. As such, access models may provide the most accurate estimates of project-specific impact at a localized level of impact. On the other hand, it is also clear that agglomeration impacts have micro-, meso-, and macroeconomic implications that require feedback mechanisms to benefit/cost work and input-output work. This is precisely the point made by Johansson's (2007) paper, which builds access measures into an empirical framework operating on three interrelated geographic levels. It recognizes the importance of distinguishing between local and distant markets, and that changes in infrastructure may affect one, the other, or both. In essence, access measures are the ties that bind local welfare impacts to macroeconomic growth impacts.

However, Johansson's work also reveals that current appraisal techniques may discount the importance of threshold effects and non-linearities when assessing the economic impact of an access improvement. He provides an example of these phenomena as they relate to the labor market. Commuting preferences are shown to vary considerably over different ranges of access to employment, and one source of these non-linearities is that labor markets are local, but are also embedded in larger functional urban regions. His work thereby demonstrates a method of incorporating the effects of agglomeration into predictive models, with the most direct application to the LUTI framework.

The work of Sue Wing et al (2007) touches on the themes raised by the other authors – in particular, the need to account for spatially mobile economic factors, and the need to expand benefit/cost work beyond its narrow view of the transport market only. General equilibrium models are, in principle, a method of doing both. The primary benefit of such models is that they provide for a wide range of economic adjustments across a broad range of markets while preserving the assumptions that underlie benefit/cost analysis. Results therefore reflect gains in consumer and producer surplus (as before), but transportation is treated not only as an isolated market, but also as having an instrumental impact on *all* markets. Despite the tremendous theoretical benefits of this approach, Sue Wing et al (2007) and Vickerman (2007) each note its limitations. In the context of the models discussed above, the most significant limitation of CGE models is that they may be impractical or invalid for analysis at small geographic scales. This limits their use to a small number of very large-scale projects, but does not assist in the vast majority of appraisals focusing on a single network link or node.

Finally, the authors reviewed here collectively raise a critical issue regarding the nature of a project's *benefit* versus its *impact*. In early *ex ante* appraisal work, this distinction was very clear (if somewhat naïve), but the evolution of methods described above has blurred it in many cases (see Alstadt and Weisbrod, 2007). The nature of this blurring follows from Sue Wing et al's (2007) discussion of the traditional benefit/cost analysis.

> "[T]he beauty of [benefit/cost analysis] lies in the theoretical argument that consumer surplus, which is a measure of travelers willingness-to-pay, captures the full range of economic benefits. For example, other measurable benefits, such as property appreciation near the improved facility, are chiefly outcomes of reduced travel time so including them in benefit calculation constitutes double-counting (p. 8)."

A benefit, therefore, is a precise outcome of a change in equilibrium in a well-defined market, as reflected by supply, demand, and internal costs (prices). But each successive improvement of travel modeling techniques has, in essence, expanded the scope of the market under consideration. LUTI models, for example, have expanded the scope by "connecting" related models together. CGE models integrate markets into a unified framework. Business access models (as do estimates of environmental impacts) separately calculate impacts external to the markets discussed above. In each case, the assumptions that underlie the model(s) indicate whether certain benefits may be redundant. The quote above indicates that in the traditional analysis, benefits in rental markets would be redundant to those in the transport market. However, for CGE models, they would not, because prices in one area a function of prices in the other, and markets clear simultaneously. For some LUTI models, the interpretation can be ambiguous, and would depend on the specific nature of how the "connected" models interact.

Moreover, as noted by Vickerman, all the appraisal methods reviewed here measure welfare impacts. Even when precise benefit/cost work is unnecessary or imprecisely determined, LUTI and economic simulation models (as well as traditional input-output models) estimate changes in personal income. Sue Wing et al (2007) have demonstrated a way of potentially reconciling potential differences between welfare as measured by benefit/cost analysis vs. changes in personal income. Namely, by introducing a time constraint on household utility, they can estimate the welfare impacts of travel time changes in the context of a macroeconomic adjustment model.

5. METHODOLOGICAL ENHANCEMENTS NEEDED FOR POLICY EVALUATION

The growing research on wider impacts of transport and multiple levels of spatial analysis is an encouraging direction, as it increases the range of methods available for transportation planning and assessment. The challenge moving forward is to enhance the ability of models to address policy issues across a broad range. To do so, four sets of issues will have to be pursued:

1. *Matching the Spatial Scales of Models and Transportation Policy Issues* – The types of benefit evaluation methods needed for large area, program-wide funding decisions are very different from those needed for local facility design and location decisions. The economic issues are at different spatial scales and the justifiable budgets for appraisal are also of different magnitudes. There are also tradeoffs in the spatial, transport and economic resolution of various models. Thus, it can be appropriate to allow different types of models to be applied to different policy contexts. Such a approach could

provide superior detail and policy sensitivity compared to attempts to develop complex mega models that try to apply the same macroeconomic processes at all possible spatial scales of study.

2. *Recognizing Non-Linear Factors* – The growing research on agglomeration economies is a start towards what are actually a much broader need to recognize non-linear factors and threshold effects that are important for decision-making. For instance, if the question is "how much public investment in infrastructure is the right amount?" then the predictive model should be able to show steep returns from new investment where the current situation is particularly deficient, but diminishing returns from over-building. If the question is "how can a new highway affect the local economy?" then the predictive model should be able to show potentially dramatic impacts from reducing isolation and improving system connectivity, but trigger little impact from small, incremental savings in average travel times even if they affect a large population. Many current models that have constant response elasticities are ill equipped to differentiate these non-linear factors. However, policy makers become suspect when economic models with linear responses purport to show wage rates and population migration shifts occurring from small improvements in transportation conditions.
3. *Recognizing Multi-modal and Inter-modal Factors* – With growing globalization of products, services and supply chains, economic growth is becoming more sensitive to multi-modal freight transportation performance and inter-modal transportation connections. Many current economic models that purport to address returns from transportation investment are actually focused just on highway system performance. Even those that also include rail transport costs often do not capture the special economic consequences of constraining global trade growth and reducing freight reliability due to congestion at marine ports, airports and intermodal rail terminals. For such facilities, the issue is often not high transport costs, but actually decreasing reliability and outright growth constraints. The economic consequences can also be particularly sever for those transportation facilities that serve particularly important gateway and network connectivity functions.
4. *Modeling Policies Affecting Service Quality and Economic Feasibility* – Many economic impact and benefit-cost models represent changes in travel times, safety, frequency, reliability and even market access as changes in a generalized transport cost. Many regional location models represent transport access in terms of time distances. While there is a theoretical clarity to these simplifications, such approaches can be inappropriate for transport projects that are designed to enable activities that were previously not economically feasible due to insufficient market size or insufficient service frequency or deficient service quality. This is most aptly illustrated by cases where transport improvements enable just-in-time production processes that were previously not even possible. In effect, such projects may be changing basic characteristics of available transport modes, or they may be changing the location options for economic growth in certain industries. Failure to allow for such impacts can lead to under-statement of the economic value of associated transportation investments.

The four general classes of issues that were described here represent common concerns of economic developers – that transportation policies can affect multiple modes of travel, the service quality attributes of locations, the feasible of economic activities and threshold effects that can preclude or enable particular forms of economic activity. Ultimately, a common accounting framework is needed to span the wide range of economic impact and benefit-cost studies, making it possible to include recognition of the potential for wider economic benefits while avoiding the pitfall of double-counting. That, in turn, can promote greater convergence of perspectives between transport economists and economic developers. The end result can be an enhanced relevance of models for decision-making, and an enhanced capability for transportation investments to be designed and implemented in ways that maximize productivity and job growth.

NOTES

1. Examples of North American business organizations funding research to emphasis freight issues missed by traditional transportation planning models include the Oregon Business Council, Chicago Metropolis 2020 and Vancouver (BC) Gateway Council.

2. Integrated land use and transportation models vary in their features. Examples include MEPLAN (e.g. Echenique 1994), PECAS (Hunt and Abraham, 2005) and TELUM (Pignataro, 2000).

3. CGE models vary in features and spatial breadth. Examples include the integrated transport-network-multiregional CGE model for Korea (Kim and Hewings, 2003) and PINGO, a spatial CGE model for Norway (Ivanova, 2004).

4. Examples of dynamic simulation models operating at sub-national regional zones include ASTRA (Cambridge Econometrics, 2003) and REMI Policy Insight Model (Treyz, 1993).

BIBLIOGRAPHY

Alstadt, B., and G. Weisbrod (2007). "A generalized approach for assessing the direct user impacts of transportation projects." *Transportation Research Board Annual Meeting,* 2008; publication forthcoming. *http://www.edrgroup.com/edr1/bm~doc/a-generalized-approach-fo.pdf.*

Cambridge Econometrics (2003). *Transport Infrastructure and Policy Macroeconomic Analysis for the EU*, European Commission, 2003.

Cohen, Jeffrey P. (2007). "Wider economic benefits of investment in transport infrastructure," paper prepared for the Roundtable on Macro-, Meso- and Micro-Infrastructure Planning and Assessment Tools, Organization for Economic Co-operation and Development.

Echenique, Marcial H. (1994). "Urban and Regional Studies at the Martine Centre: Its Origin, Its Present, Its Future", *Environment and Planning B*: *Planning and Design*, Volume 21, pp. 157-533.

Graham, Daniel J. (2007). "Agglomeration economies and transport investment," paper prepared for the Roundtable on Macro-, Meso- and Micro-Infrastructure Planning and Assessment Tools, Organization for Economic Co-operation and Development.

Hunt, J.D. and J.E. Abraham (2005). "Design and implementation of PECAS: A generalized system for the allocation of economic production, exchange and consumption quantities"; Chapter 11 in *Foundations of Integrated Land-Use and Transportation Models: Assumptions and New Conceptual Frameworks,* Elsevier, London, pp. 217-238.

Ivanova, Olga (2004). "Evaluation of infrastructure welfare benefits in the Spatial Computable General Equilibrium (SCGE) Framework," Department of Economics, University of Oslo. *http://www.oekonomi.uio.no/seminar/torsdag-v03/ivanova.doc.*

Johansson, Börje (2007). "Transport Infrastructure Inside and Across Urban Regions: Models and Assessment Tools," paper prepared for the Roundtable on Macro-, Meso- and Micro-Infrastructure Planning and Assessment Tools, Organization for Economic Co-operation and Development.

Kim, Euijune and Geoffrey Hewings (2003). "An Application of Integrated Transport Network-Multiregional CGW Model," presented at the 42nd Meeting of the Southern Regional Science Association.

Pignataro, Louis J. et al. (1998). "Transportation Economic and Land Use System", *Transportation Research Record,* #1617, Transportation Research Board.

Sue Win, I., W. Anderson, and T. Lakshmanan (2007). "The broader benefits of transportation infrastructure," paper prepared for the Roundtable on Macro-, Meso- and Micro-Infrastructure Planning and Assessment Tools, Organization for Economic Co-operation and Development.

Treyz, George (1993). *Regional Economic Modeling: A Systematic Approach to Economic Forecasting and Policy Analysis*, Kluwer Academic Publishers.

Vickerman, Roger (2007). "Recent evolution into the wider economic benefits of transport infrastructure investments," paper prepared for the Roundtable on Macro-, Meso- and Micro-Infrastructure Planning and Assessment Tools, Organization for Economic Co-operation and Development.

Weisbrod, Glen (2007). Models to predict the economic development impact of transportation projects: historical experience and new applications. *Annals of Regional Science*, forthcoming. *http://www.edrgroup.com/edr1/bm~doc/models-to-predict-the-eco.pdf.*

LIST OF PARTICIPANTS

Professor T.R. LAKSHMANAN **Chairman**
Director
University of Boston
Center for Transportation Studies
675 Commonwealth ave., 4th Floor
BOSTON, MA 02215
USA

Professor Roger VICKERMAN **Rapporteur**
Director
University of Kent
Centre for European, Regional and
Transport Economics
Keynes College
GB- CANTERBURY, CT2 7NP
United Kingdom

Professor Jeffrey P. COHEN **Rapporteur**
University of Hartford
Barney School of Business
200 Bloomfield Ave
WEST HARTFORD, CT 06117
USA

Dr. Daniel GRAHAM **Rapporteur**
Senior Research Fellow
University of London
Centre for Transport Studies
Civil and Environmental Engineering
Imperial College London
GB- LONDON SW7 2BU
United Kingdom

Prof. Börje JOHANSSON **Rapporteur**
Jönköping University
Jönköping International Business School
PO Box 1026
S-551 11 JÖNKÖPING
Sweden

Professor William P. ANDERSSON **Co-Rapporteur**
University of Boston
Center for Transportation Studies
675 Commonwealth ave., 4th Floor
BOSTON, MA 02215
USA

Mr. Ian SUE WING **Co-Rapporteur**
University of Boston
Center for Transportation Studies
675 Commonwealth ave., 4th Floor
BOSTON, MA 02215
USA

Mr. Brian Baird ALSTADT **Co-Rapporteur**
Economist
Economic Development Research Group, Inc.
2 Oliver St, FL9,
BOSTON, MA 02109
USA

Mr. Glen WEISBROD **Co-Rapporteur**
Economic Development Research Group, Inc.
2 Oliver St, FL9,
BOSTON, MA 02109
USA

Prof. Alex ANAS
Professor of Economics
State University of New York at Buffalo
Dept. of Economics
405 Fronczak Hall
AMHERST, NY 14260
USA

Professor Joseph BERECHMAN
Chairman, Department of Economics
The City College, The City University of New York
160 Convent Ave., NA 5/144
NEW YORK, NY 10031
USA

Prof.Dr. Ulrich BLUM
President
Institut für Wirtschaftsforschung Halle
Kleine Märkerstrasse 8
D-06108 HALLE (Saale)
Germany

Professeur Yves CROZET
Laboratoire d'Economie des Transports (LET)
Université Lumière Lyon 2
MRASH
14 avenue Berthelot
F-69363 LYON Cedex 07
France

Mr. Bruno DE BORGER
University of Antwerp
Prinsstraat 13
B-2000 ANTWERP
Belgium

Mr. Alim DEMCHUK
Head of Department
Ministry of Transport and Communications
Financial Regulations and Social Policy
14 av. Peremogy
UKR-01135 KIEV
Ukraine

Mr. Andrew HAUGHWOUT
Assistant Vice President
Microeconomic and Regional Studies Function
Federal Reserve Bank of New York
33 Liberty Street
NEW YORK, NY 10045
USA

Mr. Gunnar ISACSSON
TEK/VTI
Box 760
S-781 27 BORLÄNGE
Sweden

Mr. Ronald F. KIRBY
Director of Transportation Planning
Metropolitan Washington Council of Governments
777 North Capitol Street, N.E., Suite 300
WASHINGTON, DC 20002-4239
USA

Prof. Kiyoshi KOBAYASHI
Kyoto University
Graduate School of Management
Yoshidahonmachi, Sakyo-ku
J-606-8501 KYOTO
Japan

Professor Peter MACKIE
University of Leeds
Institute for Transport Studies
36 University Road
GB- LEEDS, LS2 9JT
United Kingdom

Ms Ganna MAZUR
Deputy Head, Unit for the bilateral cooperation,
CIS Organizations and International Agreements,
Department for Foreign Economic Relations
Ministry of Transport and Communications
14 av. Peremogy
UKR-01135 KIEV
Ukraine

Professor Michael D. MEYER
Georgia Institute of Technology
School of Civil and Environmental Engineering
790 Atlantic Drive
ATLANTA, Georgia 30332-0355
USA

Professor Catherine J. MORRISON PAUL
University of California, Davis
Department of Agricultural and Resource Economics
One Shields Avenue
DAVIS, CA 95616
USA

Prof. Jan OOSTERHAVEN
University of Groningen
Faculty of Economics
PO Box 800
NL-9700 AB GRONINGEN
The Netherlands

Dr. Wolfgang SCHADE
Sustainability and Infrastructures
Fraunhofer Institute for Systems
and Innovations Research ISI
Breslauer Strasse 48
D-76139 KARLSRUHE
Germany

Mr. Derek SWEET
Transportation Research Board (TRB)
500 5th Street NW
20001 WASHINGTON
USA

Dr. Mary Lynn TISCHER
Director, Commonwealth's Multimodal
Transportation Planning Office
1401 E. Broad Street
RICHMOND, Virginia 23219
USA

Mr. Martin WEISS
Office of Planning, Environment, and Realty
Federal Highway Administration
1200 New Jersey Ave., SE
WASHINGTON, DC 20590
USA

Dr. Karen WHITE
Economist
Federal Highway Administration
1200 New Jersey Avenue, SE, mailstop E83-431
WASHINGTON, DC 20590
USA

OECD-INTERNATIONAL TRANSPORT FORUM SECRETARIAT

JOINT TRANSPORT RESEARCH CENTRE

Mr. Stephen PERKINS
Head of Joint Transport Research Centre of
the OECD and the International Transport Forum
2 rue André Pascal
F-75775 PARIS CEDEX 16
France

Dr. Kurt VAN DENDER
Joint Transport Research Centre of
the OECD and the International Transport Forum
2 rue André Pascal
F-75775 PARIS CEDEX 16
France

Ms. Françoise ROULLET
Joint Transport Research Centre of
the OECD and the International Transport Forum
2 rue André Pascal
F-75775 PARIS CEDEX 16
France

ALSO AVAILABLE

Transport and International Trade. Series ECMT – Round Table 130 (2006)
(75 2006 13 1 P) ISBN 92-821-1338-8

Transport and Decentralisation. Series ECMT – Round Table 131 (2006)
(75 2006 12 1 P) ISBN 92-821-1342-6

Transport Infrastructure Investment and Economic Productivity. Series ECMT – Round Table 132 (2007)
(74 2007 04 1 P) ISBN 978-92-821-0124-7

(De)Regulation of the Taxi Industry. Series ECMT – Round Table 133 (2007)
(74 2007 02 1 P) ISBN 978-92-821-0112-4

Market Access, Trade in Transport Services and Trade Facilitation. Series ECMT – Round Table 134 (2007)
(74 2007 05 1 P) ISBN 978-92-821-0146-9

Transport Infrastructure Charges and Capacity Choice : Self-financing Road Maintenance and Construction. Series ECMT – Round Table 135 (2007)
(74 2007 02 1 P) ISBN 978-92-821-0108-7

Estimation and Evaluation of Transport Costs. Series ECMT – Round Table 136 (2007)
(74 2007 06 1 P) ISBN 978-92-821-0151-3

Transport, Urban Form and Economic Growth. Series ECMT – Round Table 137 (2007)
(74 2007 07 1 P) ISBN 978-92-821-0164-3

Biofuels: Linking Support to Performance. Series ECMT – Round Table 138 (2008)
(75 2008 02 1 P) ISBN 978-92-82-10179-7

Oil Dependence: Is Transport Running out at Affordable Fuel? Series ECMT – Round Table 139 (2008)
(74 2008 03 1 P) ISBN 978-92-82-10121-6

To register for information by email about new OECD publications: www.oecd.org/OECDdirect
For orders on line: www.oecd.org/bookshop
For further information about ITF: www.internationaltransportforum.org

OECD PUBLICATIONS, 2, rue André-Pascal, 75775 PARIS CEDEX 16
PRINTED IN FRANCE
(74 2008 04 1 P) ISBN 978-92-821-0160-5 – No. 56291 2008

140

THE WIDER ECONOMIC BENEFITS OF TRANSPORT

MACRO-, MESO- AND MICRO-ECONOMIC TRANSPORT PLANNING AND INVESTMENT TOOLS

The standard cost-benefit analysis of transport infrastructure investment projects weighs a project's costs against users' benefits. This approach has been challenged on the grounds that it ignores wider economic impacts of such projects. Since there is empirical evidence that these effects can be substantial, relying on the standard approach potentially produces misleading results.

At the International Transport Forum Round Table, leading academics and practitioners addressed these concerns and examined a range of potential approaches for evaluating wider impacts – negative as well as positive. They concluded that for smaller projects, it is better to focus on timely availability of results, even if this means forgoing sophisticated analysis of wider impacts. For larger projects or investment programs, customized analysis of these effects is more easily justifiable. Creating consistent appraisal procedures is a research priority.



www.internationaltransportforum.org

OECD*publishing*
www.oecd.org/publishing

(74 2008 04 1 P1) ISBN 978-92-821-0160-5

