A Dissertation

entitled

Evaluating Efficiency of Transportation Infrastructure:
Effects and Implications for the Spatial Economy

by

Jeffrey J. Eloff

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the
Doctor of Philosophy Degree in Spatially Integrated Social Sciences

Dr. Oleg A. Smirnov, Committee Chair

Dr. Peter S. Lindquist, Committee Member

Dr. David J. Nemeth, Committee Member

Dr. David C. Black, Committee Member

Dr. Mark A. Vonderembse, Committee Member

Dr. Patricia R. Komuniecki, Dean
College of Graduate Studies

The University of Toledo
December 2014
An Abstract of
Evaluating Efficiency of Transportation Infrastructure: Effects and Implications for the Spatial Economy
by
Jeffrey J. Eloff
Submitted to the Graduate Faculty as partial fulfillment of the requirements for the Doctor of Philosophy Degree in Spatially Integrated Social Sciences
The University of Toledo
December 2014

This dissertation is concerned with the effects of publicly provisioned transportation infrastructure investments on the spatial economy. Evaluating the efficiency of these investments has important policy implications for the determination of how funds can best be apportioned. An evaluation framework is proposed that consists of two main parts: the identification of evaluation criteria and the development of practical methods for measuring the effects of transportation infrastructure. The formulated evaluation criteria are related to the goals of economic growth and the allocation of economic activity in space. The implementation of practical methods is concerned with the measurement of difficult-to-estimate effects on the spatial economy. The proposed framework is then applied to various portions of the US economy. First, an examination into the domestic manufacturing industry at the state level is undertaken. Attention is then turned towards the effects of transportation investment on household welfare in the Mid-West. The final portion of the dissertation examines the domestic U.S. aviation industry and provides a methodological framework to valuing infrastructure investment in light of network effects.
For Kirra.
Acknowledgments

This work was completed with the support and partnership of the National Center for Freight and Infrastructure Research and Education (CFIRE), a Tier 1 University Transportation Center (UTC) funded by the US Department of Transportation (USDOT) Research and Innovative Technology Administration (RITA). Additional assistance was provided by the Federal Aviation Administration, via the Airports Cooperative Research Program which is administered by the Transportation Research Board of the National Academies.

This is the culmination of the last few years of my life. None of the exciting work you are about to have the pleasure of reading would have come to fruition without the exceptional guidance of Oleg A. Smirnov and Peter S. Lindquist. The work benefitted a great deal from many hours of their time. Many thanks to you both for your time and dedication. I would also like to thank my committee members for their thoughtful input and assistance throughout this process.
Contents

Abstract iii

Acknowledgments v

Contents vi

List of Tables ix

List of Figures x

1 Introduction 1

1.1 Background . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

1.2 Objective and Scope of Work . . . . . . . . . . . . . . . . . . . . . 4

2 Literature 9

2.1 Public Capital . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

2.1.1 Capital Stock Data Creation . . . . . . . . . . . . . . . . . . . . 16

2.1.2 Production Function Literature . . . . . . . . . . . . . . . . . . 18

2.1.3 Cost Function Literature . . . . . . . . . . . . . . . . . . . . . . . 19

2.1.4 Spatial Dependence . . . . . . . . . . . . . . . . . . . . . . . . . . 20

3 Transportation Infrastructure, Industrial Productivity and Return on Investment: A Spatial Spillover Approach 21

3.1 Overview . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 21

3.2 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 22
List of Tables

3.1 Summary Statistics of Profit Function Input Data. ...................... 38
3.2 Profit Function Estimation Results. .................................... 39
3.3 Profit Function Heteroskedasticity Test ................................. 41
3.4 State Specific Elasticities of $\pi$ with respect to $G$ and $\bar{G}$ ............ 45
3.5 Production Function Summary Statistics .................................. 45
3.6 Production Function Estimation Results ................................ 46
3.7 Production Function Heteroskedasticity Test ............................. 46
4.1 Cost Function Estimation ..................................................... 58
4.2 Cost Function Elasticities .................................................... 59
5.1 Household Dataset Summary ............................................... 74
5.2 Household Estimation Results ............................................. 75
6.1 Yearly Mean By Hub-Size ................................................... 90
6.2 Airport Operations Estimation Results ................................... 97
6.3 Average Direct Effects by Hub Size (OPS) ............................... 98
6.4 Average Total Effects by Hub Size (OPS) ............................... 99
6.5 Average Total Effects by Census Regions (OPS) ....................... 99
6.6 On-Time Percentage Estimation Results .................................. 100
6.7 Average Total Effects by Hub Size (OTP) ............................... 100
List of Figures

3-1 Mean Profit Margin .............................................. 27
3-2 Mean Gross State Product 1997-2010 .......................... 30
3-3 Intermediate Cross-Section Analysis .......................... 35
3-4 Highway Construction Price Index ............................ 36
3-5 Mean Elasticity of $G$ over time. ................................. 40
3-6 $Y$ (Profit Margin) vs. $\hat{Y}$ (Predicted Values) ............. 41
3-7 State Specific Transportation Infrastructure Elasticities. ........ 43
3-8 State-Specific Neighboring Benefits Based on Own-State Investments 44
3-9 $Y$ (Output) vs $\hat{Y}$ (Predicted Output) ....................... 47
3-10 Neighboring-State Transportation Infrastructure Benefits Relative to Own-State 48
3-11 Ratio of Residuals to Transportation Infrastructure Effects ....... 49
4-1 Elasticity of VC with respect to Own-State Transportation Infrastructure 60
4-2 Elasticity of VC with respect to Neighboring-State Transportation Infrastructure .................... 61
4-3 Elasticity of Own-State Transportation Investment with respect to Neighboring-State Transportation Investment 61
4-4 Elasticity of Private Capital with respect to Neighboring-State Transportation Investment 62
4-5 Demand for Non-Production Workers with respect to Own-State Transportation Investment 63
4-6 Demand for Non-Production Workers with respect to Neighboring-State Transportation Investment ........................................ 64

5-1 Preliminary Model Prediction ........................................... 72
5-2 Average Real Per Capita Income Distribution Map ................. 76
5-3 Variance of Residuals Model 5.2 ..................................... 77
5-4 Variance of Residuals Model 5.3 ..................................... 78
5-5 Weighted AADT, 2008 .................................................. 80
5-6 Elasticity of AADT with respect to Per Capita Income ............... 81

6-1 On-Time to T-100 Operations Ratio .................................. 90
6-2 Airport Hub-Size Distribution ........................................ 91
6-3 Mean ln(Capital Stock) ................................................ 93
6-4 Annual Mean Spatial Multiplier ....................................... 94
6-5 Annual Mean Flight - Percent Full .................................. 94
6-6 U.S. Airports With Complete Observations ......................... 96
6-7 Total OTP Effects Flow By Hub-Size ................................ 101
6-8 Distribution of $\bar{OTP}$ Multipliers by Hub Size ..................... 102
6-9 Operations Model Diagnostics ....................................... 103
6-10 $\bar{OTP}$ Model Diagnostics ........................................ 104
6-11 Spatial Multipliers – CDF – Operations & $\bar{OTP}$ .................. 106
6-12 Man-Made Network Connectivity Matrix View .................... 108
6-13 Geo-Spatial Based Connectivity Matrix View ..................... 109
6-14 Total Operations Multiplier Cumulative Probability Distribution (GEV)
  versus Normal .......................................................... 110
6-15 Geo-Spatial Based Multiplier Cumulative Probability Distribution .... 111
6-16 Total Operations Multiplier Density Distribution .................... 112
6-17 Geo-Spatial Based Multiplier Density Distribution .................. 112
C-1 Transportation Infrastructure, Industrial Productivity and Return on Investment: A Spatial Spillover Approach. .......................... 162
C-2 Transportation Infrastructure Usage and Household Welfare. ............ 163
Chapter 1

Introduction

This dissertation studies the effects that transportation infrastructure investments have on various portions of the spatial economy. The contributions of this work stem from an explicit recognition of the spatial variations present in economic benefits associated with transport infrastructure investment and are three-fold in nature. The first examination, represented in chapters 3 and 4, explores interactions between spatially separated economies. This investigation hypothesizes that spatially segregated locations interact economically and that the networks providing the linkage between the interacting economies are important for economic sustainability and growth. By exploring the effects of neighboring economies’ transportation networks, deeper insights into these spatial interactions can be made. Conclusions have the potential to aid policy makers in further developing an understanding of the various implications that decisions related to transportation infrastructure have on the broader economy. Insights from the results of data analysis point to a re-thinking of certain aspects of the funding apportionment process as well as the potential for increased efficiencies arising from strategic interaction at the planning level. Subsequent analysis undertaken in chapter 5 serves as an effort to differentiate the effects of transportation infrastructure related investments received by households as well as industry. The results related to the effects of transportation infrastructure investments on specific industries of
interest and at what geographic scale these productivity benefits may be expected to spread are presented. Accounting for interactions between agents in geographic space via transport infrastructure provides a rich understanding of the complexities associated with the intertwined effects on the broader economy. The document outlines methods to measure these difficult-to-estimate effects between transportation infrastructure and household welfare, as well as between infrastructure investment and private industry. A framework aimed at accounting for the true value of proposed investment projects (in light of network effects) is also outlined. This methodology is applied to airport related infrastructure investments across the U.S. in an attempt to illustrate the benefits of coordinated investment and strategic allocation type policies. By lifting from ground-based transportation networks and focusing on man-made networks, which are intended to directly connect spatially separated locations, further conclusions as to the differences between man-made and naturally occurring networks are made. A method to identify natural versus man-made networks is also outlined and explored.

The work presented herein provides an analysis of robust models focused on the effects of transportation infrastructure and networks on the spatial economy. These models provide empirical measurements of the aforementioned effects as well as outline a road-map of possible analytical avenues for which researchers could follow to conduct valuation experiments. Sensitivity analyses\(^1\) and explorations of the spatial and temporal dynamics related to economic growth, convergence and locational shifts in industrial clusters are also examined.

The dissertation is organized as follows: Section 1.1 provides background information and presents some motivating factors as to why investigations into the effects of transportation infrastructure are important. Section 1.2 details the objectives and

\(^1\)Sensitivity in terms of capital data specifications, model specifications, data transformations, etc.
Section 1.3 provides an overview of the remainder of this thesis.

1.1 Background

Public capital is a vital component of any economy. It contributes not only to the quality of life of a country’s citizens, but also to the productivity of the economy as a whole. It is without doubt that a well developed and well maintained public capital stock is necessary for any highly productive society (Treasury, 2010). In 2008, U.S. public infrastructure ranked 25th (out of 32 OECD nations) in terms of public satisfaction (Treasury, 2010). Broadly speaking, public capital has long been explored for its effects on economic growth, productivity and development, among others. Regional scientists, geographers and economists alike have had an interest in the relationship between public capital and regional output as well as economic and regional development for more than four decades (Costa et al., 1987).

A 2010 study by the Department of Treasury cites that 19 out of 20 Americans are concerned with the state of the U.S. infrastructure, while 84% “support greater investment to address infrastructure problems” (Treasury, 2010). In recent years, over half of all US state and local government annual public capital expenditures have been directed for transportation infrastructures (Bureau, b). The optimization of many business models revolve around better utilization of existing publicly provided transportation infrastructure. Many types of public infrastructure have positive effects on the performance of firms and the broader economy alike. As such, maintaining, updating and replacing an aging transportation infrastructure system is crucial for any economy’s continued success.\(^2\) The spatial variation in the levels of adequacy\(^3\) and performance of transportation infrastructure in the U.S. are of significant value.

\(^2\)The terms infrastructure and capital will be used interchangeably. The author acknowledges that differences exist, however for the purpose of this study the semantics are less than relevant.
\(^3\)Adequacy in terms of quality and quantity
Given current funding pitfalls and an antiquated funding apportionment system for surface transportation, an exploration into the locations of economic benefits arising from transportation infrastructure is important for future economic stability. In hopes of achieving higher levels of efficiency from public expenditures for transportation infrastructure, it is necessary to quantify the effects of these investments both socially and economically, as well as to provide insight into how these effects may propagate through the spatial economy. In doing-so, methods by which to uncover these difficult-to-estimate measurements are presented and applied.

1.2 Objective and Scope of Work

Evaluating the efficiency of transportation infrastructure investments has important policy implications for the determination of how funds can best be apportioned. An evaluation framework is first proposed that consists of two main parts: the identification of evaluation criteria and the development of practical methods for measuring the economic effects of transportation infrastructure. The formulated evaluation criteria are related to the goals of economic growth and the allocation of economic activity in space. The implementation of practical methods is concerned with the measurement of difficult-to-estimate effects on the spatial economy. The proposed framework is then applied to the US manufacturing industry at the state level to examine specific regional benefits in terms of these key criteria. First, the conditions, locations and variations for productive publicly provided transportation infrastructure are shown. Examinations into the dynamics of these impacts are made on industry and what implications these transportation infrastructures can and do have on economic growth. Drawing extensively on the existing literature, key departures from existing methods are shown and compared to past work. After the establishment of the productiveness of transportation infrastructure and an understanding of the spatial variations of the
benefits, further analysis is conducted to separate the effect that publicly provided transportation infrastructures have on industry and households. This section of the dissertation aims to aid in the valuation of urban networks and quantify the level of benefits received by households. This portion of the work takes a key departure from that found in the existing literature in that it examines physical measures of public infrastructure usage at the county level (in the 10 member states of the Mid-America Freight Coalition) rather than monetary investment values. Post-estimation tests are performed at each stage of the analysis and presented to illustrate additional insights uncovered throughout the process. The goals of this portion of the dissertation are to determine the existence of a link between transportation infrastructure usage and household welfare, as well as outline an approach one might take to quantify such effects. As the literature is predominantly concerned with the linkage between road supply and road demand alone, this approach expands upon previous work. While the previous literature unequivocally establishes that an increase in supply of infrastructure leads to an increase in the demand for it (Duranton and Turner, 2011), the findings presented here are indicative of the notion that this increased demand also then leads to a higher standard of living (measured via per capita income).

The final portion of the dissertation examines the interdependencies between infrastructures at spatially separated airports. As the FAA forecasts air traffic growth for U.S. carriers to increase substantially over the next twenty years, it follows that airports will also be subjected to similar increased levels of demand. One component of the solution is expected to come from investments in airport infrastructure. Given current fiscal constraints, the inherent network structure of the National Airspace System (NAS) and the fact that delays and congestions propagate throughout the system, this research proposes that it is more efficient for capital investments to be

---

4 Which, in the literature, are typically conducted at the state level.
5 By 2032, FAA (of Transportation: Federal Aviation Administration, 2012) expects a 90 percent increase in revenue passenger miles, and a 50 percent increase in the number of handled aircraft.
made in an integrated fashion - one which serves to maximize the productivity of the system as a whole as opposed to investments on an airport-by-airport basis. More simply, as traffic between airports continues to increase, if one airport substantially increases its capacity and throughput, is it best for other highly-connected airports to implement upgrades at a similar point in time? Without some congestion mitigation tactics or efficiency-improving investments elsewhere in the system, a portion of these infrastructure improvements may fail to reach their full potential if upgrades merely shift congestion to different nodes on the network. As such, it is the goal of this research to understand current airport investment strategies. By exploring decision making patterns related to infrastructure investments within the system, the first aim is to determine to what extent airports currently engage in coordinated investments, or strategic interactions. By focusing on airports in a network-based framework, inefficiencies in current valuation methods are uncovered. The results of this comprehensive approach also have policy implications; a methodological framework that values investments as they pertain to the aviation network as a whole (instead of on an airport-by-airport basis) therefore is presented.

These insights are uncovered by exploiting the network structure of the NAS in a spatial econometric modeling framework. The data necessary to determine these relationships is available from multiple sources. The U.S. Department of Transportation’s (USDOT) Bureau of Transportation Statistics (BTS) maintains two databases relevant for the study: the T-100 data on origin-destination pairs and the Airline On-Time Performance Database. BTS T-100 provides dynamic measures of connectivity while the Airline On-Time Performance Database outlines origin-destination pairings and airline on-time measures. Additionally, BTS maintains aggregate monthly reports on delay causes for 29 major airports with coverage over 15 major domestic airlines.\(^6\) This data is useful in determining which delays are congestion related, as

\(^6\)As of 2012 data.
opposed to those due to natural or technological problems. Furthermore, FAA data on airport investments provides the necessary information to determine infrastructure investment patterns.

By exploring the parameter coefficients from the results of various spatial panel data models, insights into the extent to which airports coordinate their investments in infrastructure are shown. Where investment decisions are currently being made with less than the ideal level of coordination throughout the system, the results of this research serve to provide policy makers with relevant information that can further guide investments towards more efficient outcomes. Broadly speaking, this approach is theoretically based on the spillover and resource-flow models outlined in K. (2003). The spillover model, framed as a spatial reaction function, explicitly recognizes the interdependencies between a decision variable (e.g. infrastructure stock) for one unit of observation (airport) and all other units in the system (other connected airports). Applied in this context, it allows for the determination of how one airport’s level of capital stock is determined in light of other airports’ given level of stocks. A model of this type has the ability to uncover the extent and presence of strategic interactions in infrastructure investment between connected airports. Alternatively, the resource-flow type model models the decision of one agent (airport) not directly based upon the levels chosen by other agents, but via indirect means. This case would be representative of reactionary investment policies, whereby airports do not strategically invest based upon one another’s investments, but because of the (hypothetically) increased throughput of passengers and planes through an airport due to some infrastructure improving investment.

Imagine for a moment that node A, a major airport with significant congestion issues, implements a new infrastructure improvement project that effectively allows for an increase in its throughput capabilities. Node B, which is heavily connected to node A, must now handle increased levels of traffic due to the now more efficient
infrastructure employed at node A. For node B to maintain the new status quo, it is now necessary to implement upgrades capable of (at least) handling the new capacity. However, the adjustment period during which the new investment is implemented (via these indirect and reactionary means) serves to reduce the overall return on investment from the project conducted at node A. Similarly, from a user welfare perspective, the initial investment was a waste; it does nothing to aggregate welfare until the congestion at node B is reduced and the system moves back towards efficiency. Although this example does not lay out all possible scenarios of the benefits of strategic interaction, it serves to highlight the premise of the argument. This work also provides a method for determining the differences between spatially occurring and man-made networks, as the distinction has important econometric estimation implications.

The next chapter provides an in-depth overview of the literature pertaining to public capital investment and productivity effects.

\[\text{\footnotesize\textsuperscript{7}}\,\text{For convenience, assume that this user just so happens to be engaged in travel between node A and node B.}\]
Chapter 2

Literature

Publicly provided infrastructure has received much attention in the economic literature. Much of the more recent work is concerned with the Public Capital Hypothesis; i.e. that public capital proves beneficial for the goals of economic growth and increased productivity. The underlying notion of this hypothesis is that a solid infrastructure underpinnings enables producing agents in the economy to better utilize their inputs in such a way that maximizes output. The productive capacity of public capital is thought of as being twofold: that infrastructure can be treated as intermediaries in firms’ production processes and that these investments increase the productivity of other private inputs. For transport infrastructure in particular, this notion is quite easy to grasp. Given such high levels of labor mobility in the US, it seems plausible that without sufficient and adequate levels of publicly provisioned transportation infrastructure, the phenomenon of high mobility would not exist\(^1\). Without the ability for labor to freely and easily move throughout geographic space, firms’ labor supply pool\(^2\) may contain an inadequate level of qualified workers and would therefore be under producing. Further supply-side extensions can readily be made: if a region is lacking in adequate transportation infrastructure, the transport of intermediate goods utilized in production may not be possible, forcing firm’s to relocate closer to

\(^1\)At least, not in the levels witnessed in today’s economy
\(^2\)Depending, of course, upon said firm’s location
the source of the producers of the intermediate goods.

Although the literature contains many theoretical and historical contributions (see Blum (1982), Eberts (1986), Dalenberg (1987), among others), empirical interest and findings were long absent until Aschauer’s (Aschauer, 1989) seminal work. This work, which found public capital to explain a significant portion of the US productivity growth slow-down of the 1970’s, catapulted empirical interest in the topic into the mainstream. It is worth noting, however, that a number of empirical works pre-date Aschauer (Aschauer, 1989). See, for example, Mera, Blum (1982), Ratner (1983), Eberts (1986), Dalenberg (1987), Costa et al. (1987), Deno (1988).

The theoretical foundation from which the public capital hypothesis is based is an extension of early work by development economists. Meade (1952) provides one of the earliest known theoretical foundations upon which the majority of subsequent empirical studies are based. His argument essentially lays the groundwork of the idea that unpaid factors of production\(^3\) can aid other industries (or firms) for which no price of the capital was paid. In this light, public capital can be thought of “as an input in the production process of private industry that contributes independently to firms’ output” (Deno, 1988).

Hansen (1965), among the first to directly discuss the effects of public capital on regional development, divides public capital into two parts: “social” and “economic”. Economic overhead capital (EOC), which includes roads, electricity, water treatment facilities, drainage/sewer systems, etc. can be thought of as the foundational public capital used in economic production processes. Social overhead capital (SOC), which includes schools, police/fire services, etc., is generally non-economic in nature and serves to provide support for social institutions. A further distinction between these two types of public capital is also made: consumers of EOC tend to want easy access to the goods (well developed and widely available), while agents are often more

\(^3\)Externally provisioned capital, for example.
willing to travel to utilize SOC facilities. Hansen’s objective is to uncover “possible
determinants of variation in...expenditures” (Hansen, 1965) and he hypothesizes that
economic development potential is related to two factors: the type of public capital
implemented and where in the development cycle the region exists. He determines
that EOC has larger benefits on economic growth in “intermediate” regions (as op-
posed to congested or lagging\(^4\) ones) and that increased provisions of SOC tend to
prove most beneficial when made in “lagging” regions.

Hansen’s argument for the types of investments most beneficial to intermediate
and lagging region’s is simple and intuitive: intermediate regions tend to have more
favorable conditions for development (lower costs, access to skilled labor and raw
materials). Therefore, public investments of the EOC type have the power to spur
further private investment which thereby stimulate economic growth. This growth
comes from the expansion of economic activity within the intermediate region. This
effect would serve to increase scale economies at a rate that far exceeds the incurred
social costs of the increased [economic] activity and therefore makes the most sense. A
similar argument follows for lagging regions; although access to beneficial ingredients
for private investments are much less than when compared to intermediate areas, SOC
type investments would tend to better serve as a foundation for future improvements
in the area\(^5\).

Section 2.1 provides a brief introduction into the empirical literature, while the
subsequent subsections provide in-depth background on each relevant portion of the
more current literature.

\(^4\)Congested regions imply over-developed or saturated areas, while lagging would be indicative
of underdeveloped regions

\(^5\)Hansen argues that lagging regions will never gain in stature to become an intermediate region
without the necessary social provisions to attract and prepare the locale for higher productive
industry.
2.1 Public Capital

The most renowned and traceable beginnings of the empirical research dealing with public capital is that of Aschauer (1989). Aschauer (1989) found public capital to be highly productive, specifically when investments that increased the capital stock of core infrastructure were made. By estimating a log linear Cobb-Douglas aggregate production function for the U.S. using a time series for the period 1949 to 1985, the presence of large positive effects on output arising from increases in public capital stock was found. Taking public capital as a whole, the elasticity of output (with respect to public capital investment) was found to be 0.39. When the elasticity of output was taken with respect to core infrastructure, it was determined to be 0.24. This would imply that if there was a one percentage increase in total public capital investments, a 0.39 percent increase in output would be realized (or 0.24 in the case of the core infrastructure).

Many works followed in Aschauer’s footsteps—a few of which provide support. The majority, however, fail to attain similar magnitudes [as Aschauer]. It has since been realized that the methodology associated with Aschauer and his followers is potentially plagued by a number of econometric issues. For instance, the estimation of a single equation production function is known to suffer from simultaneity bias and problems discerning causality. This fact has led to something of a great methodological debate regarding the proper estimation techniques related to the effects of public infrastructure.

Pereira and Andraz (2011) classify the literature into various subsections. They note that the body of literature can be split according to the broader model specifi-

---

6Aschauer (1989) defines core infrastructure to be highways, streets, water systems and sewers.
7Munnell (1990), Morrison and Schwartz (1996b), Nadiri and Mamuneas (1994b) all find support consistent with Aschauer at the national level. Others, for instance Hulten and Schwab (1993) and Holtz-Eakin (1994), find no evidence when utilizing regional data.
8For an overview of the broader classifications as related to model specification in the literature the reader is referred to Pereira and Andraz (2011)’s recent and extensive review of the literature.
cation employed - those which estimate the effects of public infrastructure investment in terms of a production function, cost [function] or the vector autoregressive (VAR) model. One can further classify the studies according to the scale at which each analysis has been conducted. In addition to the model and the scale divisions, these studies can be further divided based upon the measure of public capital employed in the study. Many of the studies which follow Aschauer’s methodology of the single-equation production function have varied the level of aggregation contained in the data in search of similarly large effects. See, for examples of such, Holtz-Eakin (1988), Munnell (1990), Tatom (1991). Varying the scale of analysis has brought to light new technical issues as well. Many studies conducted at the regional level suggest that only about 20% of “the aggregate effects of public investment in highways in the U.S. are captured by the direct effects on each state output of public investment in the state itself” - the other 80% are related to the spillover effects that arise from public investment outside the state (Pereira and Andraz, 2003).

In recognition of the econometric issues associated with single-equation production functions, the literature then moves towards the estimation of “dual cost functions” (Pereira and Andraz, 2011). While more advantageous than the production function approach, depending upon the goal of the study, the estimation of cost functions has its own set of issues. Cost function studies, like those estimating production functions, also take public capital to be exogenous, and which, while taking care of the simultaneity bias, leave issues associated with causality unaddressed. For an overview of the direction of causality regarding employment and investment see Jiwattanakulpaisarn et al. (2009). Additionally, dual cost function estimation methods found in the literature fail to adequately account for the dynamic nature of the [economic] system. Nonstationarity, spurious correlation and non-cointegration remain a concern (Pereira and Andraz, 2011).

The problems associated with the production and cost function approaches sub-
sequently led to the introduction of the dynamic multivariate vector autoregressive (VAR) models. Models of this type attempt to address some of the major shortcomings of the previous frameworks, however, not without introducing new issues of their own. The utilization and proper implementation [of VAR models] allows for “comprehensive feedbacks between private inputs and public capital as well as” feedbacks between the inputs themselves (Pereira and Andraz, 2011). This is not to say that the VAR model is free from problems as a model of this type is essentially a reduced form of the production, demand and policy functions of the greater economy which hence fail to identify external shocks to the system and are, thus, easy to mis-specify. Furthermore, the VAR model complicates comparison of relative performance between previous literature and VAR estimation results - something commonly overlooked by researchers. The elasticities computed within a VAR framework are not *ceteris paribus* estimates. They are dynamic, and “reflect the total accumulated long-term changes... due to an initial shock in public capital” (Pereira and Andraz, 2011). There are also ways in which to incorporate feedback effects into the dual framework. Allowing for spatial structures in the error terms, and spatial lags in the independent variables, it is possible to allow for interactions between geographic areas and, thus, spatially displaced firms.

Some of the more common measures of public capital found are public capital stock, infrastructure capital stock, highways, telephones, electricity, public investment, the net stock of non-military capital, among others. The majority of studies utilizing the broader measure of public capital stock are created using the perpetual inventory method. This method involves the use of gross investment in constant dollars, an initial benchmark for the capital stock and a depreciation rate or service life estimate (Zegeye, 2000). This is to say that, at minimum, information on investment and asset deterioration are necessary. There are many measures of deterioration employed; one-hoss shay, 0.9 hyperbolic, straight-line and geometric depreciation are
common. For an in-depth discussion see Fraumeni (1999).

Many of the studies following the perpetual inventory method benchmark and rectify/apportion their (often less aggregated) estimates of the measure to aggregate (national) BEA capital stock estimates (which assume a straight-line or geometric measure of depreciation). According to Fraumeni (1999) “productive capital stock is the appropriate concept for estimating the productivity of capital stock or measuring the contribution of capital stock to economic growth” while wealth capital stock “is the appropriate measure of the market value of capital.” Many researchers have either failed to make this important distinction when constructing a measure of capital stock, or simply utilized already assembled data which employs an improper variable construction methodology. Fraumeni further notes that the BEA estimates are of the wealth capital stock type. BEA capital stock data, available from 1925 to 1996 “is a measure of tangible wealth and not productive capital stock,” as the capital stock data are compiled using information from a multitude of sources (Fraumeni, 1999). The majority of the empirical studies have utilized data that has been benchmarked and apportioned from the BEA wealth capital stocks.

The majority of the studies concerned with productivity benefits associated with investment in public infrastructure have constructed their own measure of public capital stock. As the perpetual inventory method involves some subjective assumptions, many of these capital stock variables are different (and as such cannot be directly compared). For example, the Munnell (1990) study created a public capital stock variable for the 48 continental states beginning in 1958 and ending in 1988. The estimates for each state were created by utilizing data from the Commerce Department’s Government Finance publications. These data, combined with BEA assumptions of retirement, lifetime and depreciation, aided in what Munnell determined to be an adequate estimate of public capital stock. In her study, capital outlay was defined as “direct expenditures for the construction of buildings, roads, and other improvements,
including additions, replacements, and major alterations to fixed works and structures, whether contracted privately or built directly by the government” (Munnell, 1990). Munnell’s measure is of the wealth capital stock type which is inadequate for the purpose of determining productivity impacts [as wealth stock does not provide a measure of productive capital].

The final classification of the methodology found in the literature and relevant to this study lies in the scale at which previous investigations have been conducted: some at the international, national, regional and even sectoral levels. It is hypothesized that the magnitude and significance of the results of public capital impacts on productivity vary based upon the level of [geographic] aggregation due to the presence of spatial spillovers being captured in the data. Examples of such can be found in Hulten and Schwab (1993) and Holtz-Eakin (1994). The fact that such variation in results appears to be related to the level of aggregation has lead researchers to theorize that the aggregate (national) studies show relatively large productivity impacts from public capital due to the nature of the data - that aggregation captures geographic spillovers (Holtz-Eakin and Schwartz, 1995). While Holtz-Eakin and Schwartz (1995) find no evidence of spillovers at the state level, subsequent studies employing the use of spatial econometric methods have. It is possible that the same principle of variation at varying scales exists in studies that control for the spatial element. See, for example, Cohen and Paul (2004) and Jiwattanakulpaisarn et al. (2009).

The next section serves to provide an overview of the most common methods associated with converting expenditure (flow) data into capital stocks. This is appropriate to determine the usefulness of capital, as the conversion of a one time expenditure into a physical capital good is necessary to capture the future benefits of an investment. This service life of capital is important to note as it is longer than otherwise implied by assuming a one-time use level.

---

9Alternatively, this notion could be better described as the service life of an asset.
2.1.1 Capital Stock Data Creation

Fraumeni (1999) provides an in-depth overview of the differences in proper capital stock creation techniques. The goal of the study is to “improve measures of highway capital stock, and thus further refine economic analysis of the stock’s contribution to economic growth.” She begins by providing an overview of the concept of capital stocks; explaining efficiency, deterioration, lifetimes and retirements as well as quality changes. She further distinguishes between productive versus wealth capital stock types. The most common technique employed empirically in the conversion of investment (flow) data into stock data is through the perpetual inventory method. This technique necessitates investment data and, at minimum, asset deterioration information.

Creation of the capital stock data series has proven quite problematic as evidenced by the recurring discussion found at various points throughout the literature. For examples of such see Ball (1985), Paul (1999), Berndt (1990), Hulten and Wykoff (1981) or Hansson (1988). As the categorical division of items constituting capital [stock] are not well defined, a rather large amount of subjectivity is introduced that reduces the comparability between various series constructed by different authors. While buildings and structures, plants and equipment and motor vehicles are typically included in an estimate of a firm or industry’s capital stock, other spending line-items may be relevant. Some, such as R&D, have the concept of a ‘stock’ implicit to them and are often included in such a measure. Further, R&D often times proves instrumental to the production process - therefore rendering itself a component of the productive stock of capital employed. Additional variation\textsuperscript{10} exists around whether to include maintenance outlays (or not) as mitigating the reduction in production capacity can also be argued as the equivalent of an increase to the productive capacity

\textsuperscript{10}Discrepancies
of capital employed in the transformative process.

Additional considerations to the creation of the capital stock data series involve the distinction between the economic and the accountant’s definitions of capital. An accountant, and by extension, a firm, will depreciate the value of capital over a long term time horizon. This ignores the idea of opportunity cost, implicit to an economist’s idea of capital. This so-called opportunity cost is an important portion of the capital stock as it includes the “user cost” of capital (Paul, 1999) (depreciation in productive capacity) as well as the returns foregone from the funds spent acquiring the capital entity.

### 2.1.2 Production Function Literature

By and large, the most widely employed method of empirically uncovering the effects of public capital on private industry and productivity measures have been conducted utilizing an aggregate production function. While the geographic scale and the data varies depending upon the study in question, the method of analysis across studies of this type have remained similar throughout the literature. The earliest\(^{11}\) empirical study examining the role of public capital in production is that of Mera (1973). Mera’s motivation for examining the role of public capital (what he terms, social capital) is as an explanation for inequality in income and economic growth among varying geographical regions. As a way to understand the “operational characteristics of agglomeration economies” Mera undertakes an empirical study to quantitatively determine the role of infrastructure investment on aiding convergence in income among the 46 prefectures that make up Japan over the period 1954 to 1963. Mera separates the effects by exploring multiple sectoral levels and differentiating social capital stock into four classifications. Social capital is found to be significant

---

\(^{11}\)Known to this author.
in the production process for all sectors considered, with an elasticity of around 0.2. A number of subsequent studies follow in Mera’s (Mera, 1973) footsteps and pre-date Aschauer’s work, though with a more strict focus on regional development. Blum (1982), Ratner (1983), Eberts (1986), Dalenberg (1987), Costa et al. (1987) and Deno (1988) all examine the effects of various forms of public capital on regional growth and development in a production function setting.

The literature surrounding the production function approach is vast and long debated; estimates are contradictory and little consensus has been reached as to the magnitude of public capital productivity. Many researchers interests into the topic are unrelated, yet all provide compelling reasons for examination. For instance, Costa et al. (1987) state that their interest stems from the fact that investments are made unevenly, in a non-linear fashion. Their study, which examines state level public capital data, finds public capital to have a significant and positive effect on value added industries. Additionally they find that “scale elasticities are greater for rich than for poorer states” (Costa et al., 1987) and that public capital and labor were complements. They conclude that public capital is best placed in richer regions because industry is more widely available to extract the benefits of the increased infrastructure, a notion somewhat in agreement with Hansen’s earlier theoretical work.

After Aschauer’s initial work (Aschauer, 1989), the focus in the production function literature shifted greatly away from regional development and growth in favor of industrial productivity and economic growth as a whole. Many of the early studies advocate greater levels of investment into public capital as a way to jump-start productivity growth. For instance, one of the conclusions reached by Aschauer (1989) is a stimulus, as this would serve to increase the return to private capital. From his analysis he found, in 1987, that movements in public capital expenditures generated “four-to-seven times larger private sector expenditures.” Munnell and others find similar results and recommend large stimulus expenses as these should lead to pro-
ductivity growth which thereby generates higher levels of sustained economic growth.

2.1.3 Cost Function Literature

Cost function studies that examine the role of public capital on private industry differ [from production function studies] in two ways: the cost function implicitly assumes cost minimization (whereas the production function does not) and the approach directly examines determinants of and impacts on production costs. These models tend to employ highly flexible functional forms that preserve typical properties associated with firm behavior. The fact that most studies employ a form that is twice differentiable also allows for the deduction of a wide array of impacts. For instance, under cost functions of a transcendental logarithmic (translog) or a Generalized Leontief (GL) flexible form, impacts of public capital on private capital, output and labor (among others) are directly obtainable. Furthermore, this form allows one to uncover estimates of a firm’s willingness to pay for additional factors of production as well the interplay between different factors. Cost studies provide direct estimates of cost impacts associated with changes in different factor inputs, which indirectly provide insight into input demand. Among the earliest studies that employ the estimation of cost functions that include some aggregate measure of public capital are those carried out by Berndt and Hansson (1992) and Morrison and Schwartz (1992). Similar studies have been undertaken by Nadiri and Mamuneas (1994a), Morrison and Schwartz (1996a), Cohen and Paul (2003) and Cohen and Paul (2004). Morrison and Schwartz (1996a) directly uncover the effects of public capital on firm costs and patterns of substitution between internal and external input factors.

\footnote{For example, complementarity/substitutability and increases in productivity, etc.}
2.1.4 Spatial Dependence

Spatial dependence\textsuperscript{13}, defined as the presence of statistical dependence in some random variables that are associated with geographical locations, forms an important component of this research. As discussed, a significant amount of work has been conducted exploring the relationship between economic performance indicators and transportation infrastructure investment. Few studies have explored this relationship in light of the presence of spatial dependence. As transportation infrastructure serves to enhance connections of spatially separated regions, it is relevant to examine the effects of investment on spatially connected, yet separate, locations.

The next chapter presents the first of three major studies that examine the effects of transportation infrastructure investment on industrial and economic performance.

\textsuperscript{13}Also known as spatial autocorrelation
Chapter 3

Transportation Infrastructure, Industrial Productivity and Return on Investment: A Spatial Spillover Approach

3.1 Overview

This section examines the North American Industrial Classification System (NAICS) based Manufacturing Industry (NAICS 31-33) from 1997-2010 in a cost based framework. It begins first by the construction and estimation of both profit and production function models for the manufacturing industry (as a whole) at the state level that allow for spatial spillovers and interactions. The utilization of profit and production provides an alternative approach to the dual cost function. Via the inclusion of infrastructure spending data elasticities associated with infrastructure investment and industry total costs are determined. The results of the spatial econometric models and the computed elasticities are then delivered in a Geographic Information System. This study examines the lower 48 states of the continental U.S. in terms of the
broader effects that state-specific transportation infrastructure investments have on own-state industry as well as surrounding (connected) states. By employing spatial panel data methods, difficult problems associated with consistent and efficient estimation (due to heterogeneity and endogeneity) are adequately treated. The measure of transportation infrastructure comes from all public expenditures on transportation related projects for air, water and road based infrastructure investments thereby understanding the increasingly important role that transportation plays on US industry.

3.2 Introduction

Public capital is a vital component of any economy. It contributes not only to the quality of life of a country’s citizens, but also to the productivity of the economy as a whole. It is without doubt that a well developed and well maintained public capital stock is necessary for any highly productive society (Treasury, 2010). In 2008, U.S. public infrastructure ranked 25th (out of 32 OECD nations) in terms of public satisfaction (Treasury, 2010). Broadly speaking, public capital has long been explored for its effects on economic growth, productivity and development, among others. Regional scientists, geographers and economists alike have had an interest in the relationship between public capital and regional output as well as economic and regional development for more than four decades (Costa et al., 1987).

A 2010 study by the Department of Treasury cites that 19 out of 20 Americans are concerned with the state of the U.S. infrastructure, while 84% “support greater investment to address infrastructure problems.” In recent years, over half of all US state and local government annual public capital expenditures have been directed for transportation infrastructures (U.S. Census Bureau). The optimization of many business models revolves around better utilization of existing publicly provided transportation infrastructure. As such, maintaining, updating and replacing an aging transportation
infrastructure system is crucial for any economy’s continued success. This section is focused on the spatial variation in benefits to the US manufacturing industry that arise from transportation infrastructure investments. Given fiscal constraints and the current condition of the nation’s aging infrastructure, it is of critical importance to determine not only the magnitude and extent of benefits arising from investments in transportation infrastructure but also the specific locations that provide higher rates of return on investment. That is, which geographic areas, when invested in, prove most beneficial to the broader economy?

This portion of the dissertation is organized as follows: section 3.3 provides an overview of the literature regarding the productivity of public capital investment, section 3.4 details the theory and models employed; section 3.5 describes the data used and discusses the estimation procedures. Section 3.6 presents the results as well as discusses the implications associated with the methods employed, while section 3.7 concludes.

3.3 Literature Review

Empirical work examining the productivity of infrastructure was scarce until the late 1980’s when, with Aschauer’s seminal work, the literature investigating these effects exploded. (Aschauer, 1989) found public expenditures to be highly productive, specifically when investments that increased the capital stock of core infrastructure were made.

Many works have followed in Aschauer’s footsteps; see Munnell (1990), Morrison and Schwartz (1996b), Nadiri and Mamuneas (1994b), Hulten and Schwab (1993), Holtz-Eakin (1994) for more details. This has led to something of a great methodological debate as it is said that the estimation of a single equation production function is known to suffer from simultaneity bias and problems discerning causality (Baird,
Pereira and Andraz (2011) classify the literature into various subsections. They note that the body of literature can be split according to the broader model specification employed - those which estimate the effects of public infrastructure investment in terms of a production function, cost function or the vector autoregressive (VAR) model. One can further classify the studies according to the scale at which each analysis has been conducted. In addition to the model and the scale divisions, these studies can be further divided based upon the measure of public capital employed in the study. Many of the studies which follow Aschauer’s methodology of the single-equation production function have varied the level of aggregation contained in the data in search of similarly large effects. The reader is referred to Holtz-Eakin (1988), Munnell (1990), Tatom (1991) for more in-depth specifics related to the various aggregations contained in the literature. Varying the scale of analysis has brought to light new technical issues as well. Many studies conducted at the regional level suggest that only about 20% of “the aggregate effects of public investment in highways in the U.S. are captured by the direct effects on each state output of public investment in the state itself” - the other 80% are related to the spillover effects that arise from public investment outside the state (Pereira and Andraz, 2003).

Some of the more common measures of publicly provided capital are public capital stock, infrastructure capital stock, highway spending, public investment, the net stock of non-military capital, among others. The majority of studies measure public capital stock using the perpetual inventory method. This method involves the use of “gross investment in constant dollars, an initial benchmark for the capital stock and a depreciation rate or service life estimate” (Zegeye, 2000). This is to say that, at minimum, information on investment and asset deterioration are necessary. There are many measures of deterioration employed; one-hoss shay, 0.9 hyperbolic, straight-line and geometric depreciation are common (Fraumeni, 1999).

A final classification of the methodology found in the literature and relevant to this
study lies in the scale at which previous investigations have been conducted: some at the international, national, regional and even sectoral levels. It is hypothesized that the magnitude and significance of the results of public capital impacts on productivity vary based upon the level of geographic aggregation due to the presence of spatial spillovers being captured in the data (Hulten and Schwab (1993), Holtz-Eakin (1994)). The fact that such variation in results appears to be related to the level of aggregation has led researchers to theorize that the aggregate (national) studies show relatively large productivity impacts from public capital due to the nature of the data - that aggregation captures geographic spillovers Holtz-Eakin and Schwartz (1995). While Holtz-Eakin and Schwartz (1995) find no evidence of spillovers at the state level, subsequent studies employing the use of spatial econometric methods have. As mentioned in Cohen and Paul (2004) and Jiwattanakulpaisarn et al. (2009), it is possible that the same principle of variation at varying scales exists in studies that control for the spatial element.

3.4 Theory and Approach

The goal of the first portion of this study is to examine productivity effects associated with investments in publicly provided transportation infrastructure. This is relevant from a policy perspective as more state departments of transportation (DOTs) and metropolitan planning organizations (MPOs) align their transportation plans with state economic development goals in the interest of attracting new business by leveraging existing industrial clusters and infrastructures. This approach can provide significant insight into where investments may propagate further geographically and economically through the target industry (or industries) in question. Examining the cost structure of a given industry sheds light on two aspects of productivity – each relative to any firm (or industry), whose main objective is to maximize profits.
Results are presented in Section 3.6 and show the effects that transportation infrastructure investments have on the U.S Manufacturing industry at the state level over the period 1997 to 2010. This portion of the study proceeds in two separate yet related manners. The first examines the effect that public transportation infrastructure expenditures have on the manufacturing industry’s profit margin. Although the industry in each state contains a multitude of firms, due to data availability, examination can only be made in such a way that each state is thought of as a homogeneous collection of firms. The study then proceeds to the estimation of an aggregate production function which, in conjunction with the profit function model, provides an alternative approach to examining industrial cost and productivity benefits associated with publicly provided transportation infrastructure investments.

3.4.1 Profit Margin

It is assumed that all firms engaging in the industry are profit-maximizers. A profit-maximizing firm chooses the level of inputs and outputs that maximize economic profit. As industrial activity takes place in space, it is expected that there exist some market interactions, agglomeration effects and, thus, that the data contain spatial dependence. This section of the thesis is concerned with deducing what effects, if any, transportation infrastructure has on manufacturers. The data utilized is discussed in more detail in Section 3.5

As the primary intention is to determine if there is a measurable effect of transportation infrastructure on manufacturer’s profit margins, an examination of industry wide profit margins for the time period under study serves as the starting point of the analysis.
Therefore, the model takes the form:

\[
\pi_{i,t} = \alpha + \beta_1 \bar{E}_{i,t} + \beta_2 P_{L,i,t} + \beta_3 P_{M,i,t} + \beta_4 \sigma_{PF,t} + \beta_5 \bar{P}_{F,t} + \beta_6 G_{i,t} + \beta_7 \bar{G}_{i,t} + \beta_8 K_{i,t} + u_{i,t}
\]

\[
u_{i,t} = \rho(I_T \otimes W_N)u_{i,t} + \epsilon_{i,t}
\]

(3.1)

\[
\epsilon_{i,t} = (e_T \otimes I_N)\mu_i + \nu_{i,t}
\]

where all right hand side variables are in natural log form; \(\pi_{i,t}\) is defined as \(1 - \frac{\text{total cost}}{\text{total revenue}}\) and represents the profit margin for the industry in state \(i \in N\) at time \(t \in T\); \(\bar{E}_{i,t}\) is state \(i\)'s average firm size engaged in the industry at time \(t\); \(\bar{E}_{i,t}\) is the average number of employees per firm for each state \(i\) and each year \(t\); \(P_{L,i,t}\) and \(P_{M,i,t}\) are labor and materials prices in each state \(i\) and each year \(t\) respectively; \(\sigma_{PF,t}\) is the average standard deviation of national weekly average fuel prices for year \(t\), chosen as it is hypothesized that volatility in the oil markets has an effect on firm’s decisions and therefore their bottom line; \(P_{F,t}\) is the national average retail gasoline price in year \(t\) and included in the model because manufacturers are heavily dependent upon fossil fuels in the production process. \(G_{i,t}\) represents the sum of transportation capital expenditures on highways, water ports and airports by local and state governments in state \(i\) for each year, \(t\); \(\bar{G}_{i,t}\) represents the spatial lag of these transportation expenditures. This is to say \(\bar{G} = WG\) where \(W\) is the row-normalized first order queen contiguity spatial weights matrix constructed consistently with the spatial econometric literature. \(K_{i,t}\) represents private capital expenditures for the industry. Elasticities are also computed at the means for the dataset as well as for each state and presented in Section 3.6. The elasticities are computed as \((\frac{\partial \pi}{\partial X}) \cdot (\frac{X}{\pi})\).

### 3.4.2 Aggregate Production Function

Utilizing the theory of production and estimating the industry’s production function it becomes possible to examine the effects that various inputs for production have on output. For manufacturing firms, this is especially relevant as the act of manufacturing is defined to be the transformation of raw input goods to consumable units of output. The production
function offers a way of modeling the physical relationship between these inputs and outputs. At the aggregate level it allows researchers to uncover deeper trends about the decision making that takes place in the industry under study. Aside from raw-inputs, modeling in this way effectively allows for an examination of the effects of exogenous (and free) inputs. In this case, our measure of transportation infrastructure is not purchased directly by the firm – firms do not pay an explicit price to utilize these public capital, nor do they determine the amount of capital endowed to specific geographic regions. The aggregate production function takes the following form:

$$\ln Y_{i,t} = \alpha + \rho W \ln Y_{i,t} + \beta_1 \ln G_{i,t} + \beta_2 \ln \bar{G}_{i,t} + \beta_3 \ln K_{i,t} + \beta_4 \ln P_{L,i,t} + \beta_5 \ln P_{L,i,t} + \beta_6 D_{06} + \epsilon_{i,t}$$

$$\epsilon_{i,t} = \mu_i + \nu_{i,t} \quad (3.2)$$

where all previously defined variables are the same as in (3.1). $Y$ is gross value of shipments, $K$ is state specific private capital expenditures, $t\ln P_L$ is a labor price-time interaction variable and $D_{06}$ is a dummy for the year 2006\(^1\). Additionally, the coefficients ($\beta$'s) in (3.2) are interpreted as elasticities. The spatial lag model is employed in (3.2), above, as the theory of economic geography states that firms engaged in similar industries will have a tendency to locate near one another. This spatial clustering implies spatial dependence and due to firm co-location, the presence of endogeneity. The results are also reported in Section 3.6.

3.4.3 Cost Functions

While production functions provide insight into the conditions related to marginal productivity, no further (e.g. second order) relationships can be understood or examined. An alternative approach to the production function that provides such insights is based on the dual cost function. The cost function implicitly assumes cost minimization (whereas the production function does not) as the optimization problem faced by firms. As such, to min-

\(^1\)This dummy is included due to preliminary cross section analyses showing a shock to the price of labor in 2006.
Figure 3-2: Mean Gross State Product 1997-2010
imize cost, firms are required to choose inputs (based on input price levels), output levels and the form of production. The solution of this optimization (minimization) problem is the dual to the production function.

There are a number of differences between production and cost functions concerning what can be determined by each system. Under a dual cost function, both input quantities and production costs are endogenous while prices and output are exogenously determined. This is nearly opposite of production function determinants; as output is endogenous and input quantities are exogenous (Berndt, 1991). For these reasons, when output and input prices are safely assumed to be exogenous, cost function estimation is preferable to production functions (Zellner et al., 1966). It is worth noting, however, that cost functions based upon a Cobb-Douglas production function can be analytically derived (as is shown in the following subsection). For an investigation into these effects under the dual, see the following chapter.

3.4.3.1 Analytical Cost Function

Examining productivity and cost benefits due to public capital investments are traditionally examined in a [dual] cost function-based framework. Through the combination of the previously outlined profit and production functions one can analytically derive a cost function, as total cost is defined to be \(1 - (\frac{\text{total revenue}}{\pi})\). Substituting the profit and production functions into this expression yields our cost function:

\[
\pi_{i,t} = 1 - \frac{TC_{i,t}}{Y_{i,t}} \tag{3.3}
\]

\[
TC_{i,t} = Y_{i,t} - Y_{i,t} \times \pi_{i,t} \tag{3.4}
\]

where \(TC_{i,t}\) is state \(i\)'s total cost in year \(t\). Since the production function is expressed in natural log form, it must be transformed. Utilizing Taylor series, limiting the analysis to first order terms, removing the subscripts (for brevity) and dropping the stochastic component, yields:
\[ \exp(lnY) = \exp(\alpha + \rho WlnY + \beta_1 lnG_{i,t} + \beta_2 ln\bar{G}_{i,t} + \beta_3 lnK_{i,t} + \beta_4 lnP_{L,i,t} + \beta_5 tlnP_{L,i,t} + \beta_6 D_{06}) \]

\[ Y \approx 1 + \alpha + \rho WlnY + \beta_1 lnG_{i,t} + \beta_2 ln\bar{G}_{i,t} + \beta_3 lnK_{i,t} + \beta_4 lnP_{L,i,t} + \beta_5 tlnP_{L,i,t} + \beta_6 D_{06} \]

Denoting the function for \( Y \) in (3.5) as \( Y(\cdot) \) and substituting into (4) yields:

\[ TC = Y(\cdot) - Y(\cdot)\pi \]

Substituting the profit function from (3.1) for \( \pi \) in (3.6) allows for an analysis of the industry’s cost structure as well as the impacts that transportation infrastructure investments in own state and neighboring states have on the manufacturing industry. The analysis is carried out in Section 3.6. Analyzing cost in this light yields a number of departures from the traditional methodology of explicit estimation. For example, in this framework output levels are endogenized which allows for a more robust understanding of the effects of public capital as output varies.

### 3.5 Data

The data utilized for this study comes from several sources. Manufacturing (MFG) data is from the U.S. Department of Commerce’s Annual Survey of Manufacturers (ASM) Geographic Area Statistics Bureau (a) over the period 1997 - 2010. The public capital data comes from the U.S. Department of Commerce’s Government Finances series Bureau (b). Further analysis to separate the effects received by households and by industry from investments in transportation industry is carried out based on Federal Highway Administration (FHWA) Highway Performance Monitoring System (HPMS) Data. The socio-economic data utilized in conjunction with these physical
infrastructure measures are taken from the Bureau of Economic Analysis’ Regional Economic Information System (REIS) as well as the U.S. Department of Commerce’s USA Counties Database.

3.5.1 Industry

Industry measures are taken from the Annual Survey of Manufactures. The level of industrial analysis to determine benefits to the largest freight users of infrastructure is conducted at two digits North American Industrial Classification codes 31-33. These statistics are collected every year and contain data on the number of firms, private capital investments (flow), non-production as well as production labor, labor wages, intermediate material prices, value-added and the gross value of shipments in current dollars, among others. Due to data-withholding requirements these data are obtained at the state level. Although the industry in each state contains a multitude of firms, due to data availability, examination can only be made in such a way that each state is thought of as a homogeneous collection of firms.

The value of shipments from the ASM is taken to be output \((Y)\) in the model specification given in equations 3.1-3.6. Value of shipments is commonly viewed as a better proxy for real output of an industry than value-added Zegeye (2000). The ASM presents total payroll for each state as well as number of employees. The price of labor, \(P_L\) is computed by creating a price index of the annual payroll divided by number of employees, normalized to the base year, 2005. As the ASM gives data in current dollars, a price deflator is necessary to convert all observations into real terms. A state specific price deflator for manufacturing output is created by taking the ratio of nominal MFG Gross State Product to real MFG Gross State Product (GSP). This is also used as the price for materials. Private capital data is employed in two ways. As the ASM releases expenditure data for each year, this flow investment data is utilized first as a proxy for capital stock and also to create a capital stock
data series utilizing the perpetual inventory method\textsuperscript{2}. In creating a capital stock series, a deflator is needed. The capital deflator comes from the Bureau of Labor Statistics, Office of Productivity and Technology. The price of capital is constructed as: 

\[(i_t + d_t) \times q_{K,t}\left[\frac{1}{1 - \text{taxrate}_t}\right]\]

where \(d_t\) is the depreciation rate, \(i_t\) is the Moody’s Baa corporate bond rate (obtained from the Economic Report of the President), \(q_{K,t}\) is the investment deflator and \(\text{taxrate}_t\) is the corporate tax rate from Bureau of Economic Analysis’ (BEA) National Income and Productivity Accounts (NIPA) table. Figures 3-1 and 3-2 display state average profit margins and output, respectively. Notice the substantial clustering in levels present in both maps in various regions throughout the continental US. It is interesting to note that the highest profit margin states tend to border the most densely populated and highest output states - a motivating factor in the examination of neighboring state effects. For instance, Nevada, Oregon, Idaho, and Arizona exhibit profit margins a magnitude larger than the large market states of Washington and California. New Mexico and Lousiana are among the highest margin states, perhaps owing in part to their geographic relation to Texas, which Figure 3-2 clearly establishes as a very large market. Similar patterns can be found when focusing on the East coast also. This pattern is worth noting, as deeper investigations reveal some interesting patterns related to spatial distribution.

Preliminary analysis guided the model selection process. Conducting both production and profit function models in a cross-sectional setting (single time period) yielded interesting results. Figure 3-3 depicts annual estimates of variables of interest. This figure serves as a stationarity check as well as an attempt to determine how stable the time period under investigation is with respect to the variables of interest. The results where \(X=10\) is the primary motivation for the inclusion of a year 2006 dummy in the model specifications employed. All parameter estimates are fairly consistent in magnitude throughout the years except for 2006.

\textsuperscript{2}As outlined in Fraumeni (1999).
Figure 3-3: Intermediate Cross-Section Analysis
Further examination shows substantial jumps in other sources of data in 2006 as well. Figure 3-4 shows a rapidly rising trend in the cost of highway construction beginning in 2004.

### 3.5.2 Capital Measures

All transportation related expenditures on highways, waterways and airports at the state and local government level are aggregated (for each state) to create a transportation capital expenditure variable, $G$.

Due to the difficulties associated with valuation of capital and a general lack of consensus on proper depreciation rates, these capital expenditures are utilized as a proxy for total capital stock in each location $j$ at time $t$. Each model estimation procedure is taken with public and private capital data in the form of expenditures due to the subjectivity of such stock data creation. This is a means of robustness to the past results found in the literature. Each study utilizing the perpetual inventory method to examine the impacts of public capital either creates the stock data variable differently (from one another) or fails to make the distinction between wealth stock and productive stock. As such, running the model on both variable types (proxy and stock), for both private and public capital types is a necessary step to understanding
the sensitivity of previous results to the technique used in the formation of the capital stock variable. As previously mentioned, $WG$, in equations (3.1-3.9) is taken as the neighboring levels of public capital stock where neighbors are defined in the spatial weights matrix $W$.

### 3.5.3 Estimation Routines

The profit function in (1) is estimated utilizing Kelejian, Kapoor and Prucha’s (2007) Panel Data General Method of Moment estimation routine (Kapoor et al., 2007). This estimation routine has been implemented in Python\(^3\). As the production function follows a spatial lag panel data model, maximum likelihood estimation is employed and carried out as outlined in Elhorst, 2003 Elhorst (2003). Tests for heteroskedasticity and normality are also performed utilizing transformed data. The results from the Breusch-Pagan heteroskedasticity test for both models are presented in Section 3.6. The cost-based analysis is done analytically utilizing the estimates from both the profit and production functions as detailed in Section 3.4.

In addition to the spatial econometric estimation routines, residuals and mean elasticities are computed state and time-wise and mapped using Geographic Information Systems (GIS). The mapping allows for various GIS analyses and also provides a visualization tool. Visualizing these spillover effects, as well as the productivity effects associated with a number of the variables of interest, allows for a clearer picture in terms of which states receive the largest benefits from transportation infrastructure investment. This method also helps bring the sometimes abstract estimation routines back to reality, as the estimation results provide the input data to create the maps.

\(^3\)Relevant portions of the software code are included in the appendix.
3.6 Data Analysis Procedures

3.6.1 Profit Margin Function

Summary statistics of the data utilized in the modeling procedure are outlined in Table 3.1.

Table 3.1: Summary Statistics of Profit Function Input Data.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi$</td>
<td>672</td>
<td>0.34676</td>
<td>0.08539</td>
<td>-0.32868</td>
<td>0.74601</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>672</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\bar{E}$</td>
<td>672</td>
<td>42.8854</td>
<td>13.5457</td>
<td>14</td>
<td>83</td>
</tr>
<tr>
<td>$\ln(P_L)$</td>
<td>672</td>
<td>4.67448</td>
<td>0.19338</td>
<td>4.10162</td>
<td>5.88615</td>
</tr>
<tr>
<td>$P_M$</td>
<td>672</td>
<td>1.036824</td>
<td>0.16267</td>
<td>0.70155</td>
<td>3.03907</td>
</tr>
<tr>
<td>$\sigma_{P_{Fuel}}$</td>
<td>672</td>
<td>-3.87263</td>
<td>1.65149</td>
<td>-6.85469</td>
<td>-0.83584</td>
</tr>
<tr>
<td>$P_{Fuel}$</td>
<td>672</td>
<td>1.89875</td>
<td>0.69658</td>
<td>1.01665</td>
<td>3.26567</td>
</tr>
<tr>
<td>$\ln(G)$</td>
<td>672</td>
<td>7.63742</td>
<td>0.88268</td>
<td>5.79313</td>
<td>9.83131</td>
</tr>
<tr>
<td>$\ln(\bar{G})$</td>
<td>672</td>
<td>7.91257</td>
<td>0.52704</td>
<td>6.17822</td>
<td>8.79303</td>
</tr>
<tr>
<td>$\ln(K)$</td>
<td>672</td>
<td>14.1135</td>
<td>1.21385</td>
<td>0.69315</td>
<td>16.7357</td>
</tr>
</tbody>
</table>

The results of the estimation of the profit function model are presented in Table 3.2. The signs, magnitudes and significance of the variables follow economic intuition and suggest a promising model. Intermediate material prices and the variance of fuel prices are the only variables shown not to be statistically different from zero. It is plausible to assume that the average yearly fuel price captures the information contained in the intermediate material prices ($P_M$) variable. The results of the profit function estimation show positive and significant impacts of publicly provided transportation infrastructure investments on manufacturer’s profit margins. On average, those states containing firms with more employees are less profitable. On a different level, though, the results show that as the price of labor increases, so does the industry’s profit margin. This is likely due to the fact that more expensive labor yields higher skill-sets and that the productivity benefits (and therefore the profitability
Table 3.2: Profit Function Estimation Results.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std-Error</th>
<th>P-Value</th>
<th>Elasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>-0.39205</td>
<td>0.13899</td>
<td>0.00493*</td>
<td>-1.13401</td>
</tr>
<tr>
<td>$\bar{E}$</td>
<td>-0.00145</td>
<td>0.00046</td>
<td>0.00159*</td>
<td>-0.17971</td>
</tr>
<tr>
<td>ln($P_L$)</td>
<td>0.11827</td>
<td>0.01589</td>
<td>0.00000*</td>
<td>1.59916</td>
</tr>
<tr>
<td>ln($P_M$)</td>
<td>0.02461</td>
<td>0.01554</td>
<td>0.11385</td>
<td>0.07380</td>
</tr>
<tr>
<td>$\sigma_{P_{fuel}}$</td>
<td>0.00249</td>
<td>0.00325</td>
<td>0.44225</td>
<td>-0.02799</td>
</tr>
<tr>
<td>$P_{Fuel}$</td>
<td>-0.05433</td>
<td>0.00765</td>
<td>0.00000*</td>
<td>-0.29839</td>
</tr>
<tr>
<td>ln($G$)</td>
<td>0.01879</td>
<td>0.00709</td>
<td>0.00825*</td>
<td>0.41529</td>
</tr>
<tr>
<td>ln($\bar{G}$)</td>
<td>0.04915</td>
<td>0.01183</td>
<td>0.000003*</td>
<td>1.12499</td>
</tr>
<tr>
<td>ln($K$)</td>
<td>-0.01404</td>
<td>0.00369</td>
<td>0.00016*</td>
<td>-0.57316</td>
</tr>
<tr>
<td>$\sigma_{FGLS}^2$</td>
<td>0.06042</td>
<td>0.42818</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N$</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T$</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.85708</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

of the firm/industry) more than offset the additional cost of the [relatively] higher priced (and more skilled) labor. This is to say that the wage rate is in equilibrium and therefore equal to the marginal product of labor (MPL). Since a considerable portion of manufacturing firms’ intermediate expenditures are on energy needs, it is not surprising that as the price of fuel increases, manufacturers' profit margins fall. Private capital detracts from profit margins, something likely related to the short run scope of our analysis.

Own-state transportation infrastructure and neighboring state transportation infrastructure investments are both statistically significant and positive; increases in either measure yield higher profit margins for the industry. A 1 percent increase in own-state transportation infrastructure expenditures yields, on average, a 0.4 percent increase in manufacturer’s profit margins. Similarly, on average a 1 percent increase in neighboring state transportation infrastructure expenditures yields a 1.12 percent increase in profit margins. Since these impacts are averaged over the entire sample over time, Figure 3-7 presents state specific elasticities averaged over time.
Neighboring transportation infrastructure expenditures are profitable for the manufacturing industry. A possible explanation for this lies in agglomeration effects and the increasing size of market areas over time.

As market areas become more geographically dispersed, utilization of transportation systems increase to overcome the spatial separations between suppliers and buyers. Spatial spillovers are present and appear to be increasing over time. Computing mean elasticities annually for neighboring transportation infrastructure and regressing these on a time variable, $t$, where $t = 1, 2, ..., 14$ confirms this as the time trend is positive and statistically significant at the 1 percent level. This provides evidence in support of increasing spatial spillovers arising from transportation systems and is depicted in Figure 3-5. Since transportation systems are inherently spatial this result not only makes sense, but also leads one to believe that the benefits of transportation infrastructure are measurable in more than just public use scenarios - that they are a significant determinant in the success of private industry. Figure 3-6 presents a scatter plot of the model predicted profit margin versus actual observed [profit margin]. The specification test results of the model are also shown in Table 3.3. Evidence of heteroskedasticity is non-existent.

As our focus lies in the geographical component of returns on investment, the spatial component of the model is also of interest. As expected, the model shows significant spatial dependence (denoted in the model by $\rho$). This implies that more profitable states are clustered - that similarly profitable firms locate near one another. This could be due to one of (or) several reasons related to agglomeration effects. As previously discussed, firms tend to locate near one another to leverage the skill-set of the labor force. Also of importance is the relatively high level of factor mobility present in the US. Additionally, firms (especially in the manufacturing industry) co-locate due to the high levels of vertical integration present in the industry.
Figure 3-5: Mean Elasticity of $\bar{G}$ over time.

Table 3.3: Profit Function Heteroskedasticity Test

<table>
<thead>
<tr>
<th>Breusch-Pagan Heteroskedasticity Test Results</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Breusch-Pagan LM-Statistic</td>
<td>2.77e^{-10}</td>
</tr>
<tr>
<td>Chi-Squared Probability</td>
<td>0.99</td>
</tr>
<tr>
<td>Degrees of Freedom</td>
<td>8</td>
</tr>
</tbody>
</table>

Figure 3-7 depicts the elasticity values for own-state transportation infrastructure expenditures. The map visually presents the ratio of the estimated impact of own-state transportation infrastructure to the predicted value of the state’s profit margin. This term is best expressed as $\frac{\beta \ln(G_j)}{\bar{\pi}_j}$, which is the elasticity of own-state transportation infrastructure with respect to own-state profit margins. The visual representation allows for a further understanding of which areas receive the larger benefits from increased transportation infrastructures. Examining Figure 3-7 reveals a fairly homogeneous belt running through the Mid-West/Great Lakes region of the continental US where high concentrations of manufacturing are present. Looking west to Nevada, Arizona and New Mexico it can be seen that this region suffers from rel-
atively lower benefits to increased expenditures. This is likely due to a combination of the region’s geographic structure and high concentrations of economic activity in small areas.

Figure 3-8 presents the benefits of neighboring-state infrastructure expenditures relative to own-state expenditures and is represented by \( \frac{\beta \ln(G_{1i})}{\ln(G_{1i})} \). This map shows that the biggest beneficiaries of neighboring state transportation infrastructure tend to be those states which border states with much higher [transportation] infrastructure expenditures than their own. For example, Nevada, Delaware, Vermont, and Wyoming show the biggest benefits with respect to neighboring states’ investments. Nevada (which borders California) Vermont (New York), Delaware (Maryland) and Wyoming (Colorado, Nevada, Idaho and Montana) all experience large benefits due to the size of their neighboring state’s expenditures. The fact that neighboring state expenditures have higher impacts than own state spending provides support to the notion that neighboring transportation infrastructure is truly “free”, at least from a taxation point of view; the use of neighboring state infrastructure does not have fees in the form of taxes (assuming fee-based infrastructure use is exempt from the
Figure 3-7: State Specific Transportation Infrastructure Elasticities.
Table 3.4: State Specific Elasticities of $\pi$ with respect to $G$ and $\bar{G}$

<table>
<thead>
<tr>
<th>State</th>
<th>$\epsilon_{\pi,G}$</th>
<th>$\epsilon_{\pi,\bar{G}}$</th>
<th>State</th>
<th>$\epsilon_{\pi,G}$</th>
<th>$\epsilon_{\pi,\bar{G}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AL</td>
<td>0.4217</td>
<td>1.1378</td>
<td>NE</td>
<td>0.4827</td>
<td>1.3023</td>
</tr>
<tr>
<td>AZ</td>
<td>0.3850</td>
<td>1.0387</td>
<td>NV</td>
<td>0.3352</td>
<td>0.9044</td>
</tr>
<tr>
<td>AR</td>
<td>0.4530</td>
<td>1.2222</td>
<td>NH</td>
<td>0.4690</td>
<td>1.2653</td>
</tr>
<tr>
<td>CA</td>
<td>0.4095</td>
<td>1.1047</td>
<td>NJ</td>
<td>0.3841</td>
<td>1.0363</td>
</tr>
<tr>
<td>CO</td>
<td>0.4261</td>
<td>1.1496</td>
<td>NM</td>
<td>0.3426</td>
<td>0.9244</td>
</tr>
<tr>
<td>CT</td>
<td>0.3916</td>
<td>1.0566</td>
<td>NY</td>
<td>0.3917</td>
<td>1.0568</td>
</tr>
<tr>
<td>DE</td>
<td>0.4282</td>
<td>1.1554</td>
<td>NC</td>
<td>0.4646</td>
<td>1.2535</td>
</tr>
<tr>
<td>FL</td>
<td>0.3675</td>
<td>0.9917</td>
<td>ND</td>
<td>0.4175</td>
<td>1.1265</td>
</tr>
<tr>
<td>GA</td>
<td>0.4192</td>
<td>1.1311</td>
<td>OH</td>
<td>0.4330</td>
<td>1.1683</td>
</tr>
<tr>
<td>ID</td>
<td>0.3878</td>
<td>1.0462</td>
<td>OK</td>
<td>0.4142</td>
<td>1.1176</td>
</tr>
<tr>
<td>IL</td>
<td>0.4236</td>
<td>1.1430</td>
<td>OR</td>
<td>0.3599</td>
<td>0.9710</td>
</tr>
<tr>
<td>IN</td>
<td>0.4493</td>
<td>1.2123</td>
<td>PA</td>
<td>0.4159</td>
<td>1.1221</td>
</tr>
<tr>
<td>IA</td>
<td>0.4499</td>
<td>1.2139</td>
<td>RI</td>
<td>0.3779</td>
<td>1.0196</td>
</tr>
<tr>
<td>KS</td>
<td>0.4817</td>
<td>1.2997</td>
<td>SC</td>
<td>0.4813</td>
<td>1.2987</td>
</tr>
<tr>
<td>KY</td>
<td>0.4666</td>
<td>1.2590</td>
<td>SD</td>
<td>0.4554</td>
<td>1.2286</td>
</tr>
<tr>
<td>LA</td>
<td>0.3576</td>
<td>0.9648</td>
<td>TN</td>
<td>0.4649</td>
<td>1.2542</td>
</tr>
<tr>
<td>ME</td>
<td>0.5360</td>
<td>1.4463</td>
<td>TX</td>
<td>0.4411</td>
<td>1.1901</td>
</tr>
<tr>
<td>MD</td>
<td>0.4142</td>
<td>1.1175</td>
<td>UT</td>
<td>0.4580</td>
<td>1.2357</td>
</tr>
<tr>
<td>MA</td>
<td>0.4161</td>
<td>1.1066</td>
<td>VT</td>
<td>0.3825</td>
<td>1.0321</td>
</tr>
<tr>
<td>MI</td>
<td>0.4178</td>
<td>1.1273</td>
<td>VA</td>
<td>0.4487</td>
<td>1.2107</td>
</tr>
<tr>
<td>MN</td>
<td>0.4274</td>
<td>1.1530</td>
<td>WA</td>
<td>0.4553</td>
<td>1.2283</td>
</tr>
<tr>
<td>MS</td>
<td>0.5251</td>
<td>1.4168</td>
<td>WV</td>
<td>0.4126</td>
<td>1.1134</td>
</tr>
<tr>
<td>MO</td>
<td>0.4430</td>
<td>1.1953</td>
<td>WI</td>
<td>0.4241</td>
<td>1.1443</td>
</tr>
<tr>
<td>MT</td>
<td>0.4775</td>
<td>1.2883</td>
<td>WY</td>
<td>0.4294</td>
<td>1.1586</td>
</tr>
</tbody>
</table>

Table 3.4 presents the mean elasticities for $G$ and $\bar{G}$ at the means for each state.

### 3.6.2 Aggregate Production Function

Table 3.5 displays the summary statistics for the dataset employed in estimation. The results listed in Table 3.6 are directly comparable to previous literature. An elasticity measure of 0.207 for transportation infrastructure is found with respect to output. This is to say that a 1 percent increase in own-state transportation infrastructure yields a 0.2 percent increase in output. Neighboring state transpor-
Figure 3-8: State-Specific Neighboring Benefits Based on Own-State Investments
Table 3.5: Production Function Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>lnY</td>
<td>672</td>
<td>17.90451</td>
<td>1.103903</td>
<td>15.04272</td>
<td>20.42754</td>
</tr>
<tr>
<td>α</td>
<td>672</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>lnG</td>
<td>672</td>
<td>7.637415</td>
<td>0.882682</td>
<td>5.793134</td>
<td>9.83131</td>
</tr>
<tr>
<td>ln</td>
<td>672</td>
<td>7.91257</td>
<td>0.527038</td>
<td>6.178221</td>
<td>8.793034</td>
</tr>
<tr>
<td>lnK</td>
<td>672</td>
<td>14.1135</td>
<td>1.213851</td>
<td>0.693147</td>
<td>16.73567</td>
</tr>
<tr>
<td>lnP_L</td>
<td>672</td>
<td>4.674479</td>
<td>0.193385</td>
<td>4.101624</td>
<td>5.886145</td>
</tr>
<tr>
<td>tlnP_L</td>
<td>672</td>
<td>34.94539</td>
<td>18.87802</td>
<td>4.429191</td>
<td>74.11072</td>
</tr>
<tr>
<td>D_{06}</td>
<td>672</td>
<td>0.071429</td>
<td>0.257731</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Private capital, labor prices and spatial dependence are all present, highly significant and follow economic theory in terms of their direction and magnitudes. More private capital and more expensive labor (and hence more skilled labor) have positive effects on output. Spatial dependence is also similar in magnitude to that found in the profit function estimation.

While the magnitudes of spatial dependence between the two models are very similar, they represent two very different types of spatial dependence. Under the profit function model, spatial dependence is contained in the error term and not modeled explicitly, while in the production function model location is a determinant of production itself. Production exhibits a different form of spatial dependence in that higher output states (which tend to be located near other high output states) have higher output due to their higher producing neighbors. This is to say that there is a measurable feedback effect present in production. For example, State A’s output increases because output in State B (which borders State A) increases and State B’s output increases because of State A’s increase. Again Breusch-Pagan heteroskedasticity results are included in the estimation output and shown to be relatively insignificant.
Table 3.6: Production Function Estimation Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.04264</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\ln G$</td>
<td>0.20765</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\ln G$</td>
<td>0.46304</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\ln K$</td>
<td>0.05184</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\ln P_L$</td>
<td>0.77131</td>
<td>0.0000</td>
</tr>
<tr>
<td>$t\ln P_L$</td>
<td>-0.00033</td>
<td>0.2544</td>
</tr>
<tr>
<td>D$_{06}$</td>
<td>0.03464</td>
<td>0.0751</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.45398</td>
<td>0.0000</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.9868</td>
<td></td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.0161</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>14</td>
<td></td>
</tr>
</tbody>
</table>

Breusch-Pagan LM-Statistic: 1.8610
Chi-Squared Probability: 0.9320
Degrees of Freedom: 6

Table 3.7: Production Function Heteroskedasticity Test

<table>
<thead>
<tr>
<th>Breusch-Pagan Heteroskedasticity Test Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breusch-Pagan LM-Statistic: 1.8610</td>
</tr>
<tr>
<td>Chi-Squared Probability: 0.9320</td>
</tr>
<tr>
<td>Degrees of Freedom: 6</td>
</tr>
</tbody>
</table>
Furthermore, Figure 3-9 is indicative of a strong model as the predicted values versus actual follow the general trend without an appearance of over-fitting.

Figure 3-10, in the same fashion as Figure 3-8, depicts the results of $\frac{\beta_2 \ln G}{\beta_1 \ln G}$. The same phenomenon of larger expenditure states having a larger, positive effect on the smaller neighboring states is also present when examining production (output). However, under the production function model specification, the results of $\frac{\beta_2 \ln G}{\beta_1 \ln G}$ exhibit less variation and higher magnitudes than that of the profit function. This leads one to conclude that transportation infrastructure has a larger spatial spillover component with respect to output (than profit). This follows intuition as the movement of the industry’s goods are what actually depend upon the infrastructure. Indeed, moving goods farther and for less money contributes to increased profits but it is the ability to move a larger volume of goods further and faster that has a more direct effect on output.

Figure 3-11 displays the ratio of production function residuals to transportation infrastructure investment’s output contribution $\left(\frac{\text{residuals}}{\beta_1 \ln G}\right)$. It illustrates that the ma-
Figure 3-10: Neighboring-State Transportation Infrastructure Benefits Relative to Own-State
Figure 3-11: Ratio of Residuals to Transportation Infrastructure Effects

Majority of the states experience similar contributions to manufacturing from transportation infrastructure. In many cases there are certain qualities unique to specific places that make a model of this general type difficult to draw conclusions from. This map is used to quell concerns that the results of the preceding analysis fall victim to overbearing statements. The range within which $\frac{\text{residuals}}{\beta_1 \ln G}$ fall is so narrow that it can safely be assumed that transport infrastructure has similar importance across the entire study area and that there are no place-specific phenomenon of consequence at play skewing the results of this experiment.

3.6.3 Analytical Cost Function Results

As shown in Section 3.4, the cost function for the manufacturing industry is analytically derived according to (3.6). The coefficient of public capital expenditures from
Table 3.2, which represents the marginal effect of $G$ on profit margin, is 0.01879. The coefficient of transportation infrastructure expenditures from Table 3.6 represents the marginal effect of $G$ on output and is 0.20765. Therefore the effect of an increase in own-state transportation infrastructure expenditures on industry total costs is $-0.196$. Similarly, a corresponding increase in $\bar{G}$ has a marginal effect of 0.04915 on profit margins and a marginal effect of 0.46304 on output. This represents an industrial total cost reduction by a factor of $-0.44$ associated with an increase of neighboring state transportation infrastructure expenditures.

The cost structure of the model given by (3.6) shows that an increase in $G$ or $\bar{G}$ has a beneficial effect on both output and profit margins. Neighboring state transportation infrastructure expenditures have around twice the effect on both production and profit margins when compared to own-state transportation infrastructure expenditures. This is related to the fact that an increase in own-state transportation infrastructure expenditures is due to a corresponding increase in taxes, something not experienced by out-of-state firms. This makes the increased infrastructure more beneficial to out of state firms as they maintain the ability to use the new infrastructure without a change in price.

### 3.7 Conclusions

As is evident from this study, transportation infrastructure continues to play a role in reducing manufacturing production costs. Spatial spillovers are present and appear to be increasing over time. Own-state transportation infrastructure and neighboring state’s transportation infrastructure reduces production costs and increases output as well as industrial profitability. This study took a departure from the literature in its derivation and estimation of the cost function. This method allows output to vary and provides results of other common points of interest (profit and production) for
firms. Additionally, through the utilization of Geographic Information Systems, the results are presented in map form. This allows for quick visual comparisons and a rich analysis in terms of locations that receive relatively higher benefits from increases in levels of transportation infrastructure.

This type of approach has relevance for policy makers and transportation planners. While the results show that strategic cooperation among transportation planning agencies across state borders offers a place for additional returns on investment, the method by which these effects and benefits are extrapolated can also prove useful. Several examples of infrastructures and policies that benefit neighboring states are possible: the Tri-State Tollway in Illinois, which has significant benefits not only for Illinois, but also Wisconsin and Indiana. Similar policy can be found in Michigan’s neighboring states; Ohio has strategic routes termed ‘Michigan Legal Weight’ - whereby Ohio weight limits for trucks can, via a permit, be raised to equal Michigan’s [legal limit] - a measure enacted in the 1980’s to increase the port of Toledo’s competitiveness4 (Ohio Department of Transportation). As economic activity does not follow strict political boundaries, more strategic planning and interaction can prove beneficial to improve the cohesiveness of transportation systems. Analyzing the locational component of benefits for specific industries can uncover areas that experience higher rate of returns. The availability of more disaggregated geographic data may help fine-tune the findings of our analysis. As recent transportation policy mandates increasing the alignment of transportation projects to broader national and economic goals, this approach can aid in the determination of areas which satisfy such requirements under various industries of interest (Moving Ahead for Progress in the 21st Century). An analysis that explores how the locations of benefits change under the modern economy’s shift towards more service-oriented activities or based on the changing industrial mix may also prove fruitful in terms of policy.

4Wisconsin also has a similar policy in place for Michigan weight limits.
Chapter 4

Manufacturer’s Cost Function Analysis

4.1 Background Information

A number of studies concerned with productivity effects of infrastructure have examined benefits and productivity enhancements in terms of the [dual] cost function. As discussed in Section 2.1, this setting has a number of convenient exploitable characteristics. It relaxes a number of assumptions embedded in a single production function framework by explicitly accounting for behavioral responses by firms interested in minimizing costs of production. Further, it allows the researcher to uncover firm decision making patterns regarding input substitution and complementarity as well as endogenizing output to the system (something not done in a single equation production function framework). As a great number of the previously reviewed studies suggest that public capital has far-reaching benefits that potentially surpass the benefits of private capital, further investigation is conducted here. The goal is to explicitly account for the spatial interactions between firms and industries within the US via transportation infrastructure, and include this infrastructure as an input in the production process. In a similar fashion to the previous study, internal state
and neighboring state infrastructure is expected to play a role in industrial production characteristics. As transportation infrastructure is a major component of access to markets, it is expected that those firms in locations with ease of access to substantial levels of transportation infrastructure will experience significant benefits from transportation related investments as it pertains to their profitability, production and overall firm success.

Cost function studies that examine the role of public capital on private industry differ [from production function studies] in two ways: the cost function implicitly assumes cost minimization (whereas the production function does not) and the approach directly examines determinants of and impacts on production costs. These models tend to employ highly flexible functional forms that preserve typical properties associated with firm’s behavior. The fact that most studies employ a form that is twice differentiable also allows for the deduction of a wide array of impacts. For instance, under cost functions of a transcendental logarithmic (translog) or a Generalized Leontief (GL) flexible form, impacts of public capital on private capital, output and labor (among others) are directly obtainable. Furthermore, this form allows one to uncover estimates of firm’s willingness to pay for additional factors of production as well the interplay between different factors\(^1\). Cost studies provide direct estimates of cost impacts associated with changes in different factor inputs, which indirectly provide insight into input demand. Among the earliest studies that employ the estimation of cost functions that include some aggregate measure of public capital are those carried out by Berndt and Hansson (1992) and Morrison and Schwartz (1992). Similar studies have been undertaken by Nadiri and Mamuneas (1994a), Morrison and Schwartz (1996a), Cohen and Paul (2003) and Cohen and Paul (2004). Morrison and Schwartz (1996a) directly uncover the effects of public capital on firm costs and patterns of substitution between internal and external input factors.

\(^1\)For example, complementarity/substitutability and increases in productivity, etc.
4.2 Data

The data used to estimate the cost function econometrically, as in Chapter 3, comes from the ASM and covers the same time period (1997-2010). Labor is separated into production labor, $P_{LP}$ and non production labor, $P_{LNP}$. A short run cost function is estimated and as such relies on variable costs, $VC$ – taken as the product of the prices of nonproduction and production labor times their respective wage bills in addition to the price of materials times total material expenditures for each state $i$ in each year $t$.

The capital stock creation is carried out via the perpetual inventory method to convert the expenditure (flow) data into stocks. The benchmark (initial level of capital) is taken as an average of transportation expenditures (for each mode) over the years 1990-1994 multiplied by the modal service life. As the transportation capital stock variable is composed of highways, airports and waterports, the three components are computed separately such that: highway service life is taken to be 35 years; airports 25 years; and waterports 20 years. The depreciation rates come from the Federal Highway Administration and are applied to the expenditure series.

Private capital stock variables are created in a similar fashion based on the perpetual inventory method. According to theory, this is necessary as utility is had from the entire usable capital, rather than that of only newly created capital. It is, however, quite subjectively determined by the researcher, as no solid and concrete data exists on the truly generalized depreciation and service lives of public capital. As previously noted, the creation of capital stocks require at minimum, gross investment, an initial benchmark and a depreciation rate\(^2\). The methodology employed here closely resembles that of Zegeye (2000), Cohen and Paul (2004) as well as many others.

\(^2\)Or, service life
4.3 Dual Cost Functions

Implicit to the cost-function model choice is the assumption that firms minimize costs through the optimal combination of inputs (and their prices), demand, capacity and technological as well as environmental conditions (Cohen and Paul, 2004). In general form we estimate the cost function resembling:

\[
TC = VC_{i,t}(Y_{i,t}, K_{i,t}, p_{i,t}, r_{i,t}) + p_{i,t}K_{i,t}
\]

where \(i = 1, 2, \cdots, 48; t = 1997, \cdots, 2010\); \(TC_{i,t}\) is the total cost in state \(i\) at time \(t\), \(VC_{i,t}(\cdot)\) represents the restricted cost function in state \(i\) at time \(t\) which incorporates short-run constraints from fixed capital stocks in each period \(t\). \(Y_{i,t}\) is the aggregate output in state \(i\) at time \(t\) while \(K_{i,t}\) represents fixed private capital in state \(i\) at time \(t\).\(^3\) \(p_{i,t}\) is the vector of variable input prices in state \(i\) at time \(t\) and \(r_{i,t}\) is the vector representative of external shift factors for state \(i\) at time \(t\). Variable input prices, \(p\), are comprised of both production and non production labor - denoted \(L_{i,t}^P\) and \(L_{i,t}^{NP}\) respectively - as well as input material, \(P_{M_{i,t}}\). Similarly, the prices for each are given as \(P_{i,t}^{LP}, P_{i,t}^{LNP}\) and \(P_{i,t}^M\) such that \(p_{i,t} = (P_{i,t}^{LP}, P_{i,t}^{LNP}, P_{i,t}^M)\). The external shift factor is expressed with a time trend which is representative of technical change, \(t\), public infrastructure stock \(G_{i,t}\) in state \(i\) and public infrastructure stocks of neighboring states, \(\bar{G}_{i,t}\); hence, \(r_{i,t} = (t, G_{i,t}, \bar{G}_{i,t})\).

The allowance for a spatial autoregressive error structure is specified so as to allow for spatial and temporal correlation. The public and private components of the restricted cost function are created utilizing the perpetual inventory method. Estimation is carried out utilizing Kapoor, Kelejian and Prucha (KKP) 2007 panel data General Method of Moments (GMM) estimation on a transcendental logarithmic Cost Function. The model takes the

\(^3\)Computed private stock levels are for the end of each time period, \(t\).
form:

\[ VC_{i,t} = V_{i,t}(Y_{i,t}, K_{i,t}, p_{i,t}, G_{i,t}, \bar{G}_{i,t}, t) \]

\[
\ln\left(\frac{VC}{P_M}\right) \equiv \alpha_0 + \beta_{LP} \ln\left(\frac{PLP}{PM}\right) + \beta_{LNP} \ln\left(\frac{PLNP}{PM}\right) + \beta_Y \ln(Y) + \beta_K \ln(K) + \beta_G \ln(G)
\]

\[
+ \beta_{G} \ln(\bar{G}) + \beta_t + \frac{1}{2} [\gamma_{LP} \ln^2(\frac{PLP}{PM}) + \gamma_{LNP} \ln^2(\frac{PLNP}{PM}) + \gamma_{YY} \ln^2(Y)]
\]

\[
+ \gamma_{KK} \ln^2(K) + \gamma_{GG} \ln^2(G) + \gamma_{GG} \ln^2(\bar{G}) + \gamma_{tt} \ln^2(t) + \gamma_{LP} \ln^2(\frac{PLP}{PM}) + \gamma_{LNP} \ln^2(\frac{PLNP}{PM})
\]

\[
+ \gamma_{LP} Y \ln(\frac{PLP}{PM}) \ln(Y) + \gamma_{LP} K \ln(\frac{PLP}{PM}) \ln(K) + \gamma_{LP} G \ln(\frac{PLP}{PM}) \ln(G)
\]

\[
+ \gamma_{LP} G \ln(\frac{PLP}{PM}) \ln(G) + \gamma_{LNP} t \ln(\frac{PLNP}{PM}) \ln(Y) + \gamma_{LNP} G \ln(\frac{PLNP}{PM}) \ln(G)
\]

\[
+ \gamma_{LNP} K \ln(\frac{PLNP}{PM}) \ln(K) + \gamma_{LNP} G \ln(\frac{PLNP}{PM}) \ln(G) + \gamma_{LP} \ln(Y) \ln(K) + \gamma_{LP} \ln(Y) \ln(G) + \gamma_{LP} \ln(Y) \ln(\bar{G})
\]

\[
+ \gamma_{LNP} K \ln(Y) \ln(K) + \gamma_{LNP} G \ln(Y) \ln(\bar{G}) + \gamma_{LNP} K \ln(K) t
\]

\[
+ \gamma_{LNP} G \ln(G) \ln(\bar{G}) + \gamma_{LP} \ln(G) t + \gamma_{LNP} \ln(G) t + u_N
\]

where

\[
u_N = \rho(I_T \otimes W_N)u_N + \epsilon_N \quad (4.2b)
\]

\[
\epsilon_N = (eT \otimes N)\mu_N + \nu_N \quad (4.2c)
\]

and \( \nu_{i,t} \sim N(0, \sigma^2) \); \( w_{i,j} \in W \) is the weight that state \( j \) has on state \( i \) with \( w_{i,i} = 0 \), \(-1 < \rho < 1; n, m = (L^P, L^{NP}, M), n \neq m; i, j = (1, 2, \cdots, 46); t = (1997, \cdots, 2010). \) \( \epsilon_N \) & \( \nu_N \) are each assumed to be \( i.i.d \) and independent of one another. Estimation utilizes the cost function (4.2a–c) as well as demand equations of each input. The equations for the demand of each variable cost input are computed according to Shephard’s Lemma. That is, \( \gamma_{i,t} = \frac{\partial VC_{i,t}}{\partial p_{i,t}} \) where \( n = (L^P, L^{NP}, M) \). They take the following form:

58
\[ q_{n,i,t} = \gamma_{n,i,t}(Y_{i,t}, K_{i,t}, P_{i,t}, G_{i,t}, \bar{G}_{i,t}, t) \equiv \frac{\partial VC(\cdot)_{i,t}}{\partial p_{n,i,t}} \] (3.9)

\[ S_L = \gamma_{PA} + \gamma_{PAP} P_{A} \ln(\frac{P_{PA}}{P_{M}}) + \gamma_{PAP} P_{LN} \ln(\frac{P_{LN}}{P_{M}}) + \gamma_{PAP} Y \ln(Y) \] (3.9a)

\[ + \gamma_{PAP} K \ln(K) + \gamma_{PAP} G \ln(G) + \gamma_{PAP} \bar{G} \ln(\bar{G}) + \gamma_{PAP} t + u \]

\[ S_{LN} = \gamma_{PAP} + \gamma_{PAP} P_{LN} \ln(\frac{P_{LN}}{P_{M}}) + \gamma_{PAP} P_{LN} Y \ln(Y) + \gamma_{PAP} K \ln(K) + \gamma_{PAP} G \ln(G) \] (3.9b)

\[ + \gamma_{PAP} \bar{G} \ln(\bar{G}) + \gamma_{PAP} t + u \]

\[ S_M = 1 - S_L - S_{LN} \] (3.9c)

where

\[ u_N = \rho (I_T \otimes W_N) u_N + \epsilon_N \] (3.9d)

\[ \epsilon_N = (e_T \otimes N) \mu_N + \nu_N \] (3.9e)

\[ \nu_{n,i,t} \sim N(0, \sigma^2_N); \ q_{n,i,t} \] is the demand for input \( n \) in state \( i \) in year \( t \), and as before, \(-1 < \rho < 1\). Elements of \( \nu_{n,i,t} \) are \( i.i.d. \), the elements of \( \epsilon_{n,i,t} \) are \( i.i.d. \) and \( \nu_{n,i,t} \) and \( \epsilon_{n,i,t} \) are independent. Combinations of \( \nu_{i,t}^{LP}, \nu_{i,t}^{LN} \) and \( \nu_{i,t}^{M} \) also remain independent. With estimation complete it becomes possible to uncover firm behavior with respect to changes in variable inputs and various capital measures. This second order analysis provides notions of firm’s willingness to pay for additional inputs and insight into the substitutability or complementarity of inputs. It is also possible to derive estimates of how exogenous inputs affect endogenous inputs productivity\(^4\).

\(^4\)For example, it is possible to compute how an increase in neighboring transportation infrastructure may impact the demands for private capital or labor inputs, etc.
4.4 Econometric Cost Function Results

By imposing the constraints as outlined above, the number of parameters to estimate is reduced substantially. The estimation is carried out in two ways. The first uses maximum likelihood and ignores spatial autocorrelation. Examining the residuals shows the presence of spatial dependence and motivates the use of the Kapoor et al. (2007) General Method of Moments routine. Results are presented below.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coeff</th>
<th>Std. Err</th>
<th>Variable</th>
<th>Coeff</th>
<th>Std. Err</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{VC}$</td>
<td>10.9646</td>
<td>1.7660</td>
<td>$\gamma_{PLPK}$</td>
<td>-0.0200</td>
<td>0.0311</td>
</tr>
<tr>
<td>$\alpha_{LP}$</td>
<td>0.1428</td>
<td>0.0145</td>
<td>$\gamma_{PLPG}$</td>
<td>0.0887</td>
<td>0.0294</td>
</tr>
<tr>
<td>$\alpha_{LNP}$</td>
<td>0.1118</td>
<td>0.0144</td>
<td>$\gamma_{PLPG}$</td>
<td>0.0612</td>
<td>0.0272</td>
</tr>
<tr>
<td>$\beta_{LP}$</td>
<td>1.9143</td>
<td>0.5292</td>
<td>$\gamma_{PLPt}$</td>
<td>-0.0365</td>
<td>0.0057</td>
</tr>
<tr>
<td>$\beta_{LNP}$</td>
<td>-0.2147</td>
<td>0.5316</td>
<td>$\gamma_{PYNP}$</td>
<td>-0.0253</td>
<td>0.0342</td>
</tr>
<tr>
<td>$\beta_{Y}$</td>
<td>3.9365</td>
<td>0.5872</td>
<td>$\gamma_{PLNK}$</td>
<td>0.0165</td>
<td>0.0327</td>
</tr>
<tr>
<td>$\beta_{K}$</td>
<td>-2.5743</td>
<td>0.4136</td>
<td>$\gamma_{PLNG}$</td>
<td>0.0488</td>
<td>0.0298</td>
</tr>
<tr>
<td>$\beta_{G}$</td>
<td>-0.3306</td>
<td>0.5124</td>
<td>$\gamma_{PLNG}$</td>
<td>0.0797</td>
<td>0.0269</td>
</tr>
<tr>
<td>$\beta_{\bar{G}}$</td>
<td>-1.5365</td>
<td>0.3971</td>
<td>$\gamma_{PLNt}$</td>
<td>-0.0060</td>
<td>0.0057</td>
</tr>
<tr>
<td>$\beta_{t}$</td>
<td>-0.0678</td>
<td>0.0561</td>
<td>$\gamma_{YG}$</td>
<td>0.2015</td>
<td>0.0317</td>
</tr>
<tr>
<td>$\gamma_{PLPPLP}$</td>
<td>0.1184</td>
<td>0.0343</td>
<td>$\gamma_{Y}\bar{G}$</td>
<td>0.0813</td>
<td>0.0297</td>
</tr>
<tr>
<td>$\gamma_{PYNP}$</td>
<td>-0.1643</td>
<td>0.0387</td>
<td>$\gamma_{Y}\bar{G}$</td>
<td>0.0813</td>
<td>0.0297</td>
</tr>
<tr>
<td>$\gamma_{YY}$</td>
<td>-0.2700</td>
<td>0.0421</td>
<td>$\gamma_{Yt}$</td>
<td>0.0004</td>
<td>0.0043</td>
</tr>
<tr>
<td>$\gamma_{KK}$</td>
<td>-0.0075</td>
<td>0.0056</td>
<td>$\gamma_{KG}$</td>
<td>-0.0514</td>
<td>0.0294</td>
</tr>
<tr>
<td>$\gamma_{GG}$</td>
<td>-0.2167</td>
<td>0.0330</td>
<td>$\gamma_{K\bar{G}}$</td>
<td>0.1186</td>
<td>0.0259</td>
</tr>
<tr>
<td>$\gamma_{G\bar{G}}$</td>
<td>-0.0231</td>
<td>0.0349</td>
<td>$\gamma_{K\bar{t}}$</td>
<td>-0.0009</td>
<td>0.0043</td>
</tr>
<tr>
<td>$\gamma_{tt}$</td>
<td>0.0014</td>
<td>0.0013</td>
<td>$\gamma_{G\bar{G}}$</td>
<td>-0.2782</td>
<td>0.0307</td>
</tr>
<tr>
<td>$\gamma_{PLPPLNP}$</td>
<td>0.1227</td>
<td>0.0128</td>
<td>$\gamma_{Gt}$</td>
<td>-0.0018</td>
<td>0.0031</td>
</tr>
<tr>
<td>$\gamma_{PLPYN}$</td>
<td>-0.0898</td>
<td>0.0331</td>
<td>$\gamma_{Gt}$</td>
<td>0.0101</td>
<td>0.0029</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.5195</td>
<td>N: 48</td>
<td>$R^2$</td>
<td>.9998</td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{FGLS}$</td>
<td>0.5464</td>
<td>T: 14</td>
<td>AIC</td>
<td>-8720.0559</td>
<td></td>
</tr>
</tbody>
</table>

While the parameters in Table 4.1 are not directly interpretable, they are utilized to compute shadow values and shadow value elasticities. Shadow values allow for the evaluation of the various effects that own-state public infrastructure ($G$) and neighboring state infrastructure has on manufacturers variable cost as well as the relationship between public capital and demand for various factors of production. Shadow values are derived as the first
derivatives of the cost function and implicitly provide information about the value of these various factors to firms. For instance, to understand the effect that own-state transportation infrastructure has on manufacturers’ variable cost, in elasticity form, $\epsilon_{VC,G} = \frac{\partial \ln VC}{\partial \ln G}$. If this value is negative, it implies that an increase in own-state transportation infrastructure provides a reduction in variable cost to firms in the same state. Similar computations apply for other external factors of production. Table 4.2 displays various computations of the effects of own-state and neighboring-state transportation infrastructure on manufacturers’ profit margins.

Table 4.2: Cost Function Elasticities

| $\epsilon_{VC,G}$ | -1.14061 |
| $\epsilon_{P,L,K}$ | -0.17924 |
| $\epsilon_{VC,K}$ | -0.26539 |
| $\epsilon_{P,LNP,Y}$ | -0.17031 |
| $\epsilon_{VC,P,LNP}$ | -0.01143 |
| $\epsilon_{P,LNP,G}$ | 0.11789 |
| $\epsilon_{VC,t}$ | 0.03653 |
| $\epsilon_{P,L,G}$ | -0.02366 |
| $\epsilon_{P,LNP,G}$ | 0.75882 |
| $\epsilon_{P,LNP,K}$ | -0.27098 |

As evidenced by Table 4.2 own-state and neighboring state investments are very cost-reducing for the manufacturing industry. Similarly, increasing private capital and non-production labor is associated with a reduction in variable cost. This is suggestive of under-investment in capital structures and under-utilization of non-productive labor. Public capital, $G$ and neighboring state public capital $\bar{G}$ are both associated with reductions in the demand for production labor – suggesting some level of substitutibility, similar to the relationship commonly observed between labor and private capital. Interestingly, increases in neighboring-state infrastructure is correlated with an increased demand for non-production labor. Increases in neighboring state infrastructure drastically reduces own-state investment. The theme of free-riding, in terms of state-level investments is further reinforced in the following graphics. As in Chapter 3, it can once again be seen that those states bordering large market/high investment states (e.g. Nevada and Arizon to California, or Oklahoma, New Mexico to Texas) receive comparatively large benefits from their larger
neighbors.

### 4.4.1 Cost Function Maps

Figure 4-1 depicts this notion well in terms of own-state investment. It can be posited that those states experiencing the largest benefits in terms of cost reduction from own-state infrastructure investments do so as these investments increase the ease of access to the larger border-state markets. This clearly would reduce firms costs associated with transporting to markets as well as increase their available market size. This is easily seen for all states bordering California as well as those states in the North East bordering New York.

The map in Figure 4-2, depicting own-state manufacturers reduction to variable costs associated with increases in neighboring state infrastructure follows a similar pattern as before in that relatively underdeveloped markets (Montana, Wyoming, Nevada, North & South Dakota’s as well as New Mexico and others) experience greater benefits than those more developed states.
Figure 4-2: Elasticity of VC with respect to Neighboring-State Transportation Infrastructure

Figure 4-3: Elasticity of Own-State Transportation Investment with respect to Neighboring-State Transportation Investment
Figure 4-4: Elasticity of Private Capital with respect to Neighboring-State Transportation Investment

Figure 4-3 perhaps reinforces the free-riding notion best as the pattern is the same as previously noted. It displays that neighbors to states with large & well developed markets (and transportation networks) not only experience greater benefits from their neighbors, but also invest less as their more developed neighbors invest more.

Figure 4-4 tells a somewhat different story, although with a very similar pattern as before. The manufacturing industry in those states bordering the large market states experience a substantial increase in demand for private capital as their large market neighbors increase their levels of public capital. This is suggestive of the flight of manufacturers in search of lower costs. Firms are better off locating next to large market states without the increased costs associated with being within those market states. See for instance, Oregon, Nevada, Arizona, Louisiana, Indiana and Connecticut. All states have experienced increases in population and labor force levels in recent years as their more expensive and increasingly congested neighbors (California & Washington for Oregon, Nevada and Arizona; Texas for Louisiana; Illinois for Indiana and Massachusetts and New York for Connecticut). The rel-
ative level of influence is seemingly one directional – from large neighbor to all smaller neighbors.

Figure 4-5 tells a similar story from the perspective of the large market states. Large market states are among those that experience the largest decreases in demand for non-production labor when increased levels of public capital investment are experienced. Figure 4-6 again reinforces the idea of the substantial benefits experienced by smaller, less developed neighbors of large market states. Increases in neighboring state public capital (perhaps indicative of stronger markets) are associated with an increase in out of state non-production labor demand.
Figure 4-6: Demand for Non-Production Workers with respect to Neighboring-State Transportation Investment
Chapter 5

Transportation Infrastructure: Household Value

As shown by the results in Section 3.6, private industry receives significant benefits from transportation infrastructure investments. On its own, the results provide a compelling argument as to the extent of significant economic benefits related to transportation infrastructure. However, the preceding analysis leading to those conclusions fails to account for citizen use. A significant portion of transport infrastructure exists with the intention of allowing people to move about freely throughout geographic space. Conceptually, this implies that there are some effects associated with transportation infrastructure that are unaccounted for in the private-sector productivity-related studies. Since the household is not accounted for previously it must be established whether the potential for bias exists. If proven to be significant, this potential bias can then be remedied. As such, this portion of the study is concerned with the interplay between transportation infrastructure related investments and households. In examining the spatial variation in benefits to private industry it becomes clear that there are some locations which experience relatively large differences (in magnitude of the benefit of transportation infrastructure) and as such, it is believed that uncovering the household effect may shed some light on this phenomenon.

Furthermore, while data availability limits the geographic scale of the private industry
analysis, household data is widely available at significantly larger scales\(^1\). The finer resolution data provides further insights into how various types of surface transportation are useful for households. As this is not something directly observed, instrumental variables are utilized as a way to uncover these effects.

Transportation infrastructure is an increasingly important component of the modern spatial economy. The typical US household spends approximately 18% of its annual income on road transportation, while over half of all state and local government expenditures are related to transportation infrastructure (Bureau, b). Over the period 1980 to 2010, while the miles of public roads in the US increased by only six percent, the annual average vehicle miles traveled by Americans nearly doubled. In 1982, a typical US citizen spent 14 hours per year in congested traffic; by 2010 this figure grew to 34 hours per year (the Facts USA: The George Washington University). Given the substantial amount of time (estimated to be 161 person-minutes per day in 2001 (Duranton and Turner, 2011)) and money Americans allocate to passenger vehicle travel each year, it is hard to deny the social and economic importance of transportation infrastructure in the US. While a vast amount of work has been conducted that explores the relationship between infrastructure investment, industrial productivity and employment (Pereira and Andraz, 2011), few studies have focused on the linkage between transportation-related infrastructure and non-industry-related groups (e.g. households or individuals).

The purpose of this portion of the study is to establish whether a positive and statistically significant relationship exists between transportation infrastructure usage and the welfare of households in the 10-state Mid-America Freight Coalition Region. This work builds on the previous work which finds transportation infrastructure investment to be of value to industry and that neighboring state transportation infrastructure investments are more beneficial than own-state investments. On its own, Eloff et al. (2013) provides a compelling argument as to the extent of significant economic benefits related to transportation infrastructure. However, the analysis leading to those conclusions left citizen use outside of

\(^1\)When compared to the preceding private sector studies which take place at the state level, household data and related transport infrastructure has been obtained at the county level.
the scope. Assuming individual users experience positive and measurable effects from the use of transportation infrastructure, it is implied that there are some effects associated with transportation infrastructure that are unaccounted for in the private-sector productivity-related studies. Therefore, a quantitative examination is necessary to determine the spatial variation in benefits, if any, citizens receive from their usage of passenger vehicle related transportation infrastructure. Investigating the spatial variation in benefits to private industry (as in Eloff et al. (2013)) makes it clear that some locations experience relatively larger differences in magnitude of the benefit of transportation infrastructure. Extending this spatially focused examination to individuals has additional policy implications in the identification and understanding of ‘lagging’ regions. It serves as a tool for policy makers to understand what additional benefits (outside of industry) can be expected. By utilizing a finer geographic level of data that encompasses a homogeneous region of the United States, the data provides us with far more observations than is typical of past experiments as well as a more localized understanding of these effects.

This section of the dissertation is organized as follows: section 5.1 provides an overview of the literature regarding the economic effects of public capital, section 5.2 outlines the model and estimation; section 5.3 describes the data. Section 5.4 presents the results and section 5.5 offers some concluding remarks.

5.1 Background

The role of public capital has received much attention in the economic literature. The majority of the work is concerned with the Public Capital Hypothesis; that public capital proves beneficial for the goals of economic growth and increased productivity. The underlying notion of this hypothesis is that a solid infrastructure underpinning enables producing agents in the economy to better utilize their inputs to maximize output. The productive capacity of public capital is thought of as being twofold: that infrastructure can be treated as intermediaries in firms’ production processes and that these investments increase the productivity of other private inputs. For transport infrastructure in particular, this no-
tion is quite easy to grasp. Given such high levels of labor mobility in the US, it seems plausible that without sufficient and adequate levels of publicly provisioned transportation infrastructure, the phenomenon of high mobility would not exist, at least not at the levels witnessed today.

Although the literature contains many theoretical and historical contributions (Blum (1982), Eberts (1986), Dalenberg (1987), among others), empirical interest and findings were long absent until Aschauer’s (1989) seminal work. This work, which found public capital to explain a significant portion of the US productivity growth slow-down of the 1970’s, catapulted empirical interest in the topic into the mainstream. It is worth noting, however, that a number of empirical works pre-date Aschauer (1989). See, for example Blum (1982), Eberts (1986), Dalenberg (1987), Mera (1973), Ratner (1983), Costa et al. (1987), and Deno (1988).

5.1.1 Early Theoretical Foundations

The theoretical foundation from which the public capital hypothesis is based is an extension of early work by development economists. Meade (1952) provides one of the earliest known theoretical foundations of which the majority of subsequent empirical studies follow. His argument essentially lays the groundwork for the idea of unpaid factors of production (e.g. publicly provisioned capital) aiding other industries (or firms) for which no price of the capital was paid. In this light, public capital can be thought of “as an input in the production process of private industry that contributes independently to firms’ output” (Deno, 1988).

Hansen (1965) is among the first to directly discuss the effects of public capital on regional development. He divides public capital into two parts: ‘social’ and ‘economic’. Economic overhead capital (EOC), which includes roads, electricity, water treatment facilities, drainage/sewer systems, etc., can be thought of as the foundational public capital used in economic production processes. Social overhead capital (SOC), which includes schools, police/fire services, etc., is generally non-economic in nature and serves to provide support
for social institutions. A further distinction between these two types of public capital is also made: consumers of EOC tend to want easy access to the goods (well developed and widely available), while agents are often more willing to travel to utilize SOC facilities. Hansen's objective is to uncover “possible determinants of variation in...expenditures” (Hansen, 1965) and he hypothesizes that economic development potential is related to two factors: the type of public capital implemented and where in the development cycle the region exists. He determines that EOC has larger benefits on economic growth in “intermediate” regions (as opposed to congested or lagging ones) and that increased provisions of SOC tend to prove most beneficial when made in “lagging” regions.

Hansen’s argument for the types of investments most beneficial to intermediate and lagging region’s is simple and intuitive: intermediate regions tend to have more favorable conditions for development (lower costs, access to skilled labor and raw materials). Therefore, public investments of the EOC type have the power to spur further private investments which thereby stimulate economic growth. This growth comes from the expansion of economic activity within the intermediate region. This effect would serve to increase scale economies at a rate that far exceeds the incurred social costs of the increased [economic] activity and therefore projects of this type make the most sense to undertake. A similar argument follows for lagging regions; although access to beneficial ingredients for private investments are much less than when compared to intermediate areas, SOC type investments would tend to better serve as a foundation for future improvements in the area. Without the necessary social provisions to attract and prepare the locale for higher productive industry, Hansen argues that lagging regions will never gain in stature to become an intermediate region.

5.1.2 Differentiating Benefit Types

Much of the empirical literature reaches conclusions that are in-line with Hansen’s premise of EOC and SOC types benefiting regions differently. However a scant few explicitly recognize that focusing on one type of benefit (and failing to examine the other) is
potentially misleading. Duffy-Deno and Eberts (1991) and Dalenberg and Partidge (1997) are notable exceptions in their intent to separate the effects between industry and households. Duffy-Deno and Eberts are among the first to combine each side of the relationship between regional growth and public capital investment in a unified framework. Their study utilizes a two-stage least-squares estimation procedure to mitigate the potential bias due to errors in the estimates of public capital stock and as a fix to the notable problem of spurious correlation. They find that public capital stock has a positive and statistically significant effect on per capita income. However, their model serves only to differentiate the effects of public capital stocks and expenditures on per capita income. They are able to determine that public capital has a long lasting effect in that the benefits last longer than just the construction period (expenditure) - that is to say that there are long term benefits arising from the continual use of newly created structures, and that this benefit is twice as large as the immediate one time benefit from the expenditure itself.

Dalenberg and Partridge draw on work from the wages and rents literature. A substantial problem faced by this strand of the literature lies in the difficult-to-understand equilibrium conditions for land and labor market clearing conditions. The problem of causality is central and drawing from the wage-rent model laid forth by Roback (1982) attempts to remedy this. Dalenberg and Partridge recognize that the benefits from public capital arise in two ways; the previously mentioned unpaid factors of production and the fact that public capital is a household amenity.

5.2 Model

The benefits of transportation infrastructure usage to households are measured in terms of per capita income. It is expected that there exists endogeneity as the linkage between transportation infrastructure usage and income flows in both directions. That is, higher income receiving areas will have higher levels of transportation infrastructure usage and that having higher levels of infrastructure usage can further lead to increases in income through increased mobility, knowledge transfers and other agglomeration effects.
To account for the issue of endogeneity in our model, a set of spatial autoregressive models are employed in a panel data setting. As panel data models provide one way to control for individual heterogeneity and spatial autoregressive models account for spatial endogeneity, the combination of the two has the potential to further reduce bias that arises from measurement errors contained in our variables (Duffy-Deno and Eberts, 1991). The results in section 5.4 present the effects that the use of transportation infrastructure have on per capita income at the county level in our study area over the period 1980 to 2008.

It is assumed that per capita income in county $i$ in year $t$ is directly related to per capita income in neighboring counties and the weighted average of annual average daily traffic. As the data contain spatial dependence and due to the problem of endogeneity the model takes the form:

\[
PCI_{i,t} = \alpha + \rho WPCI_{i,t} + \beta_1 AADT_{i,t} + \epsilon_{i,t}
\]

\[
\epsilon_{i,t} = \mu_i + \nu_{i,t}
\] (5.1)

where $PCI_{i,t}$ is real per capita income in county $i$ in year $t$; $WPCI_{i,t}$ is the spatial lag of per capita income and $AADT_{i,t}$ is the weighted average of annual average daily traffic (AADT) in county $i$ in year $t$.

Initial data analysis procedures were undertaken to determine the proper model specification. Failing to account for spatial dependence and ignoring the panel structure of the data set provides alarming results. Figure 5-1 displays substantial evidence of heteroskedasticity as the widening of the prediction interval is indicative of such non-constant variances of the estimates.

Two additional model specifications are estimated that parse the effects of traffic utilization by population density. Three measures of AADT are constructed based upon the rural-urban designation contained in the Highway Performance Monitoring System (HPMS) data, denoted below by the AADT subscripts of $R$, $S$ and $U$. These roughly correspond to rural, suburban and urban areas as classified according to the population distinctions made by the Federal Highway Administration (FHWA) when computing the variable. Model (5.3)
also contains a labor to population ratio variable to help identify urban commute traffic and rectify intermediate spatial residual analysis.

$$PCI_{i,t} = \alpha + \rho W PCI_{i,t} + \beta_1 AADT_{R,j,t} + \beta_2 AADT_{S,j,t} + \beta_3 AADT_{U,i,t} + \epsilon_{i,t}$$

$$\epsilon_{i,t} = \mu_i + \nu_{i,t}$$  \hspace{1cm} (5.2)

$$PCI_{i,t} = \alpha + \rho W PCI_{i,t} + \beta_1 AADT_{R,j,t} + \beta_2 AADT_{S,j,t} + \beta_3 AADT_{U,i,t} + \frac{\text{Labor}}{\text{Population}_{i,t}} + \epsilon_{i,t}$$

$$\epsilon_{i,t} = \mu_i + \nu_{i,t}$$  \hspace{1cm} (5.3)

As the random effects model specification is a more flexible and less demanding specification when compared to fixed effects, it is the model of choice. It is, however, worth noting that both specifications produce nearly the same parameter estimates. $W$ represents the row-normalized first order queen contiguity spatial weights matrix that contains the usual properties as outlined in the spatial econometric literature Smirnov and Anselin (2001).
5.3 Data

The study area examined herein contains 962 counties in the 10 Mid-America Freight Coalition member states. These states are: Illinois, Indiana, Iowa, Kansas, Kentucky, Michigan, Minnesota, Missouri, Ohio and Wisconsin. The infrastructure data comes from Federal Highway Administration’s (FHWA) Highw ay Performance Monitoring System (HPMS) database. This database is available for all 50 states and territories of the United States from 1980-2008. The data are collected by each state and uploaded to FHWA for compilation. The database provides two annual files - the Universe and Sample. The Universe file includes information for interstate highway and other major and minor roads while the Sample includes additional information about the segments in the universe file. Each segment represented in HPMS is representative of a larger set of similar segments. Each segment then includes an expansion factor “which relates the length of the segment described in the data to the length of sample it represents” Duranton and Turner (2011). HPMS data includes records on the number of lanes, section length, annual average daily traffic (AADT), pavement rating, percent trucks among others. Our variable of interest, which is included in the sample file, is AADT. The HPMS sample file is a “stratified random sample of physical roadway sections” Administration. Each record of the sample contains section performance characteristics as well as locational data and an expansion factor. This data is grouped by county and weighted average measures of traffic volumes are constructed (with section length as the weighting scheme). These are used as input to model (1). Further distinctions are made by recomputing this measure based upon the rural-urban designator. The $AADT_R$ is applied to those routes termed rural by the HPMS (population less than 5,000). $AADT_S$ is computed for those routes through areas containing less than 50,000 persons (but more than 5,000). $AADT_U$ denotes traffic through areas with greater than 50,000 persons. There are 963 counties in our study region, of which 962 contain complete data. The county with missing data is Pope County, Illinois and is excluded to maintain a balanced panel. The panel dataset covers 29 years for a total of 27,898 observations.

Socio-economic data is obtained from the Bureau of Economic Analysis’s (BEA) Re-
Regional Economic Information System (REIS). This database covers all political boundaries (counties, metropolitan areas, etc) within the US from 1969 to 2008. Data related to population, per capita income and employment are selected for all counties in the study region.

Summary statistics of the data-set are presented below in Table 5.1.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCI</td>
<td>27898</td>
<td>18982.58</td>
<td>7857.22</td>
<td>0</td>
<td>60052</td>
</tr>
<tr>
<td>WPCIi</td>
<td>27898</td>
<td>18980.06</td>
<td>7301.656</td>
<td>0</td>
<td>46106.33</td>
</tr>
<tr>
<td>AADTRl</td>
<td>27898</td>
<td>3486.625</td>
<td>4739.983</td>
<td>0</td>
<td>66930</td>
</tr>
<tr>
<td>AADTS</td>
<td>27898</td>
<td>2790.31</td>
<td>4345.191</td>
<td>0</td>
<td>90277.62</td>
</tr>
<tr>
<td>AADTU</td>
<td>27898</td>
<td>2013.783</td>
<td>5094.255</td>
<td>0</td>
<td>96070</td>
</tr>
<tr>
<td>labor/700pop</td>
<td>27898</td>
<td>1.350343</td>
<td>0.795321</td>
<td>0.105516</td>
<td>14.77438</td>
</tr>
</tbody>
</table>

5.4 Results

The models outlined in section 5.2 are estimated using the combined HPMS and BEA dataset. A positive and statistically significant relationship is determined to exist between traffic activity and per capita income. Although the magnitude of the coefficient of traffic activity is relatively small after accounting for endogeneity, its positive significance reinforces the idea of more mobile and active locales earning higher incomes. Higher levels of traffic imply more activity that is both economic and non-economic in nature.

The results of the models outlined in section 5.2 are presented in Table 5.2. The signs, magnitudes and significance of the variables follow economic intuition. Plots of the estimated residuals appear to be normally distributed. There is a high degree of correlation between neighboring counties levels of per capita income. That is, neighboring counties tend to be very similar in their income levels. This notion is further reinforced by the parameter $\rho$ in the models and in Figure 5-2, which displays the mean per capita income distribution throughout the study area over the time period 1980-2008. $N$ denotes our sample size and $T$ represents the number of time periods. As expected, those counties containing and
Table 5.2: Household Estimation Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model (1)</th>
<th>Model (2)</th>
<th>Model (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>$-44.1484^{*}$</td>
<td>$1946.061^{*}$</td>
<td>$732.3984^{*}$</td>
</tr>
<tr>
<td></td>
<td>(93.5150)</td>
<td>(95.4139)</td>
<td>(73.6450)</td>
</tr>
<tr>
<td>$WPCI$</td>
<td>$0.9874^{*}$</td>
<td>$0.8188^{*}$</td>
<td>$0.7119^{*}$</td>
</tr>
<tr>
<td></td>
<td>(0.0022)</td>
<td>(0.0037)</td>
<td>(0.0033)</td>
</tr>
<tr>
<td>$AADT$</td>
<td>$0.0739^{*}$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td>(0.0055)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$AADT_R$</td>
<td>$-$</td>
<td>$0.1659^{*}$</td>
<td>$0.0978^{*}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0054)</td>
<td>(0.0042)</td>
</tr>
<tr>
<td>$AADT_S$</td>
<td>$-$</td>
<td>$0.0996^{*}$</td>
<td>$0.0733^{*}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0054)</td>
<td>(0.0043)</td>
</tr>
<tr>
<td>$AADT_U$</td>
<td>$-$</td>
<td>$0.3176^{*}$</td>
<td>$0.2227^{*}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.00659)</td>
<td>(0.0055)</td>
</tr>
<tr>
<td>labor</td>
<td>$-$</td>
<td>$-$</td>
<td>$2772.68^{*}$</td>
</tr>
<tr>
<td>pop</td>
<td></td>
<td></td>
<td>(0.5669)</td>
</tr>
<tr>
<td>$N$</td>
<td>962</td>
<td>962</td>
<td>962</td>
</tr>
<tr>
<td>$T$</td>
<td>29</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.9638</td>
<td>0.9181</td>
<td>0.9354</td>
</tr>
</tbody>
</table>

bordering major cities have historically had the highest levels of per capita income. The levels tend to diminish as one moves farther from these agglomeration centers. Perhaps a bit surprising is the uniform and stable levels of income in Iowa and Kansas.

The estimated coefficient for AADT of 0.0739 in (5.1) implies that an increase in annual average daily traffic yields a statistically significant increase in per capita income. The same is prevalent in models (5.2) and (5.3), albeit at differing levels depending upon population. Model (5.3) maintains a similar relationship between the magnitudes of AADT by region type as estimated in (5.2), though at reduced levels. The addition of the labor-population ratio variable also shows that larger labor markets, relative to the population levels earn significantly more. Urban areas receive the highest benefits to per capita income from increases in traffic, followed by rural and suburban areas. The inclusion of the relative labor market size variable serves to stabilize the estimates of the AADT coefficients in (5.3). This is evidenced by the maps displaying the variance of the residuals in Figures 5-3 and 5-4.

Figure 5-5 presents the annual average daily traffic for the year 2008 in the region. Highest traffic volumes occur again in and around major cities but also along interstates and along corridors between cities. This is best witnessed examining the band around Lake Michigan, from north of Milwaukee, south through Chicago and east and north all the way
Figure 5-2: Average Real Per Capita Income Distribution Map
Figure 5-3: Variance of Residuals Model 5.2
Figure 5-4: Variance of Residuals Model 5.3
to Detroit. Similarly following I-75 South from Detroit to the state of Kentucky displays
the same heightened levels of traffic volumes.

It is interesting to note that smaller cities located along major interstate corridors tend
to experience higher changes in income from increased traffic, as shown by Figure 5-6.
Figure 5-6 displays the elasticity of the constructed AADT measure (from (1)) with respect
to per capita income. It is the mean elasticity for each county throughout the time period.
The measure is constructed as \( \left( \frac{\partial PCI}{\partial AADT_i} \right) \cdot \left( \frac{AADT_i}{PCI} \right) \) which is better expressed as \( \frac{\beta_i(AADT_i)}{PCI_i} \).
Following both I-80 and I-94 east from Chicago reveal that the majority of the counties
along the route experience an increase in income as AADT rises. The northern states
bordering Wisconsin experience similar effects, as do those counties in Ohio along I-75 and
Ohio State-Route 23 (which leads to Columbus from Toledo).

5.5 Conclusions

As the results of this study indicate, transportation infrastructure usage does provide
a measurable benefit to individual standard of living (in terms of income). The statistical
significance of the positive benefit to households reinforces the notion that this difficult-
to-measure phenomenon is present. By employing spatial analytical tools and techniques
we are able to visualize the magnitude and variation of the benefits throughout the study
area. As expected, major cities experience relatively larger increases in income as AADT
increases. Perhaps more interesting is the fact that those counties located along major con-
necting routes/corridors also experience a higher increase in income (relative to non-corridor
counties) with increased traffic volumes. These results suggest that increased transportation
infrastructure usage contributes to a higher standard of living. Further, the visualization
of the elasticities aides in understanding regions that appear to stand to gain more from
increased usage. While usage alone is not the determining factor, in line with Hansen’s ar-
gument, those locations of low utilization stand to gain from increased development. This
notion is further reinforced in Duranton and Turner (2011) which details the fundamental
law of road congestion - that “if you build it, they will come.”
Figure 5-5: Weighted AADT, 2008
Figure 5-6: Elasticity of AADT with respect to Per Capita Income
Employing transportation infrastructure usage as a proxy for transportation infrastructure demand allows for us to determine the benefit to individuals that the observed demand for infrastructure has on standards of living (measured by per-capita-income). Given the presence of a measurable benefit to individuals from infrastructure usage, there is a significant role for policy. As Eloff et al. (2013) found that increased levels of strategic cooperation among transportation planners and policy makers could serve to increase return on investment, the same holds true here. Improving the cohesiveness of transportation systems can provide benefits to both industry and individual users alike. Given recent changes in national transportation policy, the method of analysis can provide an additional tool in identifying ‘lagging’ regions set to benefit the most from increased investment.

The experiment undertaken here illustrates the power of utilizing different measures of activity as well as transportation networks. By segmenting the national highway network into its various use scenarios (rural, suburban and urban) significant insight into the different purposes and implications each type of the network have on not only localized economies, but also the general impact that these types of transportation networks have on household welfare.
Chapter 6

Aviation

6.1 Objective

The main objective of this chapter is to determine whether airports engage in strategic interactions related to infrastructure investment or if policy could improve the efficiency of National Airspace System (NAS) by strategically allocating capital meant for infrastructure improvements. By examining data related to airport connectivity, delays and infrastructure investments it is possible to understand if (or which) airports invest in a coordinated manner that benefits the NAS as a whole. In the event that certain highly connected routes have not engaged in strategic interactions in the investment process, this research will aid policy makers in highlighting the existence of potential areas for further efficiency gains to the system to be realized. The contributions from this work are two-fold: first, a method of accounting for the network effects of operational changes is outlined; and second, a method for the identification of man-made versus naturally occuring networks is presented. The distinction is important as man-made networks are endogenous to the system examined and as such necessitate the utilization of panel-data methods. Furthermore, as [the] man-made network is endogenous to the system under examination the method for identification is important for identifying the generalibility of inferential statistics – the estimates arising from the examination of a man-made transportation network are subject to change in the event of a network modification that alters the connectivity thereof.
6.2 Introduction

Aviation is a vital component of the United States; contributing significantly to both the domestic and global economy,\(^1\) national defense, and recreation among others. Maintaining efficiency in the system is imperative to U.S. national interests. As the Federal Aviation Administration (FAA) forecasts air traffic growth for U.S. carriers to increase substantially over the next twenty years,\(^2\) it follows that airports will also be subjected to similar increased levels of demand. One component of the solution will necessarily come from further investments in airport infrastructure Corporation (2007). Given current fiscal constraints, the inherent network structure of the National Airspace System (NAS) and the fact that any one airport’s operations affect other airports within the system, capital investments should be made in such a way that serve to maximize the productivity of the system as a whole. More simply, as traffic between airports continues to increase, if one airport substantially increases its capacity and throughput, it may be best for other highly-connected airports to implement upgrades at a similar point in time. Without some congestion mitigation tactics or efficiency-improving investments elsewhere in the system, a portion of these infrastructure improvements may fail to reach their full potential if upgrades merely shift congestion to different nodes on the network. As such, it is the goal of this research to understand current airport interactions as well as to provide a framework for quantifying how these interactions spread throughout the network. This will further aid policy makers in understanding how far-reaching investments or operational changes [at various nodes in the network] may be when accounting for the connectivity of the system. With a better understanding, funds can better be targeted towards projects that have (relatively) larger network enhancing effects.

This notion of accounting for external benefits is not new. Many examples of policy discussions calling for the inclusion of ‘total effects’ in investment analysis can be found. For example, White House Circular No. A-94 (House), “Guidelines and Discount Rates for

---

\(^1\)Approximately 5% of U.S. GDP for Transportation (2013)

\(^2\)By 2032, FAA of Transportation: Federal Aviation Administration (2012) expects a 90 percent increase in revenue passenger miles, and a 50 percent increase in the number of handled aircraft.
Benefit-Cost Analysis of Federal Programs,” states (in section 6) that “Social net benefits, and not the benefits and costs to the Federal Government, should be the basis for evaluating...where actions by one party impose benefits or costs on other groups.” Following this Circular, the FAA issued a notice that required the inclusion of a Benefit Cost Analysis (BCA) for all proposals requesting the use of discretionary Airport Improvement Program (AIP) funds for capacity-related projects (of Transportation, 1999). When comparing competing investments, the utilization of system-wide multipliers can aid in ensuring that the most beneficial projects are chosen during the decision-making process. Accounting for the total effect of an implementation and quantifying the expected impact that a change may have is crucial from a policy planning level. Measuring the total effect of an investment ensures that the right projects are chosen that meet the expected performance goals needed for efficient outcomes at the network level well into the future. Focusing explicitly on expected outcomes at the node a project is to be undertaken potentially understates the value of benefits. By accurately measuring the operational spillovers, the expected benefits of some projects may increase by enough so as to surpass their costs. Failure to measure the true value of benefits to the system as a whole results in inefficient outcomes.

While certain multiplier effects have become frequently utilized in FAA Benefit Cost Analyses, they have historically been limited to the idea of a delay multiplier; which assumes that a project expected to reduce delay at the airport of consideration yields an increased benefit by the amount of the delay propagation multiplier (Churchill, 2007). It is the goal of this research to provide a framework that exposes how changes at any given airport can be expected to spread throughout the network.

6.3 Research Approach

By exploring the parameter coefficients from the results of various spatial panel data models, insights into the magnitude and extent to which changes in one airport’s level of capital stock can be expected to have on other connected airports can be uncovered. Broadly speaking, this approach is theoretically based on the spillover and resource-flow
models outlined in K. (2003). The spillover model, framed as a spatial reaction function, explicitly recognizes the interdependencies between a decision variable for one unit of observation (airport) and all other units in the system (other connected airports). Applied in this context, it allows for the determination of how one airport’s outcomes (operations, on-time departures, etc.) are determined in light of outcomes at other airports given their operational characteristics. This could be especially relevant when conducting project specific Benefit Cost Analyses for discretionary Airport Improvement Program funds.

A model of this type has the ability to uncover the extent and presence of strategic interactions in infrastructure investment between connected airports. Alternatively, the resource-flow type model models the decision of one agent (airport) not directly based upon the levels chosen by other agents, but via indirect means. This case would be representative of reactionary investment policies, whereby airports do not strategically invest based upon one another’s investments, but because of the (hypothetically) increased throughput of passengers and planes through an airport due to some infrastructure improving investment. In practice, these two models are observationally equivalent and represent the ‘inverse problem.’ So while we are unable to determine the cause, we can observe whether there is substantial dependence between airport investment practices. Imagine for a moment that node A, a major airport with significant congestion issues, implements a new infrastructure improvement project that effectively allows for an increase in its throughput capabilities. Node B, which is heavily connected to node A, must now handle increased levels of traffic due to the now more efficient infrastructure employed at node A. For node B to maintain the new status quo, it is now necessary to implement upgrades capable of (at least) handling the new capacity. However, the adjustment period during which the new investment is implemented (via these indirect and reactionary means) serves to reduce the overall return on investment from the project conducted at node A. Similarly, from a user welfare perspective, the initial investment was a waste; it does nothing to aggregate welfare until the congestion at node B is reduced and the system moves back towards efficiency. Although

\[\text{For convenience, this user just so happens to be engaged in travel between node A and node B.}\]
this example does not lay out all possible scenarios of the benefits of strategic interaction, it serves to highlight the premise of the argument.

6.3.1 Models

Two models are estimated to illustrate potential uses of the proposed framework. By accounting for the network effects present within the system, ‘spatial’ multipliers of different types can be uncovered. The first model:

\[
\ln Y_{i,t} = \rho W \ln Y_{i,t} + \alpha + \beta_1 \ln K_{i,t-1} + \beta_2 \ln OPT_{i,t-1} + \beta_3 \ln POP_{i,t-1} + \beta_4 SM_{i,t} + \beta_5 MED_{i,t} + \beta_6 LG_i, t + \epsilon_{i,t} \tag{6.1}
\]

\[
\epsilon_{i,t} = \mu_i + \nu_t
\]

presents airport i’s operations, Y, in year t as a function of connected airport operations, WlnY; own airport capital stock, K, in year t – 1; population, POP, in the three counties bordering the county containing airport i in year t – 1; while small (SM), medium (MED) and large (LG) hub binary variables denote airport i’s hubsize. The estimation procedure provides spatial multipliers for all connected airports. This can be thought of as the effect that one airport, j, exerts on each connected airport, i. Aggregate multipliers for all airports, j, can be computed individually as well as grouped according to various criterion (such as hub-size, regional location, etc). This spatial multiplier is a representation of the total effects that all airports have on one another after accounting for feedback loops and is the aggregate of the direct and indirect effects. This total multiplier measure can also be decomposed into direct and indirect effects - failure to account for the indirect effects yields lower estimated effects. This is important to note as there are perhaps many beneficial projects that have not been undertaken as the direct benefit may not outweigh the project’s costs – utilizing the total benefit measure may paint a different picture, and in the face of funding reductions, projects that contribute more to the network as a whole provide one
way to be more effective with less investment.

Similarly, if one were interested in calculating the effect of departure delays, a model of the form above can be estimated - in this case, such that:

\[ OTP_{i,t} = \rho WOTP_{i,t} + \alpha + \beta_1 SM_{i,t} + \beta_2 MED_{i,t} + \beta_3 LG_i, t + \beta_4 OTP_{i,t-1} + \beta_5 lnK_{i,t-1} + \epsilon_{i,t} \]

\[ \epsilon_{i,t} = \mu_i + \nu_t \]

where \( OTP \) is mean on-time departure percentage at airport \( i \) for year \( t \), \( WOTP \) represents the on-time departure percentage of connected airports, \( SM, MED \), and \( LG \) are hub-size dummy variables where each variable takes the value of 1 (one) if airport \( i \) is a hub of the variable-specific size and 0 (zero) otherwise. \( OTP_{i,t-1} \) represents the preceding year’s on-time departure percentage at airport \( i \); \( lnK \) represents the capital stock at airport \( i \) in year \( t - 1 \). These variables are taken in year \( t - 1 \) to further mitigate potential problems due to endogeneity.

These spatial lag models directly encompass the network effects inherent to the NAS. As one airport’s operational characteristics have both direct and indirect effects on other airports in the system, further analysis into the interplay between airports can be examined. In doing so, one can take a deeper look at how on-time departure performance at one airport affects other connected airports’ on-time departure performance. Further use of these multipliers can provide insight into how a change in an explanatory variable at one airport affects the outcome (on-time departure performance, in this case) at another. These multipliers could be utilized at the policy level to ensure that the true network-wide impact of a project under consideration is quantified.

### 6.3.2 Data

The data necessary to determine these relationships is available from multiple sources. The U.S. Department of Transportation’s (USDOT) Bureau of Transportation Statistics
(BTS) maintains two databases relevant for the study: the T-100 data on origin-destination pairs and the Airline On-Time Performance Database. BTS T-100 provides dynamic measures of connectivity while the Airline On-Time Performance Database outlines origin-destination pairings and airline on-time measures and delay causes. On-Time Performance characteristics data provides coverage for all flights occurring with one of 16 airlines: 15 of which account for “one percent of total domestic scheduled-service passenger revenue” of Transportation Statistics, while the last reports the information voluntarily. The database provides a breakdown of differing delay types for each flight that encounters a delay. These categories include delays related to weather, National Airspace System (NAS), carrier, security and late arriving aircraft as well as total departure and arrival delays (in minutes). According to the On-Time Performance dataset weather category, weather accounts for approximately 4% of delays. Upon further examination there is another category of weather that is rolled-up in the NAS category. For the year 2011, “75.5% of NAS delays were due to weather” while NAS delays comprised 24.8% of all delays. This data is also constrained to 16 airlines that have “one percent of total domestic scheduled-service passenger revenue” (Bureau of Transportation Statistics, 2014). The reduced On-Time performance dataset computed in the data processing stage contains 65% of total operations reported by the T-100 segment data. Mean quarterly averages of the coverage ratios are displayed below in Figure 6-1. Restricting the on-time data to the selected time-horizon results in information on some 67+ million flights occurring at more than 359 unique origins and 363 unique destination airports. Similarly, T-100 data provides aggregate information on 100+ million flights occurring at 2,698 unique origin and 2,786 unique destination airports during the same time period. As the estimation routines to be used in the analysis require a balanced panel dataset, those airports with entries missing from any period in the time horizon have been excluded. The result of balancing the data and performing an intersection of the two sets yields 238 unique airports with complete data. This data is used as inputs in the mod-

---

4 This is explained as avoidable weather delays – delays of the type that “could be reduced with corrective action by the airports or the Federal Aviation Administration” of Transportation Statistics.
eling process to provide information about the operational characteristics of the individual airports included in the models. Additional data from FAA outlining airport investments provides the necessary information to determine infrastructure investment patterns. The raw data is obtained from USDOT BTS covering the years 2004 to 2012.

Two spatial weights matrices are calculated - one that is a binary matrix representing whether any given pair of airports are connected (1) or not (0); the other is representative of the degree to which any pair of airports interact. That is, each element $i,j$ ($i \neq j$) represents

![Figure 6-1: On-Time to T-100 Operations Ratio](image)

Table 6.1: Yearly Mean By Hub-Size

<table>
<thead>
<tr>
<th>Hub Size</th>
<th>Flights</th>
<th>On-Time Departure Percentage</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large</td>
<td>383,440</td>
<td>60.93</td>
<td>30</td>
</tr>
<tr>
<td>Medium</td>
<td>112,920</td>
<td>64.38</td>
<td>37</td>
</tr>
<tr>
<td>Small</td>
<td>33,282</td>
<td>70.30</td>
<td>61</td>
</tr>
<tr>
<td>Non-Hub</td>
<td>11,017</td>
<td>72.97</td>
<td>90</td>
</tr>
</tbody>
</table>
Figure 6-2: Airport Hub-Size Distribution
the number of flights between airport $i$ and airport $j$. Both matrices are row-standardized such that the rows sum to one (consistent with the spatial econometric literature (Kapoor et al., 2007)).

All measures included in the modeling exercise are aggregated to annual time periods. FAA Compliance Activity Tracking System (CATS) data is obtained to create the capital stock series. The CATS database, as per the FAA Authorization Act of 1994 requires “commercial service airports to annually file financial reports with the FAA” (of Transportation: Federal Aviation Administration, 2014). These annual reports contain high level line items for income and expenses at each airport. The line item for capital expenditures is utilized in the measure of capital stock. The capital stock variable is created utilizing a benchmark for each airport and a depreciation rate of 25 years. The benchmark level is computed as an average of 2000-2003 capital expenditure data multiplied by the 25 year service life as taken from Cohen, 2003 (Cohen and Paul, 2003). This level is then depreciated and the following year’s capital expenditure level is added to the preceding capital stock figure to create the next year’s level of capital stock. This is done for each year under consideration. All monetary data has been converted to 2005 dollars using the Gross State Product (GSP) deflator to account for inflation and regional cost differentials. This data is then used, in addition to other collected variables, in the modeling process. Upon completion of the capital stock data series the panel set was re-balanced to incorporate only airports for which complete information is available. This balancing yields 218 unique airports for 9 years for a total of 1,962 observations. A map presenting the distribution of airports included in this study segmented by hub-size is shown in Figure 6-2, while the airports as well as the census regions they are segmented by are depicted in Figure 6-6. As expected, Large hubs are responsible for the largest portion of NAS flights, and experience the lowest on-time departure percentage. As evidenced in Table 6.1, which presents average operational and on-time performance data by hub-size, smaller airport’s (in terms of hub-size), on average, experience a lower number of flights and higher on-time percentages. It is also worth noting that large hubs (by count) represent less than 14% of the total number of airports examined herein.
6.3.3 Data Description

Figures 6-3, 6-4 and 6-5 present an overview of the mean annual capital stock level, spatial multiplier and the trend of commercial aviation seat capacity utilization (respectively). Figure 6-3 shows the aggregate trend of overall capital stocks in the dataset for the years 2004 to 2012. It is expected that the relatively substantial increase between 2009-2010 (X=6 and 7, respectively) is attributable to the American Reinvestment and Recovery Act. Figure 6-4 is calculated by running individual regressions for each time period and computing the average multiplier for the system. It shows a relatively dramatic increase in the feedback and interdependency effects since the end of the Great Recession – perhaps due to the significant consolidation in the industry as well as increased efficiencies. For instance, there has been a significant increase in the number of passengers per flight throughout the time period under study, as shown in Figure 6-5. In 2004, the average of all flights was nearly 66% full, whereas in 2012, the average occupancy of all flights is nearly 75%.

Figure 6-3: Mean ln(Capital Stock)
Figure 6-4: Annual Mean Spatial Multiplier

Figure 6-5: Annual Mean Flight - Percent Full

96
6.3.4 Methodology

The two models in (6.1) and (6.2) are estimated utilizing maximum likelihood as outlined in Elhorst (2003). Various statistical tests are performed to ensure consistent and robust parameter estimates. With estimates of the parameters outlined in (6.1) and (6.2), additional questions can be answered through the computation and analysis of two multipliers: the direct and the total [multiplier]. The direct multiplier is the effect that any airport $j$ has on $i$ arising from $i$ and $j$’s interactions (connections), while the total multiplier is derived from the spatial multiplier matrix. This $N \times N$ matrix, denoted $(I_N - \rho W)^{-1}$ contains elements $i,j$ that depict the relative effect that airport $j$ exerts on airport $i$ after accounting for the direct and indirect effects that spread through all nodes in the system (Anselin, 2002). This can be thought of as the sum of the direct and indirect multipliers and represents the true (full) impact of $j$ on $i$. As such the indirect multiplier for any pair can also be computed by taking the difference between the total and direct multipliers.

6.4 Results

Table 6.2 represents the results of the airport operations function from (6.1). The signs and magnitudes of the coefficient estimates are plausible, and various residual diagnostic tests suggest a sound model. While (6.1) and (6.2) are utilized as a proof-of-concept, the results are robust and tend to follow intuition. An increase in capital stock is associated with an increase in airport operations; a 1% increase in $K$ yields, on average (and all else equal) a 0.017% increase in number of operations; those airports with higher on-time departure percentages have less flights$^5$. As expected, the network exhibits substantial and significant interdependent effects - denoted by $\rho$. Tables 3 and 4 depict multipliers by hub size while Table 5 depicts the inter-regional flow of multipliers within the four US Census regions.

$^5$Population is not a statistically significant predictor of operations. While it is true that busier airports tend to be associated with more heavily populated areas, it is not the population size alone that drives these increased amounts of traffic, but rather airline network structure. Where airlines have found it advantageous to build large connecting hubs, the capital stock and air service expands beyond the needs of the local market.
Figure 6-6: U.S. Airports With Complete Observations
Table 6.2: Airport Operations Estimation Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std-Error</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>4.6952</td>
<td>0.3923</td>
<td>0.0000***</td>
</tr>
<tr>
<td>ln(K)</td>
<td>0.0172</td>
<td>0.0035</td>
<td>0.0000***</td>
</tr>
<tr>
<td>$OTP$</td>
<td>-2.4574</td>
<td>0.1995</td>
<td>0.0000***</td>
</tr>
<tr>
<td>ln(POP)</td>
<td>-0.0125</td>
<td>0.0087</td>
<td>0.1505</td>
</tr>
<tr>
<td>SM</td>
<td>0.7004</td>
<td>0.0320</td>
<td>0.0000***</td>
</tr>
<tr>
<td>MED</td>
<td>1.4896</td>
<td>0.0451</td>
<td>0.0000***</td>
</tr>
<tr>
<td>LG</td>
<td>2.5066</td>
<td>0.0574</td>
<td>0.0000***</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.3579</td>
<td>0.0268</td>
<td>0.0000***</td>
</tr>
</tbody>
</table>

Table 6.3 presents the average direct effects for inter and intra-hub-size operations. For example, focusing on the LG hub direct multiplier of 2.0455, suppose a large hub airport makes an investment that increases its capital stock (K) by 1% (in year $t-1$). On average, and all else being equal, we would expect to see operations at this airport increase by .0172% in year $t$. However, after accounting for the multiplier on all directly connected airports, the estimated effect of the 1% increase in capital stock is a 0.035% (0.0172*2.0455) increase in operations in the system as a whole. Similarly, following the LG hub total multiplier from Table 6.4, a 1% increase in capital stock at an average large hub yields an expected increase of 0.044% (0.0172*2.5748) in total system operations. This number is greater than the expected increase from the direct multiplier as it accounts for the feedback effects between airports that are both directly and indirectly connected. The same analysis can be carried out by other hub-sizes, or by the system averages given in each table. The system average provides the direct (from Table 6.3) and total (from Table reftoteffops) average multiplier uncovered in the modeled system. This is to say that on average a 1% increase in capital stock (or On-Time-Percentage) at an airport of any size yields an increase equal to the multiplier times the coefficient of the variable of interest.

Focusing on the intra and inter hub-sizes provides additional information. Examining the cell denoted LG-LG (with value 1.1081) from Table 6.3 represents the increase from
Table 6.3: Average Direct Effects by Hub Size (OPS)

<table>
<thead>
<tr>
<th>DIRECT</th>
<th>LG</th>
<th>MED</th>
<th>SM</th>
<th>GA</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>LG</td>
<td>1.1081</td>
<td>0.2084</td>
<td>0.3267</td>
<td>0.4022</td>
<td>2.0455</td>
</tr>
<tr>
<td>MED</td>
<td>0.0856</td>
<td>1.1115</td>
<td>0.1554</td>
<td>0.1786</td>
<td>1.5311</td>
</tr>
<tr>
<td>SM</td>
<td>0.0447</td>
<td>0.0385</td>
<td>1.0605</td>
<td>0.1118</td>
<td>1.2555</td>
</tr>
<tr>
<td>GA</td>
<td>0.0174</td>
<td>0.0089</td>
<td>0.0238</td>
<td>1.0740</td>
<td>1.1241</td>
</tr>
</tbody>
</table>

System Average 1.3580

An operational change at a large hub on the large hub where the change is made and all other directly connected large hubs. All other cells in Tables 6.3 and 6.4 can be interpreted the same way. The direct effect fails to account for second, third and \(n^{th}\) order effects. As such, it will always be less than the comparable total effect shown in Table 4. This is due to the fact that the ‘total effect’ measure captures these additional interactions that take place between airports via indirect means. As an example, imagine that airport (A) has flights going to airports (B) and (C), but not (D). However, airports (B) and (C) have flights going to (D) - and these flights depend on flights originating from (A). Because airport (A) is indirectly connected to airport (D), the operational characteristics of airport (A) have an indirect effect on (D). Failing to acknowledge this is akin to the direct effect measures presented in Table 6.3, while Table 6.4 depicts the total (direct and indirect effects). Assuming airports (A),(B),(C) and (D) are all LG airports and that a change affecting operations at airport (A) is made, the multiplier of 1.1081 can be broken down as follows: 1 unit (of the 1.1081) is the effect that A has on itself, while the additional 0.1081 is the impact that the one unit change at (A) has on (B) and (C). After accounting for the direct and indirect connections, this number (from Table 6.4) is found to be 1.1869, as it accounts for (A)’s effect on (B) and (C) and the effect that these changes have on (D).

When the multipliers are decomposed by hub-size it is easy to see that the magnitude of the multipliers follow intuition - large hubs exert a larger effect on medium, small and non-hub airports (GA) rather than the reverse. This is to say that multipliers arising from changes at large hubs are larger than similar changes made at a smaller hub. Additionally, by decomposing the multiplier by Census region (Table 6.5), one can see the flow of interaction.
Table 6.4: Average Total Effects by Hub Size (OPS)

<table>
<thead>
<tr>
<th>TOTAL</th>
<th>LG</th>
<th>MED</th>
<th>SM</th>
<th>GA</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>LG</td>
<td>1.1869</td>
<td>0.3014</td>
<td>0.4702</td>
<td>0.6163</td>
<td>2.5748</td>
</tr>
<tr>
<td>MED</td>
<td>0.1306</td>
<td>1.1711</td>
<td>0.2466</td>
<td>0.3094</td>
<td>1.8576</td>
</tr>
<tr>
<td>SM</td>
<td>0.0650</td>
<td>0.0655</td>
<td>1.1050</td>
<td>0.1752</td>
<td>1.4107</td>
</tr>
<tr>
<td>GA</td>
<td>0.0251</td>
<td>0.0191</td>
<td>0.0409</td>
<td>1.1047</td>
<td>1.1898</td>
</tr>
</tbody>
</table>

System Average 1.5576

between regions. These census region numbers correspond to the U.S. Census Bureau’s division naming scheme: 1 represents the Northeast; 2 the Midwest; 3 the South; and 4 the Pacific. For a map view of the groupings, see Figure 6-6. Total regional multipliers are also presented and can be further disaggregated according to hub-size within each region (for example, as a means to examine optimal location and hub-size for investment) though each region is distributed closely around the system average. This is to say that at the Census region level, each region has a relatively uniform impact on the system as a whole.

The results of the Census region grouping are interesting to note as the total multipliers essentially depict a robust network in terms of geographic dispersion – no one portion of the country dominates another. This is also interesting as it serves to show the relative robustness of the network in the event of regional network disruptions.

Table 6.5: Average Total Effects by Census Regions (OPS)

<table>
<thead>
<tr>
<th>TOTAL</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.1532</td>
<td>0.1149</td>
<td>0.2745</td>
<td>0.0854</td>
<td>1.6280</td>
</tr>
<tr>
<td>2</td>
<td>0.0521</td>
<td>1.1929</td>
<td>0.1895</td>
<td>0.1209</td>
<td>1.5554</td>
</tr>
<tr>
<td>3</td>
<td>0.0526</td>
<td>0.0891</td>
<td>1.3409</td>
<td>0.0791</td>
<td>1.5617</td>
</tr>
<tr>
<td>4</td>
<td>0.0182</td>
<td>0.0693</td>
<td>0.0932</td>
<td>1.3519</td>
<td>1.5325</td>
</tr>
</tbody>
</table>

System Average 1.5576

Table 6.6 presents the results of the estimation procedure carried out for the model in (6.2). On average, medium (MED) hubs experience 1% less on-time departures than non-hub airports while large (LG) hubs have an average on-time departure percentage that is 2.5% lower than non-hubs. Airports tend to be relatively stable in their year-to-year on-
time departure percentage, as shown by $OTP_{t-1}$. Here too, capital stock, $K$, is positively associated with on-time departure percentage. That is, a 1% increase in capital stock is, on average, associated with an increase in on-time departures by 0.1445%. This model from (6.2) exhibits a higher dependence between airport outcomes - that is airport outcomes in terms of on-time departure percentages are tightly linked. Unsurprisingly, a significant disruption in one spreads throughout the system adversely affecting others. It is worth noting that model (6.2) utilizes an entirely different composition of the weight matrix, yet yields extremely similar estimates of the various multipliers. See Figure 6-11 and the associated discussion for more details.

**Table 6.6: On-Time Percentage Estimation Results**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std-Error</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>5.6323</td>
<td>2.5746</td>
<td>0.0288**</td>
</tr>
<tr>
<td>SM</td>
<td>0.0844</td>
<td>0.3664</td>
<td>0.8178</td>
</tr>
<tr>
<td>MED</td>
<td>-0.9973</td>
<td>0.5955</td>
<td>0.0942*</td>
</tr>
<tr>
<td>LG</td>
<td>-2.4835</td>
<td>0.8014</td>
<td>0.0020***</td>
</tr>
<tr>
<td>$OTP$</td>
<td>0.6194</td>
<td>0.0169</td>
<td>0.0000***</td>
</tr>
<tr>
<td>lnK</td>
<td>0.1445</td>
<td>0.0327</td>
<td>0.0000***</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.4780</td>
<td>0.0239</td>
<td>0.0000***</td>
</tr>
<tr>
<td>N</td>
<td>218</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>.7382</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 6.7: Average Total Effects by Hub Size (OTP)**

<table>
<thead>
<tr>
<th>TOTAL</th>
<th>LG</th>
<th>MED</th>
<th>SM</th>
<th>GA</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>LG</td>
<td>1.4418</td>
<td>0.6256</td>
<td>1.0536</td>
<td>1.6619</td>
<td>4.7829</td>
</tr>
<tr>
<td>MED</td>
<td>0.1513</td>
<td>1.1807</td>
<td>0.2559</td>
<td>0.2897</td>
<td>1.8776</td>
</tr>
<tr>
<td>SM</td>
<td>0.0482</td>
<td>0.0453</td>
<td>1.0565</td>
<td>0.0822</td>
<td>1.2323</td>
</tr>
<tr>
<td>GA</td>
<td>0.0143</td>
<td>0.0113</td>
<td>0.0184</td>
<td>1.0418</td>
<td>1.0859</td>
</tr>
<tr>
<td>System Average</td>
<td>1.7730</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Like the disaggregated multiplier measures derived from model (6.1), Table 6.7 presents the estimates of the total multipliers by hub-size from model (6.2). The total multiplier for medium, small and non-hub airports follow that of the multipliers from the operations
Figure 6-7: Total OTP Effects Flow By Hub-Size

model in (6.1). Figure 6-7 presents the information contained in Table 6.7 graphically and shows, unsurprisingly, that disruptions to on-time performance at large hub airports have the greatest effect on the network as a whole (regardless of hub-size). Figure 6-8 displays the distribution of non-hub, small, medium and large airport multipliers - the presence of a few outliers are best noted by examining the large hub multiplier distribution. While the overall average multiplier for large hubs is significantly larger than the same multiplier from model (6.1), it is perhaps due to the truncated nature of the data set. In the data-set utilized in the estimation procedure, a select few airports appear to be the center of the network; something that may be offset if additional domestic and international airports were included in the analysis.
6.4.1 Model Diagnostics

Figure 6-9 presents model diagnostics related to the preceding airport operations analysis while Figure 6-10 presents model output for the on-time percentage analysis. Both sets of plots present fairly well-behaved models. The plot in the upper left corner of each Figure displays the model predictions versus actual. As the general trend is portrayed in both while still allowing for random distributions around the trend line, the results are suitable. Perhaps of most particular interest are the top right plots in each set. These depict the predicted values (on the X-axis) versus the model residuals on the Y [axis]. Both models appear randomly distributed around zero with no over-arching directional relationship, which would otherwise be indicative of some systematic bias. Similarly, the bottom left plots (entitled — residuals — vs predicted) present the absolute value of the residuals on the Y-axis against the value of the predictions on the X. An expansion to the right and to the left (spreading to both horizons) is indicative of heteroskedasticity – something not found
in either of these plots\textsuperscript{6}. Further, the QQ plots are included to ensure that the assumption of residual normality is not violated.

Figure 6-9: Operations Model Diagnostics

6.5 Conclusion

The purpose of this study has been to examine airport interactions in light of network effects. The method by which these network effects are uncovered provide the basis of a framework for quantifying the magnitude and propagation of these interaction effects. As changes in any given airport in the system have an effect on own airport performance as well

\textsuperscript{6}For reference, this relationship can be witnessed in Figure 5-1.
Figure 6-10: \( \text{OTP} \) Model Diagnostics
as directly and indirectly connected airports, gaining insight into where and how far these effects spread is of critical importance to maintain efficiency in an increasingly congested system (Corporation, 2007).

This type of analysis can be of use in examining how proposed changes to the operational characteristics at various nodes in the system can be expected to impact the system as a whole. Furthermore, by accounting for the network effects we are able to calculate multipliers associated with various characteristics of interest. This is akin to the delay propagation multiplier as outlined in Churchill (2007); however, more generalized in that the multipliers uncovered herein do not have to be strictly delay related - they can be uncovered as they relate to any potential airport characteristics of interest. This could be undertaken when analyzing grant proposals from the discretionary component of the Airport Improvement Program or other similar capacity-improvement programs whereby projects are rated upon their merit in accordance with national or system-wide performance goals. As shown in the preceding analysis, various model specifications exhibit a central tendency towards a common multiplier. This is further reinforced via Figure 6-11, below, which shows that 80% of either model’s multipliers fall between the range of 1 and 1.8.

The fact that the output from the two very different model specifications (which utilized two very different metrics of interaction) provides such similar estimates for the majority of the multipliers is suggestive of consistency in the estimates of the multipliers. Utilizing a framework as outlined here could prove beneficial in the planning stages of a policy or decision-making process that alters the characteristics of an airport in the system as uncovering the expected propagation of benefits could ensure that those projects with the highest system-wide benefit are undertaken. By accounting for total (direct and indirect) system-wide effects, as opposed to direct effects only, policy and decision-makers can be sure that the increasingly scarce resources at their disposal are invested in areas which provide the highest returns to the system. Failure to realize the interaction effects might

---

7It is implicitly assumed that if an analysis of this type were to be applied to specific project proposals, only those capital projects with an expectation of capacity or efficiency improvement would be considered to have multipliers.
Figure 6-11: Spatial Multipliers – CDF – Operations & $\bar{OTP}$
otherwise improperly prioritize projects. This approach could also serve as an additional tool by which competing changes or investments could be compared to ensure the optimal outcome for the system as a whole is reached.

6.6 Spatial Matrices: Man-made or Naturally Occurring?

The results derived in the investigation of airport network effects and airport multipliers necessitates the discussion of the implications of network type for statistical analysis. The traditional motivation for spatial models and methods is to overcome the violation of the assumption of independence between observations (in a regression setting). Connecting and allowing for the interaction between these spatially located data-points adequately accounts for this. However, it assumes that the spatial structure is exogenous to the system. Therefore it is important to understand whether the spatial structure is man-made or naturally occurring, as they have very different implications for proper analytical procedure. In the case of man-made networks, they will most likely always be endogenous to the system of interest as they are constructed in a way to fulfill a need and solve a problem; that is, they serve a purpose. Whereas most transportation networks are created to connect spatially separated regions, the manner and degree to which they are connected is based on some purpose that the network should fulfill. Most all networks will offer some form of redundancy but this redundancy must be balanced against efficiency. As such connections are subject to change. Framing this network in terms of a connectivity matrix and contrasting this against that of a geospatial connectivity matrix will yield wildly differing results. See Figures 6-12 and 6-13 for an illustration of such. Figure 6-12 represents (by blue dots) a direct connection between airports throughout the time-period 2004-2012 while Figure 6-13 depicts the naturally occurring connections between bordering counties in the 10 state Mid America Freight Coalition (MAFC) region. The visual representations are starkly different as they

---

8See Chapter 5 for more details to the MAFC
represent very different types of connectivity.

Figure 6-12: Man-Made Network Connectivity Matrix View

Understanding this has important distinctions for the modeling framework necessary to ensure consistent estimates. Under the naturally occurring spatial connection, results can be expected to hold for other areas of similar nature, while the man-made network cannot. While the aviation network is dynamic and changing throughout time (as deemed necessary for competitive and market based reasons), the structure found in Figure 6-13 will not change\(^9\). Under an endogenous spatial structure, panel data methods are required. Furthermore, if the network changes, the results will change. It is important to keep this in mind when designing an experiment examining spatial networks. If they are man-made, cross-sectional analysis results will be inconsistent, inefficient and potentially biased.

\(^9\)As long as the governmental structure stays the same. However, even if county borders no longer existed, the connections between the areas conforming to the old county boundaries would of course stay the same.
As such, a simple identification method to understand which type of network is at hand is outlined. As a simple validity check, one can compute multipliers related to the spatial weighting (connectivity) matrix under scrutiny. As the only parameter necessary to compute spatial multipliers is rho, these calculations can be carried out before undertaking a full model-building exercise (by randomly generating values for the parameter $\rho$). The spatial multiplier matrix is computed as $(I_N - \rho W)^{-1}$ where $W$ is a row-standardized matrix. Taking the sum of the columns of the multiplier matrix provides the spatial multipliers associated with each location, $j$. Calculating and plotting cumulative probability distributions of the resulting multipliers, as in Figures 6-14 and 6-15 provides the necessary information as to how to proceed (in terms of estimation technique). If the resulting distribution appears similar to that of Figure 6-14 (a Fréchet distribution), the analyst must utilize panel data methods to ensure consistency of the estimates. Further work should be done to understand whether these properties hold under other types of networks.
Figure 6-14: Total Operations Multiplier Cumulative Probability Distribution (GEV) versus Normal

For reference, cumulative probability distributions as well as density distributions for both GEV Type II and normal multipliers are presented in Figures 6-16 and 6-17 under varying values of $\rho$. 
Figure 6-15: Geo-Spatial Based Multiplier Cumulative Probability Distribution
Figure 6-16: Total Operations Multiplier Density Distribution

Figure 6-17: Geo-Spatial Based Multiplier Density Distribution
Chapter 7

Conclusions & Contribution

As is evident from this study, transportation infrastructure continues to play a role in reducing manufacturing production costs. Spatial spillovers are present and appear to be increasing over time. Own-state transportation infrastructure and neighboring state’s transportation infrastructure reduces production costs and increases output as well as industrial profitability. This study took a departure from the literature in its derivation and estimation of the cost function. This method allows output to vary and provides results of other common points of interest (profit and production) for firms. Additionally, through the utilization of Geographic Information Systems, the results are presented in map form. This allows for quick visual comparisons and a rich analysis in terms of locations that receive relatively higher benefits from increases in levels of transportation infrastructure.

This type of approach has relevance for policy makers and transportation planners. While the results show that strategic cooperation among transportation planning agencies across state borders offers a place for additional returns to investment, the method by which these effects and benefits are extrapolated can also prove useful. Several examples of infrastructures and policies that benefit neighboring states are discussed: the Tri-State Tollway in Illinois, which has significant benefits not only for Illinois, but also Wisconsin and Indiana. Similar policy can be found in Michigan’s neighboring states; Ohio has strategic routes termed ‘Michigan Legal Weight’ - whereby Ohio weight limits for trucks can, via a permit, be raised to equal Michigan’s [legal limit] - a measure enacted in the 1980’s to increase the port of Toledo’s competitiveness (Wisconsin also has a similar policy in place for
Michigan weight limits) (ODOT, 2011). As economic activity does not follow strict political boundaries, more strategic planning and interaction can prove beneficial to improve the cohesiveness of transportation systems. Analyzing the locational component of benefits for specific industries can uncover areas that experience higher rate of returns. The availability of more disaggregated geographic data may help fine-tune the findings of our analysis. As recent transportation policy mandates increasing the alignment of transportation projects to broader national and economic goals, this approach can aid in the determination of areas which satisfy such requirements under various industries of interest of the United States of America (2012). An analysis that explores how the locations of benefits change under the modern economy’s shift towards more service-oriented activities or based on the changing industrial mix may also prove fruitful in terms of policy.

As the results of this study indicate, transportation infrastructure usage does provide a measurable benefit to individual standard of living (in terms of income). The statistical significance of the positive benefit to households reinforces the notion that this difficult-to-measure phenomenon is present. By employing spatial analytical tools and techniques we are able to visualize the magnitude and variation of the benefits throughout the study area. As expected, major cities experience relatively larger increases in income as AADT increases. Perhaps more interesting is the fact that those counties located along major connecting routes/corridors also experience a higher increase in income (relative to non-corridor counties) with increased traffic volumes. These results suggest that increased transportation infrastructure usage contributes to a higher standard of living. Further, the visualization of the elasticities aides in understanding regions that appear to stand to gain more from increased usage. While usage alone is not the determining factor, in line with Hansen’s argument, those locations of low utilization stand to gain from increased development. This notion is further reinforced in Duranton and Turner (2011) which details the fundamental law of road congestion - that “if you build it, they will come.”

Employing transportation infrastructure usage as a proxy for transportation infrastructure demand allows for us to determine the benefit to individuals that the observed demand for infrastructure has on standards of living (measured by per-capita-income). Given the
presence of a measurable benefit to individuals from infrastructure usage, there is a significant role for policy. As Eloff et al. (2013) found that increased levels of strategic cooperation among transportation planners and policy makers could serve to increase return on investment, the same holds true here. Improving the cohesiveness of transportation systems can provide benefits to both industry and individual users alike. Given recent changes in national transportation policy, the method of analysis can provide an additional tool in identifying ‘lagging’ regions set to benefit the most from increased investment.

An additional facet of this study has been to concerned with the examination of airport interactions in light of network effects. The method by which these network effects are uncovered provide the basis of a framework for quantifying the magnitude and propagation of these interaction effects. As changes in any given airport in the system have an effect on own airport performance as well as directly and indirectly connected airports, gaining insight into where and how far these effects spread is of critical importance to maintain efficiency in an increasingly congested system (Corporation, 2007).

This type of analysis can be of use in examining how proposed changes to the operational characteristics at various nodes in the system can be expected to impact the system as a whole. Furthermore, by accounting for the network effects we are able to calculate multipliers associated with various characteristics of interest. This is akin to the delay propagation multiplier as outlined in Churchill (2007); however, more generalized in that the multipliers uncovered herein do not have to be strictly delay related - they can be uncovered as they relate to any potential airport characteristics of interest. This could be undertaken when analyzing grant proposals from the discretionary component of the Airport Improvement Program or other similar capacity-improvement programs whereby projects are rated upon their merit in accordance with national or system-wide performance goals. As shown in the preceding analysis, various model specifications exhibit a central tendency towards a common multiplier. This is further reinforced via Figure 4, below, which shows that 80% of

---

1It is implicitly assumed that if an analysis of this type were to be applied to specific project proposals, only those capital projects with an expectation of capacity or efficiency improvement would be considered to have multipliers.
of either model’s multipliers fall between the range of 1 and 1.8.

The fact that the output from the two very different model specifications (which utilized two very different metrics of interaction) provides such similar estimates for the majority of the multipliers is suggestive of consistency in the estimates of the multipliers. Utilizing a framework as outlined here could prove beneficial in the planning stages of a policy or decision-making process that alters the characteristics of an airport in the system as uncovering the expected propagation of benefits could ensure that those projects with the highest system-wide benefit are undertaken. By accounting for total (direct and indirect) system-wide effects, as opposed to direct effects only, policy and decision-makers can be sure that the increasingly scarce resources at their disposal are invested in areas which provide the highest returns to the system. Failure to realize the interaction effects might otherwise improperly prioritize projects. This approach could also serve as an additional tool by which competing changes or investments could be compared to ensure the optimal outcome for the system as a whole is reached.
References


The MITRE Corporation. Capacity needs in the national airspace system fact 2. 2007.


121


Bureau of Transportation Statistics. Understanding the reporting of causes of flight delays and cancellations.


Oleg Smirnov and Luc Anselin. Fast maximum likelihood estimation of very large spatial


Appendix A

Source Code Bits

A.1 Estimation Routine

The relevant sections of the Python code written to implement the Kelejian, Kapoor and Prucha GMM routine is outlined below. This procedure employs a number of free and open source packages such as NumPY, PYSAL, SCIPY and, of course, Python.

```python
import numpy as np
import pysal
import scitools.filetable, sys, math
from scipy.io import loadmat
from scipy import stats, linalg
from pylab import double, dot, identity, kron, eye, Inf,
mean, shape, ones, zeros, log, sqrt, diag

datadir = "~/home/Data/"
file = "X"
ext = ".mat"
data = loadmat(datadir+file+ext,matlab_compatible=True)

import pysal
from pysal import higher_order
shpfile=48state
shpext=.shp
W = pysal.queen_from_shapefile('datadir+shpfile+shpext')
W.id_order=[Enter the proper comma separated FID sequence which corresponds to the way in which the data is sorted in the X and Y data files]
```
Similarly one could replace 'queen_from_shapefile' with 'rook_from_shapefile'. PySAL also offers the ability to create higher order spatial weights matrices via the command $W_2 = \text{higher\_order}(W,2)$ where 2 is the order to create the higher order matrix.

```
W.transform='r' # This row standardizes the Weights Matrix
mat = pysal.open(datadir+shpfile+ext, 'w') # Tells the system where to write
mat.write(W) # Writes it
mat.close() # And closes it.
[N, junk] = shape(W);
[junk, K] = shape(X)
N=double(N) # To avoid integer math
[temp] = shape(y);
temp=double(temp)
T=temp/N
bols = dot((dot(linalg.inv(dot(X.T,X)),X.T)),y) # OLS estimates of betas
utilde = y - dot(X,bols) # Which we use to compute the residuals
IT=identity(T)
IN=identity(N)
JT=ones((T,T))
from scipy.optimize import leastsq

def residuals(a,z,v1,v2,v3):
    rho, rhosq, sigma2v = a
    err = z-(v1*rho+v2*rhosq+v3*sigma2v)
    return err

def aeval(v1,v2,v3,a):
    return v1*a[0]+v2*a[1]+v3*a[2]

a0=[−.9, .81, 0.1]
alsq, ite=leastsq(residuals, a0, args=(z,v1,v2,v3))
rho_initial=alsq[0]
rho_sq_initial=alsq[1]
sigma2vinitial=alsq[2]
check=np.Inf
tolerance=10**−10 # Set termination tolerance to obtain several digits
rho_hat=0
sigv_hat=0
sig1_hat=0
rho_limit=[−.9,.9] # Set
sigv_limit=[0.0,10.0] # search
sig1_limit=[0.0,10.0] # limits
```
\[ \rho_{\text{inc}} = (\rho_{\text{limit}[1]} - \rho_{\text{limit}[0]}) / 10 \] # Set

\[ \text{sigv}_{\text{inc}} = (\text{sigv}_{\text{limit}[1]} - \text{sigv}_{\text{limit}[0]}) / 10 \] # # # # # search

\[ \text{sig1}_{\text{inc}} = (\text{sig1}_{\text{limit}[1]} - \text{sig1}_{\text{limit}[0]}) / 10 \] # # # # # increments

while \( \text{sig1}_{\text{inc}} > \text{tolerance} \) and \( \rho_{\text{inc}} > \text{tolerance} \) and \( \text{sigv}_{\text{inc}} > \text{tolerance} \):

\[ \text{for } \rho_{\text{hat}} \text{ in } \text{np.arange}(\rho_{\text{limit}[0]}, \rho_{\text{limit}[1]}, \rho_{\text{inc}}): \]

\[ \text{for } \text{sigv}_{\text{hat}} \text{ in } \text{np.arange}(\text{sigv}_{\text{limit}[0]}, \text{sigv}_{\text{limit}[1]}, \text{sigv}_{\text{inc}}): \]

\[ \text{for } \text{sig1}_{\text{hat}} \text{ in } \text{np.arange}(\text{sig1}_{\text{limit}[0]}, \text{sig1}_{\text{limit}[1]}, \text{sig1}_{\text{inc}}): \]

\[ e = (G \ast (\text{np.matrix}([\rho_{\text{hat}}, \rho_{\text{hat}}**2, \text{sigv}_{\text{hat}}, \text{sig1}_{\text{hat}}]).T) - g) \]

\[ w_{\text{res}} = e.T \ast \text{Ups}_{\text{Inv}} \ast e \]

\[ \text{if } w_{\text{res}} < \text{check}: \ # \text{Check if sum of sq is falling} \]

\[ \text{check} = w_{\text{res}} \ # \text{If so, update our check} \]

\[ \text{newrho} = \rho_{\text{hat}} \ # \text{And assign this iterations values} \]

\[ \text{newsig}2v = \text{sigv}_{\text{hat}} \] # to be new estimates

\[ \text{newsig}2one = \text{sig1}_{\text{hat}} \]

\[ \rho_{\text{limit}[0]} = (\text{newrho} - \rho_{\text{inc}}) \] # Recompute

\[ \rho_{\text{limit}[1]} = (\text{newrho} + \rho_{\text{inc}}) \] # the

\[ \text{sigv}_{\text{limit}[0]} = \text{newsig}2v - \text{sigv}_{\text{inc}} \] # limits

\[ \text{sigv}_{\text{limit}[1]} = \text{newsig}2v + \text{sigv}_{\text{inc}} \]

\[ \text{sig1}_{\text{limit}[0]} = \text{newsig}2one - \text{sig1}_{\text{inc}} \]

\[ \text{sig1}_{\text{limit}[1]} = \text{newsig}2one + \text{sig1}_{\text{inc}} \]

\[ \rho_{\text{inc}} = (\rho_{\text{limit}[1]} - \rho_{\text{limit}[0]}) / 10 \] ## And

\[ \text{sigv}_{\text{inc}} = (\text{sigv}_{\text{limit}[1]} - \text{sigv}_{\text{limit}[0]}) / 10 \] ## the

\[ \text{sig1}_{\text{inc}} = (\text{sig1}_{\text{limit}[1]} - \text{sig1}_{\text{limit}[0]}) / 10 \] ## increments... and continue

\[ y_{\text{star}} = \text{dot}(\text{kron}(\text{IT}, (\text{IN} - \rho_{\text{full}} \ast W)), y) \]

\[ x_{\text{star}} = \text{dot}(\text{kron}(\text{IT}, (\text{IN} - \rho_{\text{full}} \ast W)), X) \]

\[ \Omega_{\text{half}} = \text{dot}((1/\text{sigma2v}_{\text{full}} ** .5), qu0) + \text{dot}((1/\text{sigma2one}_{\text{full}} ** .5), qu1) \]

\[ \Omega_{\text{inv}} = \text{dot}(\Omega_{\text{half}}, \Omega_{\text{half}}) \]

### FEASIBLE GLS

\[ \text{beta}_{\text{fgls}} = \text{dot}(\text{linalg.inv}(\text{dot}(x_{\text{star}}.T, \text{dot}(\Omega_{\text{inv}}), x_{\text{star}})), \text{dot}(x_{\text{star}}.T, \text{dot}(\text{←} \Omega_{\text{inv}}, y_{\text{star}}))) \]

### Covariance Matrix

\[ \phi = (1.0/(N*T)) \ast \text{linalg.inv}(((1.0/(N*T)) \ast \text{dot}(x_{\text{star}}.T, \text{dot}(\Omega_{\text{inv}}, x_{\text{star}})))) \]

\[ (#\text{ustar=ystar−xstarB}) \]

\[ y_{\text{star1}} = \text{dot}(\Omega_{\text{half}}, y_{\text{star}}) \]

\[ x_{\text{star1}} = \text{dot}(\Omega_{\text{half}}, x_{\text{star}}) \]

\[ \text{betas} = \text{zeros}(K) \]

for \( i \) in range(\( K \)):

\[ \text{betas}[i] = \text{beta}_{\text{fgls}}[i] \]

\[ \text{betas} = \text{np.matrix}(\text{betas}.T) \]

\[ \text{print}("\text{ Variable Name } \text{ Coefficient Estimates } \text{ Standard Errors} \)"
for i in range(K):
    print("{} + {} + {} + {} + {} + {} + {}
se[i] + " + str(t_stat[0,i]) + " + str(p_val[0,i]) + " + 
print("Sigma2_Hat_FGLS is: " + str(phi_fgls[0]))
Appendix B

Aviation Data Processing Software

B.1 Data Preparation Scripts

Attached is the PC code written for the purpose of obtaining, processing and manipulating the data from the Bureau of Transportation Statistics (BTS) Airline On-Time Performance and T-100 Segment datasets. The software requires three entirely free prerequisites: access to a MySQL database server, Python 2.7\textsuperscript{1} and the T-100 data. The script has been tested on a PC with an Intel i7-950 processor and 36 GB of ram running Debian Linux OS.\textsuperscript{2} The memory requirements are quite large as the on-time performance database contains some 67 million rows for the time horizon under investigation.\textsuperscript{3} The code is highly specific to the task at hand but serves to illustrate the approach taken while allowing for reproducibility and variation in the dataset should it be needed at a later point in time. Execution of the script requires only the T-100 data to be previously obtained - the on-time data is downloaded based on user inputs. All other processes are automated\textsuperscript{4} leaving an output file representing an aggregated merger of the two data sources in a panel type format. The code below is licensed under the GNU GPL 3 and available upon request.\textsuperscript{5}

\begin{footnotesize}
\begin{itemize}
  \item[1] And 4 libraries - NumPY, SciPY, MySQLdb and pandas
  \item[2] A few slight modifications are required for Windows, including an installation of cygwin utilities.
  \item[3] In this specific use-case, upwards of 12 gigabytes of free ram are needed in the on-time performance database SQL access portion of the data processing.
  \item[4] Such as the creation of the database and necessary tables, the import of the data and all processing tasks
  \item[5] The full text of the license can be found at: http://www.gnu.org/licenses/gpl.html
\end{itemize}
\end{footnotesize}
#!/usr/bin/env python
# -*- coding: utf-8 -*-

""" include "COPYING"
Created on Fri Jan 3 13:24:20 2014
Copyright 2014: Jeffrey J Eloff

This program is free software: you can redistribute it and/or modify
it under the terms of the GNU General Public License as published by
the Free Software Foundation, either version 3 of the License, or
(at your option) any later version.

This program is distributed in the hope that it will be useful,
but WITHOUT ANY WARRANTY; without even the implied warranty of
MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the
GNU General Public License for more details.

You should have received a copy of the GNU General Public License
along with this program. If not, see <http://www.gnu.org/licenses/>.
"""

import os

print "Welcome to the data downloader, mysql table creator, and python panel data preparation for BTS On-Time Performance and T-100 data"
begyearrange=raw_input("Please enter the first year of data to obtain (int YYYY):")
print "First year selected is: " + str(begyearrange)
endyearrange=raw_input("Please enter the last year of data to obtain (int, YYYY):")
print "You selected end year of " + str(endyearrange)
year=[]
i=int(begyearrange)
endyear=int(endyearrange)+1
while i<endyear:
    year.append(i)
i=i+1
quarter=[]
for i in range(4):
    quarter.append(i+1)
dates=[]
quarter_start={1:'01-01',2:'04-01',3:'07-01',4:'10-01'}
quarter_end={1:'03-31',2:'06-30',3:'09-30',4:'12-31'}
for j in year:
    for i in range(4):
        dates.append(str(j) + '−' + quarter_start[i+1])
        dates.append(str(j) + '−' + quarter_end[i+1])
firstperiod=raw_input("Select starting quarter of " + str(begyearrange) + " (1−4): ←")
print "Dataset begins with Quarter " + str(firstperiod) + " of " + str(begyearrange←)
    + "...OK"
endperiod=raw_input("Select end quarter of " + str(eyearrange) + " (1−4): ")
print "Dataset ends with Quarter " + str(endperiod) + " of " + str(eyearrange) + ←"...OK"
print "Time period for panel data creation runs from Quarter " + str(firstperiod) + ←" " + str(begyearrange) + " to Quarter " + str(endperiod) + ←" " + str(eyearrange)
confi=raw_input("Is this correct? (y/N)")
if confi=='y' or confi=='Y':
    print "continuing...
else:
    print "Error — more error checking can be added but hasn’t."
#need to check if dates list needs any items removed — and if so, how many
start_dates_remove=(int(firstperiod)-1)*2
if start_dates_remove>0:
    dates=dates[start_dates_remove:]
else:
    dates=dates
#check if end is in agreement
end_dates_remove=(4-int(endperiod))*2
if end_dates_remove>0:
    dates=dates[0:len(dates)-end_dates_remove]
else:
    dates=dates
print "Please specify a directory (to which you have write access)"
print "This will take a while...and a bit of space"
directory=raw_input("Complete path for data directory or press D for current")
if directory=='D' or directory=='d':
    directory=os.getcwd()
else:
    directory=directory
print "Quick check to ensure write access"
def trailingslashremoval(loc):
    if loc[-1]=='/':
loc=loc[:-1]
else:
    loc=loc
to return loc
directory=trailingslashremoval(directory)

if os.access(directory, os.W_OK)==True:
    print "write access confirmed"
else:
    print "No write access"
    print "more error control would be needed..."

fill="_"
ext=".zip"

#begyearrange has first year
begmonth=int(dates[0][5:7])
endmonth=int(dates[-1][5:7])
files=[]
filename=[]
if begmonth<>1:
    monthsoffirstyearint=12-(begmonth-1) #this is the number of months of the first→
year to download

i=year[0]
while i<year[1]:
    while monthsoffirstyearint<13:
        files.append(base+fill+str(i)+fill+str(monthsoffirstyearint)+ext)
        filename.append(str(i)+fill+str(monthsoffirstyearint)+ext)
        monthsoffirstyearint=monthsoffirstyearint+1
    i=i+1

j=1
if endmonth<>12:
    for i in year[1:-1]:
        while j<13:
            files.append(base+fill+str(i)+fill+str(j)+ext)
            filename.append(str(i)+fill+str(j)+ext)
            j=j+1
        i=year[-1]
    j=1
while i<year[-1]+1:
    while j<endmonth+1:
files.append(base+fill+str(i)+fill+str(j)+ext)
filename.append(str(i)+fill+str(j)+ext)
        j=j+1
        i=i+1
    else:
        for i in year[1:]:
            while j<13:
                files.append(base+fill+str(i)+fill+str(j)+ext)
                filename.append(str(i)+fill+str(j)+ext)
        i=i+1
fullfilename=[]
for file in filename:
    fullfilename.append("On_Time_On_Time_Performance","+file)
import numpy as np
import scipy
os.chdir(directory)
def otdler(cond):
    if cond=='y' or cond=='Y':
        zippeddatadir=raw_input("Please enter full path to directory containing ←
                                 zips")
        os.chdir(zippeddatadir)
        zippeddatadir=trailingslashremoval(zippeddatadir)
        print "Checking to ensure all files exist"
        counter=len(fullfilename)
        check=np.zeros(len(fullfilename))
        for i in range(len(fullfilename)):
            if os.path.isfile(zippeddatadir+'/'+fullfilename[i]):
                check[i]=1
            else:
                check[j]=0
        if counter==int(sum(check)):
            exiter="skipping download"
            print exiter
        else:
            exiter="skipping download"
            print exiter

```
exiter="some files missing...determining which"

print exiter

# need to complete this...not really important at the moment
else:
    newdircheck=os.path.isdir(directory+"/downloads/")
    if newdircheck==False:
        os.mkdir(directory+"/downloads/")
        os.chdir(directory+"/downloads/")
    else:
        os.chdir(directory+"/downloads/")
    print "Spawning downloader...
    for p in xrange(20):
        print "."*p
    for i in files:
        os.system('wget %s %i')
zippeddatadir=directory+"/downloads"
    exiter="download complete"
    print exiter
    return zippeddatadir
dlcheck=raw_input("Have you already obtained the ontime data? (y/N)")
if dlcheck=='y' or dlcheck=='Y':
    dlcheck2=raw_input("Have they already been unzipped? (y/N)
    if dlcheck2==='y' or dlcheck2==='Y':
        print "Good..."
    else:
        zippeddatadir=otdler(dlcheck)
        newdircheck=os.path.isdir(directory+"/orig/")
        if newdircheck==False:
            os.mkdir(directory+"/orig/")
        else:
            print "Unzipping!"
            os.chdir(zippeddatadir)
            os.system('y [y] | for i in `ls`; do unzip $i; mv $i .. /orig; done'
        else:
            zippeddatadir=otdler(dlcheck)
            newdircheck=os.path.isdir(directory+"/orig/")
            if newdircheck==False:
                os.mkdir(directory+"/orig/")
            else:
                print "Unzipping!"
            os.chdir(zippeddatadir)
```
os.system('y [y] | for i in `ls`; do unzip $i; mv $i ../orig; done')
import pandas as pd
import MySQLdb
import sys
print "Files covering time period Q\" + firstperiod + " : " + begyearrange + " to Q\" +
        + endperiod + " : " + endyearrange + " have been downloaded and unzipped."

print "Input mysql login credentials"
host=raw_input("Hostname or ip address of mysql server :")
username=raw_input("User Name : ")
import getpass
password=getpass.getpass()
#database='ontimeperf'
con= MySQLdb.connect(str(host),str(username),str(password),local_infile=1)
c=con.cursor()
print "You have two options:

def dbselect(condi):
    availabledatabase=raw_input("Press 1 to select from available databases or 2 to--
    create a new database.")
    if availabledatabase==1:
        dbgetter="\"SHOW DATABASES\""
        listofdbs=c.execute(dbgetter)
        lodbs=[]
        for i in range(int(listofdbs)):
            lodbs.append(c.fetchone())
        i=i+1

        print "Make a selection from the currently available databases"
        for i in range(int(listofdbs)):
            print i+1, lodbs[i]

    dbname=raw_input("Select a number from " + str(range(int(listofdbs))[0]+1) + "
            + " - " + str(range(int(listofdbs))[-1]+1) + " corresponding to the db
            list above :")
    if int(dbname)>(range(int(listofdbs))[0]+1) and int(dbname)<=(range(int(+
            listofdbs))[-1]+1):
        dbname=lodbs[int(dbname)-1]
        selecteddatabase=dbname[0]
    else:
        print "Invalid input"
    else:
        selecteddatabase=raw_input("Enter new database title")
dbcreation="CREATE DATABASE '%s'" % (selecteddatabase)
c.execute(dbcreation)
con.commit()
print "Successfully created new database " + str(selecteddatabase)
return selecteddatabase
def dbchange(yorn):
    dbchanger="USE %s" % (selecteddatabase)
c.execute(dcharger)
con.commit()
return dbchanger
dbcondition = raw_input("Do we need to create a database? (y/N)"")
if dbcondition == 'y' or dbcondition == 'Y':
    selecteddatabase=dbselect(dbcondition)
    dbchange(selecteddatabase)
else:
    selecteddatabase = raw_input("Enter pre-existing database name then...")
    dbchanger = dbchange(selecteddatabase)
    print "Database changed to : " + str(selecteddatabase) + " with command " + str← (dbchanger)
def tablecreator():
    ontimetablecreator="CREATE TABLE `%s` (" %
    `Year` YEAR(4) NULL DEFAULT NULL,
    `Quarter` TINYINT(1) NULL DEFAULT NULL,
    `Month` TINYINT(2) NULL DEFAULT NULL,
    `DayofMonth` TINYINT(3) NULL DEFAULT NULL,
    `DayOfWeek` TINYINT(1) NULL DEFAULT NULL,
    `FlightDate` DATE NULL DEFAULT NULL,
    `UniqueCarrier` VARCHAR(50) NULL DEFAULT NULL,
    `AirlineID` INT(10) NULL DEFAULT NULL,
    `Carrier` VARCHAR(50) NULL DEFAULT NULL,
    `TailNum` VARCHAR(50) NULL DEFAULT NULL,
    `FlightNum` INT(20) NULL DEFAULT NULL,
    `OriginAirportID` INT(10) NULL DEFAULT NULL,
    `OriginAirportSeqID` INT(10) NULL DEFAULT NULL,
    `OriginCityMarketID` INT(10) NULL DEFAULT NULL,
    `Origin` VARCHAR(10) NULL DEFAULT NULL,
    `OriginCityName` VARCHAR(50) NULL DEFAULT NULL,
    `OriginState` VARCHAR(10) NULL DEFAULT NULL,
    `OriginStateFips` TINYINT(2) NULL DEFAULT NULL,
    `OriginStateName` VARCHAR(50) NULL DEFAULT NULL,
    `OriginWac` TINYINT(10) NULL DEFAULT NULL,
<table>
<thead>
<tr>
<th>Field</th>
<th>Type</th>
<th>Default</th>
</tr>
</thead>
<tbody>
<tr>
<td>DestAirportID</td>
<td>INT(10) NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>DestAirportSeqID</td>
<td>INT(10) NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>DestCityMarketID</td>
<td>INT(10) NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>Dest</td>
<td>VARCHAR(10)</td>
<td>NULL DEFAULT</td>
</tr>
<tr>
<td>DestCityName</td>
<td>VARCHAR(50)</td>
<td>NULL DEFAULT</td>
</tr>
<tr>
<td>DestState</td>
<td>VARCHAR(10)</td>
<td>NULL DEFAULT</td>
</tr>
<tr>
<td>DestStateFips</td>
<td>TINYINT(2)</td>
<td>NULL DEFAULT</td>
</tr>
<tr>
<td>DestStateName</td>
<td>VARCHAR(50)</td>
<td>NULL DEFAULT</td>
</tr>
<tr>
<td>DestWac</td>
<td>TINYINT(10)</td>
<td>NULL DEFAULT</td>
</tr>
<tr>
<td>CRSDepTime</td>
<td>TIME NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>DepTime</td>
<td>TIME NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>DepDelay</td>
<td>FLOAT NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>DepDelayMinutes</td>
<td>FLOAT NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>DepDel15</td>
<td>FLOAT NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>DepartureDelayGroups</td>
<td>INT(10) NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>DepTimeBk</td>
<td>VARCHAR(10)</td>
<td>NULL DEFAULT</td>
</tr>
<tr>
<td>TaxiOut</td>
<td>FLOAT NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>WheelsOff</td>
<td>TIME NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>WheelsOn</td>
<td>TIME NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>TaxiIn</td>
<td>FLOAT NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>CRSArrTime</td>
<td>TIME NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>ArrTime</td>
<td>TIME NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>ArrDelay</td>
<td>FLOAT NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>ArrDelayMinutes</td>
<td>FLOAT NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>ArrDel15</td>
<td>FLOAT NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>ArrivalDelayGroups</td>
<td>INT(10) NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>ArrTimeBk</td>
<td>VARCHAR(10)</td>
<td>NULL DEFAULT</td>
</tr>
<tr>
<td>Cancelled</td>
<td>FLOAT NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>CancellationCode</td>
<td>VARCHAR(10)</td>
<td>NULL DEFAULT</td>
</tr>
<tr>
<td>Diverted</td>
<td>FLOAT NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>CRSElapsedTime</td>
<td>FLOAT NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>ActualElapsedTime</td>
<td>FLOAT NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>AirTime</td>
<td>FLOAT NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>Flights</td>
<td>FLOAT NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>Distance</td>
<td>FLOAT NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>DistanceGroup</td>
<td>INT(10) NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>CarrierDelay</td>
<td>FLOAT NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>WeatherDelay</td>
<td>FLOAT NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>NASDelay</td>
<td>FLOAT NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>SecurityDelay</td>
<td>FLOAT NULL</td>
<td>DEFAULT NULL</td>
</tr>
<tr>
<td>LateAircraftDelay</td>
<td>FLOAT NULL</td>
<td>DEFAULT NULL</td>
</tr>
</tbody>
</table>
```
'FirstDepTime' VARCHAR(255) NULL DEFAULT NULL,
'TotalAddGTime' VARCHAR(255) NULL DEFAULT NULL,
'LongestAddGTime' VARCHAR(255) NULL DEFAULT NULL,
'DivAirportLandings' VARCHAR(255) NULL DEFAULT NULL,
'DivReachedDest' VARCHAR(255) NULL DEFAULT NULL,
'DivActualElapsedTime' VARCHAR(255) NULL DEFAULT NULL,
'DivArrDelay' VARCHAR(255) NULL DEFAULT NULL,
'DivDistance' VARCHAR(255) NULL DEFAULT NULL,
'Div1Airport' VARCHAR(255) NULL DEFAULT NULL,
'DivAirportID' VARCHAR(255) NULL DEFAULT NULL,
'Div1AirportSeqID' VARCHAR(255) NULL DEFAULT NULL,
'Div1WheelsOn' VARCHAR(255) NULL DEFAULT NULL,
'Div1TotalGTime' VARCHAR(255) NULL DEFAULT NULL,
'Div1LongestGTime' VARCHAR(255) NULL DEFAULT NULL,
'Div1WheelsOff' VARCHAR(255) NULL DEFAULT NULL,
'Div1TailNum' VARCHAR(255) NULL DEFAULT NULL,
'Div2Airport' VARCHAR(255) NULL DEFAULT NULL,
'Div2AirportID' VARCHAR(255) NULL DEFAULT NULL,
'Div2AirportSeqID' VARCHAR(255) NULL DEFAULT NULL,
'Div2WheelsOn' VARCHAR(255) NULL DEFAULT NULL,
'Div2TotalGTime' VARCHAR(255) NULL DEFAULT NULL,
'Div2LongestGTime' VARCHAR(255) NULL DEFAULT NULL,
'Div2WheelsOff' VARCHAR(255) NULL DEFAULT NULL,
'Div2TailNum' VARCHAR(255) NULL DEFAULT NULL,
'Div3Airport' VARCHAR(255) NULL DEFAULT NULL,
'Div3AirportID' VARCHAR(255) NULL DEFAULT NULL,
'Div3AirportSeqID' VARCHAR(255) NULL DEFAULT NULL,
'Div3WheelsOn' VARCHAR(255) NULL DEFAULT NULL,
'Div3TotalGTime' VARCHAR(255) NULL DEFAULT NULL,
'Div3LongestGTime' VARCHAR(255) NULL DEFAULT NULL,
'Div3WheelsOff' VARCHAR(255) NULL DEFAULT NULL,
'Div3TailNum' VARCHAR(255) NULL DEFAULT NULL,
'Div4Airport' VARCHAR(255) NULL DEFAULT NULL,
'Div4AirportID' VARCHAR(255) NULL DEFAULT NULL,
'Div4AirportSeqID' VARCHAR(255) NULL DEFAULT NULL,
'Div4WheelsOn' VARCHAR(255) NULL DEFAULT NULL,
'Div4TotalGTime' VARCHAR(255) NULL DEFAULT NULL,
'Div4LongestGTime' VARCHAR(255) NULL DEFAULT NULL,
'Div4WheelsOff' VARCHAR(255) NULL DEFAULT NULL,
'Div4TailNum' VARCHAR(255) NULL DEFAULT NULL,
'Div5Airport' VARCHAR(255) NULL DEFAULT NULL,
```
'Div5AirportID' VARCHAR(255) NULL DEFAULT NULL,
'Div5AirportSeqID' VARCHAR(255) NULL DEFAULT NULL,
'Div5WheelsOn' VARCHAR(255) NULL DEFAULT NULL,
'Div5TotalGTime' VARCHAR(255) NULL DEFAULT NULL,
'Div5LongestGTime' VARCHAR(255) NULL DEFAULT NULL,
'Div5WheelsOff' VARCHAR(255) NULL DEFAULT NULL,
'Div5TailNum' VARCHAR(255) NULL DEFAULT NULL
)
ENGINE=InnoDB ;

""" % (tablename)
c.execute(ontimetablecreator)
con.commit()
tableneeded=raw_input(" Shall we create the table in the database " + str(←
    selecteddatabase) + "? (y/N)"")
if tableneeded=='y' or tableneeded=='Y':
    tablename=raw_input(" Input title for table: ")
tablecreator(tablename)
else:
    tablename=raw_input(" Input previously created tablename: ")
loaddata=raw_input(" Shall we load the downloaded files into " + str(←
    selecteddatabase) + " database, table " + str(tablename) + "? (y/N)"")
if loaddata=='y' or loaddata=='Y':
    filename=[fills.replace('zip', 'csv') for fills in filename]
    indivcsvs=[]
    for i in filename:
        indivcsvs.append('On_Time_On_Time_Performance_\'+str(i))
    for i in indivcsvs:
        loader="""Load data local infile '%s' into table %s fields terminated by ↔
        ', , optionally enclosed by ' '" IGNORE 1 LINES
    (Year, Quarter, Month, DayOfMonth, DayOfWeek, FlightDate, UniqueCarrier, AirlineID, Carrier←
    , TailNum, FlightNum, OriginAirportID, OriginAirportSeqID, OriginCityMarketID, Origin←
    , OriginCityName, OriginState, OriginStateFips, OriginStateName, OriginWac,←
    DestAirportID, DestAirportSeqID, DestCityMarketID, Dest, DestCityName, DestState,←
    DestStateFips, DestStateName, DestWac, @bdCRSDepTime, @bdDepTime, DepDelay,←
    DepDelayMinutes, DepDel15, DepartureDelayGroups, DepTimeBlk, TaxiOut, @bdWheelsOff,←
    @bdWheelsOn, TaxiIn, @bdCRSArrTime, @bdArrTime, ArrDelay, ArrDelayMinutes, ArrDel15,←
    ArrivalDelayGroups, ArrTimeBlk, Cancelled, CancellationCode, Diverted,←
    CRSElapsedTime, ActualElapsedTime, AirTime, Flights, Distance, DistanceGroup,←
    CarrierDelay, WeatherDelay, NASDelay, SecurityDelay, LateAircraftDelay, FirstDepTime←
    , TotalAddGTime, LongestAddGTime, DivAirportLandings, DivReachedDest,←
DivActualElapsedTime, DivArrDelay, DivDistance, Div1Airport, Div1AirportID, ←
Div1AirportSeqID, Div1WheelsOn, Div1TotalGTime, Div1LongestGTime, Div1WheelsOff, ←
Div1TailNum, Div2Airport, Div2AirportID, Div2AirportSeqID, Div2WheelsOn, ←
Div2TotalGTime, Div2LongestGTime, Div2WheelsOff, Div2TailNum, Div3Airport, ←
Div3AirportID, Div3AirportSeqID, Div3WheelsOn, Div3TotalGTime, Div3LongestGTime, ←
Div3WheelsOff, Div3TailNum, Div4Airport, Div4AirportID, Div4AirportSeqID, ←
Div4WheelsOn, Div4TotalGTime, Div4LongestGTime, Div4WheelsOff, Div4TailNum, ←
Div5Airport, Div5AirportID, Div5AirportSeqID, Div5WheelsOn, Div5TotalGTime, ←
Div5LongestGTime, Div5WheelsOff, Div5TailNum)

SET CRSDepTime=STR_TO_DATE(@bdCRSDepTime,'%H %i'), DepTime=STR_TO_DATE(@bdDepTime,'%H %i'), WheelsOff=STR_TO_DATE(@bdWheelsOff,'%H %i'), WheelsOn=STR_TO_DATE(@bdWheelsOn,'%H %i'), CRSArrTime=STR_TO_DATE(@bdCRSArrTime,'%H %i'), ArrTime=STR_TO_DATE(@bdArrTime,'%H %i')

print "Executing load of csv file " + i
c.execute(loader)
con.commit()

print "successfully loaded table " + i
print "Now reconfiguring composite primary keys"

composite_keys="""" alter table '%s' add primary key (FlightDate, UniqueCarrier,←
AirlineID, TailNum, FlightNum, OriginAirportID, OriginAirportSeqID,←
DestAirportID, DestAirportSeqID, DepTimeBlk, ArrTimeBlk);"""" %(tablename)
c.execute(composite_keys)
con.commit()

#can re-write for flexibility – but not important

print "Now partitioning table for speedup"

partition_table="""" alter table %s partition by range ( year(FlightDate) ) ←
subpartition by hash ( month(FlightDate) ) {

partition y03 values less than (2004) {

subpartition m103,
subpartition m203,
subpartition m303,
subpartition m403,
subpartition m503,
subpartition m603,
subpartition m703,
subpartition m803,
subpartition m903,
subpartition m1003,
subpartition m1103,
subpartition m1203
},
partition y04 values less than (2005) (  
  subpartition m104,  
  subpartition m204,  
  subpartition m304,  
  subpartition m404,  
  subpartition m504,  
  subpartition m604,  
  subpartition m704,  
  subpartition m804,  
  subpartition m904,  
  subpartition m1004,  
  subpartition m1104,  
  subpartition m1204  
),
partition y05 values less than (2006) (  
  subpartition m105,  
  subpartition m205,  
  subpartition m305,  
  subpartition m405,  
  subpartition m505,  
  subpartition m605,  
  subpartition m705,  
  subpartition m805,  
  subpartition m905,  
  subpartition m1005,  
  subpartition m1105,  
  subpartition m1205  
),
partition y06 values less than (2007) (  
  subpartition m106,  
  subpartition m206,  
  subpartition m306,  
  subpartition m406,  
  subpartition m506,  
  subpartition m606,  
  subpartition m706,  
  subpartition m806,  
  subpartition m906,  
  subpartition m1006,  
  subpartition m1106,  
  subpartition m1206  
)
);  
  partition y07 values less than (2008) (  
    subpartition m107,  
    subpartition m207,  
    subpartition m307,  
    subpartition m407,  
    subpartition m507,  
    subpartition m607,  
    subpartition m707,  
    subpartition m807,  
    subpartition m907,  
    subpartition m1007,  
    subpartition m1107,  
    subpartition m1207  
  ),  
  partition y08 values less than (2009) (  
    subpartition m108,  
    subpartition m208,  
    subpartition m308,  
    subpartition m408,  
    subpartition m508,  
    subpartition m608,  
    subpartition m708,  
    subpartition m808,  
    subpartition m908,  
    subpartition m1008,  
    subpartition m1108,  
    subpartition m1208  
  ),  
  partition y09 values less than (2010) (  
    subpartition m109,  
    subpartition m209,  
    subpartition m309,  
    subpartition m409,  
    subpartition m509,  
    subpartition m609,  
    subpartition m709,  
    subpartition m809,  
    subpartition m909,  
    subpartition m1009,  
    subpartition m1109,
subpartition m1209
)

partition y10 values less than (2011) {
    subpartition m110,
    subpartition m210,
    subpartition m310,
    subpartition m410,
    subpartition m510,
    subpartition m610,
    subpartition m710,
    subpartition m810,
    subpartition m910,
    subpartition m1010,
    subpartition m1110,
    subpartition m1210
}

partition y11 values less than (2012) {
    subpartition m111,
    subpartition m211,
    subpartition m311,
    subpartition m411,
    subpartition m511,
    subpartition m611,
    subpartition m711,
    subpartition m811,
    subpartition m911,
    subpartition m1011,
    subpartition m1111,
    subpartition m1211
}

partition y12 values less than (2013) {
    subpartition m112,
    subpartition m212,
    subpartition m312,
    subpartition m412,
    subpartition m512,
    subpartition m612,
    subpartition m712,
    subpartition m812,
    subpartition m912,
    subpartition m1012,
subpartition m112,
subpartition m1212
)
partition y13 values less than (2014) (  
subpartition m131,
subpartition m213,
subpartition m313,
subpartition m413,
subpartition m513,
subpartition m613,
subpartition m713,
subpartition m813,
subpartition m913,
subpartition m1013,
subpartition m1113,
subpartition m1213
)
));""" %(tablename)
c.execute(partition_table)
con.commit()
alterfordelay=""""alter table %s ADD column delaydummy tinyint DEFAULT 0 after DepDel15, ADD column arrdelaydummy tinyint DEFAULT 0 after ArrDel15;"""" %(tablename)
c.execute(alterfordelay)
con.commit()
dummyupdates=""""update %s set delaydummy=1 where DepDelay>0;"""" %(tablename)
c.execute(dummyupdates)
con.commit()
dummyupdates1=""""update %s set arrdelaydummy=1 where ArrDelay>0;"""" %(tablename)
c.execute(dummyupdates1)
con.commit()
yeardict={}
for i in range(len(year)):
    yeardict[i+1]=year[i]
else:
    print " Apparently the data is loaded... moving on"
yeardict={}
for i in range(len(year)):
    yeardict[i+1]=year[i]
#table='ontime_q303up'
#ontimeq3_03=pd.io.sql.read_frame('''select * from ontime_q303up where flightdate >="%s" AND flightdate <="%s" ', con) %(quarterlyrangessimple [0],
    quarterryanq303up where flightdate <=)%s" AND flightdate <="%s"
quarterlyranges=zip(dates [0::2], dates [1::2])
quartersql=[]
for i,j in quarterlyranges:
    quartersql.append('''select * from %s where flightdate >="%s" AND flightdate <="%s"

    dataframenames=[]
for i in range(len(quarterlyranges)):
    for qno,qdates in quarter_start.iteritems():
        if qdates==quarterlyranges [i][0][5:]:
            idle=df'-'q'+str(qno)+'-'+str(quarterlyranges [i][0][4])
            dataframenames.append(idle)

def colnamechange(name,df):
    newnames=[]
    for i in df.columns:
        newnames.append(str(name)+str(i))
    for i in range(len(df.columns)):
        df.rename(columns={str(df.columns[i]) : str(newnames[i])},inplace=True)
    return df
newdircheck=os.path.isdir(directory+'/'+output+'/')
if newdircheck==False:
    os.mkdir(directory+'/'+output+'/')
    os.mkdir(directory+'/'+output+'/'+delay+'/')
else:
    subdir=os.path.isdir(directory+'/'+output+'/'+delay+'/')
    if subdir==False:
        os.mkdir(directory+'/'+output+'/'+delay+'/')
        print "Beginning ontimeperformance data manipulation..."
    else:
        print "Beginning ontimeperformance data manipulation..."
outputdir=directory+'/'+output+'/'+delay+'/'
origin_columns_for_mean=['DepDelayMinutes', 'delaydummy', 'DepDel15', 'TaxiOut', '-' Cancelled', 'AirTime', 'Flights', 'Distance', 'CarrierDelay', 'WeatherDelay', '-' NASDelay', 'SecurityDelay', 'LateAircraftDelay']
dest_columns_for_mean=['ArrDelayMinutes', 'arrdelaydummy', 'ArrDel15', 'TaxiIn', '-' Cancelled', 'AirTime', 'Flights', 'Distance', 'CarrierDelay', 'WeatherDelay', '-' NASDelay', 'SecurityDelay', 'LateAircraftDelay']
datecounter=['Year', 'Quarter']
def time_count(currentdate):
last_date=[]
last_date.append(currentdate)
return last_date

def pdDFmaker(outputname, df, groupitem, columnlist, perform, change, newcolname):
    outputname=df.groupby([str(groupitem)]).perform
    if change=='n':
        outputname=outputname
    else:
        outputname=colnamechange(newcolname, df)
    return outputname

aggor=np.zeros(len(quartersql))
aggde=np.zeros(len(quartersql))
for i in range(len(quartersql)):
    counter=i
    fnme=outputdir+str(dataframenames[counter])
    fills=(str(fnme+'_origin_stats.csv'))
    fills2=(str(fnme+'_dest_stats.csv'))
    if os.path.isfile(fills):
        aggor[i]=1
    else:
        aggor[i]=0
    if os.path.isfile(fills2):
        aggde[i]=1
    else:
        aggde[i]=0
if 2*int(len(dataframenames))==int(sum(aggor))+int(sum(aggde)):
    exiter="origin and dest indiv files exist"
    print exiter
    if os.path.isfile(outputdir+'deststats.csv') and os.path.isfile(outputdir+'←originstats.csv') and os.path.isfile(outputdir+'odstats.csv'):
        exiter="origin dest aggregate stats exist"
        print exiter
        ontimefilesexist=1
    else:
        exiter="some files missing...determining which"
        print exiter
        ontimefilesexist=1
else:
    exiter="some files missing...re-manipulating"
    print exiter
    ontimefilesexist=0
if ontimefilesexist==0:
    for i in range(len(quartersql)):
        #currentquarter=quarterlyrange []
        #previous_date=last_date
        print "Querying database now for time period: " + quartersql[i]
        currentdataframe=pd.io.sql.read_frame(quartersql[i],con)
        currentdataframe=pd.DataFrame(currentdataframe)
        counter=i
        #current quarter : year
datecurrently=[str(dataframenames[counter])]
        #output placement based on current period
        fnme=outputdir+str(dataframenames[counter])
        #ORIGIN DATA GROUPINGS #
        #JOINING THE SEPARATE ORIGIN BASED TABLES #
        datedf=currentdataframe.groupby(['Origin'])[datecounter].mean()
        cdfmean=currentdataframe.groupby(['Origin'])[origin_columns_for_mean].mean()
        colnamechange('Mean',cdfmean)
        cdfvar=currentdataframe.groupby(['Origin'])[origin_columns_for_mean].var()
        colnamechange('Var',cdfvar)
        cdfsum=currentdataframe.groupby(['Origin'])[origin_columns_for_mean].sum()
        colnamechange('Sum',cdfsum)
        origin_stats=datedf.join(cdfmean.join(cdfvar.join(cdfsum)))
        #saving the origin file
        origin_stats.to_csv(str(fnme+'_origin_stats.csv'))
        #DEST DATA GROUPINGS #
        #JOINING THE SEPARATE ORIGIN BASED TABLES #
        datedf=currentdataframe.groupby(['Dest'])[datecounter].mean()
        cdfmean=currentdataframe.groupby(['Dest'])[dest_columns_for_mean].mean()
        colnamechange('Mean',cdfmean)
        cdfvar=currentdataframe.groupby(['Dest'])[dest_columns_for_mean].var()
        colnamechange('Var',cdfvar)
        cdfsum=currentdataframe.groupby(['Dest'])[dest_columns_for_mean].sum()
        colnamechange('Sum',cdfsum)
        dest_stats=datedf.join(cdfmean.join(cdfvar.join(cdfsum)))
        #saving the origin file
dest_stats.to_csv(str(fnme+'.'+dest_stats.csv'))

# function returns list 'last_date'

print("Time period " + str(time_count(dataframenames[counter])) + " just processed")

ontimefileexist=1
else:
    ontimefileexist=ontimefileexist

    # reimport to create currentdataframe as concat/appension of all

    if ontimefileexist==1:
        originfilesuffix=['_origin_stats.csv']
        destfilesuffix=['_dest_stats.csv']
        originimportlist=[]
        destimportlist=[]
        for name in dataframenames:
            originimportlist.append(outputdir+name+originfilesuffix[0])
            destimportlist.append(outputdir+name+destfilesuffix[0])

        genlist=[]
        for file in originimportlist:
            currentdataframe=pd.read_csv(file,index_col=None, header=0)
            genlist.append(currentdataframe)

        currentdataframe=pd.concat(genlist)

        originstats=currentdataframe

        originstats.to_csv(outputdir+'originstats.csv')
        genlist=[]

        for file in destimportlist:
            currentdataframe=pd.read_csv(file,index_col=None, header=0)
            genlist.append(currentdataframe)

        currentdataframe=pd.concat(genlist)

        deststats=currentdataframe

        deststats.to_csv(outputdir+'deststats.csv')

        originstatsreindex=originstats.set_index(['Origin','Year','Quarter'])
        colnamechange('origin',originstatsreindex)

        deststatsreindex=deststats.set_index(['Dest','Year','Quarter'])
        colnamechange('dest',deststatsreindex)

        odstats=pd.merge(originstatsreindex,deststatsreindex,how='outer',left_index=True,right_index=True)

        odstats.to_csv(outputdir+'odstats.csv')

        ontimefileexist=100
    else:
        ontimefileexist=ontimefileexist

if ontimefileexist==100:
#extract unique airports to balance panel
#odstats=pd.read_csv(('outputdir+'odstats.csv'),sep=',')
#odstats=pd.DataFrame(odstats)
origins=pd.read_table(('outputdir+'originstats.csv', sep='',)
originiter=(set(x.split('|')) for x in origins.Origin)
oairports=sorted(set.union(*originiter))
odummies=pd.DataFrame(np.zeros((len(origins),len(oairports))),columns=oairports)

for i,gen in enumerate(origins.Origin):
    odummies.ix[i, gen.split('|')] = 1
def uniquelistmaker(sumlist):
    trackerlist=[]
    for i in range(len(sumlist)):
        if sumlist[i]==len(quarterlyranges):
            trackerlist.append(sumlist.index[i])
        else:
            continue
    return trackerlist

#this is slow but it works
sumorigins=odummies.sum()
origintracker=uniquelistmaker(sumorigins)
dests=pd.read_table(('outputdir+'deststats.csv', sep='',)
destiter=(set(x.split('|')) for x in dests.Dest)
dairports=sorted(set.union(*destiter))
ddummies=pd.DataFrame(np.zeros((len(dests),len(dairports))),columns=dairports)

for i,gen in enumerate(dests.Dest):
    ddummies.ix[i, gen.split('|')] = 1
sumdest=ddummies.sum()
desttracker=uniquelistmaker(sumdest)

print str(len(origintracker)) + " unique origin airports"
print str(len(desttracker)) + " unique dest airports"

print "Checking for Set equality"
if set(origintracker)==set(desttracker):
    print "Sets are equal"
    print "Origins match destinations --> balancing panel worked"
    ontimebalanced=origintracker
else:
    print "More work required to balance"
    ontimebalanced=set(origintracker).intersection(set(desttracker))

print "Now recombing data in balanced format"
odstats.reset_index(level=0,inplace=True)
balanceddelaypanel=odstats[odstats['level_0'].isin(ontimebalanced)]
balanceddelaypanel.to_csv(directory+'/output/>'+str(dataframenames[0][3:])+str(dataframenames[-1][2:]+'_balanced_delay_panel.csv'))
delayodstats=balanceddelaypanel

else:
    print "Something went wrong"

#T100 Date import
# This is less generalized but is fairly intuitive
print "On time performance data successfully combined and extracted"
print "Now we will gather t100 data"
datadir=raw_input("Please enter location for raw t100 data files (unable to automatically download these...): ")
if datadir=='D':
    datadir='/media/Files/Data/AVIATION/Data/segment/
else:
    datadir=datadir
ext='.csv'
prepension='t100_a'
t100files=[]
for i in year:
    t100files.append(prepension+str(i)+ext)
lof=t100files

currentcount=0
tablename='t100.'+str(year[0])+'
t100tablemaker="""create table "/%
"DEPARTURES_SCHEDULED' INT(10) NULL DEFAULT NULL ,
"DEPARTURES_PERFORMED' INT(10) NULL DEFAULT NULL ,
"PAYLOAD' INT(10) NULL DEFAULT NULL ,
"SEATS' INT(10) NULL DEFAULT NULL ,
"PASSENGERS' INT(10) NULL DEFAULT NULL ,
"FREIGHT' INT(10) NULL DEFAULT NULL ,
"MAIL' INT(10) NULL DEFAULT NULL ,
"DISTANCE' INT(10) NULL DEFAULT NULL ,
"RAMP_TO_RAMP' INT(10) NULL DEFAULT NULL ,
"AIR_TIME' INT(10) NULL DEFAULT NULL ,
"UNIQUE_CARRIER' VARCHAR(50) NULL DEFAULT NULL ,
"AIRLINE_ID' INT(10) NULL DEFAULT NULL ,
"UNIQUE_CARRIER_NAME' VARCHAR(255) NULL DEFAULT NULL ,
"UNIQUE_CARRIER_ENTITY' VARCHAR(10) NULL DEFAULT NULL ,
"""
"REGION" VARCHAR(1) NULL DEFAULT NULL,
"CARRIER" VARCHAR(5) NULL DEFAULT NULL,
"CARRIER_NAME" VARCHAR(255) NULL DEFAULT NULL,
"CARRIER_GROUP" TINYINT(1) NULL DEFAULT NULL,
"CARRIER_GROUP_NEW" TINYINT(1) NULL DEFAULT NULL,
"ORIGIN_AIRPORT_ID" INT(5) NULL DEFAULT NULL,
"ORIGIN_AIRPORT_SEQ_ID" INT(7) NULL DEFAULT NULL,
"ORIGIN_CITY_MARKET_ID" INT(5) NULL DEFAULT NULL,
"ORIGIN" VARCHAR(3) NULL DEFAULT NULL,
"ORIGIN_CITY_NAME" VARCHAR(255) NULL DEFAULT NULL,
"ORIGIN_STATE_ABR" VARCHAR(2) NULL DEFAULT NULL,
"ORIGIN_STATE_FIPS" VARCHAR(2) NULL DEFAULT NULL,
"ORIGIN_STATE_NM" VARCHAR(255) NULL DEFAULT NULL,
"ORIGIN_COUNTRY" VARCHAR(2) NULL DEFAULT NULL,
"ORIGIN_COUNTRY_NAME" VARCHAR(255) NULL DEFAULT NULL,
"ORIGIN_WAC" INT(3) NULL DEFAULT NULL,
"DEST_AIRPORT_ID" INT(5) NULL DEFAULT NULL,
"DEST_AIRPORT_SEQ_ID" INT(7) NULL DEFAULT NULL,
"DEST_CITY_MARKET_ID" INT(5) NULL DEFAULT NULL,
"DEST" VARCHAR(3) NULL DEFAULT NULL,
"DEST_CITY_NAME" VARCHAR(255) NULL DEFAULT NULL,
"DEST_STATE_ABR" VARCHAR(2) NULL DEFAULT NULL,
"DEST_STATE_FIPS" VARCHAR(2) NULL DEFAULT NULL,
"DEST_STATE_NM" VARCHAR(255) NULL DEFAULT NULL,
"DEST_COUNTRY" VARCHAR(2) NULL DEFAULT NULL,
"DEST_COUNTRY_NAME" VARCHAR(255) NULL DEFAULT NULL,
"DEST_WAC" INT(3) NULL DEFAULT NULL,
"AIRCRAFT_GROUP" TINYINT(1) NULL DEFAULT NULL,
"AIRCRAFT_TYPE" INT(3) NULL DEFAULT NULL,
"AIRCRAFT_CONFIG" TINYINT(1) NULL DEFAULT NULL,
"YEAR" YEAR(4) NULL DEFAULT NULL,
"QUARTER" TINYINT(1) NULL DEFAULT NULL,
"MONTH" TINYINT(2) NULL DEFAULT NULL,
"DISTANCE_GROUP" INT(2) NULL DEFAULT NULL,
"CLASS" VARCHAR(1) NULL DEFAULT NULL,
"DATA_SOURCE" VARCHAR(2) NULL DEFAULT NULL
)
ENGINE=InnoDB;

```c

if raw_input( "Has t100 data table been created? (y/N)" ) == 'y' or raw_input == 'Y':
```
print "good"
else:
c.execute(t100tablemaker)
con.commit()
t100switch=raw_input("Has the t100 data already been loaded into the database? (y/N)")
if t100switch=='n' or t100switch=='N':
    for i in l0f:
        loader="""Load data infile '%s' into table %s fields terminated by ',' optionally enclosed by '"' IGNORE 1 LINES"""%(datadir+i,tablename)
        print "Executing load of csv file " + i
        c.execute(loader)
        con.commit()
        print "successfully loaded table " + i
        con.commit()
        print "successfully loaded table " + i
        print "Moving on"
else:
    print "good"
print "All t100 csv's loaded!"
t100deletions=raw_input("Would you like to remove extraneous data from the database? (y/N)"
if t100deletions=='y' or t100deletions=='Y':
    print "Checking if months need removing..."
    loader="""SELECT min(quarter) from %s where year=%s"""%(tablename,begyearrange"
    c.execute(loader)
    minq=c.fetchone()
    minq=int(minq[0])
    if minq<firstperiod:
        loader="""DELETE FROM %s where year=%s and quarter<%s"""%(tablename,begyearrange,firstperiod)
        print "Removing unneeded data prior to start of time period"
        c.execute(loader)
    else:
        print "continuing"
        loader="""SELECT max(quarter) from %s where year=%s"""%(tablename,endyearrange)
        c.execute(loader)
        maxq=c.fetchone()
        maxq=int(maxq[0])
if maxq>endperiod:
    loader="""DELETE FROM %s where year=%s and quarter>=%s""" % (tablename, endyearrange, endperiod+1)
    print "Removing post-time data"
c.execute(loader)
else:
    print "continue"
else:
    print "okay... beginning extractions"
origin_columns_for_mean=['DEPARTURES.PERFORMED', 'PAYLOAD', 'SEATS', 'PASSENGERS', 'FREIGHT', 'MAIL', 'DISTANCE', 'RAMP_TO_RAMP', 'AIR_TIME']
dest_columns_for_mean=['DEPARTURES.PERFORMED', 'PAYLOAD', 'SEATS', 'PASSENGERS', 'FREIGHT', 'MAIL', 'DISTANCE', 'RAMP_TO_RAMP', 'AIR_TIME']
datecounter=['YEAR', 'QUARTER']
qrange=[]
for i in range(len(dataframenames)):
    q=dataframenames[i][4:5]
    r=dataframenames[i][6:]
    qrange.append([q,r])
quartersql=[]
for i,j in qrange:
    quartersql.append("'select * from %s where quarter=%s AND year=%s'" % (tablename, i, j))
newdircheck=os.path.isdir(directory+'/output/'+t100+'/')
if newdircheck==False:
    os.mkdir(directory+'/output/'+t100+'/')
    print "Beginning t100 data manipulation..."
else:
    print "Beginning t100 data manipulation..."
print "determining how much we have to do..."
outputdir=directory+'/output/'+t100+'/'
aggor=np.zeros(len(quartersql))
aggde=np.zeros(len(quartersql))
for i in range(len(quartersql)):
    counter=i
    fnme=outputdir+str(dataframenames[counter])
    fills=(str(fnme+'_t100_origin_stats.csv'))
    fills2=(str(fnme+'_t100_dest_stats.csv'))
    if os.path.isfile(fills):
        aggor[i]=1
else:
    aggor[i]=0
if os.path.isfile(fills2):
    aggde[i]=1
else:
    aggde[i]=0
if 2*int(len(dataframenames))==int(sum(aggor)+sum(aggde)):
    exiter="t100 origin and dest indiv files exist"
    print exiter
    if os.path.isfile(outputdir+'deststats.csv') and os.path.isfile(outputdir+'odstats.csv'):
        exiter="origin dest aggregate stats exist"
        print exiter
t100filesexist=1
else:
    exiter="some files missing...determining which"
    print exiter
t100filesexist=0
else:
    exiter="some files missing...re-manipulating"
    print exiter
t100filesexist=0
if t100filesexist==0:
    for i in range(len(quartersql)):
        currentdataframe=pd.io.sql.read_frame(quartersql[i],con)
        currentdataframe=pd.DataFrame(currentdataframe)
        counter=i
datecurrently=\[str(dataframenames[counter])]\nfnme=outputdir+str(dataframenames[counter])

############### ORIGIN T100

datedf=currentdataframe.groupby(\['ORIGIN'\])\[datecounter\].mean()
cdfmean=currentdataframe.groupby(\['ORIGIN'\])\[origin_columns_for_mean\].mean()
    colnamechange(\'Mean\',cdfmean)
cdfvar=currentdataframe.groupby(\['ORIGIN'\])\[origin_columns_for_mean\].var()
colnamechange(\'Var\',cdfvar)
cdfsun=currentdataframe.groupby(\['ORIGIN'\])\[origin_columns_for_mean\].sum()
colnamechange(\'Sum\',cdfsun)
origin_stats=datedf.join(cdfmean.join(cdfvar.join(cdfsun)))
origin_stats.to_csv(str(fnme+'\_t100\_origin\_stats.csv'))

############### DESTINATION T100
datedf = currentdataframe.groupby(["DEST"])[datecounter].mean()
cdfmean = currentdataframe.groupby(["DEST"])[dest_columns_for_mean].mean()
colnamechange('Mean', cdfmean)
cdfvar = currentdataframe.groupby(["DEST"])[dest_columns_for_mean].var()
colnamechange('Var', cdfvar)
cdfs = currentdataframe.groupby(["DEST"])[dest_columns_for_mean].sum()
colnamechange('Sum', cdfs)
dest_stats = datedf.join(cdfmean.join(cdfvar.join(cdfs)))
dest_stats.to_csv(str(fnme + 'deststats.csv'))

odcdf = currentdataframe.groupby(["ORIGIN", "DEST"])["DEPARTURES PERFORMED"]
odfull = odcdf.sum().unstack()
odfull.to_csv(str(fnme + 'odops.csv'))

print "Time period " + str(time_count(dataframenames[counter])) + " just ← processed"
t100fileexist = 1
else:
t100fileexist = t100fileexist
if t100fileexist == 1:
    originfilesuffix = ["_t100_origin_stats.csv"]
destfilesuffix = ["_t100_dest_stats.csv"]
originimportlist = []
destimportlist = []
for name in dataframenames:
    originimportlist.append(outputdir + name + originfilesuffix[0])
    destimportlist.append(outputdir + name + destfilesuffix[0])
genlist = []
for file in originimportlist:
    currentdataframe = pd.read_csv(file, index_col=None, header=0)
    genlist.append(currentdataframe)
currentdataframe = pd.concat(genlist)
originstats = currentdataframe
originstats.to_csv(outputdir + 't100_originstats.csv')
genlist = []
for file in destimportlist:
    currentdataframe = pd.read_csv(file, index_col=None, header=0)
    genlist.append(currentdataframe)
currentdataframe = pd.concat(genlist)
deststats = currentdataframe
deststats.to_csv(outputdir + 't100_deststats.csv')
originstatsreindex = originstats.set_index(["ORIGIN", "YEAR", "QUARTER"])
colnamechange('origin', originstatsreindex)
deststatsreindex=deststats.set_index(['DEST', 'YEAR', 'QUARTER'])
colnamechange('dest', deststatsreindex)
odstats=pd.merge(originstatsreindex, deststatsreindex, how='outer', left_index=True, right_index=True)
odstats.to_csv(outputdir+'t100_odstats.csv')
t100filesexist=100

else:
t100filesexist=t100filesexist
if t100filesexist==100:
    origins=pd.read_table(outputdir+'t100_originstats.csv', sep=',',)
originiter=(set(x.split('|')) for x in origins.ORIGIN)
oairports=sorted(set.union(*originiter))
odummies=pd.DataFrame(np.zeros((len(origins),len(oairports))),columns=oairports)

    for i,gen in enumerate(origins.ORIGIN):
        odummies.ix[i, gen.split('|')]=1
sumorigins=odummies.sum()
origintracker=uniquelistmker(sumorigins)
dests=pd.read_table(outputdir+'t100_deststats.csv',sep=',',)
destiter=(set(x.split('|')) for x in dests.DEST)
dairports=sorted(set.union(*destiter))
ddummies=pd.DataFrame(np.zeros((len(dests),len(dairports))),columns=dairports)

    for i,gen in enumerate(dests.DEST):
        ddummies.ix[i, gen.split('|')]=1
sumdest=ddummies.sum()
desttracker=uniquelistmker(sumdest)
print str(len(origintracker)) + " unique origin airports"
print str(len(desttracker)) + " unique dest airports"
print "Checking for Set equality"
if set(origintracker)==set(desttracker):
    print "Sets are equal"
    print "Origins match destinations --> balancing panel worked"
t100balanced=origintracker
else:
    print "More work required to balance"
    print "Doing that thing..."
    t100balanced=set(origintracker).intersection(set(desttracker))
    print "Done."
else:
    print "not sure"
print "Now recombining data in balanced format"
odstatsbak=odstats
odstats.reset_index(level=0,inplace=True)
balancedt100panel=odstats[odstats['level_0'].isin(t100balanced)]
balancedt100panel.to_csv(directory+'output/'+'str(dataframenames[0][3:])+'str(←
    dataframenames[−1][2:]+'balanced_t100_panel.csv')
t100odstats=balancedt100panel

#now to combine the two...

t1air=otair

t1airset=set(x.split('|') for x in t1airports.level_0)
dairports=sorted(set.union(*t1airset))
ddummies=pd.DataFrame(np.zeros((len(t1airports),len(dairports))),columns=←
dairports)

for i,gen in enumerate(t1airports.level_0):
    ddummies.ix[1, gen.split('|')]=1

otair=ddummies.sum()

otairset=uniqulistmker(otair)

t100airports=pd.read_table(directory+'output/'+'str(dataframenames[0][3:])+'str(←
    dataframenames[−1][2:]+'balanced_t100_panel.csv'),sep='|')
t100iter=(set(x.split('|')) for x in t100airports.level_0)
tairports=sorted(set.union(*t100iter))
t100dummies=pd.DataFrame(np.zeros((len(t100airports),len(tairports))),columns=←
tairports)

for i,gen in enumerate(t100airports.level_0):
    t100dummies.ix[1, gen.split('|')]=1
tlair=t100dummies.sum()
tlairset=uniqulistmker(tlair)

combobalance=set(otairset).intersection(set(tlairset))
t100odstats.reset_index(level=1,inplace=True)
t100odstats.reset_index(level=2,inplace=True)
clippedbalancedt100panel=t100odstats[t100odstats['level_0'].isin(combobalance)]
clippedbalancedt100panelreindex=clippedbalancedt100panel.set_index(['level_0','level_1'])

colnamechange('t100',clippedbalancedt100panelreindex)
delayodstats.reset_index(level=1,inplace=True)
delayodstats.reset_index(level=2,inplace=True)
delayodstatsreindex=delayodstats.set_index(['level_0','index','level_1'])
colnamechange('delay',delayodstatsreindex)
mergedbalance=pd.merge(clippedbalancedt100panelreindex,delayodstatsreindex,how='←
    outer',left_index=True,right_index=True)
mergedbalance.to_csv(directory)+'/output/'+str(dataframenames[0][3:])+str(→dataframenames[-1][2:]+',merged_delay_t100_balanced.csv'))

# generate file list for weight matrix creation
newdircheck=os.path.isdir(directory+'/output/wmats/)
if newdircheck==False:
    os.mkdir(directory+'/output/wmats/)
    print "Beginning weight matrix reduction..."
else:
    print "Beginning weight matrix reduction..."

odopslist=[]
for i in range(len(quartersql)):
    odopslist.append(str(outputdir)+str(dataframenames[i])+'.odops.csv')

# reduce weight matrices
redodopslist=[]
for i in range(len(odopslist)):
    odfull=pd.read_table(odopslist[i],index_col=[0],sep=',')
    # removes airport rows not matching
    odreduced=odfull[odfull.index.isin(combobalance)]
    odreduced=pd.DataFrame([odreduced[str(airport) for airport in otairset]).fillna(0)
    redodopslist.append(directory+'/output/wmats/>'+str(dataframenames[i])+'.wmat.csv')

    odreduced.to_csv(directory+'/output/wmats/>'+str(dataframenames[i])+'.wmat.csv')
mainwmat=odreduced-odreduced

# combine to form single wmat
for i in redodopslist:
    currwmat=pd.read_table(i,index_col=[0],sep=',')
    mainwmat=mainwmat+currwmat

WMAT=mainwmat

WMAT.to_csv(directory+'/output/'+str(dataframenames[0][3:])+str(dataframenames[-1][2:]+',WMAT.csv'))

dmergedpaneldf=pd.read_table(directory+'/output/'+str(dataframenames[0][3:])+str(dataframenames[-1][2:]+',merged_delay_t100_balanced.csv'),index_col=[0,1,2],sep=',')
df=dmergedpaneldf

# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
# # AGGREGATES #
##
## A G G R E G A T E S #
##

159
## AIRPORT TOTALS

\[
df['totalops']=df['t100originSumDEPARTURES_PERFORMED'] + df['t100destSumDEPARTURES_PERFORMED']
\]
\[
df['totalpayload']=df['t100originSumPAYLOAD'] + df['t100destSumPAYLOAD']
\]
\[
df['totalseats']=df['t100originSumSEATS'] + df['t100destSumSEATS']
\]
\[
df['totalpass']=df['t100originSumPASSENGERS'] + df['t100destSumPASSENGERS']
\]
\[
df['totalfreight']=df['t100originSumFREIGHT'] + df['t100destSumFREIGHT']
\]
\[
df['totalmail']=df['t100originSumMAIL'] + df['t100destSumMAIL']
\]
\[
df['totaldistance']=df['t100originSumDISTANCE'] + df['t100destSumDISTANCE']
\]
\[
df['totalr2r']=df['t100originSumRAMP_TO_RAMP'] + df['t100destSumRAMP_TO_RAMP']
\]
\[
df['totalairtime']=df['t100originSumAIR_TIME'] + df['t100destSumAIR_TIME']
\]
\[
df['meanpayloadperflight']=df['totalpayload'] / df['totalops']
\]
\[
df['meanevents']=df['totalseats'] / df['totalops']
\]
\[
df['meanevents']=df['totalpass'] / df['totalops']
\]
\[
df['meanfreight']=df['totalfreight'] / df['totalops']
\]
\[
df['meanmail']=df['totalmail'] / df['totalops']
\]
\[
df['meandistance']=df['totaldistance'] / df['totalops']
\]
\[
df['meandistance']=df['totalr2r'] / df['totalops']
\]
\[
df['meanairtime']=df['totalairtime'] / df['totalops']
\]
\[
df['meanpctfull']=df['meanevents'] / df['meanevents']
\]

## INBOUND TOTALS

\[
df['inboundops']=df['t100destSumDEPARTURES_PERFORMED']
\]
\[
df['inboundpayload']=df['t100destSumPAYLOAD']
\]
\[
df['inboundseats']=df['t100destSumSEATS']
\]
\[
df['inboundpass']=df['t100destSumPASSENGERS']
\]
\[
df['inboundfreight']=df['t100destSumFREIGHT']
\]
\[
df['inboundmail']=df['t100destSumMAIL']
\]
\[
df['inbounddistance']=df['t100destSumDISTANCE']
\]
\[
df['inboundr2r']=df['t100destSumRAMP_TO_RAMP']
\]
\[
df['inboundairtime']=df['t100destSumAIR_TIME']
\]

## OUTBOUND TOTALS
```python
# More data manipulation required before saving

df['outboundops'] = df['t100originSumDEPARTURES_PERFORMED']
df['outboundpayload'] = df['t100originSumPAYLOAD']
df['outboundseats'] = df['t100originSumSEATS']
df['outboundpass'] = df['t100originSumPASSENGERS']
df['outboundfreight'] = df['t100originSumFREIGHT']
df['outboundmail'] = df['t100originSumMAIL']
df['outbounddistance'] = df['t100originSumDISTANCE']
df['outboundr2r'] = df['t100originSumRAMP_TO_RAMP']
df['outboundairtime'] = df['t100originSumAIR_TIME']

# Mean ( )

df['meaninboundpayload'] = df['inboundpayload'] / df['inboundops']
df['meaninboundseats'] = df['inboundseats'] / df['inboundops']
df['meaninboundpass'] = df['inboundpass'] / df['inboundops']
df['meaninboundfreight'] = df['inboundfreight'] / df['inboundops']
df['meaninboundmail'] = df['inboundmail'] / df['inboundops']
df['meaninbounddistance'] = df['inbounddistance'] / df['inboundops']
df['meaninboundr2r'] = df['inboundr2r'] / df['inboundops']
df['meaninboundairtime'] = df['inboundairtime'] / df['inboundops']

# This allows for selection criteria — see http://webcache.googleusercontent.com/
#search?q=cache:rTo03b4UxNwJ:stackoverflow.com/questions/15315452/selecting-with-
#complex-criteria-from-pandas-dataframe&cd=5&hl=en&ct=clnk&gl=us&client=firefox-a

def currmean = dfc curr. loc [(dfc curr[ 'DepDelay' ] > 0)]. groupby ([ 'Origin', 'Year', 'Quarter' ]) ←
# .mean()

df['ontimeops'] = df[ 'delayoriginSumFlights' ] + df[ 'delaydestSumFlights' ]
df['totaldelayflights'] = df[ 'delayoriginSumdelaydummy' ] + df[ 'delaydestSumdelaydummy' ]
df['delayflightpercent'] = df[ 'delaydummy' ] / df[ 'ontimeops' ]
df['meanoriginNASDelay'] = df[ 'delayoriginSumNASDelay' ] / df[ 'delayoriginSumFlights' ]
```
Appendix C

Popularization of Research

C.1 Posters
Transportation Infrastructure, Industrial Productivity and Return on Investment: A Spatial Spillover Approach

Jeffrey J. Elb, Oleg A. Smirnov and Peter S. Lindquist

University of Nevada

Introduction

Public capital is a vital component of any economy, that contributes not only to the quality of life of a country’s citizens, but also to the productivity of the economy as a whole. In 2008, U.S. public infrastructure ranked 23rd out of 32 OECD countries in terms of public satisfaction (Dep’t of Treasury, 2008). Though public capital has long been explored for its effects on economic growth, productivity and development (among others), little emphasis has been placed on the spatial variation in benefits that arise from public capital investments. This study explores this phenomenon in a 3-part cost-benefit framework as it relates to the US manufacturing industry.

Given fiscal constraints and the current condition of the nation’s aging infrastructure, it is of critical importance to determine not only the magnitude and extent of benefits arising from investments in transportation infrastructure but also the specific locations that provide higher rates of return on investment. That is, which geographic areas, when invested in, prove most beneficial to the broader economy?

Empirical Framework

I. Profit Margin Model

\[ m_j = \alpha + \beta \ln E_j + \beta_1 \ln P_{adj} + \beta_2 \ln P_{adj} + \beta_3 \ln G_j + \beta_4 \ln G_j + \beta_5 \ln K_j + \beta_6 \ln K_j + u_{t,j} \]

\[ \epsilon_j = (\epsilon_{t,j} + F_{t,j}) + u_{t,j} \]

II. Production Function Model

\[ \ln Y_j = \alpha + \beta \ln E_j + \beta_1 \ln P_{adj} + \beta_2 \ln G_j + \beta_3 \ln G_j + \beta_4 \ln P_{adj} + \beta_5 \ln D_{adj} + \epsilon_{t,j} \]

Variables and Data Sources

- \( E_j \): number of employees
- \( P_{adj} \): price of labor
- \( G_j \): average standard deviation of weekly fuel prices
- \( T \): transportation infrastructure capital
- \( D_{adj} \): output (value of shipments)
- \( \epsilon_{t,j} \): spatial dependence term
- \( \alpha \): dummy for year 2006

Production Function Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-Statistics</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E_j )</td>
<td>0.03204</td>
<td>0.00008</td>
<td>0.03204</td>
<td>0.00008</td>
</tr>
<tr>
<td>( P_{adj} )</td>
<td>0.03694</td>
<td>0.00014</td>
<td>0.03694</td>
<td>0.00014</td>
</tr>
<tr>
<td>( G_j )</td>
<td>0.06561</td>
<td>0.00016</td>
<td>0.06561</td>
<td>0.00016</td>
</tr>
<tr>
<td>( T )</td>
<td>0.03464</td>
<td>0.00019</td>
<td>0.03464</td>
<td>0.00019</td>
</tr>
</tbody>
</table>

Firm data from Annual Survey of Manufacture’s, 1997-2010

Production Function Model

\[ \frac{\text{Ratio of Residuals to Transportation Infrastructure Effects}}{\text{Spatial Spillover Framework}} \]

Conclusions

- Own-state and neighboring state transportation infrastructure play positive and significant roles for manufacturers, in terms of profitability and production, and thus, cost.
- A 1 percent increase in own-state transportation infrastructure expenditures yields, on average, a 0.4 percent increase in manufacturer’s profit margins and a 0.2 percent increase in output.
- A 1 percent increase in neighboring state transportation infrastructure expenditures yields a 1.1 percent increase in profit margins and a 0.4 percent increase in output.
- When regressions are run on annual data and on state time series, spatial spillovers appear to be increasing over time.
- Increasing Strategic Cooperation among transportation planning agencies across state borders appear to have a larger impact on own-state manufacturing productivity and production than own-state investments.

Summary:

- Neighboring-state transportation infrastructure investments have a larger impact on own-state manufacturing profitability and production than own-state investments.

Figure C.1: Transportation Infrastructure: A Spatial Spillover Approach.
Transportation Infrastructure Usage and Household Welfare

Jeffrey J. Eloff, Oleg A. Smirnov and Peter S. Lindquist
University of Toledo

Introduction

This paper examines the role of road transportation infrastructure utilization in the context of household benefits. The relationship between the usage of these assets and personal wages in the 10 Midwest member states of the Midwest America Freight Coalition (MAFC) is established. As much of the literature is predominantly concerned with the productivity-related benefits of transportation infrastructure and public capital in general, this study focuses on the benefits that individual users receive from consuming transportation infrastructure. This study takes two key departures from the more general public capital literature by leveraging physical measures of roadway usage and focusing on individual users. The data is constructed at the county level using Federal Highway Administration’s (FHWA) Highway Performance Monitoring System (HPMS) and Bureau of Economic Analysis (BEA) Regional Economic Information System (REIS) data from 1980 to 2008. The results of the statistical and geospatial analyses are presented. Our findings indicate that there is a positive relationship between transportation infrastructure usage and wages.

Motivation

The figure above displays county-specific mean per capita income levels throughout the time horizon of the study (1980-2008). As can be seen, neighboring counties tend to be very similar in their income levels. This is further reinforced by the results of the estimation routines - specifically by the magnitude of the coefficient on the WPCI term. As expected, those counties containing and bordering major cities display the highest levels of income. The levels tend to diminish as one moves away from those agglomeration centers.

The figure below presents the weighted annual average daily traffic for the year 2008 in the region. Highest traffic volumes occur in and around major cities but also along interstates and along corridors between cities. This is best witnessed by examining the band around Lake Michigan, from north of Milwaukee, south through Chicago, and east and north all the way to Detroit. Similarly following I-75 south from Detroit to the state of Kentucky displays the same heightened levels of traffic volumes.

Empirical Framework

$$PCI_{i,j} = \alpha + \beta_1 WPCI_{i,j} + \beta_2 AADT_{i,j} + \epsilon_{i,j}$$

$$PCI_{i,j} = \alpha + \beta_1 WPCI_{i,j} + \beta_3 AADT_{i,j} + \epsilon_{i,j}$$

$$PCI_{i,j} = \alpha + \beta_1 WPCI_{i,j} + \beta_4 AADT_{i,j} + \epsilon_{i,j}$$

Variables and Data Sources

- **PCI**: Per capita income
- **WPCI**: Spatial lag of per capita income
- **AADT**: Annual Average Daily Traffic (AADT)
- **Suburban AADT**: Suburban AADT
- **Urban AADT**: Urban AADT
- **Population**: Population

Results

The estimated coefficient for AADT of 0.0739 in (1) implies that an increase in annual average daily traffic yields a statistically significant increase in per capita income. The same is prevalent in models (2) and (3), albeit at differing degrees depending upon population. Model (3) maintains a similar relationship between the magnitudes of AADT by region type as estimated in (2), though at reduced levels. The addition of the labor population ratio variable also shows that larger labor markets, relative to the population levels earn significantly more. Urban areas receive the highest benefits per capita income from increases in traffic, followed by rural and suburban areas.

Conclusions

- Individual income levels are positively affected by transportation infrastructure usage.
- Households located along major connecting routes experience a higher increase in income relative to non-corridor locations.
- An innovative combination of spatial and panel data estimation techniques provides a formidable framework to uncover the magnitude of this difficult-to-measure relationship.

Summary: Despite difficulties to deal with endogeneity, employing transportation infrastructure usage as a proxy for transportation infrastructure demand allows one to determine the benefit to individuals associated with the demand for infrastructure.

Figure C.2: Transportation Infrastructure Usage and Household Welfare.