4-18-2016

Essays on Retail and Regional Economics

David Christopher Vitt
dcvitt@gmail.com

DOI: 10.25148/etd.FIDC000697

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

ESSAYS ON RETAIL AND REGIONAL ECONOMICS

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ECONOMICS

by

David Vitt

2016

To: Dean John F. Stack, Jr.
Steven J. Green School of International and Public Affairs

This dissertation, written by David Vitt, and entitled Essays on Retail and Regional Economics, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

Cem Karayalcin

Sheng Guo

Florence George

Hakan Yilmazkuday, Major Professor

Date of Defense: April 18, 2016

The dissertation of David Vitt is approved.

Dean John F. Stack, Jr.
Steven J. Green School of International and Public Affairs

Dean Andrés G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2016

DEDICATION

For my wife Danielle, my mother Karen, and my late grandparents Marie and Otto.

If I succeed it is because of your loving support

ACKNOWLEDGMENTS

First and foremost I'd like to thank my advisor Hakan Yilmazkuday for all of the advice and time invested to help me through this degree and on the job market. Thanks for always having my back . I'm grateful for Alex McQuoid and Matt Cole for all the advice and friendship over the years. I'm indebted to my committee members Cem, Sheng, and Florence for their time and patience as committee members. I'm also really thankful for all the years of help from Mariela Delgado, Lorette Garcia, Mayte Rodriguez, and Mihaela Pintea. Finally, I am grateful to Florida International University for honoring me with a Dissertation Year Fellowship.

ABSTRACT OF THE DISSERTATION

ESSAYS ON RETAIL AND REGIONAL ECONOMICS

by

David Vitt

Florida International University, 2016

Miami, Florida

Professor Hakan Yilmazkuday, Major Professor

This dissertation is composed of three essays at the intersection of regional economic analysis and industrial organization. In the first chapter, I derive an estimating equation for retail market structure in order to quantify the effects of e-commerce competition on brick and mortar retail establishment and employment counts. Using a multilevel regression specification, I find that (i) e-commerce establishment count exposure results show heterogeneity in the sign of the effects across the retail sectors represented in the data (ii) the magnitude of the e-commerce exposure effect is also heterogeneous across retail sectors (iii) the heterogeneity is not purely random and correlates highly with retail industrial characteristics like the labor share of receipts and profit margins, (iv) the e-commerce exposure is passed through to intensive margins like employment.

The second essay turns to a regional focus, where I develop a multilevel difference-in-difference approach to estimate the causal effects of discontinued Shuttle launches on the industry and labor markets of Florida's Space Coast. I find strong evidence for (i) an across industry substitution effect previously unexplored in the regional literature(ii) a spike in unemployment of 17% relative to the estimated counterfactual outcome for the region (iii) a contraction in payroll of nearly 10% of regional GDP in some industries combined with a gain of 7.5% through across industry labor reallocation.

In the final essay, I focus on the relationship between the size of retail establishments and the growth of their proximate markets. In accomplishing this, I demonstrate the utility of Department of Defense satellite images of ambient night light activity as a measure of the spatial variation in economic activity, as well as a measure of economic growth. This allowed me to use a dynamic panel regression approach to test the concentrating effect of market growth on retail firms. I find evidence that (i) with an autoregressive coefficient closer to 0 than 1 ($\alpha = 0.23$), establishment size is not persistent (ii) firms adjustment contemporaneously to economic growth and discount past growth for hiring decisions (iii) a positive and significant firm size elasticity with respect to spatial variation in economic activity.

TABLE OF CONTENTS

CHAPTER PAGE

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

THE EMPIRICAL RELATIONSHIPS EBTWEEN INTERNET USE INTENSITY AND RETAIL MARKET STRUCTURE

1.1 Introduction

1.1.1 Motivation

In many ways, the growth of the Internet has effectively decreased the travel costs we face as consumers, provided we are sufficiently patient. By decreased travel costs I refer to the magnitude of the cost incurred with having to shop outside our homes. Data like that presented in (Figure 1.1) show that substitution away from local retail towards e-commerce is significant and strong. Online retail represents only one dimension of many along which retail and the Internet interact. While increasing Internet use may put certain types of brick and mortar retail establishments in closer competition with e-commerce competitors, it also decreases the costs consumers incur to learn about product characteristics like quality and local availability. As a result, certain retail industries may find Internet use beneficial and make establishment location decisions accordingly. I take increasing Internet use to be a proxy of decreasing consumer travel costs, and test the direction of the relationship between Internet use intensity and retail establishment counts while controlling for other market structure determinants.

I make several contributions to the understanding of retail industrial organization, regional commerce, and the determinants of market concentration with this investigation. The most significant contribution is the updated empirical strategy. I demonstrate the advantages of using Google Trends to measure consumers' revealed preferences over the standard survey based measures. The strategy provides both extensive and intensive measures of Internet use, showing that the latter is less important for determining retail market structure. I also demonstrate the use of a control function approach to estimating the effect pure e-commerce search intensity on local brick and mortar retail industries. I then also show that I can instrument for e-commerce intensity using an appropriately chosen keyword, and show that variation in e-commerce intensity have consistently negative effects on establishment counts and the payroll within each industry-state pair. I make a small contribution to the theory by showing that a Sutton style non-fragmentation result exists under very simple cost assumptions. This speaks to the role of quality escalation in retail, since it implies that all growth adjustments are absorbed by incumbent firms on some sort of intensive margin. The theoretical model is useful because it informs my econometric strategy to use multilevel linear and non-linear regression models to estimate the Internet and e-commerce exposure of each industry.

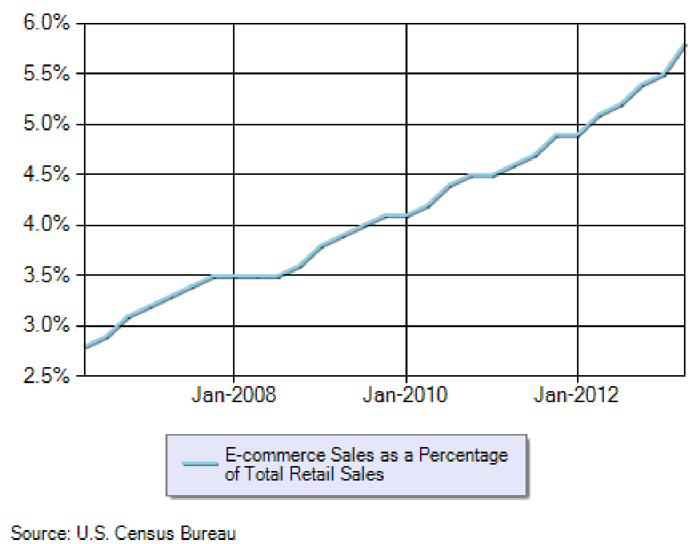


Figure 1.1: Time series of e-commerce as a fraction of total retail sales.

Prior to conducting the econometric investigation, it's not immediately clear that increased Internet use should help or hinder a given retail industry. It is reasonable to suspect that the nature of brick and mortar retail's relationship with the Internet is highly idiosyncratic to each retail industry in question. For example, retail industries with goods that are tradable (in the legal and practical sense) outside the physical retail location are likely to be more exposed competition with their Internet counterparts since shipping the good is feasible, a prime example being books and periodical industries. For some industries, like those in the business of electronic shopping, or for experiential goods, it's possible that Internet savvy consumers are targeted for an expansion of brick and mortar locations in order to "drum up" online sales. Websites managed by firms and consumers alike provide near costless access to prices. A prime example is Gas Buddy, which "crowdsources" reporting gas prices at the establishment level. This in turn allows for price dispersion investigations like (Yilmazkuday and Yilmazkuday, 2015). The presence of review and feedback websites reduce investment required to gain knowledge of product quality via discussion boards. Other retail industries are relatively isolated from e-commerce competition due to legal barriers preventing the shipment of goods via the mail, with gasoline and alcoholic beverage retail being prime examples. Variation in a stated preference data source like e-commerce use survey data, as is the standard in the literature, is insufficient at capturing the nature of the relationship between the brick and mortar retail and the Internet since it may suffer from imperfect recollection or not be truthful. To improve on this, I develop a measure that relies on temporal variation in revealed preferences by using variation in keyword search intensity as a measure of Internet use intensity.

The question of the influence of e-commerce on brick and mortar retail market structure is important on account of the implications to the real economy. On the one hand, if the Internet brings about fewer firms, this in turn may lead to a fall in regional income as consumers substitute away from local retail. On the other hand, there is scope for improvements in real income through the effect on prices associated with concentrating retail into the most cost effective firms. The strategy I develop to investigate the relationship between Internet use intensity and brick and mortar retail is to look at the role of variation in various measures of Internet use as determinants of brick and mortar retail industrial structure. The ease of online shopping activities has effectively decreased the travel costs of consumers by allowing them to substitute away from the convenience of local retail in favor of e-commerce provided they are adequately patient. Areas with consumers more willing shop online effectively have lower consumer travel costs since they are able to consume an identical bundle of goods at a lower level of expenditure. To empirically test the hypothesis I extend a framework connecting consumer travel costs to retail market structure.

Consumer travel costs have indirectly increased over the past years in spite of rising fuel prices on account of the competitive tension the Internet has provided for patient consumers. Expenditure minimizing behavior leads conscious consumers to consider a convenience-patience trade-off every time they face a major expenditure. You can imagine that, before putting the keys in the ignition, they try to weigh the expected markup, probability of stockout and cost of navigating traffic to their brick and mortar retailer against the lowest priced substitute available through e-commerce. This trade-off between the convenience of local retail goods and thrift of their e-commerce substitutes is connected to the consumer's propensity to use the Internet as well as their patience. As both dimensions increase, so too will the share of expenditures being dedicated to purchases online. The presence of this tension, along with its consequences, can be measured using intensive and extensive measures of Internet use. Extensive measures, like the percentage of the population with Internet access, reflect that some states may have better telecommunications structures, more competitive Internet service providers, and differing preferences for Internet access. This scale effect of Internet access on its own does not identify the Internet's effect on retail market structure, but does measure how well connected a state is on average. Alone, it is insufficient since it neglects any state level differences in the intensity of Internet use.

Extensive and intensive Internet use measures as a consumer travel cost proxy are developed out of necessity: there is a lack of e-commerce sales data at the industry-state level, and this measure would undoubtedly be endogenous on account of confounding factors correlated with brick and mortar retail market structure. In light of this, it is necessary first necessary to find a data source that provides information about Internet use that varies by state over time. Google Trends reports keyword search frequency on a weekly basis at the state level or higher resolution. The next step would be selecting a keyword sufficiently general as to capture the widest cross-section of the consumer base possible. For reasons discussed in further detail in the empirical section, variation in Google Trends data on searches for "pornography" across states and time was a natural choice for a variable that measures Internet use intensity. It is desirable on account of the difficulty of arguing its endogeneity within the context of the model. For example, suggesting simultaneity bias would imply that some aspect of retail market structure is in some way influencing consumers to search more frequently for this keyword, and going through the transmission mechanism from changes in retail market structure to changes in consumer attitudes towards pornography shows that any connection would be dubious. Thus, it is a highly desirable measure from an exogeneity standpoint. I also directly address the threat of e-commerce by using search intensity for "pornography" as an instrument for search intensity for "amazon.com". This keyword directly measures consumers' revealed preference not only for shopping online, but also for learning about e-commerce.

1.1.2 Related Literature

Reduced form econometric research on the determinants of retail market structure started with (Berry et al., 1962) and was replicated and further discussed in (Forbes, 1972). In these, the log of retail establishment counts are regressed on the log of population using different cross sections of MSAs, with Forbes drawing on a larger sample. Both estimate the elasticity of retail establishment counts with respect to the population; (Forbes, 1972) finds an elasticity of 0.96 while (Berry et al., 1962) finds a lower result near 0.7. A pattern of across retail industry heterogeneity in responses to population growth is also initially presented in (Forbes, 1972), though not both are estimated via simple OLS. My results will show that retail establishment counts are less and differently sensitive to population growth than these previous estimates.

The leading analysis of the relationship between retail market structure and Internet use is (Goldmanis et al., 2010). In this setup, the co-authors quantify the exposure of retail market share to variation in self reported e-commerce adoption measures for travel agencies, book stores, and new car dealers. This lays the groundwork in the field by demonstrating that there is heterogeneity in the exposure of retail industries, and that increased exposure to e-commerce displaces the least efficient firms in a manner similar to the reallocation in (Melitz, 2003a). My approach improves this strategy in that I use within state variation in keyword search intensity as revealed preference measurement of Internet use intensity, as opposed to a stated preference survey source restricted to e-commerce. A time varying measure like keyword search intensity with an appropriately chosen keyword can imply a causal relationship with variation in generic Internet use intensity, and feels like a more credible identification strategy since it does not rely on stated preferences. Additionally, instead of focusing on a few retail industries, I conduct an investigation with the entire set of retail industries represented in the U.S. Census's Statistics of U.S. Business.

My research does not stand alone in making the connection between Internet use and various costs faced by the consumer. The relationship between the Internet and reductions in transportation, communication, and search costs to near zero are argued in Shapiro and Varian (1999), Cairncross (1997), and Bakos (1997) respectively. Goolsbee (1999) suggests that the Internet reduces the importance of distance in the sense that it frequently allows consumers to avoid local taxes and therefore effectively increases real wages and consequently welfare. In a more aggregate investigation, Freund and Weinhold (2004) show that the Internet helps alleviate the influence of distance on trade. Their investigation suggests that a 10% increase in the web hosts within a country elicits a 0.2% increase in export growth.

Spatial trade models with retail sectors suggest an inverse relationship between consumer travel costs and retail establishment counts. To motivate an econometric estimating equation, I extend the (Eckel, 2009) spatial model of retail competition. Eckel's 2009 model seeks to identify how international trade affects retail market structure through the entry and exit of retail establishments. This is a valid and important question for many reasons, foremost since the retail industry employs $\approx$ 15.4 million workers, according to December 2014 BLS estimates[1]. As an industry, retail represents approximately 11% of the country's total labor force employment. It is also easily shown in Eckel (2009) that retail consolidation or fragmentation has real implications through the price effect on real wages paid to workers in the whole of the economy. The health of the retail sector is also a concern for participants in the financial markets, since retail firms compose a substantive part of the NASDAQ and nearly one quarter of the firms in the Dow Jones Industrial Average. The industry is largely referenced as a leading indicator of the health of the macroeconomy, so understanding the dynamics in this sector gives perspective on prospective states of the economy as a whole.

[1] http://www.bls.gov/news.release/empsit.t17.htm

1.2 Theoretical Model

1.2.1 Preliminaries

First I will show there is a bound to market fragmentation in a spatial retail competition model that generates a natural oligopoly retail market structure. This model has the advantage of reaching the same results regarding properties of retail market structure while assuming a simpler cost structure than is traditionally needed. Second, I show that this bound acts as an econometric specification of retail market structure in a population asymptotic limit, and I develop an econometric strategy for using the structural equation in practice. I proceed by describing the theoretical foundations on which my contribution rests, with more details left for the appendix.

I begin by summarizing the spatial retail competition model developed in Eckel (2009) upon which my contribution draws. The approach begins with a Krugman (1980) style model of monopolistic competition. As opposed to a true international investigation, I take the "world" to be the 50 United States in order to use subnational data, which are free from potentially confounding trade policies. On the demand side, consumers with CES preferences have a taste for variety as in (Dixit and Stiglitz, 1977), along with an iceberg style travel costs associated with visiting retailers. Since consumers incur a cost to shopping, their utility maximizing decision is to make "one stop" shopping trips at their local retailer.

A graphical representation of the spatial setup taken directly from (Eckel, 2009) is found in (Figure 1.2). Consumers and retailers interact in a spatial competition setup a la Salop (1979). Manufacturing is centrally located, as in the monocentric city model in Alonso (1964). Each manufacturer produces a single good. Firms and consumers are distributed uniformly on the Salop circle, whose center is the "manufacturing hub". Retailers, free to locate anywhere on the circle, simultaneously decide on

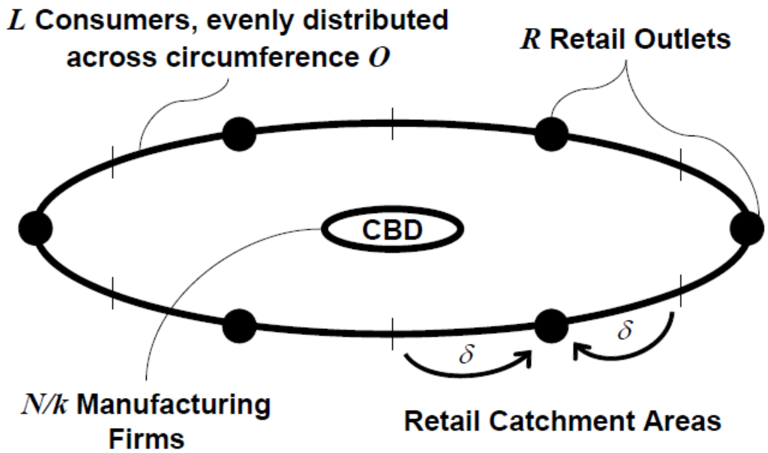


Figure 1.2: Retail Equilibrium on the Salop Circle

entry, mark-ups μ over manufacturer's wholesale prices , and the degree of product variety. They compete with each other for "catchment areas", which are given by the region between a retailer and the farthest consumer who just prefers that retailer over the next closest retailer in the opposing direction. Each consumer provides 1 unit of labor inelastically to the retail and manufacturing sectors.

To connect Internet use and travel costs in this setup, I take a change in Internet use intensity to be a change in consumer travel costs. A decrease in absolute and marginal travel costs (an increase in consumer mobility) in this setup shifts the manufacturing zero profit line in a way to increase the equilibrium number of manufacturers, and shifts retailing zero profit to reduce the number of retailers. The relatively more mobile consumers have higher price elasticities of demand, leading to lower mark-ups at retail. This leads to consolidation until the decrease in margins is met by increased catchment areas. Welfare effects of the change in mobility are consequently unambiguous: since mark-ups fall and product variety increases, welfare rises.

This direct relationship between consumer travel costs and retail market markups is not unique to the Eckel model. Another influential model of spatial competition and the implications of transportation costs to market structure follows from Vogel

(2011). The focus in that investigation is characterizing an asymmetric equilibrium in a Hotelling model, and the implications of changes in transportation costs are the same. As consumers become more mobile, they are more easily able to substitute away from high cost firms, placing these high cost retailers at a disadvantage. This notion reinforces the relationship between transportation cost and margins in retailing in geographic trade models: price conscious, highly mobile consumers are the concentrating force in the retail industry. As the consumer becomes more price savvy and increasingly patient, any individual retailer's market power falls.

1.2.2 General Equilibrium

Parameter descriptions are provided in (Table 1.1). I leave many of the details from (Eckel, 2009) in the appendix. The general equilibrium therein is characterized by simultaneous equilibrium in retail and manufacturing by way of a zero profit condition for each sector, as well as clearing of labor markets. The zero profit condition for manufacturing firms depends on the number of retailers through the retail markup, since demand for manufacturing goods is solely from retailers who provide the good at a markup to consumers. Substituting the optimal retail markup, $\mu = \tau\frac{\Omega}{R}$, into the zero profit condition for manufacturing allows a solution for the number of manufacturing firms gives

$$kL = \alpha(1 + \tau\frac{\Omega}{R})(\sigma N - \sigma + 1) \tag{1.1}$$

The next step is to solve the manufacturing zero profit condition (1.1) for N, the number of manufacturing firms as a function of the endogenous retail establishment count R and the exogenous parameters $k, \alpha, \sigma, \tau, \Omega$. Doing so gives an equation for the number of manufacturing firms that depends on the exogenous parameters and the endogenous number of retailers, an intermediate step in the process of finding a

Parameter Symbol	Description
k	# regions
Ω	circumference of circle
L	# representative agents
τ	adjusts consumer's absolute/marginal travel costs
σ	consumer's elasticity of substitution between varieties
N	# manufacturers
α	manufacturer fixed cost
β	manufacturer marginal cost
R	# retailers
γ	retail marginal cost of variety
p	retail price
p_w	price retailers face from manufactuers
μ	retail markup $p = (1+\mu)p_w$

Table 1.1: Parameter descriptions

fully reduced equation:

$$N(k, \alpha, \sigma, \tau, \Omega; R) = \frac{kLR + \alpha(\sigma - 1)(R + \tau\Omega)}{\alpha\sigma(R + \tau\Omega)} \tag{1.2}$$

To find the equilibrium retail establishment count, I substitute the right hand side of (1.2) for N in the retail zero profit condition derived on lines (3.13)-(3.15) in the appendix. Doing so gives an expression quadratic in R :

$$-\underbrace{\gamma(\alpha(\sigma-1)+kL)R^2}_{A*R^2} - \underbrace{\alpha\gamma(\sigma-1)\tau\Omega R}_{B*R} + \underbrace{\alpha L\sigma\tau\Omega}_{\text{C}} = 0 \tag{1.3}$$

It's a good practice to check the discriminant of quadratic expressions like this, in order to make statements about the properties of the solution.

$$\begin{aligned} B^2 - 4AC &= (-\alpha\gamma(\sigma-1)\tau\Omega)^2 - 4(-\gamma(\alpha(\sigma-1)+kL))(\alpha L\sigma\tau\Omega) \\ &= \underbrace{\alpha\gamma\tau\Omega}_{+}\left(\underbrace{4kL^2\sigma}_{+} + \underbrace{\alpha(\sigma-1)(\gamma(\sigma-1)\tau\Omega + 4L\sigma)}_{+}\right) \end{aligned} \tag{1.4}$$

Since (1.4) is strictly positive for permissible values of the parameters, there exists a single positive root corresponding to the number of retail establishments in equilibrium. Solving this quadratic expression yields the fully reduced equation governing retail market structure, R^*, in terms of exogenous parameters, after using the quadratic formula. Since retail establishment counts are non-negative integers, I discard the negative root of the equilibrium establishment count, arriving at (3.1):

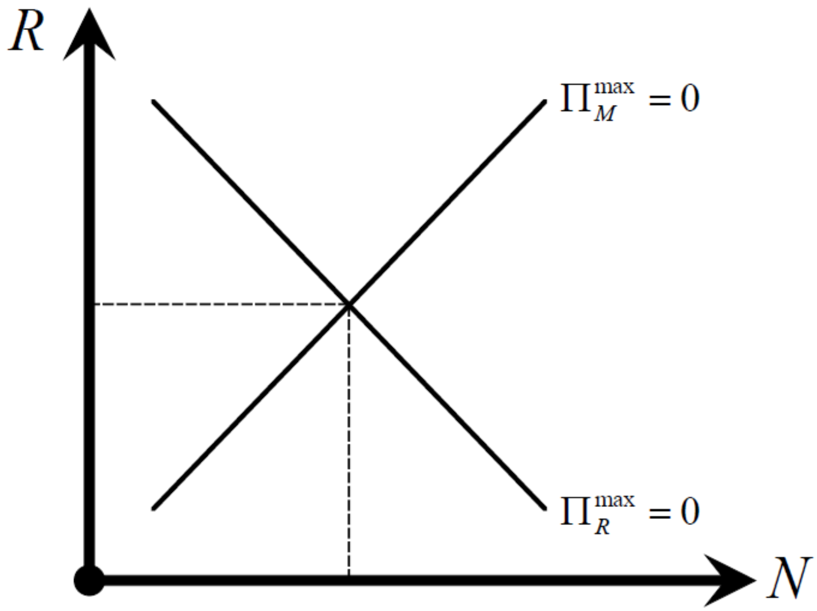
$$R^* = \frac{-\alpha\gamma(\sigma-1)\tau\Omega + \sqrt{\alpha\gamma\tau\Omega(4kL^2\sigma + \alpha(\sigma-1)(4L\sigma + \gamma(\sigma-1)\tau\Omega))}}{2\gamma(kL + \alpha(\sigma-1))} \tag{1.5}$$

(Figure 1.3) gives a graphical representation of the general equilibrium. Their intersection pins down an equilibrium number of retailers R^* and manufacturers N^*. Comparative statics of this general equilibrium with respect to both consumer travel costs and to population growth are provided in the Appendix. Verification of (3.1) is available using the snippets of Mathematica code provided in the appendix.

Motivating an Estimating Equation

To make a retail market structure measure like (3.1) tractable for an empirical investigation, it is advantageous to take an asymptotic approach similar to the retail model in Ellickson (2006). In this approach, the total revenue of each firm is evaluated in the limit by allowing an isoelastic demand parameter to make market revenue

Figure 1.3: Symmetric General Equilibrium



grow without bound. To replicate the asymptotic approach with a different class of models, I let $L \to \infty$ and examine the market structure equation in this limit. This asymptotic result is a variant where the Salop circle is extremely (infinitely) densely populated. Allowing the circle to become densely populated is the only way of allowing total market revenue in the Salop circle to grow without bound while preserving a finite price. Evaluating the market structure equation (3.1) in this limit, I arrive at (1.6):

$$\begin{aligned} \textit{Let } \bar{R} &= \lim_{L\to\infty} R^* \\ &= \frac{\sqrt{\alpha\sigma\tau\Omega}}{\sqrt{k\gamma}} \\ &= \alpha^{0.5}\sigma^{0.5}\tau^{0.5}\Omega^{0.5}k^{-0.5}\gamma^{0.5} \end{aligned} \tag{1.6}$$

Letting tildes represent natural log transformations, $\widetilde{R}^* = \ln(\bar{R})$, I arrive at (1.7), the asymptotic establishment count and bound to market fragmentation:

$$\widetilde{R}^*_{ist} = \frac{1}{2}\left(\underbrace{\widetilde{\Omega} - \widetilde{k}}_{\text{Time invariant}} + \underbrace{\widetilde{\alpha_{is}}}_{\text{industry-state fixed effect}} + \underbrace{\widetilde{\tau_{st}}}_{\text{proxy with Internet measures}} - \underbrace{\widetilde{\gamma_{ist}}}_{\text{payroll data}} + \underbrace{\widetilde{\sigma_{ist}}}_{\text{Unobserved}}\right) \tag{1.7}$$

From left to right in (1.7), the bound to market fragmentation in the asymptotic limit is governed by factors that are invariant over time and common to all industry-state pairs such as Ω, the circumference of the "world" in the model, and k the state/region count. Other retail market structure determinants are assumed to be time invariant for a given industry-state pair in a short run investigation, and therefore are captured either by $industry - state$ fixed effects or by industry fixed effects if further assumptions regarding cross-state heterogeneity are made. One example of this is α_{is} which represents the fixed cost of production for manufacturers in human capital terms. If these fixed costs are assumed to be common to an industry, which is the same as saying they are symmetric across states, then the effects will be captured in an empirical specification with industry fixed effects. There are market structure determinants that vary over time and are common to all industries in a state, the only example in the asymptotic result being τ_{st} which represents all exogenous influences on consumer travel costs. Labor supply, L_{st} in the unrestricted retail establishment count equation (3.1) is another example. I will capture variation in travel costs using state level extensive and intensive measures of internet use over time, discussed further in the empirical investigation section. Finally, there are idiosyncratic determinants that vary over time for each $industry - state$ pair. Two examples include σ_{ist}, the elasticity of substitution between varieties assumed to be greater than 1, and γ_{ist}, the marginal cost of the labor input to retail production.

Proposition 1. *The marginal effects of any retail market structure determinant vary at the industry-state-year level in the non-asymptotic model of retail market structure* *in (3.1). This differs from the asymptotic equation governing retail market structure in (1.7), which has constant and symmetric elasticities due to the log-log specification.*

Proof: One way of showing this is by taking the partial derivative of (3.1) with respect to any market structure determinant. Using L as a motivating example, this partial derivative would be interpreted as the marginal change in the number of retail establishments for a small change in the population (or market size), holding all other variables constant. Let ϕ_{ist} represent this partial derivative, reproduced below.

$$\begin{aligned}
\phi_{ist} &= \frac{\partial R^*}{\partial L} \\
&= [(\alpha_{is}(\sigma_{ist}-1)\tau_{st}\Omega) \\
&* \big(2\alpha_{is}^2(\sigma_{ist}-1)\sigma_{ist} + k\,((\alpha_{is}\gamma_{ist}\tau_{st}\Omega)(1-\sigma_{ist}) \\
&+ \sqrt{\alpha_{is}\gamma_{ist}\tau_{st}\Omega\,(4kL^2\sigma_{ist} + \alpha_{is}(\sigma_{ist}-1)(\gamma_{ist}(\sigma_{ist}-1)\tau_{st}\Omega + 4L\sigma_{ist}))} + 2\alpha_{is}L\sigma_{ist})) \\
&* \left[\frac{1}{2(\alpha_{is}(\sigma_{ist}-1)+kL)^2\sqrt{\alpha_{is}\gamma_{ist}\tau_{st}\Omega\,(4kL^2\sigma_{ist} + \alpha_{is}(\sigma_{ist}-1)(\gamma_{ist}(\sigma_{ist}-1)\tau_{st}\Omega + 4L\sigma_{ist}))}}\right]
\end{aligned} \tag{1.8}$$

Notice that the partial derivative (1.9) depends on variables that vary at the industry, state and year levels. This contrasts with the elasticities resulting from the asymptotic model, which are found by taking the partial derivative of (1.7) with respect to a variable of interest.

Proposition 2. *In the asymptotic approach as $L \to \infty$, the elasticity of retail establishment counts with respect to retail market structure determinants are symmetric and constant.*

Proof: Let $\widetilde{R^*}$ represent the natural log of retail establishments, and $\widetilde{x}$ represent the natural log of determinant x. Take the partial derivative of (1.7) with respect to any of the exogenous determinants gives either $\pm\frac{1}{2}$.

$$\frac{\partial \widetilde{R^*}}{\partial \widetilde{\tau}} = \frac{1}{2} * 1 = \frac{1}{2} \tag{1.9}$$

$$\frac{\partial \widetilde{R^*}}{\partial \widetilde{\gamma}} = -\frac{1}{2} * 1 = -\frac{1}{2} \tag{1.10}$$

which are constant and symmetric across industries.

From an econometric standpoint, the asymptotic result reduces the non-linear equation governing retail establishment counts in (3.1) to a log separable candidate specification. Distinguishing which model, asymptotic or unrestricted, best describes the nature of competition between retail establishments can occur one of many ways. Foremost, the asymptotic specification suggests that the elasticity of establishment counts with respect to changes in the market structure determinants are constant and symmetric across industries. If you can entertain the idea that it is economically reasonable for industries to be differently sensitive to changes in establishment count determinants, then you could reject the asymptotic result in favor of the unrestricted result by finding evidence that effects vary across the retail industries represented in the data.

Natural Oligopoly Result

The fact that $\widetilde{R^*}$ is finite connects Eckel (2009) to Ellickson (2006) and Shaked and Sutton (1983). In the former, the asymptotic number of retail firms, determined by allowing the total revenue of retail firms to approach infinity, is also a finite number. This result is the "natural oligopoly" outcome as described in Shaked and Sutton (1983). My extension of the (Eckel, 2009) model in preceding discussion preserves this bound on market structure in a model with constant returns in the retail industry. This is not immediately intuitive, since the standard way of generating a bound to fragmentation is through introducing retailers with increasing returns to scale. These increasing returns can be generated for retail via endogenous fixed (sunk) costs as in

Sutton (1991), or in a simple manner by adding a fixed labor cost like those facing manufacturers.

The asymptotic approach of examining the retail market structure as labor grows without bound is not immediately intuitive. In a symmetric approach, the increase in L represents both a large population density in the home region as well as the "foreign" regions that populate the Salop circle. In this sense, it is similar to the market supply and demand conditions under perfect competition, since the number of consumers in the world grows without bound. However, it differs from perfect competition on account o the fragmentation bound allowing retailers to maintain a positive price-cost margin.

Proposition 3. *There exists a bound to retail market fragmentation in this model with constant returns to scale in retail. with the bound being similar to the market concentration lower bound as discussed in Sutton (1991). As markets grow (in revenue terms), the number of retailers converges to a finite number as opposed to also growing without bound.*

Proof: There are two different ways of demonstrating this, one way is by taking the limit of (3.20) with respect to L, which evaluates to zero after application of L'Hospital's rule. This suggests that in the limit the market structure in retail is unresponsive to population growth, suggesting that all increased growth is absorbed by incumbent establishments. This means the industry adjusts to internal growth along an extensive margin like employment. A second way of demonstrating this bound is showing the market structure equation itself has limiting behavior that does not depend on L. This will be demonstrated in the section below in order to motivate an estimating equation.

Why does this type of bound matter in practice? A bound on market concentration suggests that there are "critical points" in the growth of the region beyond which all of the growth in demand is absorbed by incumbent firms. This means that firms are adjusting along an extensive margin like employment, as is the typical story in a short run investigation. If an econometric investigation is conducting a short-run investigation, as might be done with panel data using large N and small T asymptotics, this means that it is normal for there to be industries with insignificant population effect estimates, since the retail establishments may be adjusting to the growth with increased demand for labor as opposed to capital investment.

1.3 Empirical Investigation

1.3.1 Data

Table 1.2: Summary statistics for the 4 digit NAICS retail industries.

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**
Establishments	784.7208	1165.2038	2	12251
Establishments per capita	0.0001	0.0001	0	0.0008
Internet Access	62.0516	53.8532	0	769
Internet Use Intensity	51.7103	13.7055	24.4	85.6200
N		6885		

Due to data availability constraints, the time period of the investigation is from 2008 until 2012. For this period, I construct a panel of all the 4 digit retail industries represented in the North American Industry Classification System (NAICS hereafter) across the 50 United States. For each panel, I collected the number of establishments, employment count, and payroll within the industry-state pair in all represented NAICS categories from the U.S. Census Statistics of U.S. Business. From the descriptive statistics in (Table 3.1), a minimal value of 2 establishments in a 4 digit industry state pair come from "Other motor vehicle dealers" in the District of Columbia. The retail industry group with the largest establishment count is the "Clothing Store" industry in California. The extreme value for retail establishments per capita in this category implies that the most fragmented industry, i.e. that with the most number of establishments per citizen, has approximately 1 establishment for every 1250 people.

Dependent Variables

I use establishment counts as an absolute market structure measure, given by the count of establishments in a given industry-state pair through time. Additionally, I construct an "adjusted" market structure measure given by the establishment count per capita $\frac{Establishments_{ist}}{Population_{st}}$. Examples of both measures are provided in (Figure 1.4) on p. 20, which plots the within industry-state variation in these measures for the period of interest. Notice in (Figure 1.4) there is substantial within industry variation in both the absolute and adjusted market structure measures. Note the two types of variation in this measure: there is across state variation, represented by how each state's time series in (Figure 1.4) has a different intercept, and within industry heterogeneity, reflected by the variation in the establishment counts over time. The "within" estimator I plan to use will discard the cross-sectional variation in favor of the variation over time in order to identify the effects of varying Internet use intensity.

Additionally, I will explain the determinants of variation in employment counts both across and within disaggregated retail industries. Observations of employment in each industry-state-year triplet come from the Statistics of U.S. Business. Examples of the variation in employment is plotted in (Figure 1.5) for NAICS 4512 "Books, Periodicals, and Music Stores". Notice that (Figure 1.5) also presents two types of variation, across state variation s, and variation over time within each industry-state pair. A fixed effects approach will only use the variation within each industry over time to estimate the marginal effect of any variable of interest.

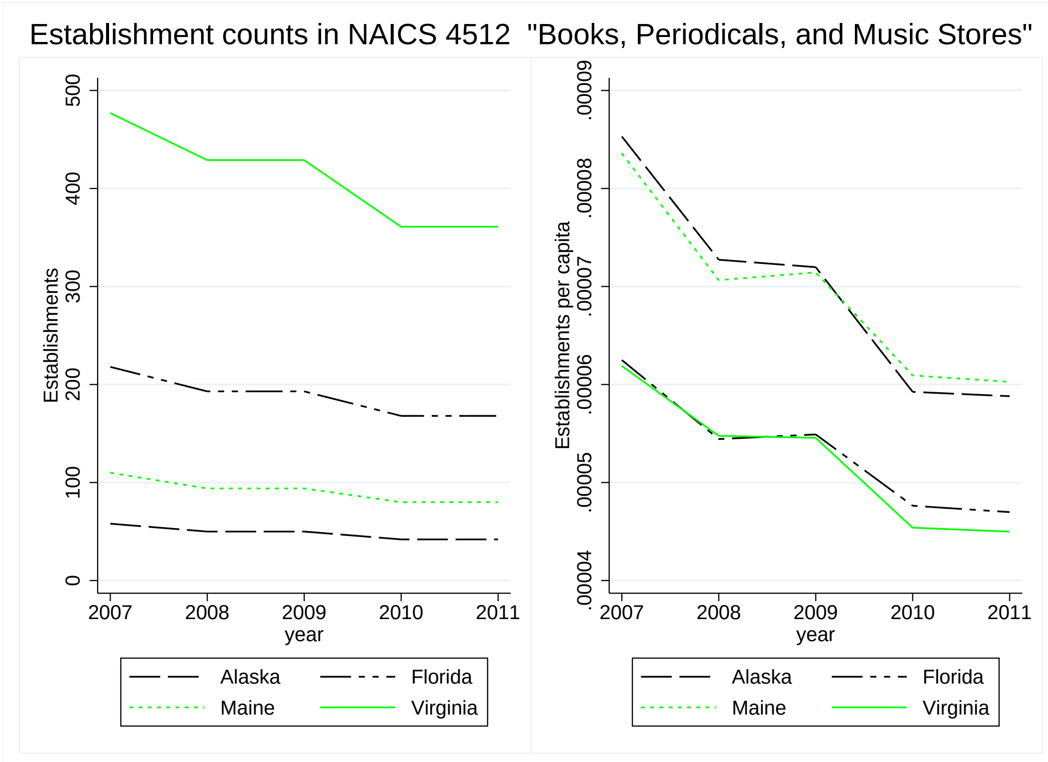


Figure 1.4: Variation in retail market structure measures over time.

Independent Variables

The effect of interest is how variation in Internet use intensity influences retail establishment counts. I propose measuring Internet use along two dimensions, one reflecting the extensiveness of Internet connectivity, another representing the intensity with which consumers use the Internet. All previous research in the area use less reliable survey data, which provide binary indicators of purchasing a good or service online. For an extensive Internet use measure within a state (labeled "Internet Access Rates" in the summary statistics), I combined the "Computer and Internet Use Survey" from the Census with data from the National Telecommunications & Information Administration (NTIA). Both of these sources provide state level data for the percentage of households who report at least one individual using the Internet from home. Finding state level measures for this variable for the year 2008 is proves

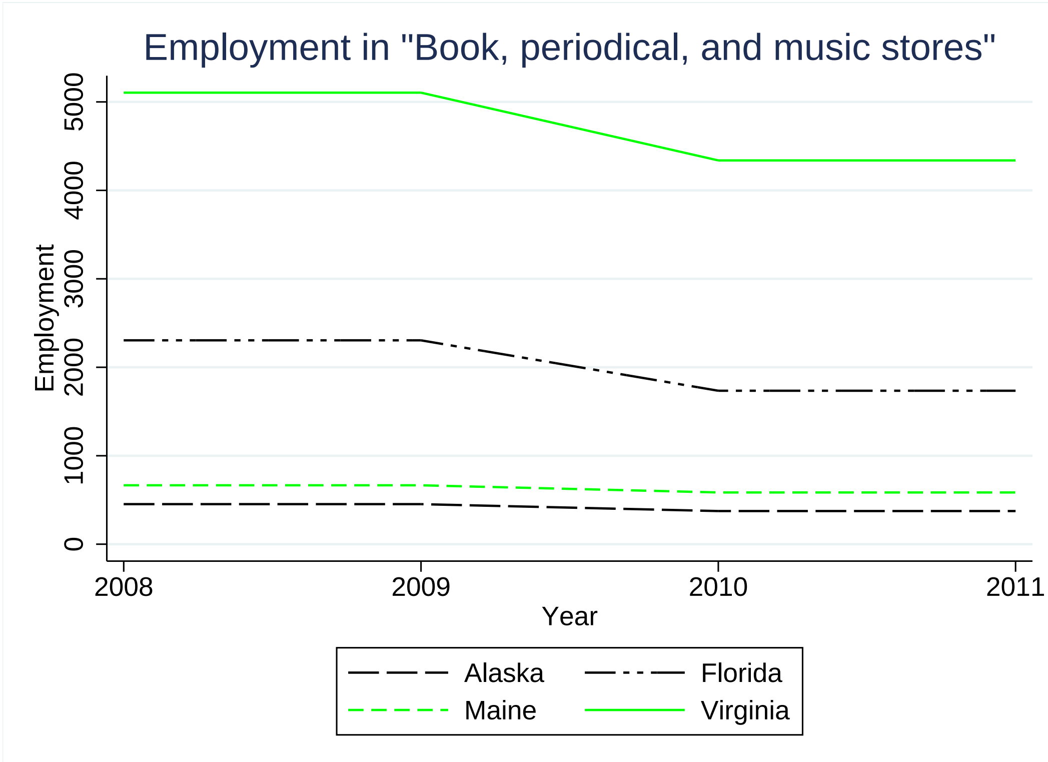


Figure 1.5: Variation in retail employment counts over time.

difficult. Until remedied, I assume that the observations of this covariate for the year 2008 are approximated by the average of the values in given in 2009 and 2007.

Inspecting the histogram of Internet access across the states (Figure 3.9), there is evidence for a high degree of heterogeneity in access rates across states as well as substantial temporal variation in these access rates. These differences reflect heterogeneous telecommunications structures, varying Internet service provider market conditions, and asymmetric preferences for Internet access across states. Mississippi drags behind the rest of the states in this dimension, with the lowest access rates at 60.9%, far behind New Hampshire's lead at $\approx 87\%$. Extensive measures alone do not capture the dimensions along which the Internet can influence retail market structure. Consider two hypothetical adjacent states, both of which have a population of 1 Internet user with identical preferences with one exception: one of the two is not savvy enough to use the Internet for shopping. Both would have the highest measure of Internet access rates (100%), yet would differ in how exposed their retail industries are to e-commerce competition. Omitting propensity measures will bias estimates of the marginal effect of Internet access upwards, since it is assumed the propensity to use Internet is highly positively correlated with the access rates measure.

One can think of many reasons there is observed differences in this propensity within and across states. Some geographical areas may have large groups of agents who strongly prefer locally sourced and assembled goods more than supporting "foreign" products, and similar spatial clustering of preferences. Anecdotally, anyone with older parents can attest to some sort of struggle or resistance regarding learning to use the Internet. Clearly, states with relatively left skewed age distributions will have effectively less intensity of Internet use, and therefore lower propensity to shop online due a lack of familiarity with the procedures involved. I assume that the age distributions of states is fixed in a short run investigation of this sort, and is therefore captured by the fixed effects.

A suitable Internet propensity measure is uncorrelated with retail market structure determinants while managing to capture variation in propensity to use the Internet. Ideally, the measure would have negligible correlation with determinants of retail market structure in order to mitigate both proxy variable bias and potential endogeneity from simultaneous/reverse causality. Additionally, the propensity measure should have the goal of capturing the largest cross section of Internet users available. With both of these caveats in mind, I use Google search frequency for the keyword "pornography" at the state-year level meets these needs adequately since it captures a vast cross-section of the population. I assume that anyone familiar enough with the Internet for this explicit purpose is identically willing to shop online. The reader may be concerned that this particular keyword introduces bias on account of 72%

of visitors to pornography sites being male[2]. While the keyword may favor male Internet activity, it is fair to assume that females in the state are just as able and willing to use the Internet as their male counterparts. The alternative, that they are any less propense to use the Internet, seems like an equally unlikely outcome. With this assumption it is sufficient to adopt a measure that may place more weight on the male population, since it reflects the propensity of both sexes to "hop online" to accomplish tasks.

A second point of my investigation is to confirm that different industries should be differently sensitive to varying Internet use intensity. There are industries that should largely be insulated from e-commerce competition, primarily those for which online shopping is a poor substitute or for which an online substitute fails to exist for legal and practical reasons. For example, consider any effects of variation in an intensive Internet use measure on the number of gasoline stations. Such an industry should be perfectly isolated from online retail competition since there are no substitutes available through online retailers, and furthermore it is illegal to ship flammable liquids of this sort. Another example of a retail industry where domestic shipping regulations preclude online is the "intoxicating liquors" industry, defined to be beverages with 0.5% or more alcohol by volume, associated with the 4452XX industries. In contrast to these examples, consider how the availability of news and online content has shaped consumer purchasing habits in the periodicals segment. This industry is in face-to-face competition with essentially homogeneous products provided online that are easily accessed via smart phone for free, so it should be that this industry is highly sensitive to variation in Internet propensity within a state. A later extension will investigate this heterogeneity in detail, characterizing it in a more systematic manner than ad hoc examples industry by industry.

[2]See http://www.familysafemedia.com/pornography_statistics.html for this statistic and similar

I use SUSB data to construct $wage_{is}$ to proxy the parameter γ, which represented the marginal labor cost of added variety. The constructed value is given by dividing annual payroll in is by the employment in is. By assuming the asymptotic form of R^* I implicitly am saying that labor supply is extremely large. Since all markets clear in an equilibrium, this would imply that labor demand is also large enough to meet supply, which would approximate perfectly competitive labor market conditions. For empirical purposes, it is assumed that labor markets are competitive and therefore that wages identify marginal labor costs. This competitive labor market assumption is a very realistic description of the retail industry as a whole, given that the average job posting in the sector (outside of upper and middle management roles) tend to have low education requirements[3]. Retail labor markets are characterized by an abundance in supply due to the relatively lax qualifications required. If the study was concerned with occupations outside of retail, take academia as an example, there may be concern regarding negotiating power being reflected in wages. Retail is a broad and relatively low-skill sector (reflected empirically in its low wages) to such a degree that this assumption makes sense in the context of both retail labor markets and retail as an industry. Since this is a short run investigation, I believe this assumption is reasonable. In a long run investigation where the brick and mortar structure itself is a variable input it would be possible to have capital be an adjustment margin for added variety, though that is not in the interest of this investigation.

1.3.2 Model Specifications and Selection

In this section I will pit the structurally motivated log-log specification of market structure against reduced form linear-linear and non-linear specifications in order to determine which empirical model best approximates the "true" process governing

[3]BLS, http://www.bls.gov/ooh/sales/retail-sales-workers.htm

retail market structure. I take two approaches regarding selection, first using information criteria based metrics like those in (Akaike, 1973), (Akaike, 1981) or (Takeuchi, 1976) when possible for model comparison. A naive approach would be to compare the AIC of log-log specifications with the linear-linear and non-linear specifications, which is not a best practice according to (Burnham and Anderson, 2002). For model comparisons based on an information criteria, the response variables in the candidate set must be measured in the same way. Further, since order of comparison does not matter, I begin by determining the best of the log-log specifications using information criteria. Then I determine the best among specifications with dependent variables entering in levels (as opposed to log). The selection strategy is summarized graphically in (Figure 3.12). I rely on a few strategies to compare models where the responses are different measurements of the same variable and to supplement the formal information criteria based approach. I perform a sequence of 10 fold cross validation trials to compare out of sample of sample prediction accuracy.

Control Function Approach

In this section, I focus purely on the relationship between e-commerce and brick and mortar retail. My strategy is to use relative search frequency for “amazon.com” at Google within the state as a determinant of brick and mortar establishment and employment counts to proxy the intensity with which firms compete with e-commerce substitutes. It’s likely the case that consumer willingness to substitute away from local retail towards e-commerce is in part determined by factors which also act as determinants of retail market structure. One example would be how there is potential for small markets to be under-represented by retail variety, therefore making e-commerce a popular option out of necessity and love for variety.

Recognizing this as a potential endogeneity problem, I propose the following control function approach and identification strategy to determining the relationship between e-commerce and brick and mortar retail. I use within state variation in relative search frequency for "porn" at Google as an instrument for within state variation in the search frequency for "amazon.com". Search frequency for "porn" certainly meets the relevancy criteria of an ideal instrument since it measures the innate willingness to hop online to accomplish a specific task. I also argue that the instrument is excludable, since preference for pornography is likely a function of consumer characteristics like religious sentiments and personal tastes, and is unlikely to share any time varying unobserved components with retail market structure. The first stage regression appears in (1.11).

$$amazon\ intensity_{st} = \pi_1 porn\ intensity_{st} + X_{st}\Pi + v_{ist} \tag{1.11}$$

In (1.11) X_{st} is a (1×4) vector that includes the control variables for Internet access rates, market size proxy, and wages along with a constant. Π is a (4×1) vector of parameters to be estimated. Estimation results from from (1.11) appear in (Table 3.5) in the appendix. The F-statistic from estimating (1.11) is ≈ 153, and each instrument has a t-statistic much larger than 3, suggesting that the instruments are not weak. I construct the residuals from (1.11) to create the control variable $\hat{v_{ist}}$. The fixed effect specification in (1.18) is then augmented with the control variable $\hat{v_{ist}}$ from estimation of (1.11) to form the control function:

$$\begin{aligned} R^*_{ist} = & \underbrace{\beta_i^{FEIV}\ amazon\ intensity_{st}}_{\text{exposure effect * e-commerce intensity}} + \underbrace{\xi \hat{v}_{ist}}_{\text{1st stage control var}} \\ & + \underbrace{X_{ist}\Gamma_i}_{\text{exog. contol vars}} + \underbrace{(\Theta + \alpha_{is} + \lambda_{it} + \theta_t)}_{\text{fixed effects}} \\ & + \underbrace{u_{ist}}_{\text{idiosyncratic component}} \end{aligned} \tag{1.12}$$

Equation (1.12) is estimated with a fixed effect estimator to mitigate the effects of time invariant confounding variables within each state. The fixed effects in (1.12) effectively demean the observation first by the panel mean, then by the year mean, and last by the industry-year mean. As such, any confounding effects must be varying at the state or industry level through time. Econometrically, this approach is outlined in (Matyas and Balzsi, 2013), (Baltagi et al., 2003), and (Baier and Bergstrand, 2007). The necessary identification assumptions are that the idiosyncratic errors u_{ist} from (1.12) and the error in the reduced form in (1.11), v_{ist}, are orthogonal to the exogenous controls in X_{ist} as well as orthogonal to the instrument $porn\ intensity_{st}$. The estimate of ξ provides a way of testing for the endogeneity of e-commerce search intensity via simple hypothesis tests.

Linear Models

For parsimony, I begin by estimating a specification similar to the asymptotic retail market structure equation in (1.7). This equation suggests a regression of the natural log of establishment counts on the natural log of various market structure determinants, with constant marginal coefficients. For parsimony I start with a model of just the time varying covariates without any fixed effects.

$$\widetilde{R^*_{ist}} = \widetilde{\Theta} + \beta net\ \widetilde{intensity}_{st} + \widetilde{X_{ist}}\Gamma + u_{ist} \tag{1.13}$$

In (1.13), $\widetilde{R^*_{ist}}$ represents the log of the retail establishment count in industry-state is for the year t. Since both the dependent and independent variables are in log form, coefficient estimates are interpreted as elasticities. The effect of interest is β, the elasticity of establishment counts with respect to the internet use intensity measure $net\ intensity_{st}$, holding all other measured variables like access rates and population constant. $X_{ist} = (wage_{ist}, access_{st}, population_{st})$ represents a 1×3 vector of control variables and Γ_{st} the associated 3×1 vector of marginal effects motivated from (1.7). The idiosyncratic "shock" u_{ist} represents represents the influence of all time varying determinants of market structure that are not included as independent variables. Further, since I have a dependent variable at the industry-state-year dimension, and several independent variables only varying at the state-year dimension on the right hand side, I bootstrap standard errors with clustering at the state level to address the concerns in (Bertrand et al., 2004).

I introduce a new candidate model of market structure by generalizing (1.13) to have slopes that are no longer fixed and identical across industries. The literature refers to specifications of this sort by many names, including but not limited to multilevel, hierarchical, random coefficient, and random slopes models. Adding this feature allows for industry specific marginal effects, so that the Internet exposure effect for gasoline stations need not be identical to the exposure effect for book stores. In enriching the specification in this manner, I relax the assumption that marginal effects are symmetric across industries, though symmetric marginal effects remains nested as a possibility that can be formally tested via formal hypothesis tests. I add the following model to the candidate set for the log-log specification and specifications not yet introduced.

$$\widetilde{R^*_{ist}} = \widetilde{\Theta} + \beta_i net\ \widetilde{intensity}_{st} + \widetilde{X_{ist}}\Gamma_i + u_{ist} \tag{1.14}$$

The multilevel specification is nearly identical to that of (1.13), with the exception that the marginal effects are now indexed by industry i, reflecting the multilevel characteristic of this specification. Symmetric marginal effects in (1.13) remain a nested possibility in (1.14) that can be formally tested as the restriction $\beta_1 = \beta_2 = \cdots = \beta_N$. A scatter of the fitted values versus the true establishment counts is provided in (Figure 1.7). Notice that the residuals, as represented by the distance

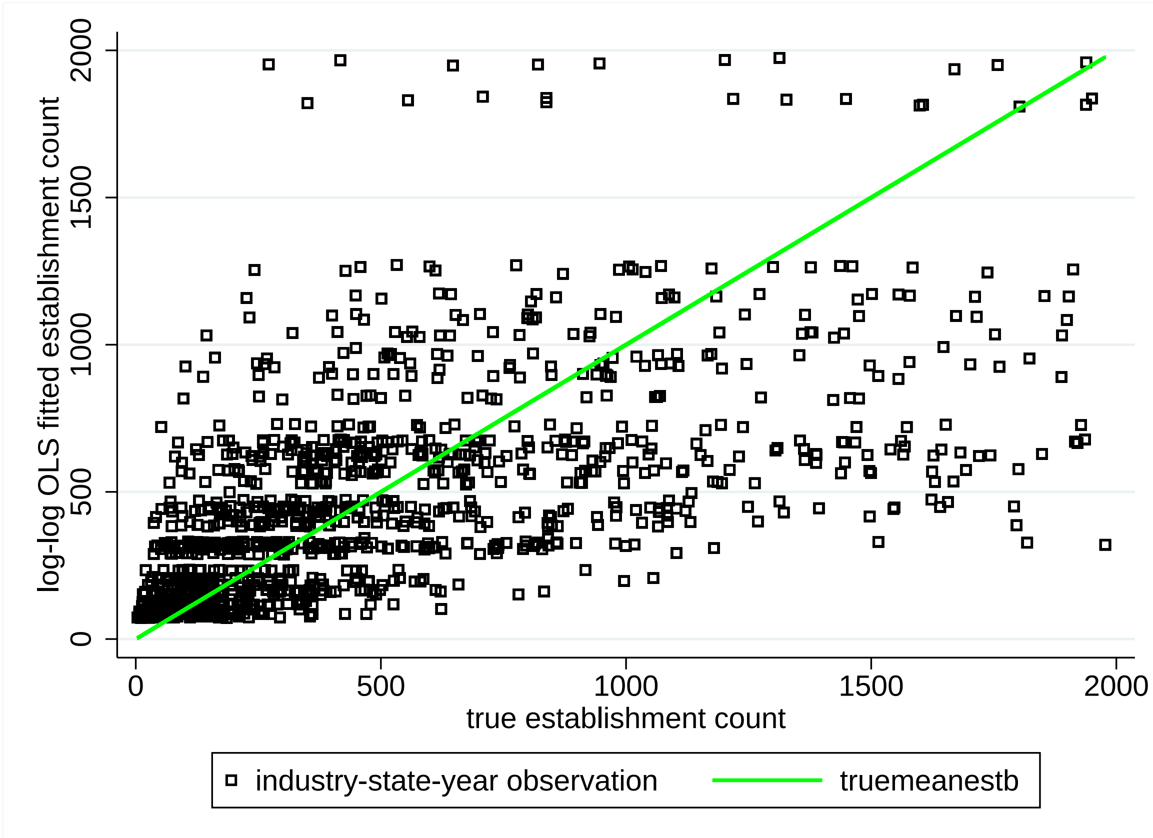


Figure 1.6: OLS log-log predicted versus observed scatter plot

between each point and the 45* line, have a relatively smaller variance than the simple model in (1.13). This is expected on account of the fact that the simple model in (1.13) presents 5 parameters to be estimated, while the multilevel approach raises this to $(4 \; independent \; variables \; \times 27 \; industries) + 1 \; constant = 109$ parameters to be estimated, so naturally the model is able to fit the data with less error.

Of the 2 models compared I compare to a classical OLS specification, the largest increase of explained variation comes from introducing multilevel effects. This can be seen by the ≈ 48 percentage point increase in R^2 in moving from column 1 to column 2 (Table 1.4).

A more credible approach for identification of the marginal effects would be to use variation within each industry-state pair over time as opposed to explaining variation across the industry-state pairs. Doing this eliminates any time invariant confounding factors, so any omitted variable bias must be from a source that varies over time. Introducing fixed effects to build a within estimator will greatly increase the number of parameters being estimated, though the strategy is identical to the time demeaning procedure discussed in (Mundlak, 1978). This added number of parameters will work against any model in an information criteria comparison due to the penalty mechanism, however the subsequent increase in the (log) likelihood of the model ought to make up for the reduction in degrees of freedom. This leads me to adopt a fixed effect specification where each panel is an industry-state pair, and is augmented with industry, year, state, and industry-year fixed effects in addition to the panel fixed effect

$$\widetilde{R^*_{ist}} = \left(\widetilde{\Theta} + \alpha_{is} + \lambda_{it} + \theta_t\right) + \beta_i net\ \widetilde{intensity}_{st} + \widetilde{X_{ist}}\gamma_i + u_{ist} \tag{1.15}$$

A discussion of the appropriate differencing achieved with a model like (1.15) is discussed in (Matyas and Balzsi, 2013). The presence of bilateral fixed effects stems from the empirical trade literature, for instance in (Baier and Bergstrand, 2007). The first fixed effect in (1.15), α_{is} is the individual effect for each panel, which absorbs the average differences in time invariant unobserved and observed across across the industry-state panels. A collection of dummy variables of this sort effectively does the within panel transformation. This makes the empirical strategy such that any confounding effects must be varying through time. To address this, I include all combination of fixed effects that do not coincide with the dimension of my dependent variables.

One of these is the year fixed effects θ_t accounts for macroeconomic factors common to all industry state pairs in a given year, like that of generic decreases in aggregate demand for retail products,or nationwide trends in Internet diffusion. In addition to year fixed effects are the bilateral industry-year fixed effects λ_{it}, which captures

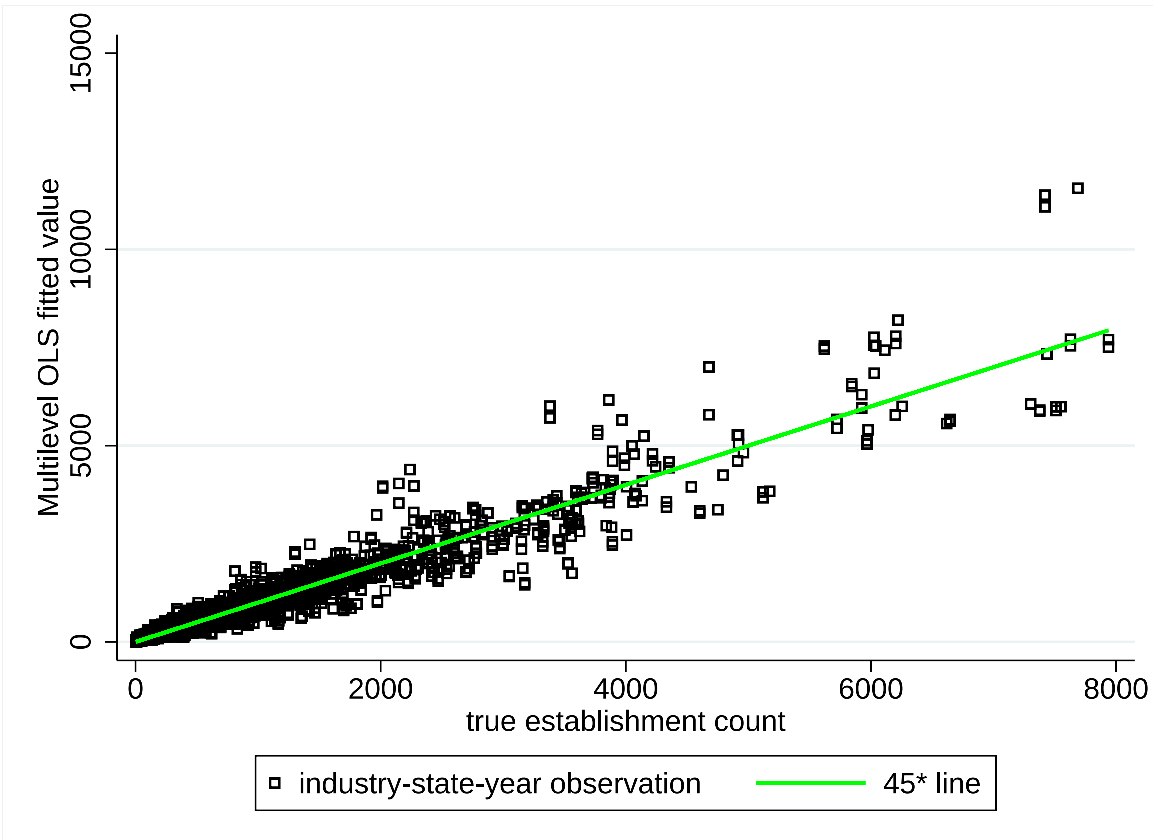


Figure 1.7: OLS log-log multilevel predicted versus fit scatter

issues that are specific to each industry across states in each period. For instance, these fixed effects may capture industry technological progress if we believe that retail R&D shocks are common to all locations where the industry operates. In the trade literature, a parameter of this sort is assumed to capture the business cycle effects. Error terms in the specification are given by u_{ist} and include the influence of factors such as the consumer's inter-variety substitution habits or other unobserved market structure determinants which are idiosyncratic to both the industry-state and year and assumed to be uncorrelated with the population, wages, Internet access, and Internet use intensity measures.

Notice that in each fixed effects model, the slope coefficient is indexed by i, reflecting the multilevel characteristic of the panel models. I accomplish these industry specific marginal effects by interacting the respective continuous variable with a set of industry indicator variables. A specification of this sort is a form of generalized linear models, as such OLS can be viewed as a restriction of the model such that each industry has a symmetric response to changes in these variables. Since the data has a large number of observations, I have the freedom to relax the symmetry restriction and adopt a specification that allows for coefficient heterogeneity. By estimating industry specific coefficients, I allow for the data to provide evidence of a symmetric response by industries to changes in market structure determinants.

Model selection metrics for all of the log-log specifications are provided in (Table 1.3) on page 33. Provided in this table are the Akaike information criteria, in addition to an adjusted R^2 measure. The table presents these measure from left to right in order of increasing number of parameters. Of course, as parameters are added each model will better fit the data, so to avoid overfitting I only include model selection metrics that penalize added parameters. From this table, the classic OLS specification without multilevel slope coefficients explains $\approx 58\%$ of the variation in establishment counts.

Model/Features	OLS log-log	OLS log-log	OLS log-log
Fixed Effects	no	no	is, t,it
Multilevel Slope Coefficients	no	industry specific	industry specific
# parameters	5	109	192
AIC	2.3	0.51	-2.97
Adj. R^2 (R^2)	0.582 (0.582)	0.936 (0.936)	0.998 (0.998)
10 fold cross validation avg. MSE	2635	2652	2618

Table 1.3: Selection metrics for log-log specifications.

Adding varying slope coefficients increases the fit of the model by explaining $\approx$ 93% of the variation in establishment counts, an improvement over the classical OLS specification of about 35 percentage points. Comparing this increase from constant to varying slopes to the increase from adding fixed effects, the majority of the increase in model fit is coming from relaxing the restrictive assumption of constant and identical marginal effects across industries.

The candidate model achieving both the lowest value of the Akaike information criteria, and the highest value of explained variation is the traditional panel fixed effects model with industry specific slope coefficients. This model is able to explain almost all of the variation in establishment counts, though this does not necessarily translate to ideal out of sample predictive power. Whether this ordering holds in out of sample prediction will be addressed using 10 fold cross validation.

I repeat this entire exercise using dependent and independent variables that enter the specification as levels, as opposed to the log-log specification. I repeat the model selection procedure with three candidate specifications: a classical OLS specification without varying slope coefficients, a specification with varying slope coefficients (multilevel), and a model with fixed effects and varying slope coefficients. Each is listed below, this time without the tildes to denote that the variables enter in levels.

$$R^*_{ist} = \Theta + \beta \; net\ intensity_{st} + X_{ist}\Gamma + u_{ist} \tag{1.16}$$

Model/Features	OLS linear-linear	OLS linear-linear	OLS linear-linear
Fixed Effects	no	no	is, t, it
Multilevel Slope Coefficients	no	industry specific	industry specific
# parameters	5	109	1467
AIC	16.3	14.3	10.8
Adj. R^2 (R^2)	.454 (0.454)	0.930 (0.931)	0.998 (0.998)
10 fold cross validation avg. MSE	2570	2554	1340

Table 1.4: linear-linear model selection metrics

$$R^*_{ist} = \Theta + \beta_i \; net \; intensity_{st} + X_{ist}\Gamma_i + u_{ist} \tag{1.17}$$

$$R^*_{ist} = (\Theta + \alpha_{is} + \lambda_{it} + \theta_t) + \beta_i \; net \; intensity_{st} + X_{ist}\Gamma_i + u_{ist} \tag{1.18}$$

A comparison of the fit of each of these linear models is provided in (Figure 1.8). In this figure, as you move from left to right and top to bottom the models are increasing in generality from the classical OLS specification in (1.16) to the multilevel panel specification in (1.18). Notice that the in sample errors improve greatly with each added generalization, which can be seen by the decrease in spread from the 45 degree line. Formal model selection metrics are provided in (Table 1.4). Though the panel fixed effects specification adds 14 times as many parameters to be estimated as the next simplest model, the added increase in explained variation makes the panel fixed effect specification an appealing candidate model.

Non-linear models

In this section, I take a reduced form approach of modeling retail establishment counts with count data models, since many of the SUSB indicators are inherently of this nature. The largest problem with using a traditional panel count data model with fixed effects, as in the methods in Hausman et al. (1984) or Cameron and Trivedi (2007) is the fact that fixed effects are are integrated out due to concerns regarding inconsistent estimation of the slope parameters. An alternative approach uses an

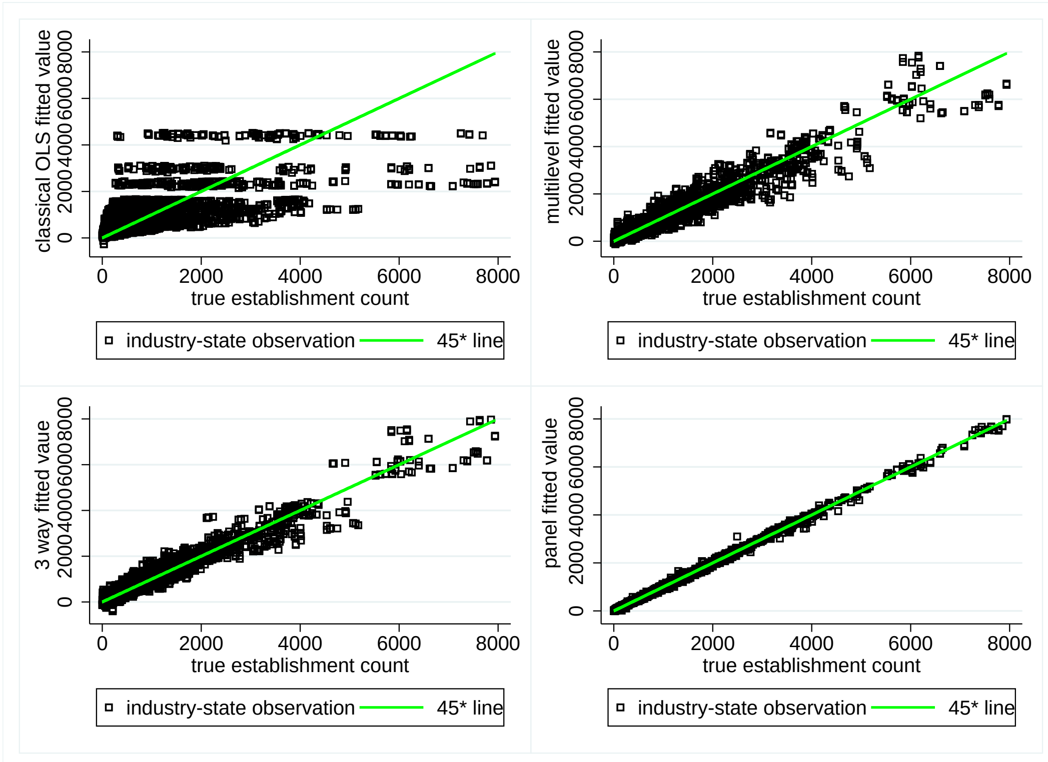


Figure 1.8: OLS linear-linear models predicted vs observed plot

unconditional strategy, as discussed in (Allison and Waterman, 2002), by estimating the fixed effects that appear in the conditional mean. Linear approaches are also not able to sufficiently capture the overdispersion that is present in the data. I use this as an opportunity to show that my results are not being driven by the linearity assumption.

An interesting pattern in the data is the presence varying degrees of dispersion in the retail establishment counts for a given industry-state pair, as presented in (Figure 1.9). Dispersion is typically defined as the ratio of the dependent variable's variance to it's mean within a group or panel. Notice that most of the extensive margins for industry-state pairs are characterized by overdispersion. In the context of this model, overdispersion refers to how the aforementioned variance seems to grow rapidly with the within panel mean.

Economically, overdispersion is a curious characteristic, since it implies that there are characteristics of larger retail markets that provide an unstable environment for retail firms to compete in, as if something churns in the shadows preventing a stable equilibrium. Graphical evidence of this overdispersion is apparent in consideration of (Figure 1.9). In this figure, notice that most of the observations fall above the $Mean_{is} = Variance_{is}$ line, which is a 45 degree line that appears distorted since the axes of this figure differ in scale. Overdispersion is a characteristic of the data on which to begin the model selection process when comparing non-nested models before doing more formal testing as in the methods discussed in (Vuong, 1989).

Consideration of overdispersion is important in selecting the proper econometric models of retail market structure, since ignoring the influence of dispersion means that the typical standard errors are not estimated correctly as explained in (McCullagh and Nelder, 1989). A basic understanding of OLS and GLS estimators suggests that they may not be able to replicate the overdispersion properties seen in the retail SUSB data. This is on account of the fact that in a typical ordinary least squares specification (for instance $y_i \sim N(x_i^T\beta, \sigma^2)$, the variance σ^2 is estimated in a manner independent of the mean function $x_i^T\beta$, and thus does not vary with the mean establishment count. In order to improve on this shortcoming, I make an appeal to the negative binomial regression technique, since in this specification the conditional mean enters directly in the functional form of the conditional variance, explicitly allowing for overdispersion.

Less formal sources claim that count data far from zero can be treated like a continuous random variable, and that the only drawback to giving it the formal count data treatment is "computational intensity". The ease of statistical software programming and speed with which it executes no longer make "computational intensity" a valid drawback of count data methods. By employing count data techniques, I hope to be better able to better approximate the true data generating process for retail establishment counts, a prominent feature of which includes the presence of overdis-

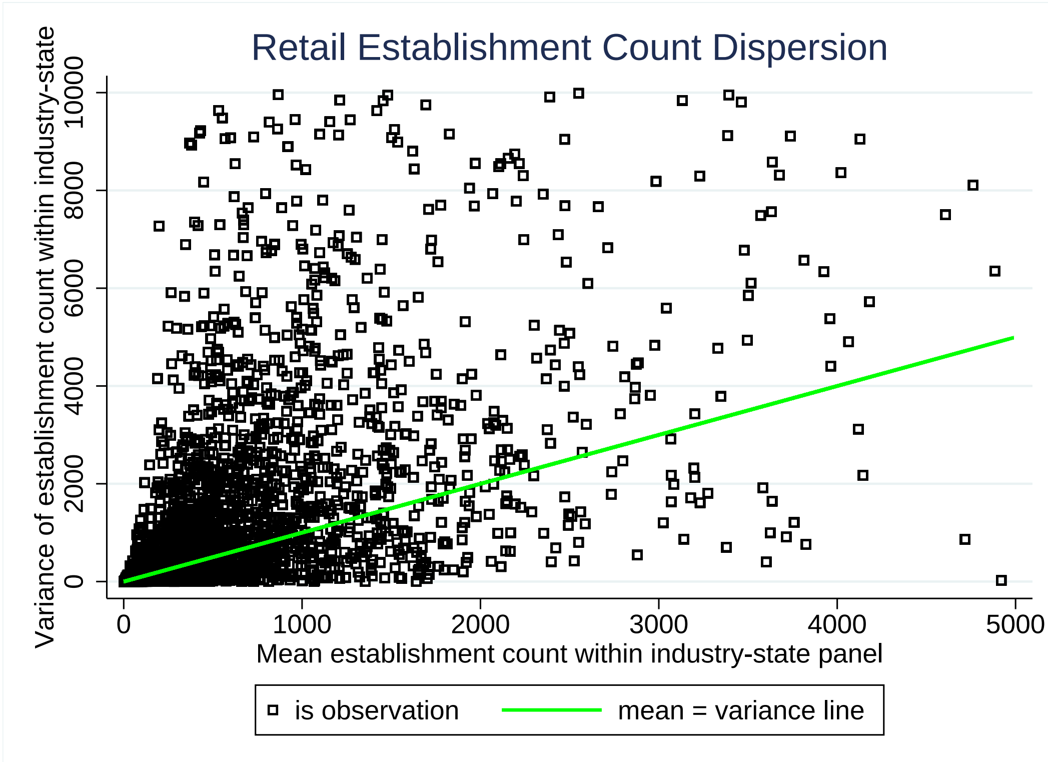


Figure 1.9: Overdispersion in establishment count data

persion. The procedure of selecting an appropriate count data model comes with the caveat that different integer distributions have different implications for the presence of dispersion. For instance, a pure Poisson specification of the likelihood function will present equidispersion since the mean of a Poisson distribution is identical to its variance. Not accounting for overdispersion may yield inconsistent estimates and grossly deflated standard errors as described in (Cameron and Trivedi, 2007).

Descriptive statistics in (Table 3.4) in the Appendix show that overdispersion is a concern in the 4 digit NAICS County Business Pattern data. Pooling all industry-state pairs, the average variance-to-mean ratio in the cross section of industry state pairs (described in Table 3.4 in the) is 2.61 with a maximum of 97.3 occurring in California's "Electronic shopping and mail order houses" industry group. In this cross-section, the minimal variance-to-mean ratio of 0 comes from the "Lawn and garden equipment" retail industry in the District of Columbia.

For a formal specification, let y_{ist} represent the market structure for industry-state is in year t. I introduce the same candidate models as the previous section, starting with a simple specification absent of persistent effects, then introduce across-industry heterogeneity in slope coefficients, followed by a specification with (industry, state, year) fixed effects, and last a traditional panel specification with (industry-state, year) fixed effects:

$$y_{ist} \sim Negative\ Binomial(\frac{\alpha e^{\beta net\ intensity_{st} + X_{ist}\Gamma}}{1+\alpha e}, \frac{1}{\alpha}) \tag{1.19}$$

$$y_{ist} \sim Negative\ Binomial(\frac{\alpha e^{\beta_i net\ intensity_{st} + X_{ist}\Gamma_i}}{1+\alpha e^{\beta_i net\ intensity_{st} + X_{ist}\Gamma_i}}, \frac{1}{\alpha}) \tag{1.20}$$

$$y_{ist} \sim Negative\ Binomial(\frac{\alpha e^{\beta_i net\ intensity_{st} + X_{ist}\Gamma + \alpha_{is} + \lambda_{it} + \theta_t}}{1+\alpha e^{\beta_i net\ intensity_{st} + X_{ist}\Gamma + \alpha_{is} + \lambda_{it} + \theta_t}}, \frac{1}{\alpha}) \tag{1.21}$$

The given parameterization in (1.21) gives

The parameter α, to be estimated, is associated with overdispersion evident in (Figure 1.9). If $\alpha = 0$, then the conditional mean is equal to the conditional variance and there is equidispersion. For $\alpha > 0$, there is evidence of overdispersion. Notice that this specification may do a better job of replicating the overdispersion pattern since the mean function $e^{X_{ist}\beta + \alpha_{is} + \lambda_{it} + \theta_t}$ appears directly in the functional form of the conditional variance. It is worth noting that marginal effects are slightly different in a negative binomial setup than in a linear regression. The marginal effects are proportional to and share the sign of each slope parameter. Each model is estimated with robust standard errors clustered at the state level since the variance is inherently heteroskedastic on account of it being a function of the covariates X_{ist}, and are not necessarily independent across industries in the same state.

Model/Features	neg. binomial 1	neg. binomial 2	neg. binomial 4
Fixed Effects	no	no	is,t,it
Multilevel Slope Coefficients	no	industry specific	industry specific
# parameters	6	110	1468
AIC	14.6	13.7	9.00
McFadden Adj. R^2	.045	.110	0.41
10 fold cross validation avg. MSE	4402	3982	2517

Table 1.5: Negative Binomial model selection metrics

Formal model selection metrics for comparing the negative binomial models are presented in (Table 1.5). Note that the negative binomial model with panel fixed effects is an ideal candidate according to a comparison made on Akaike information criteria, as well as the crude comparison based on adjusted R^2. Following the selection strategy outlined in (Figure 3.12) in the Appendix, I compare the best of the linear in parameters models to the best non-linear model based on a 10 fold cross validation comparison. I find the following stylized facts:

Proposition 4. *Of the candidate models comparable by information criteria measures, the negative binomial specification (1.21) is the ideal candidate.*

Through a cross-validation comparison with (1.15), I find that the models only differ by a few units of mean squared error in the average across a sequence of 10 fold cross validation trials. This suggests that count data strategies yield a small but significant advantage in out of sample forecasting. Likewise, when comparing the ability of the various models to match various moments in the data, as is done in (Table 3.4) and (Table 3.3) in the Appendix, I find that the negative binomial models are better able to replicate the extreme statistics and variance of the data.

1.3.3 Results

Control Function Results

The first stage results of regressing “amazon.com” search frequency on “porn” search frequency along with the exogenous control variables gives the control function in (1.23).

$$\begin{aligned}\hat{v}_{ist} &= amazon\ search\ intensity_{st} \\ &\quad - E(amazon\ search\ intensity_{st} | porn\ search\ intensity_{st}, X_{ist}) \\ &= amazon\ search\ intensity_{st} - (\underset{(7.82)}{0.095} porn\ search\ intensity_{st} + X_{ist}\hat{\Pi})\end{aligned} \tag{1.23}$$

where

$$X_{ist} = (net\ access_{st}, population_{st}, wage_{st}, \pi_0)$$

$$\hat{\Pi} = (\underset{(7.37)}{0.013}, \underset{(23.20)}{.0000004}, \underset{(3.69)}{.0551}, \underset{(36.62)}{32.87})$$

The t-statistic of the first stage parameter estimate appears in parenthesis below the estimate.This is used in the estimation of (1.12) via fixed effects instrumental variables (FEIV) in order to produce estimates of the β_i^{FEIV}.

At the 4 digit NAICS level there are 27 categories that classify retail, with each of the 27 having an amazon exposure effect $\beta_i^{\hat{FEIV}} < 0$ at the 95% confidence level. The graph in (Figure 1.10) presents the β_i^{FEIV} with the NAICS category running on the horizontal axis and the point estimate running along the vertical axis. Examining (Figure 1.10), a clear pattern appears: increasing relative search frequency for “amazon.com” within the state is associated with net exit in all retail categories. This pattern does not exclude electronic shopping and mail order houses, which shows that Amazon is a competitive threat within the domain that is e-commerce. The exit pattern extends to industries like gasoline retail that should be insulated from direct e-commerce competition, yet depend on the e-commerce habits in an indirect manner. The mechanism goes as such: an increase in consumer e-commerce intensity means that the consumer is able to substitute away from local retail in favor of e-commerce and therefore demands less gasoline to travel the distance to a retailer.

Repeating the exercise by replacing establishment counts with employment counts does not yield significant result. This null results is robust in the sense that it occurs for the 27 industries at the 4 digit NAICS level. Many things could explain this insignificant result, including but not limited to the efficiency of an IV/control function approach, insufficient variation in the first place, a small or negligible effect, or another margin of adjustment. In this last case, I suggest that there is the possibility that e-commerce has made retail employers use their existing employee stock in ways that are more efficient for the firm and not necessarily for the employee. A first example would be to lower working hours allotted per employee. Another more anecdotal example of this would be the dynamic scheduling system that many retailers use to schedule part time employees, which anecdotally provides little consistency in work time from week to week and changes the mix of full time to part time employees in the process. This inconsistency can make holding multiple part time jobs simultaneously an added challenge.

An adjustment of worker hours can be indirectly tested by replacing the left hand side in (1.23) with annual payroll in each panel-year ist. Results from this appear in (Figure 1.11). For the 27 retail industries represented at the 4 digit NAICS level, I estimate 19 negative industry effects and 8 insignificant effects, presented in (Figure 1.11) as a function of average operating expenses within the industry. This suggests that the variation in e-commerce intensity has a highly significant and negative effect on retail payroll. If you combine these two facts, that e-commerce intensity brings

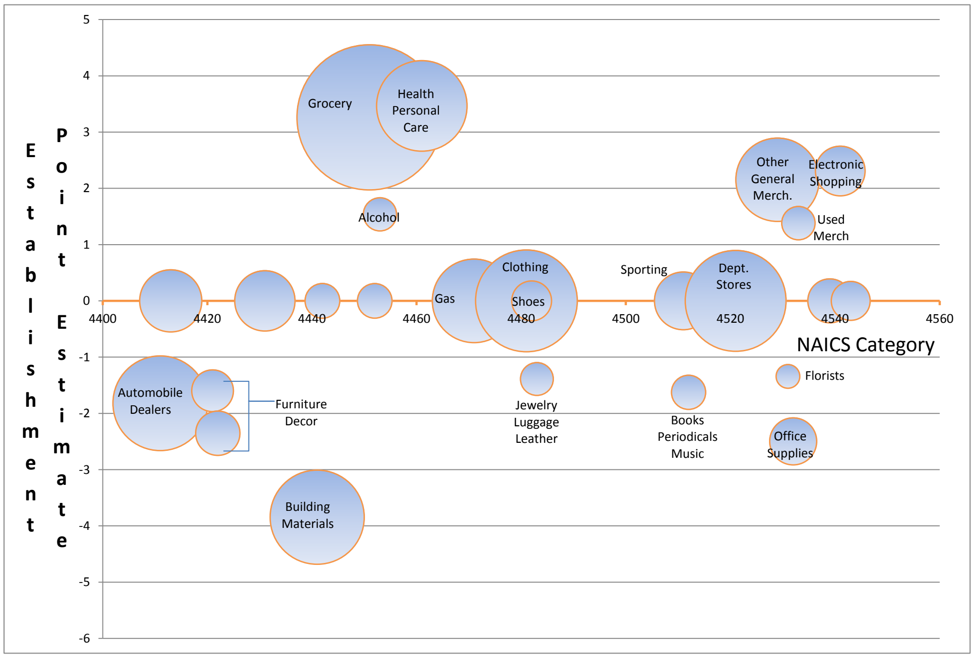


Figure 1.10: Graphical representation of e-commerce establishment count exposure effect

exit in retail, no changes in employment levels, and consistent decreases in payroll, it speaks to the evidence of quality escalation in retail.

Market Structure Results

Estimating a multilevel model gives a breadth of results that doesn't give appealing way of presenting a table of results due to the fact that now, there as many partial effects for a single independent variable as there are groups over which the effect is allowed to vary. A graphical representation of the estimates makes for a much better presentation of the across industry heterogeneity in sensitivity to changes in Internet use intensity, as is presented in (Figure 1.12). This bubble graph has the NAICS index running on the horizontal axis, and the point estimate of the industry Internet use intensity coefficient β_i given by the vertical axis coordinate of the midpoint of

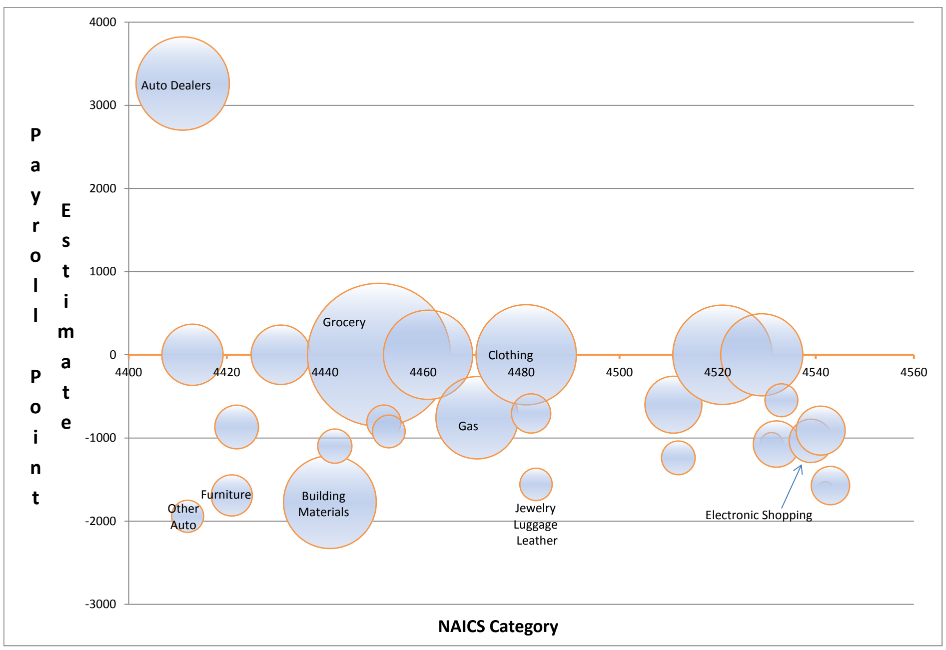


Figure 1.11: Graphical representation of e-commerce payroll exposure effect

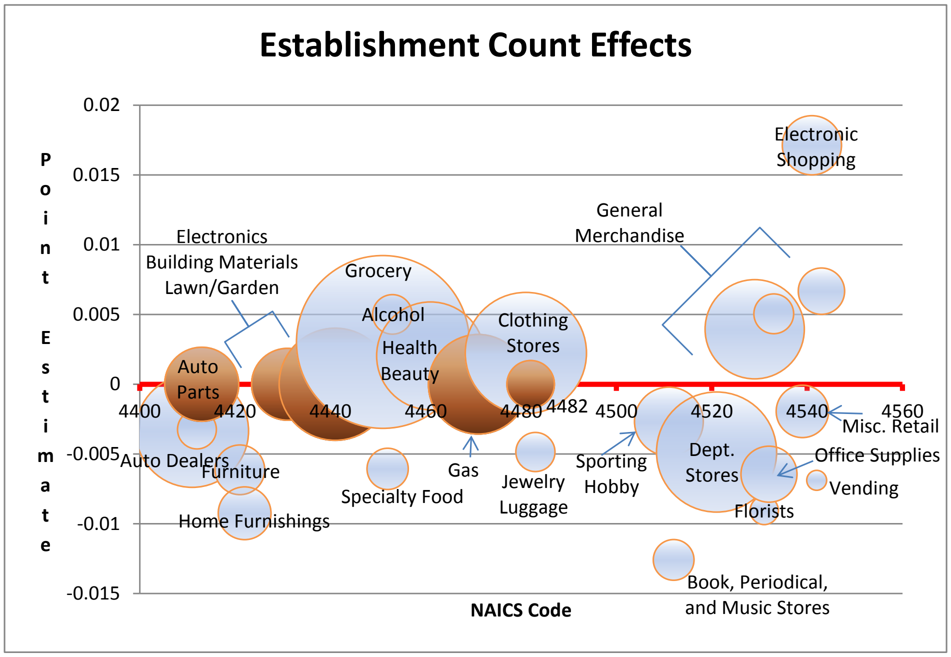


Figure 1.12: Graphical representation of Internet exposure effect in non-linear models.

the bubble. Since NAICS is a categorical variable with a nested structure, there is little meaning to the values or order of appearance on the horizontal axis, outside the fact that industries that share the same first 3 digits may have some characteristics in common. Each bubble's area is proportional to its representative share of total employment in the 44-45 digit retail categories. (Figure 1.12) gives a striking representation of the asymmetry across industries to changes in Internet use intensity, which I discuss in further detail below.

A first glance at the Internet exposure results in (Figure 3.13) show a few patterns in the results. One of the most apparent is the across industry heterogeneity in the exposure to Internet use, made evident by the differences in the center of each bubble. This heterogeneity provides justification for the multilevel approach of estimating industry-specific slope parameters. On its own, adding varying slope coefficients increases the explained variation in establishment counts by 35-48 percentage points, as seen in (Table 1.3) and (Table1.4). The alternative approach, a symmetric marginal effect for all industries, biases the true effect for each industry by presenting only a single parameter estimate that is a weighted average of effects across industries.

It is reasonable to expect that the effect of increased Internet use intensity will be different across retail industries. Not all NAICS categories are as "exposed" as others, in the sense that online shopping is not as easily substitutable for brick and mortar retail. Take grocery stores (NAICS 4451) as an example. Grocers combine labor and distribution in combination with food products to produce the final food products we browse for consumption. This is a retail segment with product substitutes that are not as readily available online in most locales due to the challenges associated with shipping fresh products that cannot be frozen. It is difficult to imagine that Internet retailers of fresh foods are able to compete against brick and mortar grocers, since the cost of refrigerated shipping is likely to be prohibitive. Furthermore, brick and mortar retail chains develop highly efficient distribution networks, giving them a cost advantage over fringe competitors. Rather, proliferation of the Internet has changed the competition in this industry by allowing consumers to observe prices at a lower cost. So it is more likely that intensive Internet use should help grocery retailers that are highly efficient at the cost of inefficient firms. (Figure 1.12) shows that grocers were among the many industries that experience pressure for entry with increasing Internet use intensity. With firms like Shipt entering and introducing online grocery ordering systems, some regions currently do have a way to substitute away from visiting a physical grocery store. Revenue still passes through to the store, so these firms really just weigh your opportunity cost of time against the added delivery fee.

Leading the set of industries that derive harm from variation in Internet use intensity are those represented by NAICS 4512, which consists of Book Stores, Music Stores, and News Dealers. Perhaps the most common characteristic of establishments in this industry is how the Internet allows near costless distribution of their products.

Periodical vendors are in closer competition with Internet substitutes than any other retail segment on account of the ease and near zero cost of acquiring news online. This effect obviously is amplified with the proliferation of cell phones, in that cell phones have given us the ability to find free news in our pocket nearly anywhere during our day to day business, including times we formerly dedicated to visiting the news stand. Similar logic applies to book retailers from the substitution toward electronic devices to read e-books purchased online. Thus, as a state's citizens use the Internet with more intensity, there is a substitution away from the products offered in this retail segment, driving away profits and forcing exit of establishments. Furthermore, the ease of finding pirated copies of books and music make the Internet even more of a threat to establishments in this segment. It's hard to think of a more topical example of how the Internet has caused retail industrial "churning" than to look at book, periodical, and music stores. Many other industries that would be expected to feel the pinch of consolidation from increased Internet use intensity actually do consolidate.

It is little surprise that many industries with highly tradable products already offered by big e-commerce retailers like Amazon have negative point estimates. These include florists (4531), office supplies (4532), specialty foods, as well as the catch all "other miscellaneous store retailers" (4539) that are likely to be in direct competition with Amazon and similar online retailers. Department stores have the second largest share of employment in the represented NAICS retail industries and also had a point estimate with a negative sign, suggesting exit by establishments in this category with increasing Internet use intensity.Another capacity in which the Internet serves consumers is a matching mechanism, particularly in the markets for used goods like automobiles. The results in (Figure 1.12) suggest that used automobile dealers experience a net exit with increased Internet use intensity, which is of little surprise when you consider the plethora of sites that match used car owners to potential buyers (eBay and Craigslist being prime examples).

Employment Results

I duplicate the entire model selection procedure discussed above replacing establishment counts with employment counts. With this outcome variable, there is little guidance for an ideal econometric specification as was the case with establishment counts. Equilibrium in the labor market is characterized by an intimidating looking polynomial, given by (3.19) in the Appendix. Unlike the establishment count results, there is no structural motivation for the equation governing labor. Various selection metrics across the best of the log-log, linear-linear, and negative binomial specifications appear in (Table 1.6). I cannot do selection based on comparing AIC across all the candidate models, since they differ in their measurement of the same dependent variable. All the models have extremely similar and high adjusted R^2 values, presumably on account of the large number of parameters fit by each model. I default to selecting an ideal candidate based on out of sample prediction powers through a sequence of 10 fold cross validation trials. Using this selection metric leads me to choosing the negative binomial specification, since it has the lowest out of sample prediction errors.

Graphical representation of the estimation results appear in (Figure 1.13). Interpretation of this graph is the same as in the market structure results, the midpoint of each bubble corresponds to both a NAICS category on the horizontal axis and a point estimate on the vertical axis. Note that marginal effects in a count data regression do not coincide with the coefficient estimates themselves, however marginal effects are proportional to the coefficient estimates and share the same sign. The employment count results share many patterns with the market structure results. There is a high degree of heterogeneity in employment results, represented by the occurrence of point

Model/Features	log-log spec	linear-linear spec	negative binomial spec
Fixed Effects	is, t	is, t, it	is, t, it
Slope Coefficients	industry specific	industry specific	industry specific
# parameters	1466	1601	1467
AIC	-2.1 (not comparable)	17.4	14.4
Adj. R^2	.997	.996	0.998
10 fold cross validation avg. MSE	37560	35491	37454

Table 1.6: Employment specification selection metrics

estimates above and below the origin. Of the 27 industries represented at the 4 digit NAICS level, 9 experience relatively increased employment with higher Internet use intensity, and 11 experience a relative decrease in employment with increased Internet use intensity, leaving 7 industries with an effect indistinguishable from zero.

With greater Internet use intensity comes not only potential threat from e-commerce, but more direct matching of buyers and sellers without a retail middleman. Substitution away from used automobile dealers is evident both in the negative effect on establishment counts and the subsequent decrease in employment seen in (Figure 1.13). Exit by retail establishments implies decreased employment at that establishment. Whether the workers substitute for retail within the industry, in a another retail industry, or in a different industrial sector altogether is highly specific to each individual and market.

Similar to the market structure results of the previous section, the pinch from e-commerce is evident within retail industries with highly tradeable products. With the near zero cost and ease of transmitting and copying digital music and books, it is little surprise to see that the exit in NAICS 4512 (Books, Periodicals, Music Stores) is associated with a reduction in employment within this category. Once again, establishments in this category have employees whose jobs are more exposed to the Internet than any other industry represented in the SUSB.

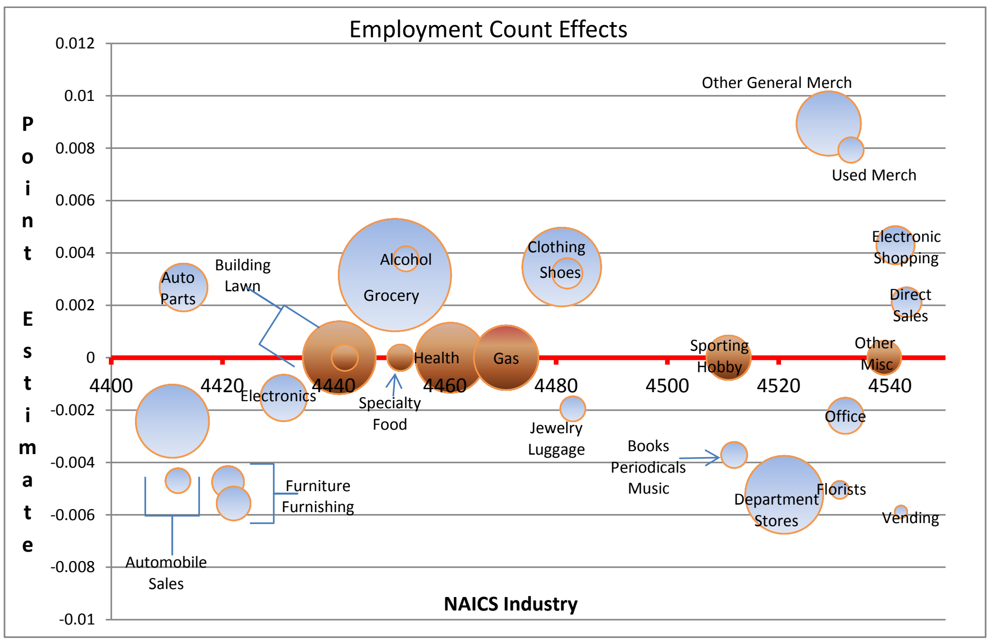


Figure 1.13: Graphical representation of Internet exposure effect on employment

Second Level Analysis

In this section, I explain the across industry heterogeneity in Internet exposure effects by examining their partial correlations with various retail characteristics for which data are available at the industry level. Since the dependent variable in this investigation is an industry specific marginal effect that varies only on the industry dimension, all independent variables must also vary on the industry dimension. The census provides the Annual Retail Trade Survey, which provides measures of operating costs and margins at the 4 digit NAICS level for retail industries. I use linear regression with a bootstrap procedure to estimate standard errors in order to non-parametrically estimate the expectation of the Internet exposure effect, conditional on an industrial characteristic:

$$E(\beta_i|characteristic_i) = \alpha + \phi characteristic_i$$

One prime characteristic would be the share of non-tradable inputs in retail output. Using the 2007 SUSB data, I can approximate this for each industry by taking the ratio of payroll to receipts within each industry state, and then averaging over states. Specifically, Let $\overline{nont}_{is} = \frac{Payroll_{ist}}{Receipts i,s,t}$ for all industry state pairs is. I aggregate $\overline{nont}_i s$ to the industry level $\overline{nont}_i$ by taking the industry average across states.

$$\overline{nont}_i = \frac{1}{50} \sum_{s \in S} nont_{is} \tag{1.24}$$

I hypothesize that retail industries with goods that have inputs which are largely non-tradable are relatively more insulated from e-commerce competition. Unfortunately there is not a large amount of variation across industries in this proxy that measures the share of the non-traded inputs in retail. This is made evident by inspecting (Figure 1.14). Examples of such industries include gasoline and alcohol retail, both of which are prohibited for domestic shipment. In the same manner, retailers

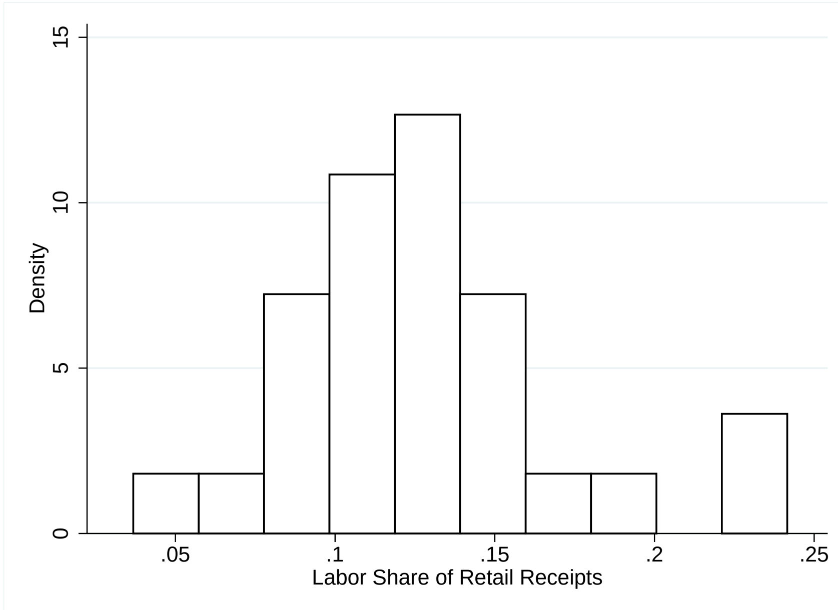


Figure 1.14: Labor input share of receipts histogram

of quickly perishable products cannot ship their products in a cost effective manner. For a crude measure of product tradability, I generate an indicator variable that takes the value of one for retail industries that purvey goods on the "Standard Prohibited and Restricted Items" list of the United States Postal Service. At the 4 digit NAICS level, there are only 4 of 27 industries that have items in these categories, which I suspect will be insufficient variation to detect an effect.

Another possible determinant of Internet exposure is the degree to which fixed/sunk costs are requisites for competing in the retail segment. Establishing an e-commerce business is relatively easier when these expenses are low and when the Internet allows for a decrease in distribution costs, as discussed in (Peitz et al., 2012). Thus, retail industries with relatively low distributive expenses (a component of operating expenses) face the largest threat by entrants. The Census provides annual data on operating expenses for the majority of 4 digit NAICS retail industries. I use this data to construct a "relative operating expense" measure, defined as the average operating expense in industry j relative to operating expenses for electronic shopping retailers: $relative\ operating\ expenses_j = \frac{E(operating\ expenses_j)}{E(operating\ expenses_{electronic})}$. I would expect that as relative operating expenses increase for an industry, that there are greater barriers to entering as an e-commerce competitor, leading to a higher degree of insulation from e-commerce and relatively lower Internet exposure. I supplement this with the average operating expense, defined to be the average of the operating expenses within each industry for the sample period 2008-2012.

Last, since profit is an ultimate motive for entry in a retail segment, I included the gross margin as a possible determinant of industry Internet exposure via the entry incentive that profit margins may represent. From an account perspective, gross margins are revenues minus cost of good sold. Thus gross margins provide an upper bound to the profits the establishments in that industry can realize, since some of the margins must be allocated to capital, labor, marketing costs, etc. I like to consider it from the perspective of entrepreneurs that represent the competitive fringe. The lower the gross margin, the lower the potential profit margin, the less attractive the industry is for potential entrant firms.

$$\begin{aligned} E(\beta_i|restricted\ shipping_i) &= \alpha + \phi * restricted\ shipping_i \\ E(\beta_i|\overline{nont_i}) &= \alpha + \phi * \overline{nont_i} \\ E(\beta_i|relative\ operating\ expenses_i) &= \alpha + \phi * relative\ operating\ expenses_i \\ E(\beta_i|average\ operating\ expenses_i) &= \alpha + \phi * average\ operating\ expenses_i \\ E(\beta_i|proft\ margin_i) &= \alpha + \phi * profit\ margin_i \end{aligned} \tag{1.25}$$

Results of estimating the partial effect of restricted shipping, increased operating cost, and average margins on Internet exposure estimates are presented in (Table 1.7). The restricted shipping indicator is not statistically different from zero, presumably

	exposure β_i	exposureβ_i	exposure β_i
$restricted\ shipping_i$	-0.00005		
	(0.03)		
$relative\ operating\ expenses_i$		0.00246	
		(2.16)*	
$profit\ margin_i$			-0.00005
			(1.20)
_cons (α)	-0.00059	-0.00211	0.00126
	(0.96)	(2.82)**	(0.83)
R^2	0.00	0.18	0.03
N	27	26	27

* $p < 0.05$; ** $p < 0.01$

Table 1.7: Second level analysis regression results

from insufficient variation. The relative operating expenses carries a positive sign, suggesting that industries with relatively large operating expenses derived a relatively larger benefit from variation in Internet use intensity. Across industry differences in relative operating expenses explained about 18% of the industry heterogeneity in Internet exposure. Economically this result is significant since it this means that Average profit margins were not statistically significant as determinants of the Internet exposure effect.

Focusing on the determinants of the e-commerce exposure effect, I plot the $B_i^{\hat{FEIV}}$ as a function of average operating expenses in (Figure 1.15) and (Figure 1.16). Both of these graphs show a positive relationship between the establishment or payroll e-commerce exposure and the average operating expense within the industry. Each of these graphs is strong empirical evidence that industries with low overhead are in closest competition with e-commerce , while the largest industries may be insulated or sufficiently adapted.

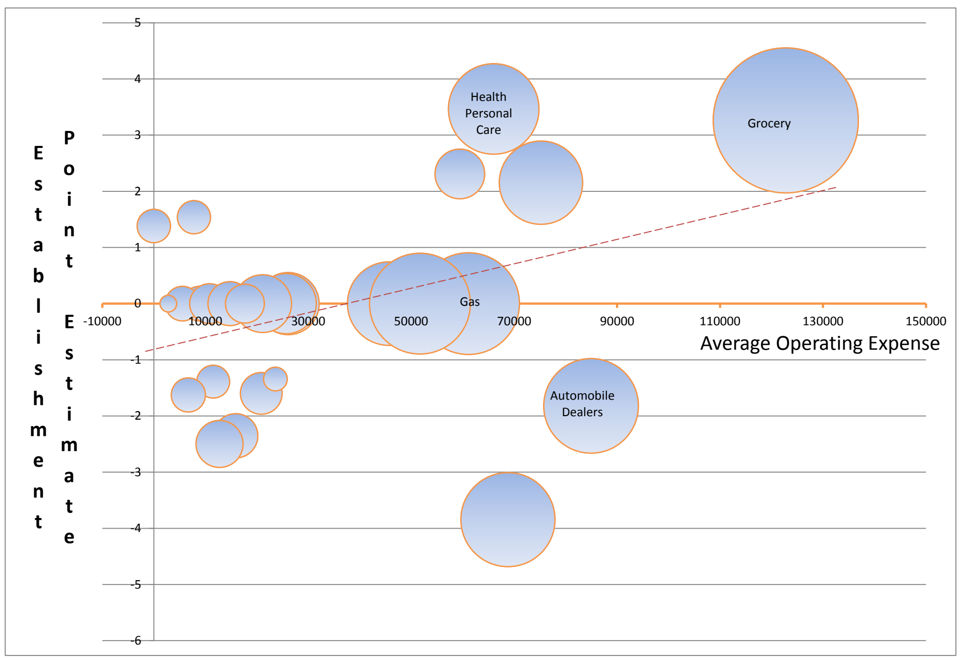


Figure 1.15: Second level analysis of e-commerce establishment count exposure effect

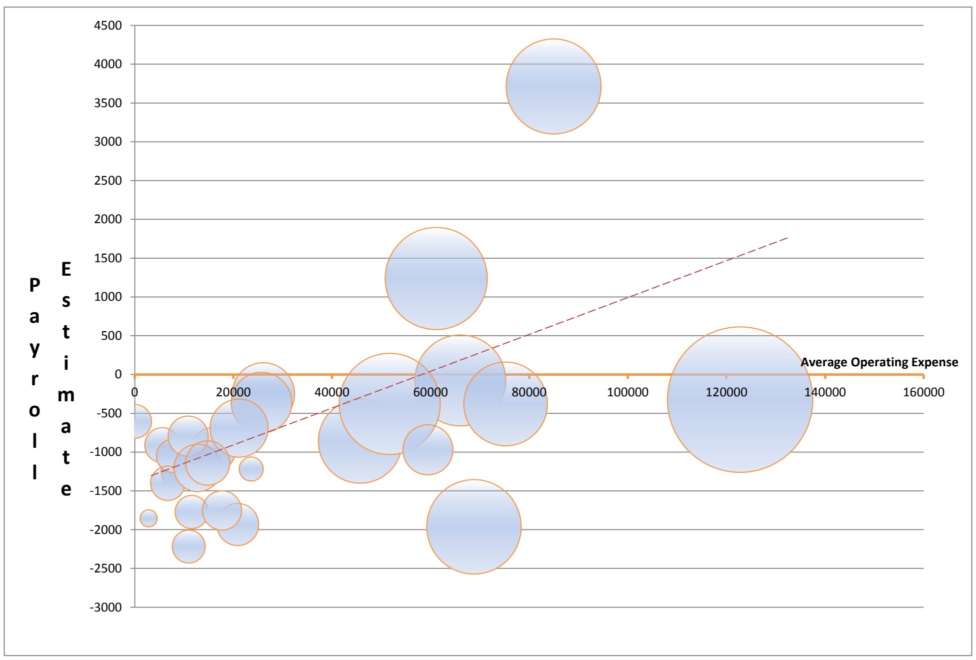


Figure 1.16: Second level analysis of e-commerce payroll exposure effect

1.4 Conclusions

In summary, this paper investigates the ambivalent relationship between the Internet and retail establishment and employment counts. I developed a spatial model of retail competition to motivate the features of an ideal econometric specification explaining entry and exit in retail. The effect of interest was the degree to which increased Internet use intensity is associated with entry and exit in retail, and the corresponding fluctuations in retail employment. Exogenous variation in consumer Internet use intensity was measured using relative search frequency for “porn” at Google, which proxies the propensity of a state’s citizens to hop online to accomplish tasks. I used this measure both on its own and as an instrument for e-commerce intensity, measured by frequency of searches for “amazon.com”. From my model selection procedures, I demonstrated that reduced form specifications are more effective than the structurally motivated estimating equation. This did not invalidate the structural model I developed since it prescribed the use of a multilevel model that drastically improves the fit of empirical models and prevents biased inference. After model selection, estimation results suggested that there is a high degree of heterogeneity across retail industries in the relative help or hindrance provided by the Internet. Naturally, some retail industries were more insulated than others from these changes, and the consistency of this isolation helped validate the results as a whole. The empirical observation that industries are differently sensitive to changes in consumer mobility was explored by allowing for the Internet use intensity coefficients to be industry specific, allowing a secondary analysis to characterize the coefficient heterogeneity.

The first immediate extension that I believe is fruitful is to investigate further use of the Google search intensity data. Google’s interface allows for one to retrieve search intensity for any keyword provided it is sufficiently popular. It must have a small and unspecified amount of search volume to be reported in the first place, so it’s best not to get too specific. Investigating the sharing economy that has sprung up in recent years could be another avenue for research. For instance, I could just as easily use the Statistics of U.S. Business or County Business Pattern data to explore the connection between consumer sentiments for ride sharing through Uber or Lyft, and attempt to detect any subsequent effect on the traditionally licensed transportation service industries.

Future empirical investigation in this work should focus on the relationship between retail and population growth. Most coefficients on population measures were not different from zero, suggesting that most retail industries do not adjust to population growth by opening new establishments. The robustness of the insignificant internal growth result from the econometric investigation leads me to believe that the population-market structure marginal effect has the appropriate sign. It must be the case that many industries are approaching the market fragmentation bound so that the majority of growth is being absorbed on intensive margins instead of on extensive margins (entry/exit). Anecdotally, it seems to be the case that there are critical points in the growth of a city beyond which big box retailers like Ikea or Walmart are enticed to open shop. It is entirely possible that when these large retailers open, they put several small retailers out of business. If this is the case, then looking at changes in the establishment count may not be the appropriate margin of adjustment in the retail industry, and we should also expect our estimates of the partial correlation between population and establishment counts to be null or negative.

There are multiple natural extensions to the theoretical side of this line of research. At some point in future work with this model, I should consider the role of endogenous fixed costs as in Sutton (1991). Retailing's comparative advantage as an industry is distribution and the incorporation of local inputs into the production process that transforms manufacturer's intermediates into the finished product we see on shelves. There is little doubt the distribution associated capital and physical brick and mortar components are not exogenously given, but rather depend critically on the optimal choice of variety. The constant return cost structure of retailers in (Eckel, 2009) and my extension completely ignore the role of fixed costs. Usually, a bound to market fragmentation like that found in my contribution are generated through these endogenous fixed cost structures. Thus, it is worthwhile to develop a dynamic model in order to allow for fixed cost decisions in one period to influence cost and revenue functions in later periods.

CHAPTER 2

REGIONAL EFFECTS OF INDUSTRY PRIVATIZATION: EVIDENCE FROM THE SPACE COAST AND THE SHUTTLE PROGRAM

2.1 Introduction

2.1.1 Motivation

I investigate the regional response to the sectoral shock of discontinuing the public provision of space transportation. The shock manifests as a large negative labor and capital demand shock to the highly specialized "allied" industries that provide inputs to the Shuttle program. In light of retiring the Shuttle fleet, there are many pertinent regional questions that must be addressed. Should there be an expectation that specialized labor will "chase industry" outside the region? Will there be entry by specialized firms to take advantage of the relative abundance of the specialized capital and labor?

These questions and their answers are important for many reasons, since they have real implications for regional income and the subsequent multiplier effect on regional business. A thorough understanding of the answers will aid in the prescription of municipal policy in the region to ameliorate the transition to a new steady state for the region experiencing the shock. Considering the transmission mechanism of the privatization is a prerequisite for prescribing policy, since the optimal response obviously hinges on the dimensions in which the county responds to the "shock" of discontinued funding of the STS (Space Transportation System) program. I will investigate the response of employment, unemployment, payroll, and establishment counts to the discontinuation in public provision of space transportation via a difference-in-difference econometric investigation.

I make many contributions to the understanding of regional factor demand shocks with this investigation. First, the quasi-experimental nature of the discontinuation of the space program allows for clean identification of the regional response to factor demand shocks in the form of layoffs in many high productivity aerospace related contracting firms. It is the first in the literature to rely on a policy shock for identification of regional labor allocation and industrial composition responses. Thus, identification of the shock does not rely on the strength of my belief on whether an instrument is relevant and excludable. Secondly, the investigation allows for analysis of the privatization of monopolized industry in a market economy, which contrasts many previous studies of privatization in economies emerging from central planning. This is important in its own right since it allows for a view of privatization in a context where results will not be confounded with the institutional and oversight problems associated with economies in transition. Lastly, the idea of a regional response to privatization seems to be largely glossed over by the literature, in part since the literature focuses on the response in more disaggregated economic units like individual firms and their competitors as opposed to a slightly more macroeconomics focus on the regions and the industries they support.

I proceed by giving a brief introduction to the space program so the reader can understand why Florida's Space Coast was host to the program in the first place, the relative importance of the space program in the regional economy, and the nature of the decision to retire the Shuttle fleet and transition. With a clear understanding of the space program and the context of the investigation, I describe my data and the econometric investigation I will conduct. The majority of my results will be presented in a graphical format, since this is more appealing than staring at a table with 90 results for each dependent variable. Finally, I estimate the aggregate effects, discuss the implied counterfactuals, and attempt to characterize the industry heterogeneity in shock responses in a systematic manner.

2.1.2 Space Program Background

"Conventional wisdom" in the *Report of the Advisory Committee on the Future of the US Space Program* gives two clear policy goals for the Shuttle program: increased access to space (goal 1) at a reduced cost (goal 2) (Pielke, 1993). The STS program is viewed as failing to achieve these goals for three reasons, two of which are financial with the third being a suboptimal favoring of short term fixes at the cost of "longer range implications" (a failure of dynamic programming). Much of the criticism stems from poorly formed expectations regarding the capabilities of the shuttle program at the outset. As an example, early cost estimates assumed that the program would reach 50 launches a year in order to minimize the average cost per launch. The Challenger disaster struck in 1986, only one year after the busiest launch calendar in program history with 9 launches in 1985. The Columbia disaster (STS-107 re-entry, 2003) warranted some time to review procedures and re-invest in the safety of the program. These disasters were a sufficient demonstration that the program and its ageing fleet would not be able to meet the grueling expectations originally outlined.

It is reasonable to question why was the east coast of Florida chosen to be host in the first place? Due to the bureaucracy behind such a choice, selection of the host county for the program was determined by many observable strategic factors, and is therefore not a random occurrence. According to (Matson, 2009), there were 2 key criteria in determining where to optimally operate the space program: safety and launch efficiency. In order to take advantage of the earth's momentum, Shuttles would need to travel east. By launching from the east coast and traveling east over the Atlantic ocean, the chances of any operational disasters harming humans and property on the surface of earth were minimized. Additionally, the launch location needed to be as close to the equator as possible for launch efficiency. The linear velocity of the earth's rotation increases as you head toward the equator. NASA uses this velocity to its advantage to save on fuel, and desired a location as far south on the coastline as possible . To this degree, Cape Canaveral is relatively far south: Florida accounts for roughly 28% of the nation's Atlantic coastline. This makes Brevard county further south than 86% (rough approximation) of the nation's Atlantic coastline, which I feel comfortable calling "relatively far south." In addition to being in proximity to army and naval bases, 1940s Brevard was mainly composed of orange groves, so the population was not very dense, simultaneously satisfying the safety condition. All of these factors in combination made Cape Canaveral (and therefore Brevard County) the most competitive candidate for locating launch operations.

In order to fully understand the economic impact of such a shock I will briefly describe the context and economic relevancy of the program in the regional economy. STS had a budget that was large relative to estimates of the county's GDP. The STS budget fluctuated between 3.8 (2010) and 5.6 (2005) billion (in 2010 dollars), compared to the estimated 18 billion dollar Brevard County GDP(E.D.C., 2010). There is no clear way of discerning how much of this budget was dedicated to operations in the space coast. Cape Canaveral was the host to every launch for the duration of the STS program. This is despite there being a launch pad at Vandenberg Air Force Base in California and many STS landings at Edwards Air Force Base (California). There is little doubt that Brevard had the largest amount of space infrastructure in employment for STS, and also benefited the most from the tourism spillovers associated with being host to the program since it was the exclusive launch host for the entirety of the program. Since it is not possible to attribute the amount of the STS budget devoted to activities within Brevard County, I elect to use the standard binary treatment indicator in the difference in difference specification, which is discussed further below.

Using data from (Pielke, 1993) I was able to construct a graph of the running average cost per launch in (Figure 2.1). For any year t on the horizontal axis of this graph, the vertical axis measures the cumulative sum of the STS budget until that year divided by the cumulative number of launches including that year. This is therefore a graph of the running average cost per launch for the STS program. The “trend” in (Figure 2.1) is a period of decreasing average costs from the start of the program until the year before the Challenger disaster. The disaster prompted significant re-investment focused on assessing safety procedures, along with a reduction in the expected number of launches per year. A similar “bump” in the average cost function is observed around the time of the Columbia disaster in 2003. From a pure cost-benefit standpoint, it is rational that our efforts to reach space make the transition to a private setting. (Figure 2.1) clearly shows the aging STS system has exhausted economies of scale, made evident by reaching a low point on the average cost curve. Further investment would prolong the life of aging technology at a great financial and risk management costs. Though it is not always desirable to extrapolate far from the sample data, there is a belief that continued launches will drive the average cost curve in (Figure 2.1 further uphill. Rather than take this avenue, our nation acknowledges that the need for updated space transportation at higher safety standards and lower costs comes with a new transition to privatization of space travel, justifying the 2004 mandate to discontinue STS.

There are several dimensions along which Florida’s Space Coast can respond to the shock of discontinuing STS launches. Employers like Boeing and the United Space Alliance (USpA), who are associated with highly specialized mechanical and aerospace engineering pools, were forced to lay off thousands of specialized workers. Ceteris paribus, the effects of a layoff episode of this sort would decrease the income of the region, and subsequently decrease the demand for goods and services. If the shock is permanent, there is an implication for local market structure, since the market would be temporarily saturated, in turn leading to pressure on the least efficient firms. At the same time as experiencing thinning labor markets, there are

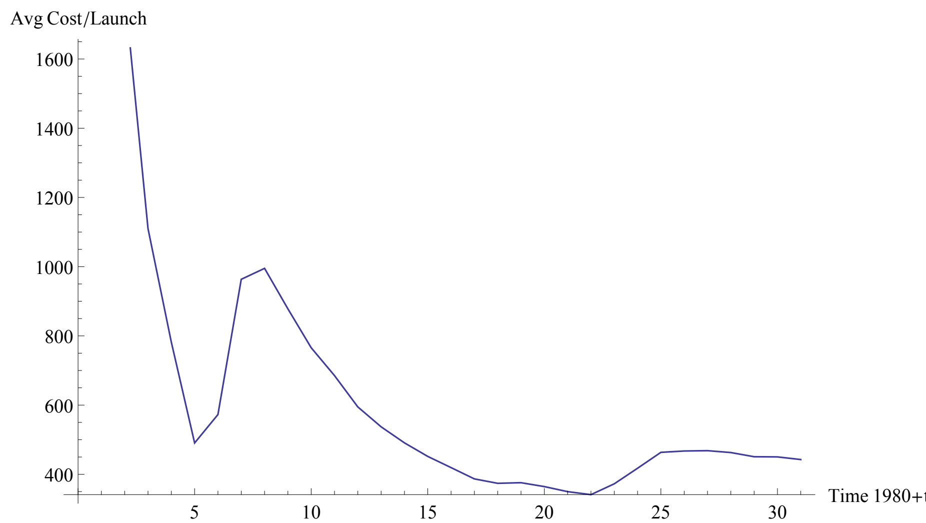


Figure 2.1: Time series of running average cost per Shuttle launch

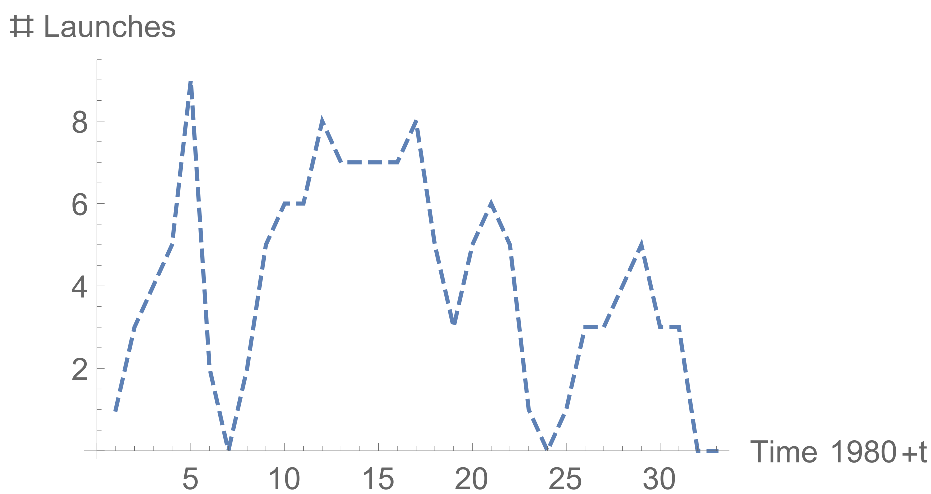


Figure 2.2: Time series of launch frequency

large stocks of highly specific and immobile capital associated with the production of air and space transportation equipment. Equipment like engine testing facilities, launch pads, and assembly buildings, all face potential lack of employment. The immobility of this capital combined with the relative abundance of high productivity labor and geographical desirability makes Brevard an attractive choice for firms in industries that could use the labor networks and aerospace infrastructure.

To determine the appropriate data to work with, I first consider the possible margins of adjustment and the time frame in which each is fixed and variable. Foremost, it's possible that the county would see a flight of labor to jobs outside the county or state. In a worse case scenario, this happens and the resulting lack of consumption spending in the region leads to an "aftershock" of businesses closing due to a lack of revenues to meet their break-even. In a more positive counterfactual world, we would see firms from outside the county/region acting opportunistically in regards to the large stocks of capital and labor experiencing unemployment, and therefore relocating to Brevard to help ease the layoff effects.

2.1.3 Related Literature

The effects of privatizing individual firms within an industry have been thoroughly investigated. There seems to be little literature regarding privatizing entire industries, where the industry generates revenue from public provision, perhaps on account of how few examples of such situations exist. Existing literature also seems to focus on allocative efficiency and cost improvements, and seem to neglect the regional economic perspective on the consequences of such a decision. With this in mind, this chapter intends to understand (i) how the region supporting the formerly public industry/firm responds to this type of shock and (ii) how will firms across all industries behave in the presence of this type of shock (iii) how does industrial composition change in light of shock. The purpose of doing such an investigation is to understand the economic transmission of the shock through the regional economy to motivate future policies when faced with similar circumstances.

My empirical strategy relies on a "natural experiment" resulting from a policy specific policy shock, with methodology similar to the approach in (Card, 1990). In the case of discontinued funding of the STS program, the shock presents as a large negative demand shock in the labor and capital markets. The shock differs in sign from the circumstance in (Card, 1990), and also differs by being a demand side shock as opposed to a supply side shock. Additionally, since industry level data was not available for the difference-in-difference investigation in (Card, 1990), I am able to use the increased degrees of freedom this dimension provides in order to estimate a multilevel model where there is industry level heterogeneity in treatment effect estimates. Allowing for this heterogeneity is important for consideration of general equilibrium effects, and allows me to describe the within and across industry adjustments to the shock.

Regional responses to employment shocks are well understood largely thanks to (Blanchard et al., 1992). Here, the hypothesis that employment rates exhibit persistence on account of a unit root, and formal statistical hypothesis testing suggests the null hypothesis of a unit root cannot be rejected. Whether employment and margins follow a unit root are important when considering the path these variables will take through time. A unit root would imply that temporary shocks to employment would have permanent effects on the employment levels for a given geographical unit. As such, a careful consideration of the best response to employment shocks is warranted. Unemployment rates do not exhibit such persistence, primarily on account of the role of unemployment as a driver of migration as in (DaVanzo, 1978).

Using more recent data, (Dao et al., 2014) conduct a follow-up replication of (Blanchard et al., 1992) and find many differences in comparison to the original paper. Foremost, Dao2014 finds the long term effect of a regional shock to be nearly half of the seminal work in (Blanchard et al., 1992). One important finding of their work for my investigation is the relatively small response in interstate net migration, which I am unable to measure due to the lack of annual data on in and out migration flows. As opposed to out-migration, they find that most workers tend to either drop out of the labor force, or remain unemployed in place of relocation. Due to the high demand associated with high skilled labor in aerospace, I would hypothesize that the workers are easily able to find employment in similar engineering or technical capacities.

Two important adjustment mechanisms come into play when considering the adjustment to a negative regional labor demand shock. The presence of sticky wages, particularly in empirically is found to be the case in (Hall, 2005), suggests adjustments to the labor input decision are more likely to occur on extensive margins (layoffs) rather than on intensive margins like wage changes as found in (Barattieri et al., 2010). Any time there are layoffs in industries where the workers have highly transferable skills one should expect the region to experience some amount of out-migration if the workers cannot get employment offers locally. Recognizing this, migration on the part of firms could be a profit maximizing decision to take advantage of the agglomeration economies. Agglomeration economies of this sort in high tech are well documented, for instance in (Henderson, 2003). Whether the in-migration effect will dominate is given by the relative strength and speed of both in the long-run. Empirically, (Blanchard et al., 1992) find that most of the adjustment to an adverse employment shock is through out-migration of labor as opposed to in-migration of firms. Presumably this is on account of individuals being more mobile than firms, and hence I suspect it will be the main transmission mechanism through which Brevard county may feel any effects from discontinuing the STS program.

(Eckel et al., 1997a) examine the privatization of a single firm, British Airways, on airfares and competitors' stock prices. They find evidence that privatization increases competition in a significant 7% fall in U.S. competitor stocks, and that the extent of the fall is proportional to the degree in which the firms compete. Additional support for a decrease in airfare was found, airfares in the international markets served by British Airways fell 14.3% relative to other transatlantic routes. In part, this line of literature demonstrates that the supply side analysis of privatization is well researched, yet leaves much to be on the part of welfare and macroeconomic analysis.

(Saal and Parker, 2000) focused on one of the "rare" cases where the industry is composed entirely of publicly held firms, all of which were privatized. They directed attention towards identifying changes in economic efficiency "in terms of technical and input price efficiencies captured in total costs" when England and Wales privatized their water industry in 1989. Marginally significant evidence is found in support of the idea that privatization had a beneficial impact on productivity growth. An indicator for a pricing review in the middle of the sample time frame suggested that increasing output price helped to slow the growth of costs for the private water producers.

As a thought exercise, applying the results of (Saal and Parker, 2000) to STS discontinuation in the Space Coast would mean an a-priori ambiguous effect. On the one hand, the private firms now providing space transportation services could locate in Brevard. On the other hand, the Space Coast could continue to enjoy the tourism and technological spillovers associated with being host to the industry, and potentially draw in firms . In another state of the world, the private firms could locate outside Brevard, and the county experiences little to none of the associated spillovers. The presence of the highly immobile aerospace capital combined with a relatively "thick" labor market for the supporting labor types makes in-migration an appealing option for firms that are sufficiently mobile or are planning for expansion.

Models of privatization that are primarily theoretical, like (Che, 2009) and (Laban and Wolf, 1993), focus on typical firm level effects. The former is more concerned with investigating why privatization may fail to improve firm performance and the role of institutional development in firm performance in the period after privatization. The latter, though not dynamic, shares a common point with the former on account of the investigations being in the context of transitioning economies like post-Soviet era Russia. It seeks to explain the slow progress of large scale privatization, where the tension is the result of differences in expectations of the returns to privatization. My investigation is distinct from these, foremost since it is free from any confounding factors associated with an economy in transition due to periods of high institutional uncertainty and erratic behavior that introduce temporal variation in confounding factors that is difficult to measure or observe.

Previous studies in firm entry and exit yield some insight on the techniques to analyzing changes in market structure within industries. In (Moretti, 2010), variation in the number of jobs in 2 industry categories are explained by variation in the number of jobs in the tradable sector. First, the author regresses the change in the number of jobs in the non-tradable sector on the change in jobs in the tradable sector. Additionally, he regresses employment changes in a random segment of the tradable sector on the change in employment for the rest of the sector. To isolate exogenous shifts in labor demand for the manufacturing center, Moretti uses a weighted average of nationwide employment growth in 77 categories in manufacturing, with the weights reflecting employment in each sector specific to each city. He finds significant and positive "local multiplier" effects of employment in tradable manufacturing on employment in the non-tradable sector, with an estimated elasticity of 0.55.

My approach to the investigation is distinct from the previously mentioned studies in many ways. Foremost, my investigation is concerned with regional outcomes as opposed to firm outcomes, as the former is well studied but the latter has significant scope for research. I am more interested in how the shock influences labor allocation and industrial composition, and less on the performance of individual firms, since the regional response is important for consumers, firm managers, and central planners alike. Additionally, the investigation I develop will not suffer from the instability associated with investigations conducted in transitioning economies as those studies in post-Soviet Russia. The retirement of the Shuttle fleet provides an example of a sectoral shock in a geographic location with stable institutions and stable contract enforcement. Conducting the investigation in the context of a stable macroeconomy makes the approach more credible in the sense that the regional adjustment is being driven by market forces instead of cronyism.

2.2 Empirical Investigation

2.2.1 Data

Variable	Mean	Std. Dev.
Employment	196040.133	195103.3
Establishments	13094.9	9199.133
Annual Payroll (1,000 U.S. $)	6902742.117	7459460.574
% firms with 50 employees or more	4.5	0.9
% firms with 100 employees or more	2.0	0.5

Table 2.1: Industrial summary statistics for Florida's space coast 2004-2012

In September 2004 George W. Bush announced plans for the retirement of the Shuttle program and the transition of space transport from public provision to private provision. Initially, the final year of launches was scheduled for 2010. Unanticipated delays in the missions during 2010 necessitated an extension of the program until the final launch in July 2011. With this in mind, for difference-in-difference investigation I take 2012 to mark the post-treatment period since it is the year after the final Shuttle launch. With a difference-in-difference approach, I will focus on comparing Brevard to its 5 neighboring counties; 2 of which are coastal: Indian River, and Volusia, as well as Orange , Osceola, and Seminole counties which are landlocked. These counties are similar to Brevard in many ways, with the exception that none were host to the Shuttle program.

The Census County Business Patterns (CBP) gives data at the industry-county-year level, where industries are divided into various resolutions with descriptions from 2-6 digits in length. Descriptions with 2 digits are sectors, while those with 3 are subsectors, 5 digits are national industries, and so on. Of the variables provided in the

CBP data, my dependent variables include the establishment and employment counts, and annual payroll. At the sub sector level, each establishment is categorized into one of 87 different 3 digit identifiers. With more disaggregation at the 5 digit level, there are somewhere near 285 different national industries. Let the set I_n be the collection of unique NAICS categories at the n digit description level. Thus, I_3 has 87 elements and I_5 has somewhere near 285. I find that econometric analysis at the 3 digit level is a healthy compromise between analytical practicality and keeping the volume of results manageable. The results are fully generalizable at more disaggregated NAICS descriptions.

Variation in the establishment counts within and across counties comes from many sources. Foremost, changes in macroeconomic conditions like aggregate demand for goods and services will have a direct effect on this measure. For Brevard, there is a dense concentration of engineering and manufacturing establishments oriented towards space transportation equipment manufacturing. Population is also a significant determinant of establishment counts. A larger population will need more firms, larger firms, or both relative to a smaller neighboring county. A larger population will also demand a larger variety of goods. As such, changes in the population may effect the firm counts in categories that are sensitive to this margin determinant. Across counties, time invariant effects like the presence of tourist attractions like Florida's beaches, or how Florida's "sunshine" climate makes it an attraction for golfers also drives some of the variation in establishment counts.

Variation in the employment of industry j within a county is foremost assumed to be the result of changes to macroeconomic conditions like aggregate demand. For instance, the decrease in consumption from the Great Recession may have lead some firms to not have sufficient demand to cover variable costs and therefore to layoff employees or discontinue production. I assume that macroeconomic effects of this sort are common to all industries and counties in the sample, and will be detected with the inclusion of time dummy variables also common to all industries and counties in the sample. Employment in Brevard certainly will vary on account of the shock of retiring the Shuttle fleet, since many employees were displaced from engineering and specialty firms like United Launch Alliance, United Space Alliance, etc. Presumably, some of these employees will substitute towards other firms in the same industry working on projects unrelated to the Shuttle program, others will supply their labor in different industries, while others will do both of these after migrating from Brevard.

Changes in employment for each industry as a result of discontinuation of STS funding can ideally broken down into two effects. One possible effect is an "across industry substitution effect" where workers in a given industry (take aerospace engineering as an example) will place into jobs in fields outside their specialty. Perhaps some will go into scientific consulting, others into teaching, etc. A second and distinct effect would be a "within industry substitution effect" where employees working in firm A in industry j move to firm B in industry j. Third, there is also a potential transition to unemployment or non-participation for the standard reasons. If estimates suggest a given industry experienced decreased employment as a result of the shock, we know that the net effect is some combination of across industry substitution, and out-migration/non-participation effects. Likewise, if estimates suggest increased employment in a given industry as a shock response, we know that this may be a combination of across industry substitution, and new hires from the unemployed and in-migrants.

As an example, if I consider the annual (log) employment change for 2011-2012 in "Professional, Scientific, and Technical services" (NAICS 54), seen in (Figure 2.3 page 73), I note that there was a large decrease in employment for a particularly high productivity sector. I would assume that the majority of this decrease is the result of the massive layoffs in the industries allied with the Shuttle program. Using a difference-in-difference approach will allow me to determine precisely how much of the change is attributable to the stop in launches by trying to estimate what the employment change would have been had the policy not changed.

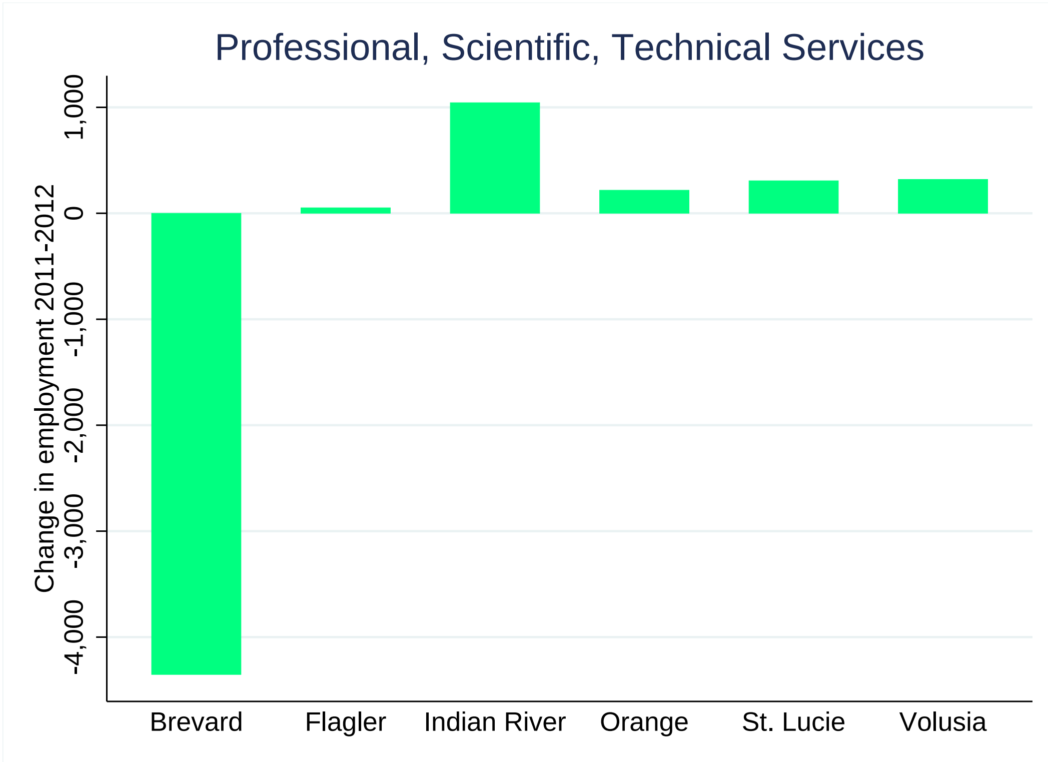


Figure 2.3: Change in logarithm of employment for Professional, Scientific, and Technical Services sector, 2011-2012 for Brevard relative to other counties

2.2.2 Specification

There are many options for specifying a difference in difference estimator for this investigation. My strategy is to use a typical difference-in-difference specification with an indicator for the "treatment" of being host to the space program. For reasons previously discussed, I take the pre-treatment period as 2004 and the post treatment period as 2012. The difference-in-difference specification follows in (2.1).

$$
\begin{aligned}
Y_{ic,t} = &\beta_0 + \beta_{i1} Host_{c,t} + \beta_{i2} D_{2012} + \beta_{i3} Host_{c,t} Post_{c,t} \\
&+ \sum^{c} \gamma_c D_c + \sum_{t=2005}^{2011} \gamma_t D_t + \sum^{i} \gamma_i D_i + \epsilon_{ic,t}
\end{aligned} \tag{2.1}
$$

I adopt i as an index of the many NAICS industries tracked in the CBP, c as an index of the 6 counties used for the investigation, and t to indicate the year in question. Here, $Y_{ic,t}$ represents the outcome variable of interest, either establishment counts, employment counts, or annual payroll. $Host_{c,t}$ is an indicator variable that is unity for Brevard county in all years except 2012 since it was the only county to host shuttle launches during these years. The indicator for the post-treatment period is given by $Post_{c,t}$, which is unity for all counties during the year 2012 and zero otherwise. In turn, $Host_{c,t}Post_{c,t}$ takes the value one only for Brevard county during the post-treatment year 2012. The specification also includes county, year, and industry fixed effects as denoted by the terms in the last three summation signs of (2.1). These fixed effects control for the average differences in observable and unobservable characteristics across industries and counties that may be determinants of the outcome variables of interest. Examples of such determinants that were previously include things like beaches and large tourist attractions like Disney World, as well as zoning laws. Sources of variation that are common to all counties are captured by the year indicators D_t, where the baseline year is taken to be 2004.

Notice that the coefficient on the treatment interaction, β_{i3}, is indexed by industry. This is to indicate that I estimate industry specific treatment effects. Implementing this flexible feature is as simple as interacting the treatment interaction $Host_{c,t}D_{2012}$ with a complete set of industry indicators. Doing so allows for a more generalized functional specification, and relaxes the strong assumption that all industries respond in a symmetric manner to changes in the treatment indicator. Additionally, this flexibility allows the model to capture more of the variance in the outcome variable while nesting the possibility of symmetric responses across industries. Finally, it will allow for a secondary analysis to characterize the across industry heterogeneity in treatment responses.

Estimation of the partial effects of discontinued funding of the STS on the various outcome variables of interest is simple. To show this, take conditional expectations of (2.1) for Brevard and any other county (I will use Indian River since it is a desirable baseline county), using the fact that $Host_{c,t} = 0$ for all counties that are not Brevard, holding the industry constant at $i = j$, and recognizing that 2004 is the baseline year:

$$\begin{aligned}
&E[Y_{ic,t}|i = j, c = Brevard, t = 2004] = \beta_0 + \beta_1 + \gamma_j \\
&E[Y_{ic,t}|i = j, c = Indian\ River, t = 2004] = \beta_0 + \gamma_j \\
&E[Y_{ic,t}|i = j, c = Brevard, t = 2012] = \beta_0 + \beta_1 + \beta_2 + \beta_{j3} + \gamma_j \\
&E[Y_{ic,t}|i = j, c = Indian\ River, t = 2012] = \beta_0 + \beta_2 + \gamma_j
\end{aligned} \tag{2.2}$$

Here, I adopt the notation $\mu_{j,c,t}$ to indicate the expected outcome for industry j in county c at time t. The treatment effect is identified with the traditional "difference-in-difference" calculation:

$$\begin{aligned}
(\mu_{j,Brevard,2012} - \mu_{j,Volusia,2012}) - (\mu_{j,Brevard,2004} - \mu_{j,Volusia,2004}) \\
&= [(\beta_0 + \beta_1 + \beta_2 + \beta_{j3} + \gamma_{3j}) - (\beta_0 + \beta_2 + \gamma_{3j})] \\
&\quad - [(\beta_0 + \beta_1 + \gamma_j) - (\beta_0 + \gamma_j)] \\
&= (\beta_1 + \beta_{j3}) - (\beta_1) \\
&= \beta_{j3}
\end{aligned} \tag{2.3}$$

2.2.3 Results and Discussion

Estimation results of (2.1) are given in the appendix. P-values are reported directly below the coefficient estimates. Standard errors are estimated with clustering at the county level, as this is the largest geographic level for which I anticipate serial correlation in the residuals. A series of bubble graphs will aid in interpreting results, the first of which presented in (Figure 2.4).The horizontal axis coordinate of each bubble corresponds to each NAICS 3 digit sub sector in the sample data. Each 3 digit integer is just a categorical representation of an industry, hence there is no meaning to their order. Subsectors (3 digit descriptions) with a common 2 digit sector description are likely to share industrial characteristics and be "close" in a competitive sense of competing together either in the product or input markets. The vertical axis coordinate of each bubble corresponds to a treatment effect estimate, $\hat{\beta_{j,3}}$, in the econometric specification in (2.1) on p. 74. Each bubble's center coordinates is an industry, treatment effect estimate pair (NAICS industry, $\hat{\beta_{j,3}}$). The area of each bubble is proportional to the industry share of total margin in question. Only the industries with treatment effects significant at the 95 % confidence level are presented in the graph. Ex-post consideration of the across industry heterogeneity in treatment effect estimates in this fashion allows for a characterization of changes in industrial composition, income, and market structure as a treatment response to stopping the STS launches.

Employment Results

Employment results were mixed. A brief tally of the signs of industrial treatment effects suggest 17 industries with relatively increased employment and 17 decreased employment. This leaves 51 industries with either no change in employment or without an estimate due to multicollinearity. A graphical representation of the results is

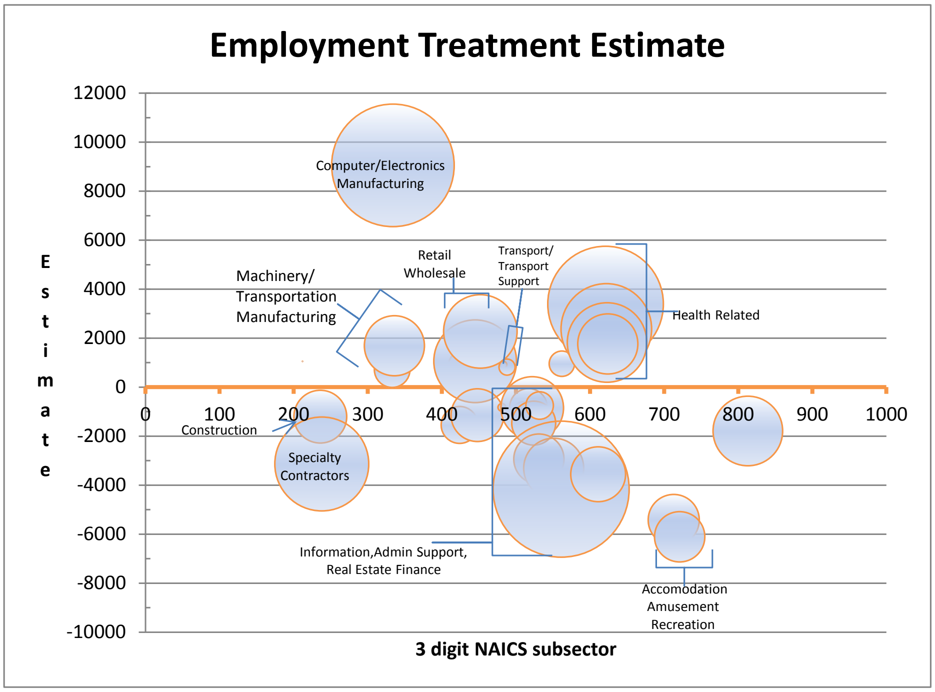


Figure 2.4: Employment treatment effect estimates

given in (Figure 2.4). Given a 5% probability of type 1 hypothesis testing error, I would expect to estimate 4-5 of the 85 sub sector (3 digit) treatment effects spuriously. In a short run investigation, where capital structure is taken as given, employment is a margin that can be adjusted freely. As such, it will paint a different picture than the establishment count results, payroll results, or firm size results.

Visualizing the treatment effect estimates as in (Figure 2.4) allows for intuitive understanding of labor allocation changes as a result of the discontinuation of STS launches. Since the outcome variable is measured as full time and part time employees, the vertical axis corresponds to how industrial employment changed in 2012 relative to a counterfactual world where STS launches continued. There are many takeaways from these results. First, note that in employment terms, many large and small industries had treatment effects that were different from zero. Second, note there is heterogeneity across subsectors in the magnitude of the effect. A second pattern to note is the heterogeneity across sectors. Some, like manufacturing, seemed to benefit, while other service oriented industries like Real estate, Management, and Administrative/Support services consistently suffered.

Consistent with immigration outflows outweighing inflows and a general decrease in Brevard's growth rate, are the effects in the Construction sector (23X). Building Construction (236) had a decrease in employment of roughly 1200 positions, and specialty contractors has an estimated loss of nearly 3 times this amount. Housing construction is viewed as a leading indicator of regional economic activity as discussed in (Stock and Watson, 1989). The strength of this variable as a leading indicator will be more clear as more years of NAICS data are released.

Searching for evidence of across industry employment substitution starts with considering the industries that would demand the specialized labor associated with Shuttle launches. Firms that provide inputs to the launch production process may compete with other STS supporting firms in the labor market. Given that much of the launch services labor is highly specialized and composes a relatively minor part of a larger transportation industry, it is expected that the labor qualified in this industry can also easily become qualified in it's allied industries like commercial aerospace transportation manufacturing and specialized computer manufacturing.

Consistent with a discontinuation in tourism spillovers are the decreased employment in accommodation (721) potentially connected to tourism. This description aggregates over travel accommodations in general, including hotels and associated overnight accommodations. Anecdotally, anyone raised on the Space Coast knows how launches were a spectacle that drew people from around the world. The closer you were to Cape Canaveral, the more people you would see pulled over to watch a launch, particularly on the causeways that connect the barrier island to the mainland. This magnitude of tourism draw is associated with consumption of travel related services, like those in accommodation and the Amusement and Recreation sub sector (713), which also had a statistically significant negative treatment effect estimate. Hence it makes sense that firms in this sector felt a pinch from the stop in big ticket launches like those in the STS program.

Another way of gaining insights from simple meta-analysis is to appeal to the binomial sign test. Each industry is represented as an individual "trial", and for each industry I take the trial to be a success if the treatment effect estimate is significantly different from zero at the $\alpha = 0.05$ level. Estimation results show 34 significant effects in 85 industries for which data are available. Suppose that the probability of a null treatment effect is equal to the probability of a statistically significant effect at $\alpha = 0.05$. More precisely, adopt:

$$\begin{aligned} H_O &: Pr(\hat{\beta_{j,3}} = 0) = Pr(\hat{\beta_{j,3}} \neq 0) = 0.5 \\ H_A &: Pr(\hat{\beta_{j,3}} = 0) \neq Pr(\hat{\beta_{j,3}} \neq 0) \end{aligned} \tag{2.4}$$

Using the data from the employment results in (Table 3.8) as the trials, the probability of estimating 34 or fewer significant estimates is approximately 4%. Prior to considering any industry characteristics, I can say that there is a large chance of any individual industry exhibiting a response to the treatment.

Repeating this meta-analysis by conditioning on the set of significant treatment effect estimates for the employment results in (Table 3.8) gives 34 "trials" to work with. Here I consider a "success" event to be a positive treatment effect estimate, $\hat{\beta_{j,3}} > 0$, for which there are 17. Overall, I can estimate a "probability of growth" since a successful event represents a treatment effect estimate with a positive effect on the outcome in question. In contrast to this probability of growth there is a negative treatment effect estimate, the "probability of contraction" $1 - p$. Given an equal probability of growth and contraction, which is specified in (2.5) estimating at most

17 positive effects in 34 trials is a common occurrence with a probability of $\approx 56\%$. With a p-value of ≈ 0.56, there is insufficient evidence to reject the null hypothesis that the probability of growth is different from the probability of contraction in the sample data.

$$\begin{aligned} H_O &: Pr(\hat{\beta_{j,3}} > 0) = Pr(\hat{\beta_{j,3}} < 0) = 0.5 \\ H_A &: Pr(\hat{\beta_{j,3}} > 0) \neq Pr(\hat{\beta_{j,3}} < 0) \end{aligned} \tag{2.5}$$

Unemployment Results

Using data from the Bureau of Labor Statistics, I was able to attain unemployment rates for a panel of Florida counties from 2004-2012. I estimate a specification similar to 2.1 with unemployment as the outcome variable. Since unemployment is an aggregate statistic and not an industry specific measure, the specification becomes

$$unemployed_{ct} = \alpha + \underbrace{I_c + I_t}_{\text{county and year fixed effects}} + \beta * \underbrace{Host_{ct}Post_{ct}}_{\text{treatment indicator}} + \epsilon_{ct} \tag{2.6}$$

The specification in (2.6) differs from the discussed specification (2.1) in that the outcome variable only varies at the county-year level, so I cannot estimate industry specific slope parameters. Select results from estimation of (2.6) appear in (Table 2.2), with full results appearing in the appendix. I bootstrap the standard errors of β in order to not rely on assumptions regarding its distribution to show its statistical significance.

Interpreting these unemployment results suggest that the shock associated with retiring the Shuttle fleet resulted in an increase in unemployment in Brevard of approximately 3,700 more workers than had the program still been in operation. This

	Unemployed
Treatment Effect	3,702.871
	(2.08)*
2005bn.year	-1,022.761
	(0.85)
2006.year	-1,490.194
	(1.16)
2007.year	-345.761
	(0.30)
2008.year	2,803.567
	(3.48)**
2009.year	8,301.164
	(6.49)**
2010.year	9,371.373
	(6.33)**
2011.year	8,058.896
	(6.63)**
2012.year	6,051.539
	(6.31)**
_cons	3,025.353
	(3.39)**
R^2	0.83
N	603

* $p < 0.05$; ** $p < 0.01$

Table 2.2: Unemployment treatment effect estimates

number is about half of the aggregate amount of layoffs by the largest two space transportation firms, suggesting that there is evidence of either out-migration or substitution to employment within and across industries. With unemployment for Brevard in 2012 at 24,794 workers, the implied counterfactual unemployment in Brevard lies somewhere near 21,000 workers. Considering the magnitude of the employment effect, unemployment in Brevard spiked nearly 17% in a single year as a result of the shock.

Establishment Count Results

Of the 87 unique 3 digit sub sectors in the NAICS classification system represented in the sample data, I estimated 12 negative coefficients and 3 positive coefficients, with 72 industries either being statistically indistinguishable from zero or dropped due to multicollinearity. Given a 5% probability of rejecting the null hypothesis of no treatment effect, $\beta_{j,3} = 0$, given that the effect is actually zero, I expect to estimate ≈ 5 industry effects as the result of type 1 error. Of all the margins of adjustment I investigate, establishment counts will have the highest persistence and be the slowest to adjust. Microeconomic theory suggests that a firm may continue to operate at a loss in the short-run, and considering only the year after discontinued STS funding (due to current data limitations) places this investigation in that time frame. I expect that this "sluggishness" of capital means there will be less colorful results for this margin of adjustment. Ideally, firms would adjust on the employment margin, and only exit if not running a profit in the long-run, which is not necessarily 2012, the year taken as the "post-treatment" period. However, increasing the specificness of the NAICS description by moving to 5 digit descriptions shows a more detailed story.

Of the industries directly connected to the space program, like those associated with "Professional, Scientific, and Technical Services" (54X) and aerospace related manufacturing, there is little evidence of a response on the establishment count margin at the 3 digit NAICS level. Notice in (Figure 2.5) the net exit within the construction industry, perhaps due to lower than expected housing growth squeezing out the relatively less efficient firms. In a richer 4 digit investigation, construction related industries includes residential and nonresidential building construction, as well as foundation, structure and exterior contractors and building equipment contractors which include electrical, plumbing, heating, and air-conditioning.

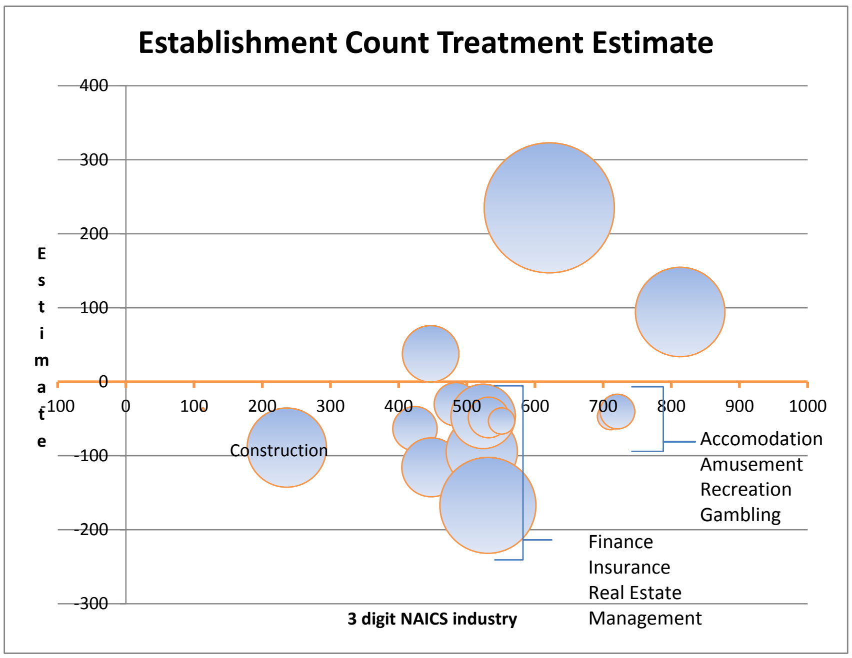


Figure 2.5: Establishment count treatment effect estimates

Net exit is not limited to firms associated with the construction side. Effects are also present on the material input side to construction through a response in many construction related manufacturing industries: plastic, gas, steel, electrical lighting, electrical equipment all relatively consolidated, along with consolidation in wholesale of hardware, plumbing and heating equipment and supplies. With a labor demand shock of this sort it is expected that there may be a slump in construction spending associated with both hesitation from current county residents as well as from potential in-migrant workers and firms who were "on the edge" regarding entry in Brevard to serve the regional markets. Part of this slump also may represent a slowing of demand for housing in Brevard due to out-migration and the across industry substitution response where the new entrants are located elsewhere in the state or nation.

Many of the estimates are consistent with generic decreases in aggregate demand and out-migration. For instance, I see that though grocery had an effect indistinguishable from zero, grocery wholesalers (4244) had a statistically significant establishment exit. On the retail side, both home furnishings (4422) and lawn and garden (4442) related retailers experienced consolidation, which could be on account of both out-migration and decreased aggregate expenditure on consumer durable goods. Also consistent with the out-migration story is the exit of firms in the storage and warehousing (4931) industry, which is surely due to decreased demand for local storage if there is a net movement out of the county.

Of the effects that are likely related to tourism, there was a significant reduction in the establishments affiliated in the transportation industry, particularly scenic and sightseeing transportation (4879), as well as taxi and limo services (4853), and urban transit systems (4851) . There was also a direct decrease in establishments affiliated with traveller accommodations, presumably on account of the launches no longer drawing spectators to the county. This category includes Hotels and Motels, as well as bed and breakfast type operations that presumably were being booked by spectators and journalists.

As previously stated, not all of the effects on the establishment margin were negative. There is clear evidence of across-industry substitution into more technical manufacturing industries like basic chemical manufacturing (3251), pharmaceutical and medical manufacturing (3254) as well as navigational, electronic, medical, and control instruments manufacturing (3345). Consistent with the hypothesis that firms may enter into business in Brevard to take advantage of the relative abundance of highly skilled and potentially unemployed labor is the positive entry within the Architectural, Engineering, and Related Services industry (5413). Since this is an industry that is likely not perfectly competitive in input or output markets, entrant firms could have bargaining power in what is an oligopsonized industry.

Payroll Results

Payroll results exhibit a high degree of across industry heterogeneity in estimates, with the estimation yielding 8 positive coefficients and 19 negative coefficients with 58 industries having coefficients either not distinct from zero or dropped due to multicollinearity. A graphical representation of the results is given in (Figure 2.6). This shares the same 4-5 potentially spurious results due to type 1 statistical error.

An obvious pattern in the payroll treatment effect estimates is the strength of across industry substitution and growth in manufacturing of computers and electronics as well as transportation equipment (336). One component of the increase in transportation equipment manufacturing payroll is the entry of Embraer, a South American regional jet manufacturer. According to their estimates in a 2010 press release, Embraer contributes over 230 engineering and support jobs to Brevard via a relatively new manufacturing plant adjacent to Melbourne International Airport. Aside from aerospace transportation equipment manufacturing, there was entry by sea trans-

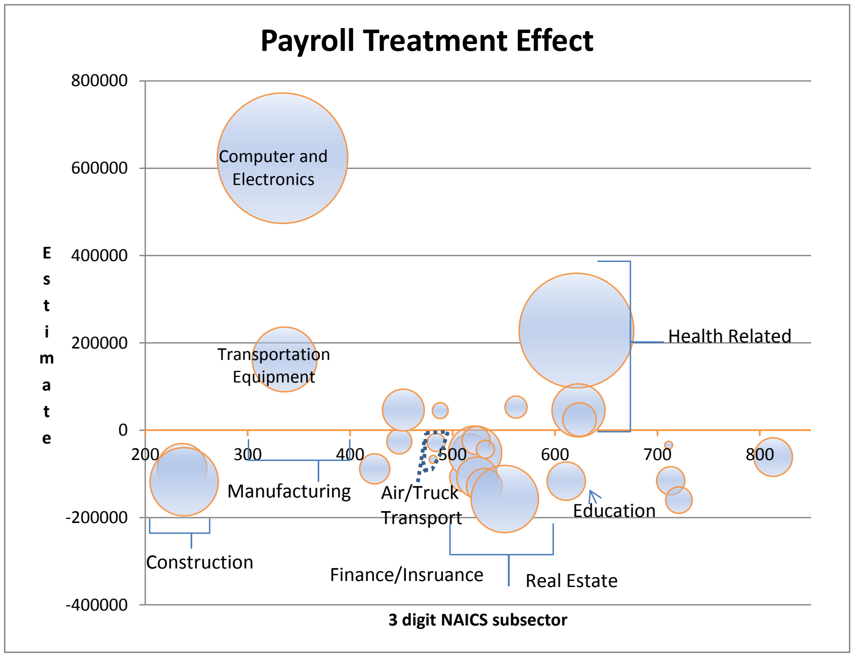


Figure 2.6: Payroll treatment effect estimates

portation equipment manufacturers, Bertram Yacht's relocation to Merritt Island being a prime example. The treatment effect estimate suggests that payroll in transportation equipment manufacturing increased by $\$162,151 * \$1,000 = \$162,151,000$ more than had the STS program continued to operate.

Consistent with a decrease in tourism spillovers are the decreases in the accommodation (72X) and Amusement (71X) sectors. For example, the payroll treatment effect estimate for Accommodation is -160075, as given in (Table 3.8) in the appendix. County Business Patterns reports payroll in thousands of dollars, so the center of the confidence interval for this estimate is $-\$165,076 * \$1000 = -\$165,076,000$ lower than the counterfactual outcome for Brevard. Given the GDP estimate in (E.D.C., 2010), this makes for a nearly 1% decrease in regional income in and of its own.

Aggregate Effect Estimates

To assess the "macro" impact of discontinued STS funding, I take a weighted average of treatment effects for each industry, where weights represent the industry share of the total outcome in question. As a concrete example, let's examine ways of determining some notion of an aggregate effect for Brevard. The estimation procedure allows for assessing these "macro" impact estimates on the outcome variables of interest. If I had to sort the outcome variables by how quickly they can be adjusted, it would be payroll $\rightarrow$ employment $\rightarrow$ establishments, and I consider them in this order.

To begin, let the set $S_k = \{i = 1, 2, ...s\}$ represent the collection of s industries for which a statistically significant treatment effect estimate exists from estimation of (2.1) at the k digit NAICS level in the empirical section as described. One notion of an aggregate effect is simply the net effect, defined as the sum of the treatment effect estimate over all industries, as specified in (2.7):

$$Net\ Effect_{outcome} = \sum_{s \in S_3} \hat{\beta_{s3}} \tag{2.7}$$

Outcome	Positive	Negative	Net Effect
Payroll	1,227,400	-1,572,909	-345,509
Employment	32,213.44	-40,158.1	-8,034.65
Establishments	367.15	-832.00	-464.815

Table 2.3: Aggregate effect estimate

The results of this procedure to determine an aggregate effect with the estimation results from (2.1) is presented in (Table 2.3). I construct this table by aggregating over industries. Payroll units represent \$1,000. Employment is measured as the count of full and part-time employees, including salaried officers and executives of corporations, who are on the payroll in the pay period including March 12. Included are employees on paid sick leave, holidays, and vacations; not included are sole proprietors and partners of unincorporated businesses. Establishments are measured as counts.

Of primary interest is the net effect on income for Florida's Space Coast, which suggest that reported income decreased by $\$345,509,000$, or just over a third of a billion dollars. This net effect is the sum of an increase in income of approximately 1.22 billion dollars for industries above the origin in (Figure 2.6) with a decrease of nearly 1.57 billion. Nearly half of the payroll increase is from growth in the Computer and Electronic Manufacturing sector. Presumably, much of the decrease in income stems from the dwindling tourism and high tech draw formerly associated with the Shuttle program. A primary policy goal in light of this figure is to identify a mechanism to minimize the loss of income to the region. The net income effect estimate of ≈ 345 million underestimates the true income effect, since it does not account for any multiplier effects. Incorporating a conservative multiplier estimate from (Beemiller and Friedenberg, 1992) of ≈ 2, the income impact to the Space Coast could be as high as a $\$700,000,000$ decrease over the coming years. On the other hand, more recent news of the location decision of privatized aerospace transportation contractors like SpaceX and Blue Horizon to operate on the Space Coast.

On an extensive margin like employment, there was a net effect of a decrease in employment of approximately 8,000 positions. This effect is the sum of a large draw from industries like Computer, Electronics, Machinery, and Transportation equipment manufacturing demanding over 11,000 employees. These subsectors alone account for $\approx 33\%$ share of the total estimated employment gain of 32,213 employees as reported in (Table 2.3). Large negative effects from the decreased tourism spillovers is dominated by a decrease in employment experienced by the tourism subsectors, which together had an estimated loss of over 11,000 employees. This tourism loss composes just over 25% of the total decrease in employment after launches stopped.

The estimated effect for the most "sluggish" margin of adjustment, establishment counts, is a decrease of approximately 464 establishments across the represented NAICS sub sectors. This net exit of 464 firms can be decomposed into entry of 367 firms across 3 sub sectors, offset by a decrease of 832 establishments across 12 sub sectors. The largest estimated establishment decreases were realized by the construction and real estate sub sectors, which is consistent with a decrease regional aggregate demand and foreshadows further regional economic contraction. Consistent with the decrease in tourism spillovers from the launches was the contraction in the accommodation sub sector, whose estimated 40 establishment contraction treatment effect estimate is approximately 5% of the total estimated contraction.

2.3 Conclusion

In this paper, I used a multilevel difference-in-difference technique to examine regional responses to the sectoral shock associated with privatizing the space program and retiring NASA's shuttle fleet. The multilevel approach allowed for estimation of within and across industry margins of adjustment previously not considered by the literature on regional economic shocks. The first of these is an across industry substitution effect, evidence of which can be found in the nearly 32,000 employee increase in some industries to nearly offset the 40,000 employee decrease in other industries, as per (Table 2.3). Without a multilevel regression specification, being able to decompose a net effect of the shock into offsetting positive and negative effects at the industry level is not possible. Without the added industry dimension, it is only possible to estimate the net effect. The small net effect on employment counts, establishment counts, and payroll, as provided in (Figure 2.4) and (Table 3.8) in the appendix, vastly understates the disturbance caused by the shock. Evidence of this reallocation is present in every margin of adjustment considered. My estimates suggest that the shock to unemployment (a near 17% increase) dwarfs the net effect on regional income.

With this in mind, a policy goal of stabilizing regional income must have a two pronged approach. First, it must minimize income losses to the region by incentivizing firms subject to the shock to avoid out-migration. To this degree, the first policy instrument available to local and state central planners would be a reduction in taxes to incentivize continued operation. This could be combined with a reduction in nominal expenses for permitting/licensing. In addition to incentivizing establishments delay or forgo exit in light of the shock, the region could incentivize establishments to in-migrate. An effort along this line starts with a vast marketing campaign to appeal both to employees and to firm managers. Following marketing appeal, central planners could again use tax incentives. An investment in these campaigns acts not only to attract visitors, but acts as a good faith investment in increasing the population draw to the region.

Future work on examining this particular type of shock clearly should focus on developing a synthetic control structure approach in order to verify the results hold under less strict assumptions. The traditional difference-in-difference approach requires that the differences between the treatment and control groups in the pre-treatment period be fixed over time. The synthetic control approach, as discussed in (Abadie et al., 2010) generalizes this by allowing unobserved confounding factors to vary over time. I could use this technique to verify the robustness of the results from the multilevel difference-in-difference approach. Synthetic control gives results that are time varying, which would add an appealing dynamic perspective. The only challenge in implementing this in practice at the moment is the lack of County Business Pattern data in the period after 2012.

CHAPTER 3

RETAIL ESTABLISHMENT SIZE AND THE SPATIAL DISTRIBUTION OF ECONOMIC ACTIVITY

3.1 Introduction

3.1.1 Motivation

Past retail research estimates a positive relationship between the population in a geographic area and retail establishment counts. The intuition is simple: larger areas demand both a greater quantity and greater variety of goods, making increased retail presence a necessity. In contrast to this, simple fixed effect estimators fit to annual data by the Census Statistics of U.S. Business data do not provide uniform evidence for this hypothesis across the nation's industries. In select cases, such investigations provide evidence to the contrary. With this research, I intend to understand the many reasons the results of today may differ from the estimates of 40 years ago. I hypothesize that in a short run investigation where the brick and mortar location of the retail establishment is held fixed, establishments adjust to the sales growth by demanding more labor as inputs to the production of retail goods. Thus, an industry-area-year level measure of an intensive margin such as employment per establishment or the proportion of firms with x many employees is a more appropriate outcome variable for measuring retail growth than participatory measures like establishment counts.

I make several contributions to the understanding of retail industrial organization and regional commerce with this investigation. First, I explain the insensitivity of retail establishment counts to market growth by examining establishment size an alternative margin of adjustment to growth. I demonstrate the utility of Department of Defense satellite data to measure economic activity and the spatial distribution of economic activity at sub-national levels. This data is an improvement on any current standard that relies on reported income alone as it is a direct measure of both reported and unreported on economic activity. To this point, this analysis is the first to use the night light data for econometric analysis in empirical industrial organization. I utilize a dynamic panel model ala (Blundell and Bond, 1998) in order to determine the effect of economic activity growth and spatial variation in economic activity on retail establishment size measures. My contribution has descriptive aspects as it is the first to estimate these partial correlations between the moments of the income distribution and the retail industry using the improved night light data and the county business pattern data.

(Figure 3.1) provides graphical evidence of the negative relationship between the number of Grocers per capita and population for a cross section of Florida counties as a motivating example in the cross section. Notice that the smaller counties actually have relatively more establishments serving each citizen. By comparison, large counties have relatively fewer establishments per capita. This negative relationship is highlighted by the fitted curve in (Figure 3.1).

Instead of relying on cross-sectional variation to identify the effect of increasing market size, I use the DMSP data to construct measures of both the mean level of economic activity within the county, as well as the dispersion of the activity. Using the (spatial) standard deviation of night light intensity will as a measure of spatial uniformity of the economic activity. Both of these variables will allow for identification of any concentrating effects of market growth and changes in the spatial distribution of income on establishment sizes by tying changes in industrial indicators to changes in economic activity within the county.

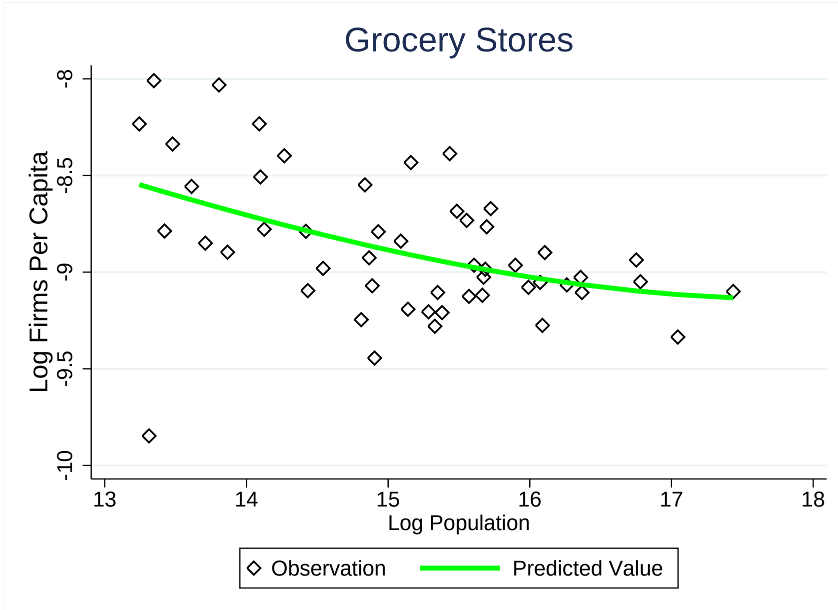


Figure 3.1: Market size and concentration relationship in Florida's grocery industry

3.1.2 Related Literature

My objective is to quantify the correlation between market size and retail establishment size. The effect of interest is the elasticity of retail establishment size with respect to changes in market size and changes in spatial variation of economic activity. If you believe the typical "Walmart effect" story, then regional market growth will be met with a concentration to larger big box type retailers that offer several higher order goods and services under a single roof. Unlike the typical "Walmart effect" story, such as,(Basker, 2005a), (Basker, 2005b), or (Stone, 1988), I do not explicitly focus on Walmart itself. Rather, I take the "Walmart effect" to mean there are certain "critical points" in the growth of a region beyond which big box retailers like Ikea or Walmart maximize profits by entering and competing in the industry-county pair. Before these critical points are reached, it is likely the case that retail is dominated by firms that do not enjoy the economies of scale/scope that characterizes big box type retail firms. These existing firms likely have higher markups than their big box counterparts. As large retailers open, they use economies of scale/scope to compete at a lower markup, and they gain market share from smaller firms with higher markups. This is the typical reallocation within industry, as described in (Melitz, 2003b), and is of benefit to the consumer since it reduces the prices they face at retail. From this perspective, retailers not in immediate competition may directly benefit from entry by big box retailers since the reduction in prices raises demand for other goods via both cross price elasticity effects as well as through income effects.

In part, this chapter is motivated from robust results in (Vitt, 2015). Results therein suggested there is either an insignificant or negative correlation between the growth of the population in a state and the number of retailers serving each citizen. This result stands in contrast to previous estimates of the relationship, as in (Clements, 1978) and (Forbes, 1972). The latter of these references suggests that a 1% increase in the population of a MSA is associated with a 0.96% increase in the number of retail establishments. It's somewhat naive to think that this estimated relationship from decades ago would extrapolate to the population levels and market structure retail faces today. Further, evidence in (Vitt, 2015), (Sutton, 1991), and (Shaked and Sutton, 1983) all suggest that in models of competition with relatively simple cost structures there exists a bound to the fragmentation in a market. A bound of this sort matters since it implies that beyond some critical market size all sales growth is absorbed by incumbent firms. The only dimensions along which existing firms can accommodate the growth is through expanding their brick and mortar structure, hiring additional employees, or both of these simultaneously. In econometric application, the existence of a bound to market fragmentation means that we cannot expect the marginal effect of population on retail establishment counts to always be non-zero and statistically significant. To ensure that this pattern in (Vitt, 2015) is not simply an artifact from relying on intercensal population estimates as proxies for market size, I demonstrate the utility of satellite imagery of night light activity as a measure of market size.

One effort to understand the determinants of firm size across industries is laid out in (Kumar et al., 1999). Judiciary efficiency, financial market development, capital intensity, and market size are all used to explain variation in establishment size. The coauthors are upfront about how their endeavors are primarily descriptive and that their estimates reflect partial correlations. They find that a 1% increase in market size, as measured by total employment within the associated 2 digit industry, is associated with a 0.19% increase in employment. Defining market size in this manner is problematic for many reasons. Foremost, it neglects the varying labor intensity across industries, skewing the true size of the market to reflect the relative labor intensity of the industry. Second, the coauthors use log population as an instrument for log market size. While log population certainly satisfies the "relevance" criteria of an instrument, the authors make no attempt to convince the reader as to why this would satisfy the exclusion restriction required of ideal instruments. Since the dependent variable is the (log of) employment in an industry divided by the establishment count in the industry, it is hard to believe that log population would not be a determinant of other establishment count determinants, violating the exclusion restriction required of instruments. For example, since large populations correlate highly with the cost of living through one of the many channels discussed in (Haworth and Rasmussen, 1973), it is likely the case that $corr(Population_{st}, wage_{st}) \neq 0$ while simultaneously $corr(wage, \#retailers) \neq 0$.

A case study of the concentrating effect of market growth towards larger "big box" and "category killer" retailers in the Toronto area is described in (Jones and Doucet, 2000). In this study, the researchers find that during the 1990s, the largest period of growth for "big box" retailing, the relative importance of retail as represented by the proportion of occupied storefronts fell from 53.7% in 1994 to 49.5% in 1997. They find that the share of total stores declined on average by 7% in response to entry by big box retailers, with the greatest number of closures occurring in office products

(-23%) and hardware (-17%). The results agree with other similar studies in the area, like (Haltiwanger et al., 2010), who find that the share of employment accounted for by big-box retail comes at the expense of single establishment and small chain stores. The common denominator between these studies is the fact that they focus on the effects of big-box retail on existing competition, as opposed to determining process of concentration into larger retailers.

3.1.3 Simple Example

To form expectations of the sign of the relationship between income growth and retail establishment size, I appeal to a comparative static exercise. Consider two hypothetical geographical areas with identical income distributions and marginal propensities to consume from income, with the exception that one area has larger and growing consumer mass. This larger area must consume a larger quantity of goods than the smaller area, since it simply has more of the identical representative agent. From this alone, big box type retailers able to compete on scale have a larger incentive to establish operations in the growing area compared to the area not experiencing growth. It is not immediately clear that increasing the number of agents in the area leads to a greater diversity of preferences on account of spatial "clustering" in preferences. A "clustering" of preferences is most apparent along the income dimension, perhaps due to sorting a la (Tiebout, 1956). Marginal consumers will only bring a diversity of preferences if their preferences are not already represented among those in the area.

Empirically, we tend to see a larger variety of goods where there is greater variation in income, since such areas likely have many areas with clustered preferences demanding different consumption baskets. Again, consider two hypothetical areas, of equal size in terms of representative agents, one of which has agents with identical incomes, and the other where there are income "clusters" as previously discussed. This relatively less homogeneous area is likely to have more diverse preferences on account of the heterogeneity and clustering in incomes. Though not explicitly divided into strata, each strata of income could demand a different set of products. Thus areas with high variation in income are desirable for retail competitors that compete on economies of scale and scope like the big box retailers.

With Florida having 63 counties, the night light data should have plenty of variation to allow me to identify the consequences of increasing the spatial variation in economic activity on local retail market structures while holding changes economic activity at a constant. It should also allow for me to isolate the pure income growth effect on local retail market structure, as well as the pure income dispersion effect. Using the night light data for measures of the changing spatial dispersion of economic activity is one of the novelties of this study, and is one of the true advantages of working with the satellite data. Data based on reported income measures are able to describe economic growth for a region, but is unable to do describe the spatial distribution of the activity.

3.2 Empirical Investigation

3.2.1 Data

Variable	Mean	Std. Dev.	Coeff. Var	Min.	Max.
Establishment Size	6.604	19.391	2.936	0	424.667
Proportion 50+	0.062	0.206	3.323	0	1
Proportion 100+	0.038	0.162	4.263	0	1
Std. Dev. Night lights	14.474	7.18	0.496	3.235	27.71
Mean Night Lights	14.787	12.615	0.853	0.940	59.572
Median Night Lights	10.373	15.497	1.494	0	63
N		31608			

Table 3.1: Establishment Size Summary Statistics

For the purposes of this investigation, I use a panel of industry-county pairs observed annually from 2004-2012 to identify the effect of economic activity growth and changes in the spatial distribution of economic activity on retail firm sizes. To measure both the changing mean and spatial dispersion of income, I will use the National Oceanographic and Atmospheric Administration's Defense Meteorological Satellite Program "night-light" data (DMSP hereafter). The "night light" data comes from the sources in the form of a picture, with each pixel reporting the amount of night time lights across in a given square kilometer block, as measured by satellites.

Prime examples of putting the night light data to work include (Henderson et al., 2012) and (Michalopoulos and Papaioannou, 2013). An image of the 2012 DMSP image of Florida is provided in (Figure 3.4). A sequence of these images enables me to collect data for each county on the mean night light intensity and the spatial distribution of the night lights as measured by the standard deviation of night light intensity (referred to as $\sigma_c t$ hereafter). DMSP indicators of night light activity are bottom coded at 0 and top coded at 63, though the summary statistics in (Table 3.1) suggest that none of the panels in my sample are consistently topcoded in this manner. In this case, OLS would still be unbiased and consistent as outlined in (Wooldridge, 2010).

County level data on the establishment counts by various employment sizes is available through the Census "Statistics of U.S.Business" (SUSB below). The economic indicators in this database are also partitioned by enterprise size and 2007 North American Industry Classification System description (NAICS below). Industries are reported at various levels of disaggregation, with 2 digit length descriptions being highly aggregated, and 6 digit descriptions being very specific national industries.

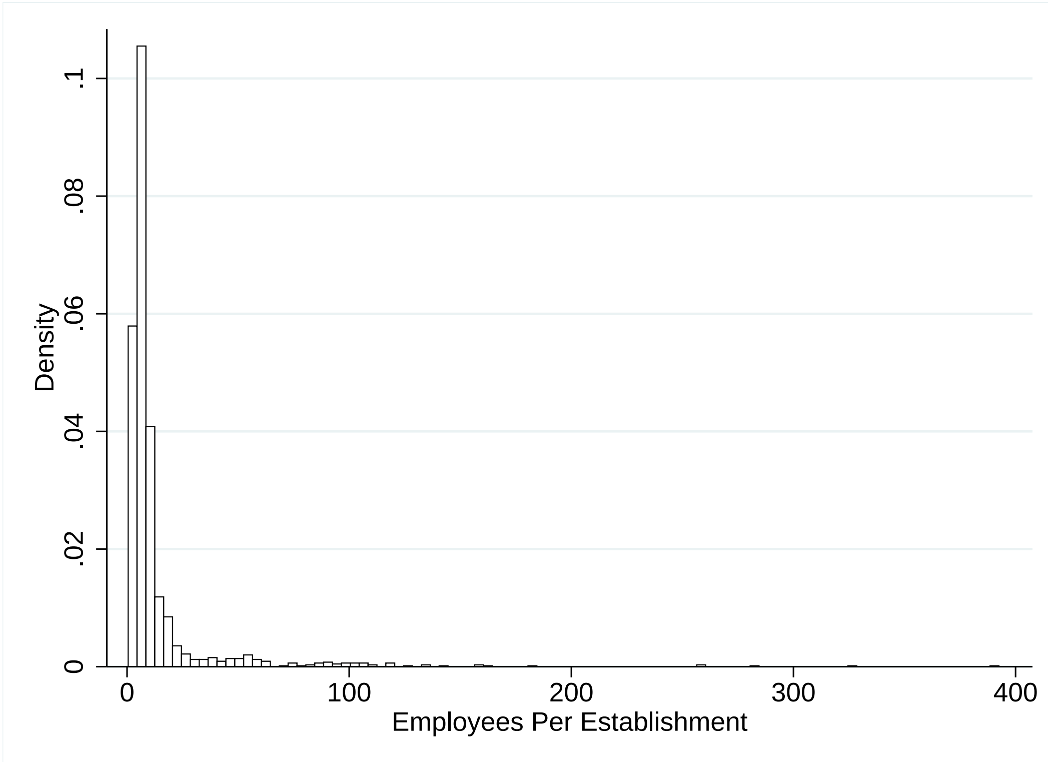


Figure 3.2: Histogram of establishment sizes across Florida's 6 digit NAICS retail industries. Each bin has a width of 4 employees

SUSB gives establishment counts for establishments with 1-4, 5-9, 10-19, 50-99, 100-249, 250-499, 500-99, and 1000 plus employees. A histogram of establishment sizes across industries and counties is given in (Figure 3.2). Inspecting this histogram, it shows that most industries have establishments with fewer than 50 employees, so I (arbitrarily) choose this level as the cutoff to classify an establishment as "large". Two of the establishment size measures I construct are the proportion of establishments with employment exceeding the values in a given SUSB bin. For example, the variable $Proportion50+$ is constructed by taking the count of establishments with 50 or more employees in a given industry-county-year, and dividing it by the total establishment count in the industry-county-year. An identical procedure is used for the variable $Proportion100+$. A graphical description of the distribution of the proportion based establishment size measures is found in the histogram in (Figure 3.3). From this

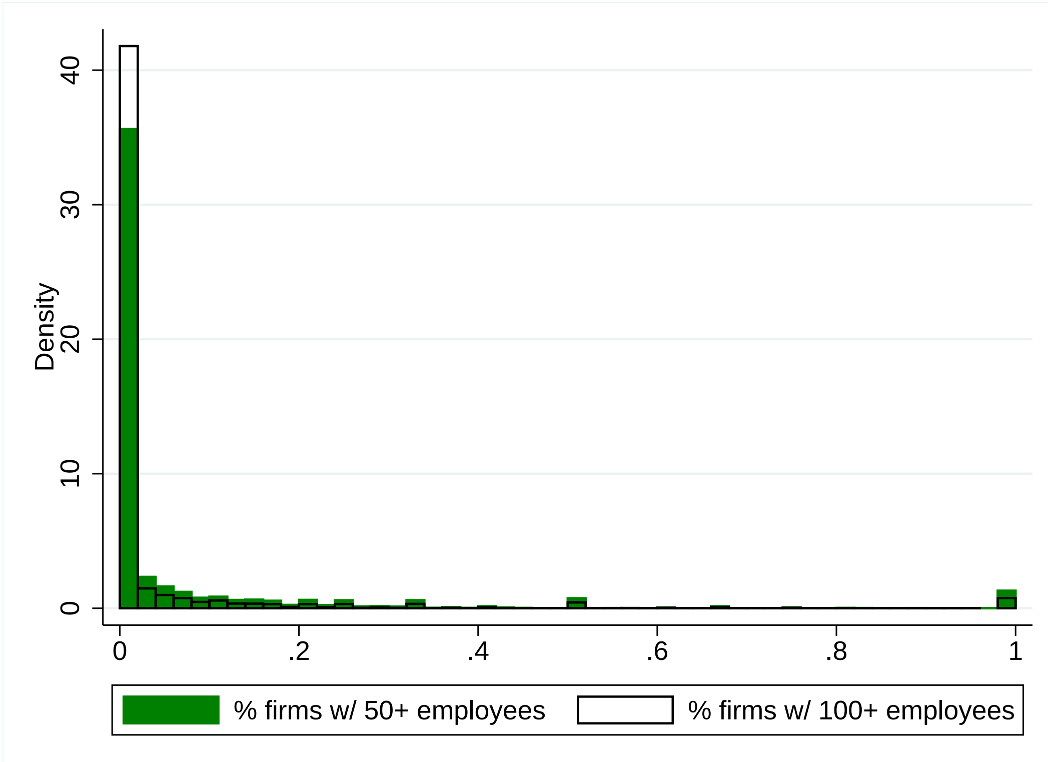


Figure 3.3: A histogram of Proportion 50+ and Proportion 100+ across Florida's 6 digit NAICS retail industries

histogram, I note that most observations have fewer than 20% of establishments falling into these "large" classification. The "fat tail" in (Figure 3.3) also shows the majority of the time large retail firms compose a relatively small share of total retail establishments, there are select occasions where they compose the vast majority of total retail. A third establishment size measure is calculated by dividing the number of employees in a given county and industry by the total number of establishments in that county and year. This variable is reported in (Table 3.1) as "Establishment Size", where the mean establishment size is approximately 7 employees. The histogram of this retail establishment size measure is given in (Figure 3.2), which again shows that the mean and median establishment size (in employment terms) is certainly less than 100 employees. In fact, the descriptive statistics in (Table 3.1) suggest that the average retail firm size has approximately 7 employees and that (supposing it behaves like a normal random variable just for argument sake) the majority of Florida's retail establishments have fewer than $6.6 + 2 * \sqrt{19.39} \approx 15$ employees.

The "Walmart effect" hypothesis would be that as counties experience growth in economic activity we should see a reallocation within retail industries towards larger establishments. Additionally, counties with more variation in economic activity may have a more diverse consumer base, and therefore are attractive to larger retail establishments that may offer a collection of higher order goods and services. As economic activity becomes distributed in a less homogeneous fashion, I expect to see an increase in the various establishment size measures. What is really being measured by the standard deviation of night light intensity? Consider (Figure 3.5), which reproduces the night light data on 6 Florida counties: 3 with the most homogeneity (lowest σ) of economic activity and 3 with the most heterogeneity (highest σ) in economic activity. It appears that the dispersion of night lights are positively correlated with the mean night light intensity. This is confirmed by the strong, almost linear relationship evidence in (Figure 3.6), which is a scatter plot of the standard deviation of night light against the mean night light intensity within all counties and years in the sample. What is clear from (Figure 3.6) is that counties with lower absolute levels of economic activity also have a more uniform distribution of economic activity, while increasing mean economic activity correlates highly with more spatial variation in the activity.

.

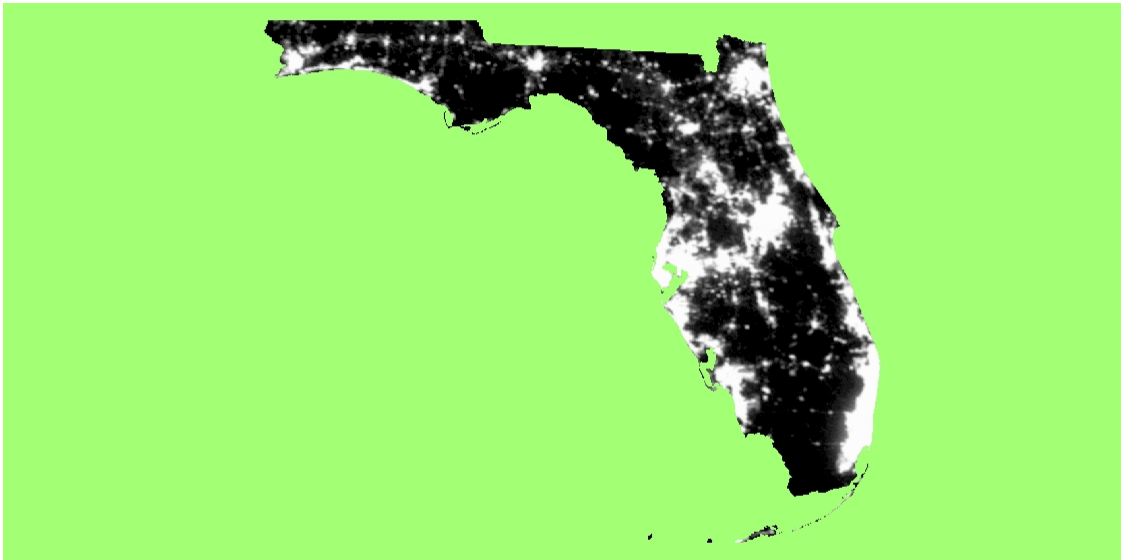


Figure 3.4: DMSP image of Florida's night light activity

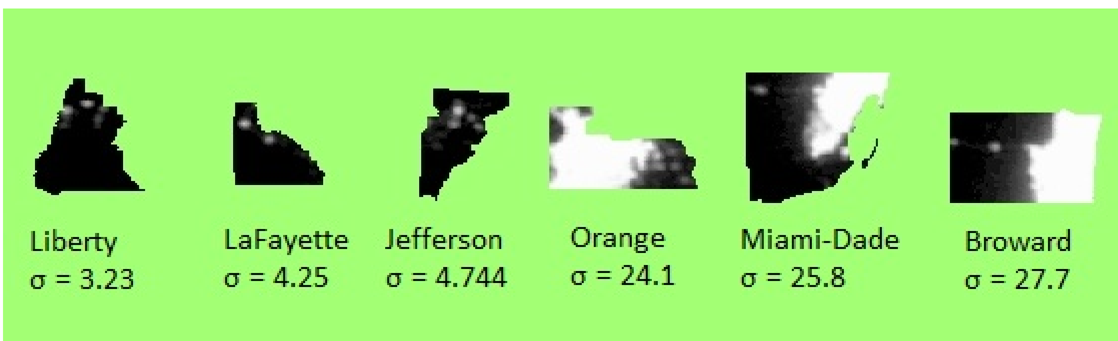


Figure 3.5: Florida's most and least uniform economic activity counties by night light activity

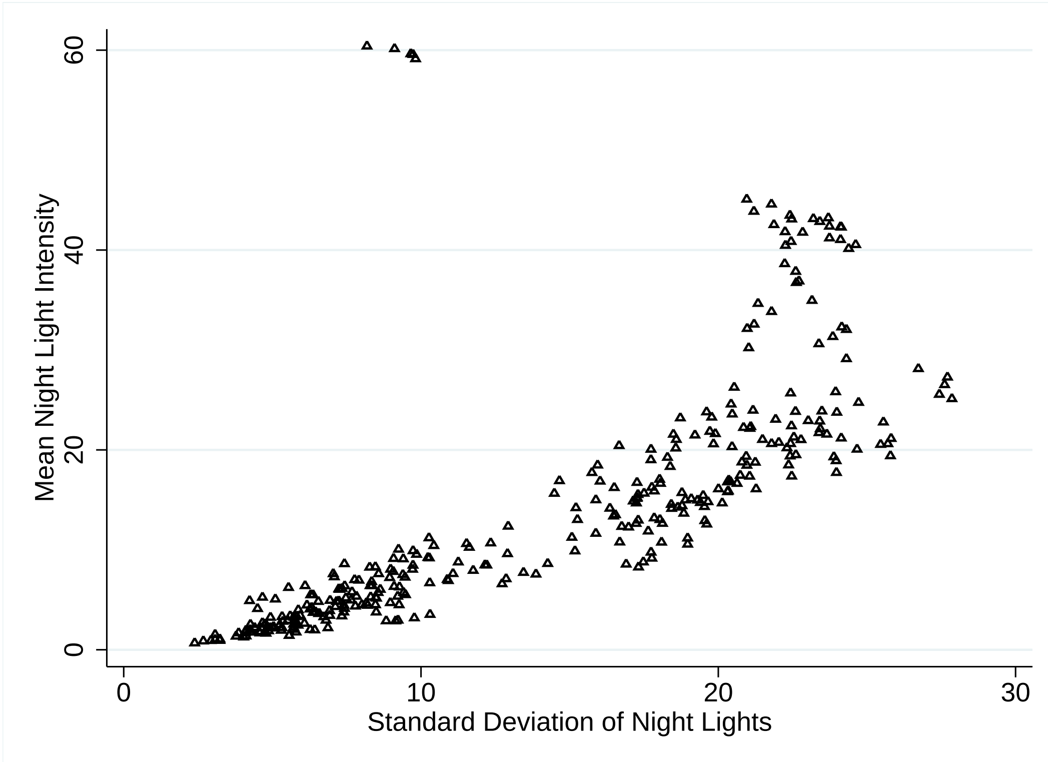


Figure 3.6: Scatter of mean night light intensity against standard deviation of night lights, county level

3.2.2 Specification

I estimate a dynamic panel model via the process described in (Blundell and Bond, 1998) and (Blundell and Bond, 1999). I consider the model

$$\begin{aligned} r_{ict} &= \alpha r_{ict-1} + x_{ct}\beta + u_{ict} \\ &= \alpha r_{ict-1} \\ &\quad + \beta_1 \Delta lights_{ct} + \beta_2 \Delta lights_{ct-1} \\ &\quad + \beta_3 \sigma_{ct} + \beta_4 \sigma_{ct-1} \\ &\quad + \theta_t + \eta_{ic} + e_{ict} \end{aligned} \tag{3.1}$$

The index ict represents retail industry i in county c during year t. Each panel is an industry-county pair ic observed through time. All of the independent variables vary at the county-year level except for the autoregressive term which varies at the industry level. The typical autoregressive process is augmented with effects like θ_t which represent a systematic shock for the year t common to all industries and counties. η_{ic} is the panel fixed effect representing one of the many reasons that OLS would yield biased estimates. Thus it will be necessary to develop an approach that only uses variation within a panel in order to mitigate any such confounding effects that are time invariant.

An immediate concern is that an economic activity indicator like the night light intensity would have a lot of persistence, perhaps even exhibit a unit root. As a precaution against spurious results, and after a regression to confirm the unit root suspicion, I take the first difference of night light activity in order to make it stationary. This was not necessary for the standard deviation measure, which does not exhibit unit root behavior. Included as controls are a full set of year indicators capturing shocks common to all industry county pairs in a given year. I supplement the traditional GMM style instruments with an interaction between 3 period lagged first difference (growth) of night lights and 3 period lagged first difference of the standard deviation of night lights ($\Delta lights_{ct-3} * \Delta\sigma_{ct-3}$).

3.3 Results and Discussion

Examining the results in (Table 3.2), the fact that the parameter on the lagged dependent variable α is statistically significant suggests that there is a dynamic adjustment process by retail establishments to regional economic growth. The estimate of the autoregressive parameter is far from 1, but still significant, suggesting that there is not a high degree of persistence in retail establishment sizes. If it is actually the case that retail labor is adjusted freely in accordance with current market conditions, a low degree of persistence in retail establishment size is expected. The OLS results in

	(1) Dynamic Panel	(2) OLS	(3) OLS
$establishment\ size_{ict-1}$	0.237***	0.770***	-
	(3.30)	(26.17)	
$\Delta lights_{ct}$	0.132*	-0.278*	-0.464**
	(2.40)	(-2.44)	(-2.82)
$\Delta lights_{ct-1}$	0.0289	-0.186*	-0.489**
	(0.59)	(-2.17)	(-3.01)
σ_{ct}	0.164***	0.00274	-0.280
	(4.03)	(0.03)	(-1.53)
σ_{ct-1}	-0.0261	0.0854	0.739***
	(-0.70)	(0.79)	(4.05)
Constant	1.204*	0.304	-0.859
	(2.15)	(1.18)	(-1.67)
N	15527	22840	22840

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.2: Dynamic Panel and OLS estimation results

column 2 of (Table 3.2) would naively suggest that retail establishment sizes exhibit much more persistence than is the case for the dynamic panel estimation results. Since the autoregressive parameter is sufficiently far from 1, it should also be the case that the GMM style instruments do not suffer from the weakness problem typically encountered with highly persistent dependent variables and small time periods.

The first significant estimate in the results suggests that holding all independent variables constant, changes in the mean night light intensity are associated with larger retail establishments. This suggests that counties with growing economic activity experience contemporaneous growth in the size of the retail establishments within the county. Evidence of this is provided in the significant and positive estimate for the parameter on the change in night light activity ($\Delta lights_{ct}$). Interpreting these results goes as follows: for a every unit increase in the growth of night light activity, there is an increase of approximately 0.1 employees per firm in the current period. Obviously this does not mean that every firm adds a fraction of a worker, but that representative retail establishments already competing in the county are growing larger.

The autocorrelation parameter estimate (first row in Table 3.2) suggests that each short run effect is approximately $\frac{1}{1-0.237} = 1.31$ times as large in the long run. Concerning the timing of adjusting labor inputs in response to market size growth, the coefficient on past economic activity growth is insignificant. A null result of this sort can have various meanings that I cannot explain in the current investigation. It could be that the effect is so small that my regression techniques have insufficient power to detect them. The fact that the contemporaneous growth effect β_1 is significant and positive suggests with the null result on the lagged effect can be easily rationalized from a business practice standpoint. Retail managers likely make scheduling and hiring decisions based on sales performance within the current year and likely place little weight on the past year's changes in economic activity. This is reasonable considering they have access to better data at higher frequency than that which is available through Census data.

Another economically significant result from (Table 3.2)is that a ceteris paribus increase in the dispersion of economic activity within the county is met with an increase in retail establishment size. The estimated partial correlation $\hat{\beta}_3 > 0$ suggests that counties experiencing increases in the spatial dispersion of economic activity are simultaneously experiencing concentration into larger retail establishments. An example of interpreting the estimated marginal effect $\hat{\beta}_3$ in an economic context would suggest that Liberty county, Florida's county with the lowest dispersion in economic activity, on average would have 4 fewer employees per retail establishment than Broward county purely on account of the effect of dispersion on retail sizes. OLS results would suggest that there is no dynamic relationship between these variables.

After estimating (3.1), I test (and reject) that the behavior of the residuals is $AR(2)$ in nature at the 95 % confidence level. I estimate standard errors via the two step process described in (Windmeijer, 2005). Looking across the columns in (Table 3.2) The biased OLS results in column 2 of (Table 3.2) would suggest that growth in economic activity within the county is associated with increasingly smaller retail establishments, which does not make economic or intuitive sense. Further, the OLS results would suggest that retail establishment size this period is relatively smaller if there was growth in economic activity in the previous period. The signs of the OLS coefficients speaks to the need for the instrumentation in the system GMM approach.

3.4 Conclusion

In this investigation I wanted to demonstrate the utility of a new source of market size data for analysis in industrial organization and regional economic analysis. This satellite data, that measures light at night at the square kilometer level, has many advantages over reported income based measures like those available through the Census or IRS. For instance, it allows for describing and measuring regional economic growth and the spatial distribution of all economic activity in an area including that which may not be reported. Empirically, I presented a model that tested a tenet of central place theory. My strategy was to use a panel of retail industry-county pairs through time to show that variation in the light activity within the county exhibits a positive partial correlation with average establishment size within the industry-county pair. There was sufficient evidence to suggest that economic growth is met with an increase in retail establishment sizes, suggesting that growth is absorbed on intensive margins.

The dynamic panel specification also allowed me to make a hypothesis about the timing of the changes in retail labor inputs to regional economic growth. Since retail firms are able to adjust their labor input very freely, only current changes in economic activity matter for establishment size (in employment terms, distant past changes in economic activity should be irrelevant for the size today. There was sufficient evidence to reject the hypothesis that past changes in economic activity, as measured by changes in night light activity in the past, have a significant partial correlation with retail

There are many things that I need to continue working on for this paper to improve in the future. The largest improvements would come with adopting better quality data, preferably at the firm level with quarterly or monthly data. This would allow a "higher resolution" investigation of the adjustment dynamics explored by this paper. It's doubtful that the night light data would be available at such high frequency. It's also very doubtful that reported income based measures are available at higher frequency than annually. Firm level data is only accessible under very close supervision.

REFERENCES

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of Californias Tobacco Control Program. *Journal of the American Statistical Association*, 105(490):493–505.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Second International Symposium on Information Theory*, pages 267–281.

Akaike, H. (1981). Likelihood of a model and information criteria. *Journal of Econometrics*, 16(1):3 – 14.

Allison, P. D. and Waterman, R. P. (2002). Fixed-effects negative binomial regression models. *Sociological methodology*, 32(1):247–265.

Alonso, W. (1964). *Location and land use: toward a general theory of land rent*. Publication of the Joint Center for Urban Studies. Harvard University Press.

Amemiya, T. (1985). *Advanced*. Harvard University Press.

Baier, S. L. and Bergstrand, J. H. (2007). Do free trade agreements actually increase members' international trade? *Journal of International Economics*, 71(1):72–95.

Bakos, J. Y. (1997). Reducing buyer search costs: Implications for electronic marketplaces. *Management Science*, 43(12):1676–1692.

Baltagi, B. H., Egger, P., and Pfaffermayr, M. (2003). A generalized design for bilateral trade flow models. *Economics Letters*, 80(3):391–397.

Banerjee, A. (1999). Panel data unit roots and cointegration: An overview. *Oxford Bulletin of Economics and Statistics*, 61(S1):607–629.

Barattieri, A., Basu, S., and Gottschalk, P. (2010). Some evidence on the importance of sticky wages. Working Paper 16130, National Bureau of Economic Research.

Barro, R. and Sala-I-Martin, X. (1991). Convergence across states and regions. *Brookings Papers on Economic Activity*, 1:107–182.

Bartik, T. J. (2015). How effects of local labor demand shocks vary with the initial local unemployment rate. *Growth and Change*, pages n/a–n/a.

Basker, E. (2005a). Job creation or destruction? labor market effects of wal-mart expansion. *Review of Economics and Statistics*, 87(1):174–183.

Basker, E. (2005b). Selling a cheaper mousetrap: Wal-Mart's effect on retail prices. *Journal of Urban Economics*, 58(2):203–229.

Beemiller, R. M. and Friedenberg, H. (1992). Regional multipliers: A user handbook for the regional input-output modeling system (rims ii). Technical report, Bureau of Economic Analysis, Washington, DC (United States). Regional Economic Analysis Div.

Bellemare, M. F., Masaki, T., and Pepinsky, T. B. (2015). Lagged Explanatory Variables and the Estimation of Causal Effects. MPRA Paper 62350, University Library of Munich, Germany.

Berry, B. J., Barnum, H. G., and Tennant, R. J. (1962). Comparative studies of central place systems. Technical report, DTIC Document.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How Much Should We Trust Differences-In-Differences Estimates? *The Quarterly Journal of Economics*, 119(1):249–275.

Blanchard, O. J., Katz, L. F., Hall, R. E., and Eichengreen, B. (1992). Regional evolutions. *Brookings Papers on Economic Activity*, 1992(1):pp. 1–75.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1):115–143.

Blundell, R. and Bond, S. (1999). GMM estimation with persistent panel data: an application to production functions. IFS Working Papers W99/04, Institute for Fiscal Studies.

Brakman, S. and Heijdra, B. J., editors (2001). *The Monopolistic Competition Revolution in Retrospect*. Cambridge University Press. Cambridge Books Online.

Brandolini, G. B. D. (1989). *Market Structure in the New Industrial Economics*, chapter Endogenous Sunk Costs and Industrial Structure. Oxford University Press.

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information- Theoretic Approach*. Springer.

Cairncross, F. (1997). *The Death of Distance: How the Communications Revolution Will Change Our Lives*. Harvard Business School Press.

Cameron, A. C. and Trivedi, P. K. (2007). *Essentials of Count Data Regression*, pages 331–348. Blackwell Publishing Ltd.

Card, D. (1990). The Impact of the Mariel Boatlift on the Miami Labor Market. *Industrial and Labor Relations Review*, 43(2):245–257.

Caselli, F. and Coleman, W. J. (2001). The u.s. structural transformation and regional convergence: A reinterpretation. *Journal of Political Economy*, 109(3):584–616.

Che, J. (2009). A dynamic model of privatization with endogenous post-privatization performance. *The Review of Economic Studies*, 76(2):pp. 563–596.

Clements, D. W. (1978). Utility of linear models in retail geography. *Economic Geography*, 54(1):pp. 17–25.

Dao, M., Furceri, D., and Loungani, P. (2014). Regional labor market adjustments in the united states and europe. *IMF Working Papers*, 14(26):1.

DaVanzo, J. (1978). Does unemployment affect migration? Evidence from micro data. *The Review of Economics and Statistics*, 60(4):504–514.

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity. *The American Economic Review*, 67(3):pp. 297–308.

Eckel, C. (2009). International trade and retailing. *Working Paper*.

Eckel, C., Eckel, D., and Singal, V. (1997a). Privatization and efficiency: Industry effects of the sale of British Airways. *Journal of Financial Economics*, 43(2):275–298.

Eckel, C., Eckel, D., and Singal, V. (1997b). Privatization and efficiency: Industry effects of the sale of british airways. *Journal of Financial Economics*, 43(2):275 – 298.

E.D.C. (2010). Space coast economic snapshot: Brevard gdp significant improvement.

Ellickson, P. B. (2006). Quality competition in retailing: A structural analysis. *International Journal of Industrial Organization*.

Ethier, W. J. (1982). National and International Returns to Scale in the Modern Theory of International Trade. *American Economic Review*, 72(3):389–405.

Forbes, J. D. (1972). Central place theory: An analytical framework for retail structure. *Land Economics*, 48(1):15–22.

Freund, C. and Weinhold, D. (2004). The effect of the internet on international trade. *Journal of International Economics*, 62(1):171 – 189.

Fuss, M. A. and Gupta, V. K. (1981). A cost function approach to the estimation of minimum efficient scale, returns to scale, and suboptimal capacity: With an application to canadian manufacturing. *European Economic Review*, 15(2):123 – 135.

Goldmanis, M., Goldmanis, M., Syverson, C., Syverson, C., Investments, P., and Investments, P. (2010). E-commerce and the Market Structure of Retail Industries. *The Economic Journal*, 120:651–682.

Goolsbee, A. (1999). In a world without borders: The impact of taxes on internet commerce. *Quarterly Journal of Economics*, 115:561–576.

Hall, R. E. (2005). Employment efficiency and sticky wages: Evidence from flows in the labor market. *Review of Economics and Statistics*, 87(3):397–407.

Haltiwanger, J., Jarmin, R., and Krizan, C. (2010). Mom-and-Pop meet Big-Box: Complements or substitutes? *Journal of Urban Economics*, 67(1):116–134.

Hausman, J., Hall, B. H., and Griliches, Z. (1984). Econometric models for count data with an application to the patents-r d relationship. *Econometrica*, 52(4):pp. 909–938.

Haworth, C. T. and Rasmussen, D. W. (1973). Determinants of Metropolitan Cost of Living Variations. *Southern Economic Journal*, 40(2):183.

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

Henderson, J. V., Storeygard, A., and Weil, D. N. (2012). Measuring economic growth from outer space. *American Economic Review*, 102(2):994–1028.

Jones, K. and Doucet, M. (2000). Big-box retailing and the urban retail structure: the case of the toronto area. *Journal of Retailing and Consumer Services*, 7(4):233 – 247.

Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. *American Economic Review*, 70(5):950–59.

Kumar, K. B., Rajan, R., and Zingales, L. (1999). What Determines Firm Size? *NBER Working Paper Series*, (July).

Laban, R. and Wolf, H. C. (1993). Large-scale privatization in transition economies. *The American Economic Review*, 83(5):pp. 1199–1210.

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics*, 95(2):391–413.

Matson, J. (2009). Why does nasa launch space shuttles from such a weather-beaten place? Scientific American. Interview w/ Dr. Roger Lanius (senior curator at the Smithsonian National Air and Space Museum).

Matyas, L. and Balzsi, L. (2013). The estimation of multi-dimensional fixed effects panel data models. (2012_2).

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models, Second Edition*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

Melitz, M. (2003a). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, pages 1695–1728.

Melitz, M. J. (2003b). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica*, 71(6):1695–1725.

Michalopoulos, S. and Papaioannou, E. (2013). Pre-colonial ethnic institutions and contemporary african development. *Econometrica*, 81(1):113–152.

Moretti, E. (2010). Local multipliers. *American Economic Review*, 100(2):373–77.

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica*, 46(1):69–85.

Peitz, M., Waldfogel, J., Lieber, E., and Syverson, C. (2012). Online versus offline competition.

Perry, D. B. M. K. (1994). Exclusive dealing in a spatial model of retail competition. *International Journal of Industrial Organization*, 12.

Phillips, P. C. and Perron, P. (1986). Testing for a unit root in time series regression. Cowles Foundation Discussion Papers 795R, Cowles Foundation for Research in Economics, Yale University.

Pielke, R. A. (1993). A reappraisal of the space shuttle programme. *Space Policy*, 9(2):133 – 157.

Raff, H. and Schmitt, N. (2009). International trade with heterogeneous retailers. *Working Paper*.

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 31(2):pp. 350–371.

Richardson, M. (2004). Trade policy and access to retail distribution. EPRU Working Paper Series 97-23, Economic Policy Research Unit (EPRU), University of Copenhagen. Department of Economics.

Rosenthal, S. S. and Strange, W. C. (2004). Chapter 49 evidence on the nature and sources of agglomeration economies. In *Cities and Geography*, volume 4 of *Handbook of Regional and Urban Economics*, pages 2119 – 2171. Elsevier.

Saal, D. S. and Parker, D. (2000). The impact of privatization and regulation on the water and sewerage industry in england and wales: a translog cost function model. *Managerial and Decision Economics*, 21(6):253–268.

Salop, S. C. (1979). Monopolistic competition with outside goods. *The Bell Journal of Economics*, 10(1):pp. 141–156.

Samuelson, P. A. (1954). The transfer problem and transport costs, ii: Analysis of effects of trade impediments. *The Economic Journal*, 64(254):pp. 264–289.

Shaffer, S. (2002). Firm size and economic growth. *Economics Letters*, 76:195–203.

Shaked, A. and Sutton, J. (1983). Natural Oligopolies. *Econometrica*, 51(5):1469–1483.

Shapiro, C. and Varian, H. (1999). *Information Rules: A Strategic Guide to the Network Economy*. Strategy/Technology / Harvard Business School Press. Harvard Business School Press.

Stevenson, B. (2008). The internet and job search. Working Paper 13886, National Bureau of Economic Research.

Stock, J. and Watson, M. (1989). New indexes of coincident and leading economic indicators. In *NBER Macroeconomics Annual 1989, Volume 4*, pages 351–409. National Bureau of Economic Research, Inc.

Stone, K. E. (1988). The effect of wal-mart stores oh businesses in host towhs and surrounding towns in iowa.

Sutton, J. (1991). *Sunk Costs and Market Structure*. The MIT Press.

Takeuchi, K. (1976). *The distribution of information statistic and the criterion of the adequacy of a model. Suri-Kagaku*, volume 153,12-18. Suri-Kagaku (Mathetmatic Sciences).

Tiebout, C. M. (1956). A Pure Theory of Local Expenditure. *Journal of Political Economy*, 64(5):416–424.

Vitt, D. C. (2015). The empirical relationships between travel costs, internet use intensity, and retail market structure. *Working Paper*.

Vogel, J. (2011). Spatial price discrimination with heterogeneous firms. *The Journal of Industrial Economics*, 59(4):661–676.

Vuong, Q. (1989). Likelihood ratio tests for model selection and non nested hypotheses. *Econometrica: Journal of the Econometric Society*, 57(2):307–333.

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics*, 126(1):25–51.

Wooldridge, J. (2002a). *Econometric Analysis of Cross Section and Panel Data*. Econometric Analysis of Cross Section and Panel Data. MIT Press.

Wooldridge, J. (2002b). *Econometric Analysis of Cross Section and Panel Data*. Econometric Analysis of Cross Section and Panel Data. MIT Press.

Wooldridge, J. (2007). Whats new in econometrics? lecture 6: Control functions and related methods.

Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data*. Econometric Analysis of Cross Section and Panel Data. MIT Press.

Yilmazkuday, D. and Yilmazkuday, H. (2015). Understanding Gasolin Price Dispersion. *Working Paper*.

APPENDICES

3.4.1 Chapter 1 Appendix

Figures

	(1) amazon intensity
porn intensity	0.0954
	(7.82)
population	0.000000433
	(23.20)
net access	0.0136
	(7.37)
wage	0.0551
	(3.69)
_cons	32.87
	(36.62)
N	5228
$F(4, 6528)$	177.53

t statistics in parentheses

Table 3.5: First stage IV results

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Market Structure (MS)	805.142	1172.527	3	12251	4007
Fitted MS- Neg. Binomial	805.142	1172.277	3.887	12092.043	4007
Fitted MS- Poisson	805.142	1172.277	3.887	12092.043	4007
Fitted MS- Linear	805.142	1172.023	-28.912	12046.549	4007
Fitted MS- log-log	811.866	1177.887	3.458	12208.912	3956

Table 3.3: Observed and fitted market structure summary statistics

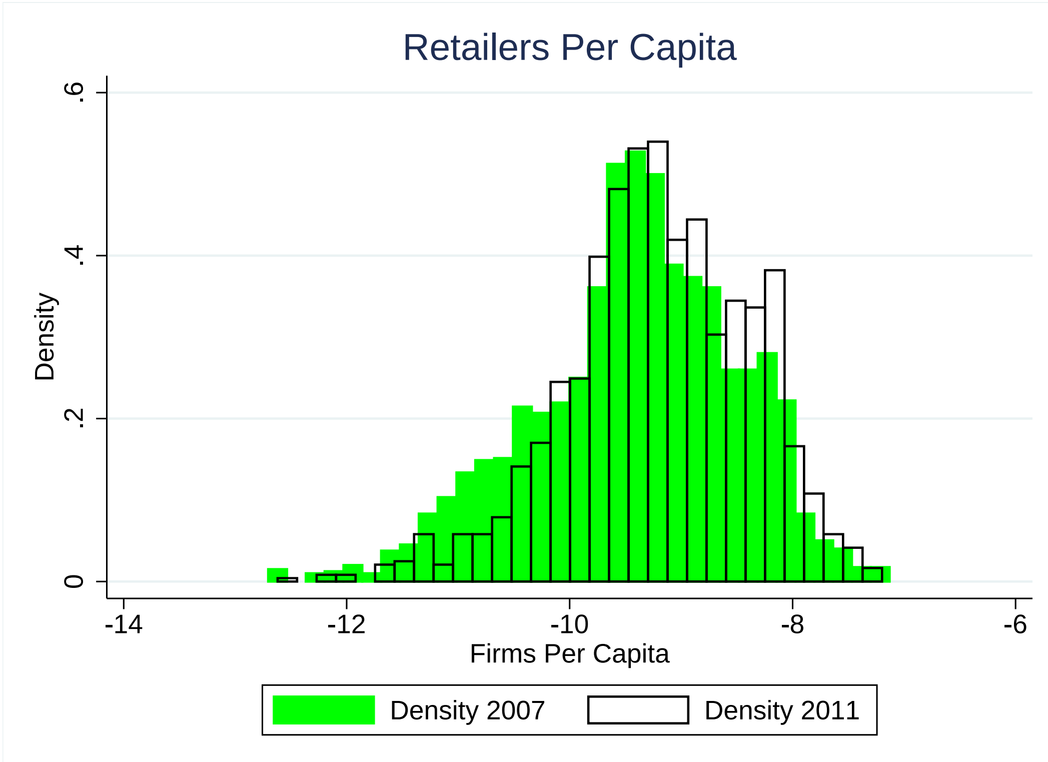


Figure 3.7: Variation in establishment counts per capita over time

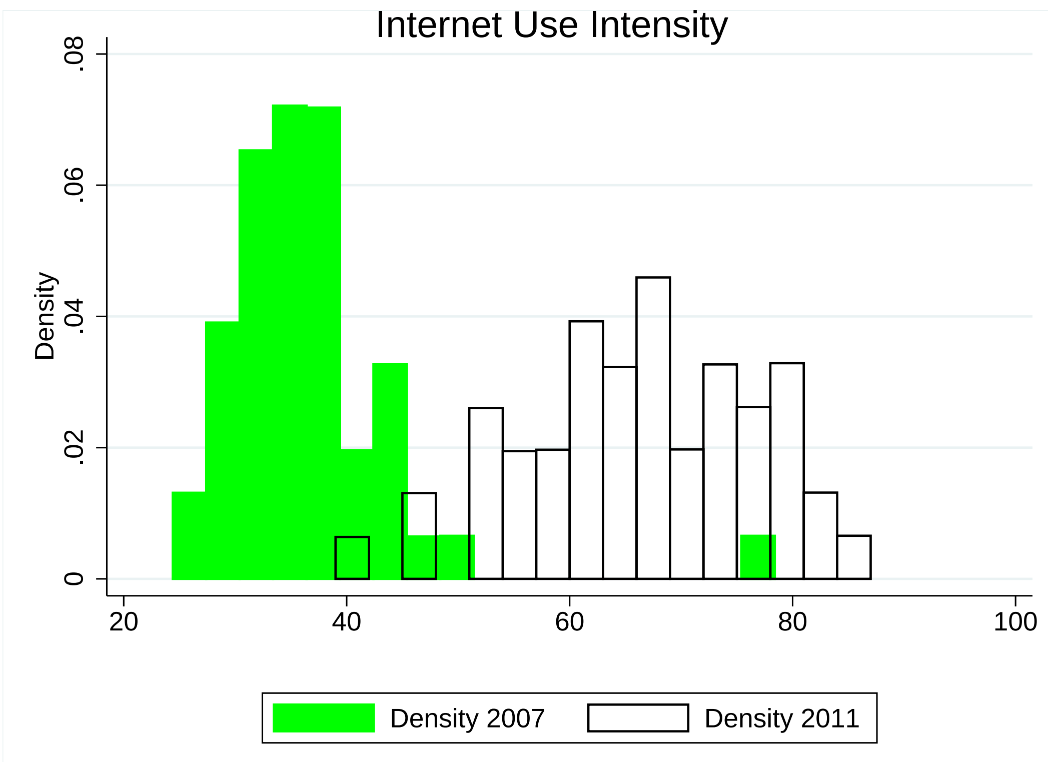


Figure 3.8: Variation in Google search frequency over time

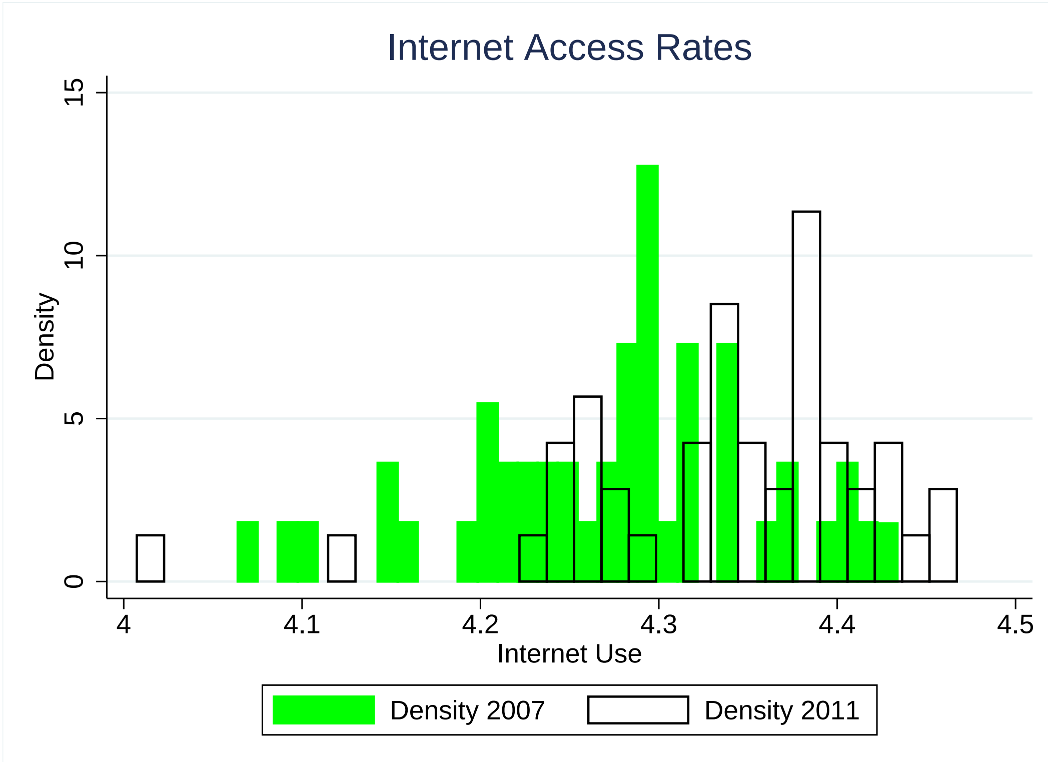


Figure 3.9: Variation in Internet access rates over time

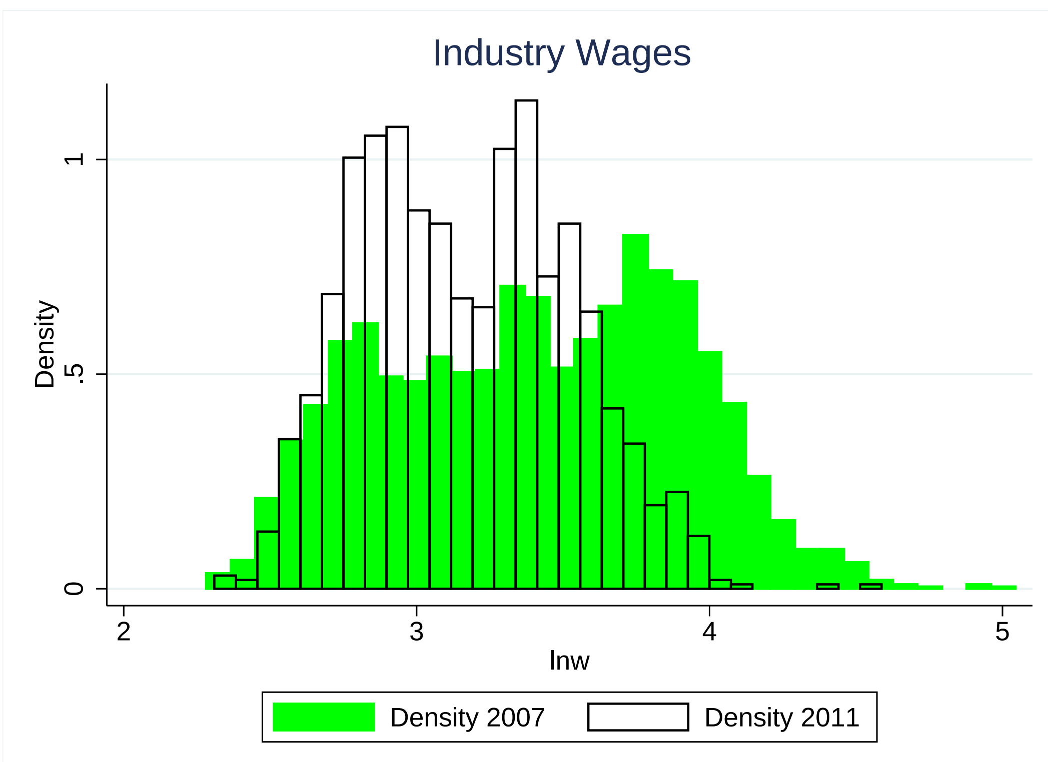


Figure 3.10: Variation in wages over time

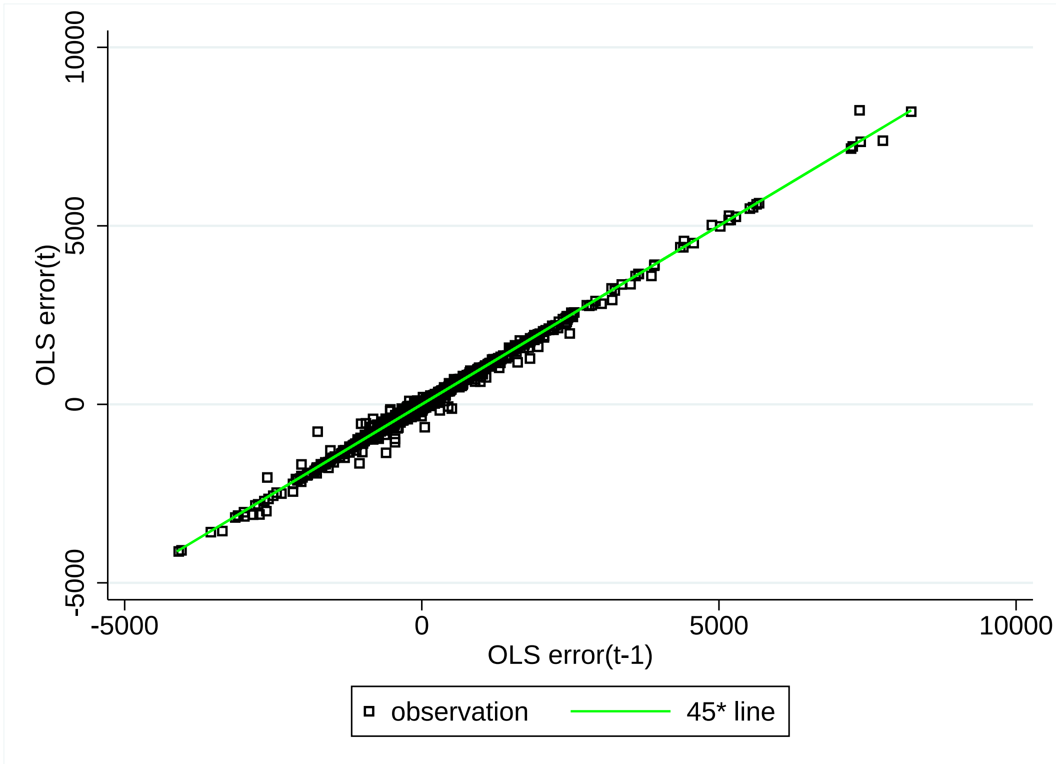


Figure 3.11: Autocorrelation in OLS errors

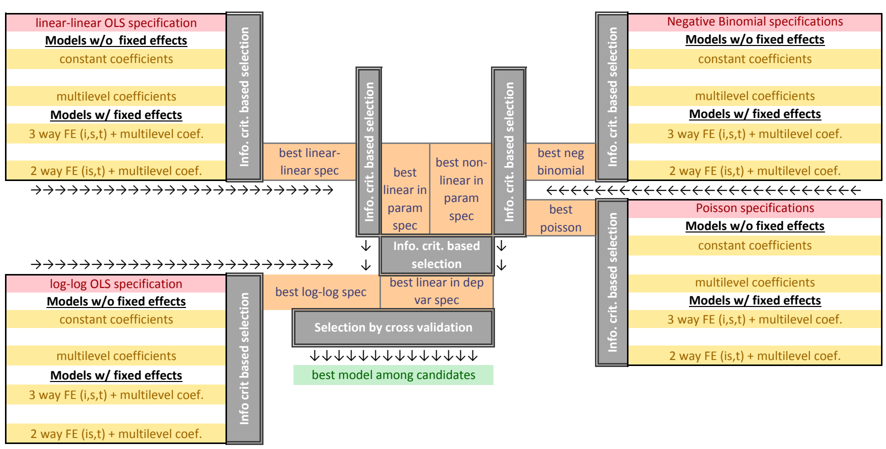


Figure 3.12: Model selection strategy

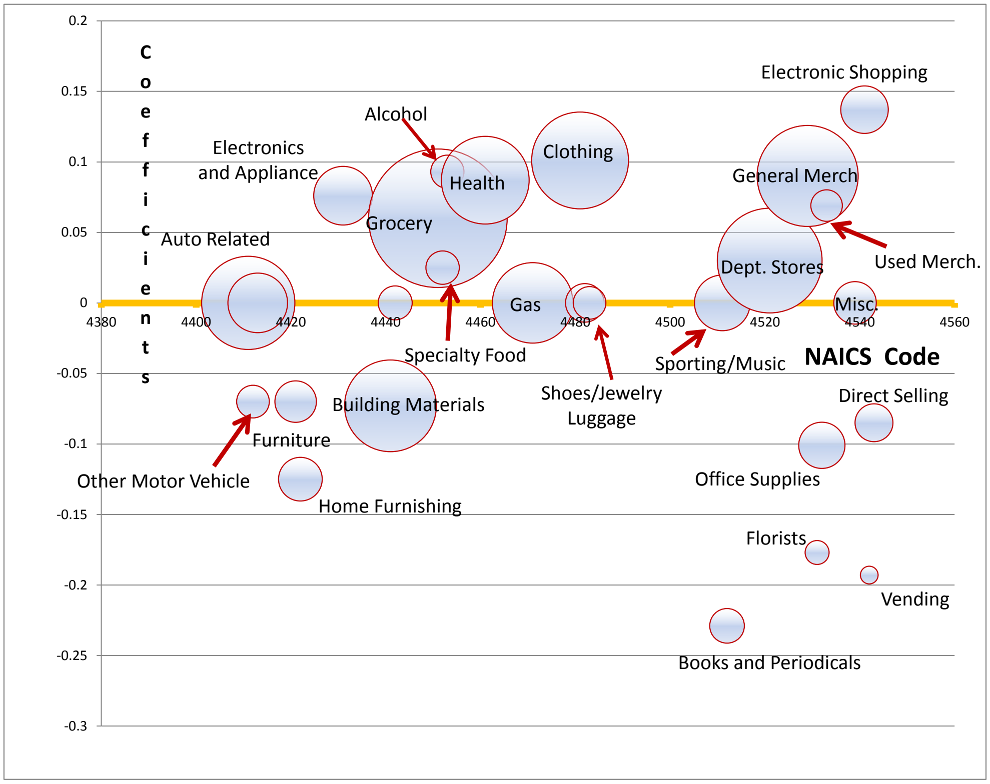


Figure 3.13: Internet exposure estimates

Variable	Mean	Std. Dev.	Min.	Max.	N
Actual Variance:Mean ratio	2.649	5.391	0	97.273	4007
Predicted Var:Mean ratio- Neg. Binomial	1.839	4.331	0	96.982	3988
Predicted Var:Mean ratio- Poisson	1.839	4.331	0	96.982	3988
Predicted Var:Mean ratio- Linear	3.742	7.25	0	110.541	3988
Predicted Var:Mean ratio- Log-Log	1.735	3.259	0	64.788	3913

Table 3.4: Dispersion summary statistics

3.4.2 Theory Appendix

As in Eckel (2009) and Salop (1979), I suppose the economy is populated by a mass of L consumers who live on a circle of circumference Ω around a business district. Consumers are distributed uniformly on this torus such that the population density is identical at all points, and given by $\frac{L}{\Omega}$. Consumers travel to one of R number of retail outlets located on the torus, and travel is costly so that consumers visit only a single outlet. Firms have endogenous "catchment" areas (δ to each side of the retailer) identified by the marginal consumer who is indifferent between two retail outlets. Catchment areas (i.e. 2δ) will be equal in a symmetric equilibrium, and the catchment area of all retailers must add up the to circumfrence of the circle. This gives the relation:

$$2\delta R = \Omega$$

The economy is characterized by N manufacturing firms across k identical states that serve the retail sector.

Consumers

Consumers are associated with the index l, and have CES preferences that include a taste for variety. Utility of each consumer l is represented by

$$U_l = \frac{1}{t_l} N^{1+\rho-\frac{\sigma}{\sigma-1}} \left(\sum_{i=1}^{N} x(i)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tag{3.2}$$

in which $x(i)$ are the i differentiated varieties of goods produced by manufacturing firms, $\sigma > 1$ represents the elasticity of substitution between varieties, $\rho \geq 1$ captures the love of variety as discussed in Brakman and Heijdra (2001). An example of how changes in the diversity preference parameter ρ is discussed in the Theory Appendix. Each consumer l is not freely mobile, he or she faces a travel cost represented by

t_l similar to the iceberg type costs described in (Samuelson, 1954). This travel cost depends on the distance δ_l between the residence of consumer l and the retailer they visit. As a result of this costly travel, the consumer will have a strong preference for shopping at a single "catch-all" destination. Thus, travel costs are convex and follow this specific form:

$$t_l = exp(\tau\delta_l)$$

The parameter τ captures exogenous influences on the travel costs to the consumer, like that of the price of fuel or public transport, and represents a general measure of travel costs and the mobility of consumers. Each household supplies one unit of labor inelastically. I take it as the numeraire so that the wage rate is normalized to one. The utility function above is maximized subject to the budget constraint $\sum_N p(i)x(i) \leq 1$ yielding an individual demand

$$x(i) = p(i)^{-\sigma} \left(\sum_{\phi=1}^{N} p(\phi)^{1-\sigma} \right)^{-1} \tag{3.3}$$

In a symmetric equilibrium, Eckel notes that the price elasticity of this demand reduces to

$$-\frac{dlnx}{dlnp} = \sigma(1 - \frac{1}{N}) + \frac{1}{N}$$

This price elasticity is a weighted average of the substitution effect characterized by σ, and the unit income effect.

Love of variety

Consumers have a "love of variety" or a preference towards a diversified consumption basket that is represented by the parameter ρ in their utility function:

$$U_l = \frac{1}{t_l} N^{1+\rho-\frac{\sigma}{\sigma-1}} \left(\sum_{i=1}^{N} x(i)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tag{3.4}$$

As stated in (Brakman and Heijdra, 2001), this was a feature of (Dixit and Stiglitz, 1977) that did not make it to the final manuscript, however was later adopted in (Ethier, 1982).

To understand how this parameter represents a love of variety, consider a consumer with utility represented by (3.4) and income m who faces $i = 1 \ldots N$ number of differentiated products with identical price p. Consider two hypothetical consumption bundles, $A = (x_1, x_2, \ldots, x_N) = (\frac{m}{p}, 0, \ldots, 0)$ and $B = (x_1, x_2, \ldots, x_N) = (\frac{1}{N}\frac{m}{p}, \frac{1}{N}\frac{m}{p}, \ldots, \frac{1}{N}\frac{m}{p})$. The first bundle differs from the second in that all income is exhausted on a single good, while in the second bundle income is split equally among all goods available. The indirect utility function V_A from bundle A is given by evaluating the utility function at the consumption vector A, symmetric price p, and income m

$$\begin{aligned} V_A(p,m) &= U(\frac{m}{p}, 0, \ldots, 0) \\ &= \frac{1}{t_l} 1^{1+\rho-\frac{\sigma}{\sigma-1}} \left(\frac{m}{p}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \\ &= \frac{1}{t_l} \left(\frac{m}{p}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \end{aligned} \tag{3.5}$$

Indirect utility for bundle B is expressed similarly:

$$
\begin{aligned}
V_B(p,m) &= U(\frac{1}{N}\frac{m}{p},\frac{1}{N}\frac{m}{p},\ldots,\frac{1}{N}\frac{m}{p}) \\
&= \frac{1}{t_l}N^{1+\rho-\frac{\sigma}{\sigma-1}}\left(\frac{m}{p}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}
\end{aligned}
\tag{3.6}
$$

Proposition 5. *The consumer preference for diversity parameter, ρ, does not appear in the expression for the consumer's marginal rate of substitution and therefore does not directly influence consumer choice.*

Proof: Let C represent the constant terms $\frac{1}{t_l}N^{1+\rho-\frac{\sigma}{\sigma-1}}$. Then the marginal utility associated with good x_i can be expressed as $C * \frac{\partial}{\partial x_i}\left(\sum_{i=1}^{N} x(i)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$ and the marginal utility of x_j expressed as $C * \frac{\partial}{\partial x_j}\left(\sum_{i=1}^{N} x(i)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$ so that the MRS between goods i and j can be expressed as the following

$$
\begin{aligned}
MRS_{ij} &= \frac{\frac{\partial U_l}{\partial x_i}}{\frac{\partial U_l}{\partial x_j}} \\
&= \frac{C * \frac{\partial}{\partial x_i}\left(\sum_{i=1}^{N} x(i)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}}{C * \frac{\partial}{\partial x_j}\left(\sum_{i=1}^{N} x(i)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}} \\
&= \frac{\frac{\partial}{\partial x_i}\left(\sum_{i=1}^{N} x(i)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}}{\frac{\partial}{\partial x_j}\left(\sum_{i=1}^{N} x(i)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}}
\end{aligned}
\tag{3.7}
$$

Since the constant terms represented by C cancel, the marginal rate of substitution does not depend on the consumer preference for diversity parameter ρ.

Proposition 6. *The ratio of consumer utility from consuming a basket with equal income share on all available goods to the utility of exhausting income on a single good is increasing can be varied independently of the elasticity of substitution between goods (represented by σ).*

Proof: Take the ratio of (3.5) to (3.6)

Manufacturing

Manufacturing firms produce individual varieties of the differentiated goods under an increasing returns structure. Profits for a manufacturing firm are given by:

$$\Pi_M = (p_w - \beta)Q - \alpha \tag{3.9}$$

where Q is world market demand given by

$$Q = x * L * k$$

which is the individual demand, x, by the L consumers in each of the k states. β and α are the variable and fixed costs facing the manufacturer, and are denominated in labor units. The wholesale price that the retailers will face from the manufacturer is represented by p_w. In the Eckel (2009), retail mark-ups are considered exogenous to the manufacturing firm, and are treated as given [1]. Profit maximization by manufacturing firms yields the optimal wholesale price that retailers face:

$$p_w = \beta \left[1 + \frac{N}{(\sigma - 1)(N - 1)} \right] \tag{3.10}$$

[1] This assumption is made for simplicity, there is scope for research where manufacturers have an option to commit their retailers to a given price by printing the price directly on the p. This is a practice that is sometimes seen with potato chips, beverages, and some baked goods. Nevertheless, relaxing it is outside the interest of the current investigation.

Manufacturing markups depend on the elasticity between varieties, represented by the parameter σ, and the variety of products available, N. Since each manufacturer only produces a single variety, this parameter also the number of manufacturing firms in the global economy. Equilibrium in the manufacturing industry is given by the free entry zero profit condition:

$$kL = \alpha(1+\mu)(\sigma N - \sigma + 1) \tag{3.11}$$

which includes the retail markup μ, consumer preference parameters σ, the firm/variety count N of differentiated product manufacturers, manufacturing fixed costs α, the world population L and the state count k. In turn, equation (3.11) can be solved for the variety/manufacturing firm count N as a function of exogenous and endogenous variables

$$N(\sigma, \alpha, k, L; \mu) = \frac{kL}{\sigma\alpha(1+\mu)} + 1 + \frac{1}{\sigma} \tag{3.12}$$

Equation (3.12) shows the nature of the relationship between retailers and manufacturers through retail markups μ as a determinant of manufacturing firm counts N.

Retail

Retailers purchase the intermediate product from manufacturers at wholesale prices p_w. They combine this intermediate product with a labor input at a martginal cost of γ per variety, which I liken to a "wage" since it is the income paid to the labor supply. This constant marginal cost implies a constant returns to scale structure for the retailers. Keeping the model parsimonious with constant returns to scale provides a puzzling market structure result discusse. In order to determine the number of retailers in an equilibrium, we constrain them to a standard zero profit condition. Profits are given by the expression:

$$\Pi_R(j) = \underbrace{\underbrace{2\delta_j \frac{L}{\Omega}}_{\text{catchment}} \left[\sum_{i=1}^{N_j} [p(i) - p_w]\, x(i) \right]}_{\text{unit profit}} - \underbrace{\gamma N_j}_{costs} \tag{3.13}$$

The first term on the right hand side of (3.13) consists of the gross profit per consumer, $\sum_{i=1}^{N_j} [p(i) - p_w]\, x(i)$, scaled by the relevant consumer count in the catchment area. To arrive at the count given in (3.13), recall each location on the circle has population density parameterized as $\frac{L}{\Omega}$, and the"length" of the catchment area $2\delta_j$. From these profits the firms must pay their labor input. The second term, γN_j is this "cost of provision", as it is the labor expenditure from providing N_j varieties of the differentiated good at a marginal cost of per variety. The equilibrium number of retailers is given by the zero profit condition of the retailing industry subject to free entry:

$$\Pi_R(j) = 0 \rightarrow \frac{\tau\Omega}{R + \tau\Omega} L = R\gamma \tag{3.14}$$

After manipulation, it becomes clear that the result of applying the zero profit condition to (3.13) as done in (3.14) results in an equation quadratic in R

$$\gamma N R^2 + \gamma\tau N \Omega R - \tau\Omega L = 0 \tag{3.15}$$

Using the quadratic equation, and taking the positive root of this quadratic expression yields the number of firms in operating in retail R as a function of the exogenous parameters γ, L, Ω, τ and the endogenous variable N.

$$R(\gamma, L, \Omega, \tau; N) = \frac{\sqrt{\tau\Omega}\sqrt{4L + \gamma\tau\Omega N}}{2\sqrt{\gamma N}} - \frac{\tau\Omega}{2} \tag{3.16}$$

Labor Market Equilibrium

Solving for equilibrium in the labor market requires showing that retail labor supply, which was assumed to be immobile between regions and exogenously given by L in each state, and is in accordance with labor demand, represented by the sum of labor demand from retailers and manufacturers. All costs for both manufacturers and retailers are denominated in retail terms, so we can express labor demand as the sum of demand from both sectors. The manufacturing sector in each state demands $(\alpha + \beta Q)$ units of labor for each of its $\frac{N}{k}$ firms, combined with the γN units of retail demanded by each of the R^* retailers.

$$\frac{N^*}{k}(\alpha + \beta Q) + R^* \gamma N^* = L \tag{3.17}$$

In the general equilibrium, the equilibrium number of manufacturers N^* and retailers R^* are functions of only the exogenously determined parameters. The former is given in (3.1), while the latter is expressed as

$$N^*(L, \alpha, \gamma, \Omega, \tau, \sigma) = \frac{(\alpha(\sigma-1) + kL)\left(\alpha\gamma(\sigma-1)\tau\Omega + \sqrt{\alpha\gamma\tau\Omega\left(4kL^2\sigma + \alpha(\sigma-1)(\gamma(\sigma-1)\tau\Omega + 4L\sigma)\right)}\right)}{\alpha\sigma\left((\sigma-1)(\alpha\gamma\tau\Omega) + \sqrt{\alpha\gamma\tau\Omega\left(4kL^2\sigma + \alpha(\sigma-1)(\gamma(\sigma-1)\tau\Omega + 4L\sigma)\right)} + 2\gamma kL\tau\Omega\right)} \tag{3.18}$$

The labor demand in the general equilibrium is given by substituting (3.1) and (3.18) in (3.17) and solving for L^* as a function of the parameters. This is akin to finding the L^* that for which there is zero excess demand for labor, given by the expression $\frac{N^*(L^*)}{k}(\alpha + \beta Q) + R^*(L^*)\gamma N^*(L^*) - L^* = 0$. Making these substitutions and simplifying, the equilibrium labor demand is the L^* that is the root of the following equation

$$
\begin{aligned}
0 = \beta k^2 L^2 x & \left(\alpha\gamma(\sigma-1)\tau\Omega + \sqrt{\alpha\gamma\tau\Omega\left(4kL^2\sigma + \alpha(\sigma-1)(\gamma(\sigma-1)\tau\Omega + 4L\sigma)\right)}\right) \\
& + \alpha^2(\sigma-1)\left(\alpha\gamma(\sigma-1)\tau\Omega + \sqrt{\alpha\gamma\tau\Omega\left(4kL^2\sigma + \alpha(\sigma-1)(\gamma(\sigma-1)\tau\Omega + 4L\sigma)\right)}\right) \\
& + \alpha kL(\sigma-1)\left((\beta x - 1)\sqrt{\alpha\gamma\tau\Omega\left(4kL^2\sigma + \alpha(\sigma-1)(\gamma(\sigma-1)\tau\Omega + 4L\sigma)\right)} + \alpha\gamma\tau\Omega(\sigma + \beta(\sigma-1)x + 1)\right) \\
& * \left(\alpha k\sigma\left(\alpha\gamma\sigma\tau\Omega - \alpha\gamma\tau\Omega + \sqrt{\alpha\gamma\tau\Omega\left(4kL^2\sigma + \alpha(\sigma-1)(\gamma(\sigma-1)\tau\Omega + 4L\sigma)\right)} + 2\gamma kL\tau\Omega\right)\right)^{-1}
\end{aligned}
\tag{3.19}
$$

Comparative Statics

To determine the general equilibrium effect of a change in consumer travel costs on the number of retailing firms, I take the partial derivative of (3.1) with respect to the travel cost parameter τ:

$$\frac{dR^*}{d\tau} = \frac{\alpha\Omega(2kL^2\sigma + (\sigma-1)(2L\alpha\sigma + \alpha\gamma\tau\Omega(\sigma-1) - \Psi)}{2(kL + \alpha(\sigma-1)\Psi)}$$

where I have made the following substitution:

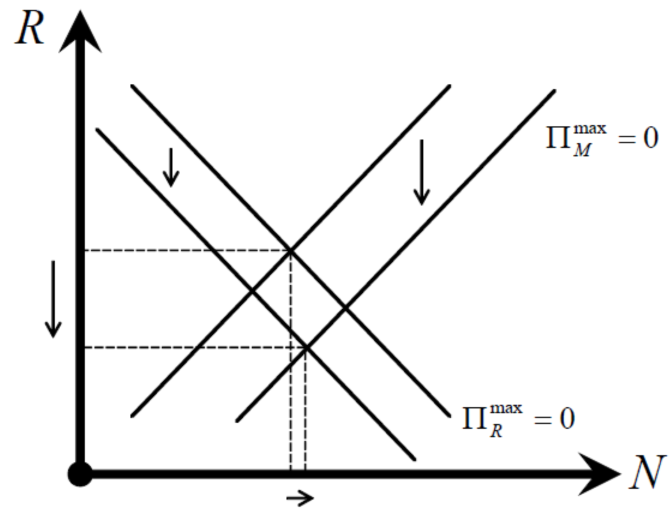
$$\Psi = \sqrt{\alpha\gamma\tau\Omega(4kL^2\sigma + \alpha(\sigma-1)(4L\sigma + \gamma(\sigma-1)\tau\Omega)}$$

In Eckel (2009) it is easily shown that

$$\frac{dR^*}{d\tau} > 0$$

which suggests that the equilibrium market structure measure for retail, R^*, move in the same direction as the change in transportation costs. Therefore, a decrease in consumer travel costs $d\tau < 0$ will be met with a consolidation in retail $dR^* < 0$. Note that this is the effect on the fully endogenized number of retailers, and the transmission mechanism is as follows: the decrease in travel costs leads to lower

Figure 3.14: Equilibrium effects of consumer travel cost (mobility) shock



markups, which increases demand for manufacturer's goods, shifting their zero profit line down as in 3.14, and shifting the retail zero profit condition downwards. Markups decrease because the now relatively more mobile consumers are less constrained to shop than before, and hence local retailers have relatively less market power over the local consumers than before. This will lead to the classic reallocation within industry, where the less productive retailers exit in favor of the more competitive retailers as in (Goldmanis et al., 2010).

It can also be shown that as the population increases so too will the number of retailers on account of increased local sales, $\frac{dR^*}{dL} > 0$:

$$\frac{dR^*}{dL} = \frac{\alpha(\sigma-1)\tau\Omega\left(2\alpha^2(\sigma-1)\sigma + k\left(-\alpha\gamma\sigma\tau\Omega + \alpha\gamma\tau\Omega + \sqrt{\alpha\gamma\tau\Omega\left(4kL^2\sigma + \alpha(\sigma-1)(\gamma(\sigma-1)\tau\Omega + 4L\sigma)\right)} + 2\alpha L\sigma\right)\right)}{2(\alpha(\sigma-1)+kL)^2\sqrt{\alpha\gamma\tau\Omega\left(4kL^2\sigma + \alpha(\sigma-1)(\gamma(\sigma-1)\tau\Omega + 4L\sigma)\right)}} \tag{3.20}$$

3.4.3 Statistical Appendix

Hypothesis Testing

See web appendix at bit.ly/davidvitt

3.4.4 Replication

Mathematica code

See web appendix at bit.ly/davidvitt

Econometric replication

See web appendix at bit.ly/davidvitt

3.4.5 Estimates

Please refer to the NAICS code list below when referencing the regression estimates below. Table 4 is the estimate of (1.7) with heteroskedasticity robust standard errors. Note that $GEOid2$ is a variable that indexes the states plus the District of Columbia as well as Puerto Rico. These correspond to the state codes as given in the CBP, and hence run from 1-72 with some values being skipped (since we really only need 52 unique values in this range). The $GEOid2$ values in ascending order corresponding to the states + DC listed in alphabetical order (so for all DC observations $GEOid2 = 1$)), with Puerto Rico last at 72.

4411	Automobile dealers
4412	Other motor vehicle dealers
4413	Auto parts, accessories, and tire stores
4421	Furniture stores
4422	Home furnishings stores
4431	Electronics and Appliance Stores
4441	Building Material and Supplies Dealers
4442	Lawn and garden equipment and supplies stores
4451	Grocery stores
4452	Specialty food stores
4453	Beer, wine, and liquor stores
4461	Health and personal care stores
4471	Gasoline stations
4481	Clothing stores
4482	Shoe stores
4483	Jewelry, luggage, and leather goods stores
4511	Sporting goods and musical instrument stores
4512	Book, periodical, and music stores
4521	Department stores
4529	Other general merchandise stores
4531	Florists
4532	Office supplies, stationery, and gift stores
4533	Used merchandise stores
4539	Other miscellaneous store retailers
4541	Electronic shopping and mail-order houses
4542	Vending machine operators
4543	Direct selling establishments

Figure 3.15: NAICS codes for regression estimates

Nonlinear estimates: See web appendix at bit.ly/davidvitt

3.4.6 Chapter 2 Appendix

3.4.7 Tables and Figures

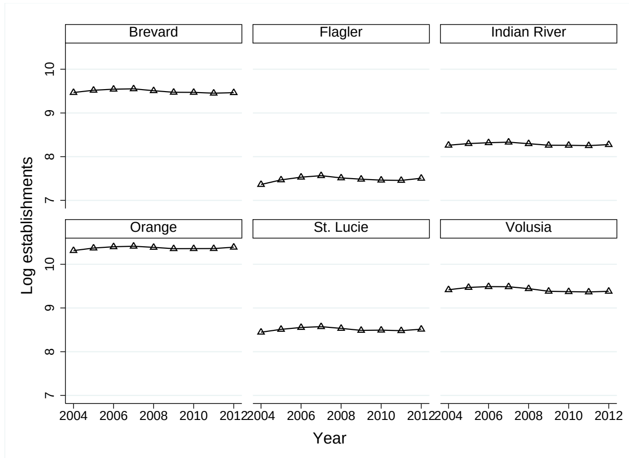


Figure 3.16: Establishment counts over time for treatment group (Brevard) compared to control counties

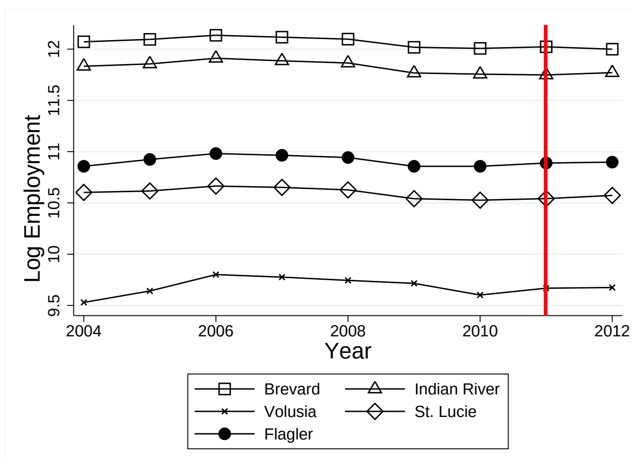


Figure 3.17: Employment counts over time

Variables	Log Establishments	Log Employment	Log Population
Log Establishments	1.000		
Log Employment	0.908	1.000	
	(0.000)		
Log Population	0.415	0.504	1.000
	(0.000)	(0.000)	
Log Shuttle Budget	0.101	0.140	0.262
	(0.001)	(0.000)	(0.000)

Table 3.6: Cross-correlation table

3.4.8 Replication

See web appendix at http://bit.ly/davidvitt

3.4.9 Unemployment Estimates

	(1) Unemployed
Treatment Effect	3702.9*
	(2.08)
2004.year	0
	(.)
2005.year	-1022.8
	(-0.85)
2006.year	-1490.2
	(-1.16)
2007.year	-345.8
	(-0.30)
2008.year	2803.6***
	(3.48)
2009.year	8301.2***
	(6.49)
2010.year	9371.4***
	(6.33)
2011.year	8058.9***
	(6.63)
2012.year	6051.5***
	(6.31)
brev	12014.2***
	(6.98)
twelve	0
	(.)
1.fipscty	0
	(.)
3.fipscty	-5779.0***
	(-3.97)
5.fipscty	-890.8

	(-1.12)
7.fipscty	-5856.1***
	(-3.95)
9.fipscty	0
	(.)
11.fipscty	54821.9***
	(7.41)
13.fipscty	-6196.6***
	(-4.04)
15.fipscty	-1210.6
	(-1.37)
17.fipscty	-2020.7*
	(-2.12)
19.fipscty	-731.1
	(-0.96)
21.fipscty	3897.1***
	(6.79)
23.fipscty	-4610.0***
	(-3.55)
27.fipscty	-5497.2***
	(-3.78)
29.fipscty	-6103.6***
	(-4.03)
31.fipscty	25148.1***
	(7.27)
33.fipscty	2717.0***
	(4.89)
35.fipscty	-3408.8**
	(-3.12)
37.fipscty	-6285.2***
	(-4.05)
39.fipscty	-5059.4***
	(-3.69)
41.fipscty	-6076.6***

	(-4.01)
43.fipscty	-6202.6***
	(-4.03)
45.fipscty	-6145.1***
	(-4.03)
47.fipscty	-6192.7***
	(-4.04)
49.fipscty	-5669.0***
	(-3.84)
51.fipscty	-4676.6***
	(-3.33)
53.fipscty	-932.7
	(-1.14)
55.fipscty	-3498.0**
	(-2.94)
57.fipscty	35652.7***
	(6.66)
59.fipscty	-6052.0***
	(-3.98)
61.fipscty	-1068.1
	(-1.22)
63.fipscty	-5326.7***
	(-3.68)
65.fipscty	-6164.2***
	(-4.03)
67.fipscty	-6395.2***
	(-4.10)
69.fipscty	2957.0***
	(5.33)
71.fipscty	14811.8***
	(6.28)
73.fipscty	1141.1
	(1.71)
75.fipscty	-5293.0***

	(-3.80)
77.fipscty	-6380.3***
	(-4.09)
79.fipscty	-5985.4***
	(-3.93)
81.fipscty	3759.0***
	(6.38)
83.fipscty	4295.2***
	(6.34)
85.fipscty	-1902.1
	(-1.94)
86.fipscty	71011.9***
	(5.41)
87.fipscty	-4468.2***
	(-3.41)
89.fipscty	-4306.7***
	(-3.52)
91.fipscty	-1716.6
	(-1.81)
93.fipscty	-5060.8***
	(-3.65)
95.fipscty	34569.9***
	(6.25)
97.fipscty	3559.0***
	(5.66)
99.fipscty	39037.8***
	(7.76)
101.fipscty	8854.3***
	(8.35)
103.fipscty	24928.6***
	(7.68)
105.fipscty	14191.8***
	(7.40)
107.fipscty	-3959.7**

	(-3.26)
109.fipscty	-1278.1
	(-1.58)
111.fipscty	4586.9***
	(7.70)
113.fipscty	-2235.8*
	(-2.22)
115.fipscty	5398.8***
	(7.85)
117.fipscty	8500.8***
	(7.96)
119.fipscty	-4575.2***
	(-3.65)
121.fipscty	-5401.6***
	(-3.83)
123.fipscty	-5889.0***
	(-3.94)
125.fipscty	-6266.1***
	(-4.06)
127.fipscty	11417.8***
	(7.72)
129.fipscty	-5763.3***
	(-3.93)
131.fipscty	-5049.3***
	(-3.72)
133.fipscty	-5866.7***
	(-3.96)
_cons	3025.4***
	(3.39)
N	603

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3.4.10 Full Estimates

On the following table, each industry treatment effect is labeled "XXX.NAICS treatment effect" where "XXX" represents a 3 digit NAICS subsector. For a full listing of the 3 digit subsectors, see the following url: http://www.dlt.ri.gov/lmi/proj/naics3indproj.htm. Each county has a fixed effect and is labeled by "X.fipscty" where "X" may be a single or two digit number that is the FIPS (Federal Information Processing Standard) code. A full listing of FIPS counties is available at https://www.census.gov/geo/reference/codes/cou.html.

	(1) Empl. Count	(2) Estab. Count	(3) Payroll
113.naics Treat Effect	-1457.4***	-126.1***	-57711.2***
	(-11.33)	(-12.72)	(-9.94)
114.naics Treat Effect	-356.6*	-31.26*	-17466.7*
	(-2.22)	(-2.55)	(-2.52)
115.naics Treat Effect	-1107.7***	-95.31***	-43740.2***
	(-8.27)	(-9.59)	(-7.32)
211.naics Treat Effect	0	0	0
	(.)	(.)	(.)
212.naics Treat Effect	-878.7***	-77.04***	-36639.0***
	(-6.06)	(-6.97)	(-5.78)
213.naics Treat Effect	-34.38	-8.902	-5366.1
	(-0.19)	(-0.65)	(-0.70)
221.naics Treat Effect	-1295.7***	-90.26***	-56298.1***
	(-10.32)	(-9.51)	(-9.91)
236.naics Treat Effect	-1363.9***	6.366	-61706.6***
	(-13.56)	(0.75)	(-11.80)
237.naics Treat Effect	-974.9***	-66.24***	-32579.8***
	(-9.37)	(-8.21)	(-6.58)
238.naics Treat Effect	-237.3	347.1***	-21118.3**

	(-1.17)	(14.49)	(-2.65)
311.naics Treat Effect	-1193.7***	-77.97***	-50317.1***
	(-10.13)	(-8.30)	(-9.09)
312.naics Treat Effect	-228.9	-19.71	-12796.8
	(-1.35)	(-1.49)	(-1.76)
313.naics Treat Effect	145.4	4.400	1734.7
	(0.78)	(0.31)	(0.22)
314.naics Treat Effect	-509.0***	-34.63**	-23124.5***
	(-3.48)	(-3.22)	(-3.56)
315.naics Treat Effect	-102.0	-13.99	-7409.6
	(-0.62)	(-1.12)	(-1.02)
316.naics Treat Effect	436.9*	25.43	13767.6
	(1.99)	(1.51)	(1.50)
321.naics Treat Effect	-1298.8***	-98.47***	-49417.9***
	(-10.23)	(-10.17)	(-8.56)
322.naics Treat Effect	122.4	7.620	47.47
	(0.66)	(0.52)	(0.01)
323.naics Treat Effect	-745.4***	-66.91***	-30467.5***
	(-6.25)	(-7.65)	(-5.49)
324.naics Treat Effect	209.6	7.420	4523.0
	(1.03)	(0.48)	(0.54)
325.naics Treat Effect	-776.2***	-57.63***	-32834.3***
	(-6.13)	(-5.52)	(-5.88)
326.naics Treat Effect	-797.5***	-56.62***	-33188.9***
	(-5.99)	(-5.35)	(-5.53)
327.naics Treat Effect	-1319.8***	-89.94***	-53685.7***
	(-10.97)	(-9.56)	(-9.69)
331.naics Treat Effect	-274.5	-25.44*	-14276.1*
	(-1.63)	(-1.99)	(-1.99)
332.naics Treat Effect	-275.9*	-32.58***	-7550.6
	(-2.44)	(-3.70)	(-1.42)
333.naics Treat Effect	-15.11	-64.33***	16318.1**
	(-0.12)	(-6.58)	(2.94)
334.naics Treat Effect	10692.0***	4.622	768960.6***

	(58.80)	(0.43)	(73.27)
335.naics Treat Effect	138.6	5.984	-1149.3
	(0.86)	(0.48)	(-0.16)
336.naics Treat Effect	1090.6***	-59.67***	155438.9***
	(8.75)	(-6.07)	(26.85)
337.naics Treat Effect	-811.3***	-66.65***	-34102.3***
	(-6.33)	(-7.11)	(-5.80)
339.naics Treat Effect	-980.9***	-39.53***	-45954.7***
	(-8.16)	(-4.44)	(-8.14)
423.naics Treat Effect	-301.6	-2.553	5063.2
	(-1.83)	(-0.14)	(0.56)
424.naics Treat Effect	-1794.1***	-92.92***	-79584.9***
	(-14.00)	(-9.46)	(-12.38)
425.naics Treat Effect	-1030.4***	-61.04***	-44610.8***
	(-8.96)	(-8.21)	(-8.39)
441.naics Treat Effect	361.9***	18.26*	13760.8**
	(3.51)	(2.56)	(2.65)
442.naics Treat Effect	-1051.8***	-66.97***	-42722.1***
	(-9.76)	(-9.01)	(-8.08)
443.naics Treat Effect	-907.4***	-57.44***	-37670.7***
	(-8.25)	(-7.43)	(-6.98)
444.naics Treat Effect	-34.70	-44.74***	-20705.9***
	(-0.35)	(-5.95)	(-4.14)
445.naics Treat Effect	1302.7***	2.841	6288.4
	(9.26)	(0.39)	(1.29)
446.naics Treat Effect	-406.8***	-18.97**	-20211.0***
	(-4.15)	(-2.67)	(-4.10)
447.naics Treat Effect	-497.2***	27.76***	-35319.2***
	(-4.53)	(3.77)	(-6.48)
448.naics Treat Effect	-541.5***	-47.27***	-42777.6***
	(-4.82)	(-4.84)	(-8.61)
451.naics Treat Effect	-772.4***	-38.42***	-37583.2***
	(-6.90)	(-4.83)	(-6.73)
452.naics Treat Effect	2955.7***	-60.65***	37020.7***

	(24.32)	(-7.32)	(7.57)
453.naics Treat Effect	-620.2***	1.355	-39676.6***
	(-5.81)	(0.19)	(-7.49)
454.naics Treat Effect	-1167.6***	-19.90**	-52605.9***
	(-10.69)	(-2.60)	(-10.16)
481.naics Treat Effect	-715.7***	-29.78**	-37638.1***
	(-5.63)	(-2.70)	(-6.35)
483.naics Treat Effect	-124.8	-3.144	-13684.3
	(-0.74)	(-0.23)	(-1.88)
484.naics Treat Effect	-1493.0***	-74.87***	-62636.6***
	(-13.20)	(-9.94)	(-11.79)
485.naics Treat Effect	-449.7***	-56.71***	-23535.9***
	(-3.52)	(-5.89)	(-3.98)
486.naics Treat Effect	93.32	-8.008	1879.5
	(0.41)	(-0.47)	(0.20)
487.naics Treat Effect	-473.4**	-42.12***	-21379.9**
	(-3.04)	(-3.68)	(-3.16)
488.naics Treat Effect	-418.1***	-57.69***	2505.6
	(-3.66)	(-7.22)	(0.48)
492.naics Treat Effect	-493.7***	-36.85***	-22310.3***
	(-4.23)	(-3.75)	(-4.07)
493.naics Treat Effect	-876.7***	-64.62***	-39136.1***
	(-6.93)	(-6.35)	(-6.86)
511.naics Treat Effect	-1728.0***	-84.22***	-85821.4***
	(-16.41)	(-9.81)	(-15.59)
512.naics Treat Effect	-881.2***	-61.94***	-31591.2***
	(-6.87)	(-6.81)	(-5.34)
515.naics Treat Effect	-822.3***	-67.88***	-31487.6***
	(-6.59)	(-6.73)	(-5.79)
516.naics Treat Effect	0	0	0
	(.)	(.)	(.)
517.naics Treat Effect	-865.3***	-66.48***	-31652.2***
	(-8.28)	(-8.45)	(-5.40)
518.naics Treat Effect	-506.2***	-52.40***	-22511.5***

	(-3.97)	(-5.40)	(-4.03)
519.naics Treat Effect	-376.4*	-30.49**	-18455.0**
	(-2.39)	(-2.61)	(-2.72)
521.naics Treat Effect	0	0	0
	(.)	(.)	(.)
522.naics Treat Effect	-550.1**	-25.50*	-30283.2**
	(-3.04)	(-2.43)	(-2.95)
523.naics Treat Effect	-1112.0***	-23.07**	-52451.3***
	(-10.18)	(-3.11)	(-8.16)
524.naics Treat Effect	-1898.5***	-4.413	-98325.5***
	(-12.62)	(-0.46)	(-10.59)
525.naics Treat Effect	327.7	14.15	7395.8
	(1.71)	(1.00)	(0.92)
531.naics Treat Effect	-1264.9***	58.76**	-58879.8***
	(-10.72)	(2.81)	(-10.29)
532.naics Treat Effect	-1258.0***	-70.20***	-51919.0***
	(-12.12)	(-9.34)	(-10.26)
533.naics Treat Effect	279.1	17.74	9258.1
	(1.50)	(1.27)	(1.18)
541.naics Treat Effect	5930.8***	618.2***	557849.5***
	(14.29)	(8.78)	(20.74)
551.naics Treat Effect	-999.5***	-76.79***	-18262.8
	(-5.10)	(-9.23)	(-1.21)
561.naics Treat Effect	3466.1***	327.7***	206596.8***
	(6.89)	(12.85)	(13.85)
562.naics Treat Effect	-167.5	-70.95***	16861.2**
	(-1.41)	(-7.59)	(3.10)
611.naics Treat Effect	-2993.9***	-20.58**	-101776.2***
	(-22.87)	(-2.88)	(-18.05)
621.naics Treat Effect	5408.0***	517.6***	381528.1***
	(18.63)	(14.31)	(23.72)
622.naics Treat Effect	3957.9***	-82.68***	129896.7***
	(15.33)	(-8.49)	(9.67)
623.naics Treat Effect	2299.9***	-35.33***	48685.6***

	(18.86)	(-4.42)	(9.71)
624.naics Treat Effect	1136.3***	-2.840	-7560.4
	(11.06)	(-0.40)	(-1.53)
711.naics Treat Effect	-894.1***	-80.69***	-53566.3***
	(-8.02)	(-10.24)	(-10.36)
712.naics Treat Effect	-787.6***	-60.45***	-32280.9***
	(-5.63)	(-5.61)	(-5.16)
713.naics Treat Effect	-1135.1***	-41.56***	-47596.7***
	(-6.10)	(-5.48)	(-8.15)
721.naics Treat Effect	-1673.4***	-72.39***	-67704.5***
	(-8.51)	(-9.01)	(-10.91)
722.naics Treat Effect	7785.3***	424.8***	75813.0***
	(16.38)	(18.24)	(10.95)
811.naics Treat Effect	-490.2***	73.86***	-24072.0***
	(-5.03)	(7.30)	(-4.90)
812.naics Treat Effect	-508.2***	86.40***	-33675.7***
	(-5.22)	(7.95)	(-6.73)
813.naics Treat Effect	413.5***	178.5***	-25765.5***
	(3.30)	(18.45)	(-5.22)
113.naics	0	0	0
	(.)	(.)	(.)
114.naics	-1100.8***	-90.86***	-40244.5***
	(-6.95)	(-7.26)	(-6.52)
115.naics	-344.8**	-28.81**	-13909.0**
	(-2.66)	(-2.85)	(-2.79)
211.naics	-2013.1***	-167.3***	-74817.0***
	(-8.80)	(-9.15)	(-8.29)
212.naics	-578.7***	-49.08***	-21072.2***
	(-4.09)	(-4.36)	(-3.87)
213.naics	-1412.0***	-114.2***	-51979.1***
	(-7.91)	(-8.23)	(-7.44)
221.naics	-161.7	-17.85	-1413.1
	(-1.34)	(-1.85)	(-0.31)
236.naics	952.5***	169.5***	49278.4***

	(10.16)	(19.77)	(12.30)
237.naics	505.5***	23.12**	24785.6***
	(5.19)	(2.81)	(6.80)
238.naics	3602.9***	483.8***	122372.1***
	(17.83)	(20.26)	(16.78)
311.naics	-99.76	-30.15**	-3636.1
	(-0.89)	(-3.15)	(-0.82)
312.naics	-1228.5***	-105.4***	-44914.4***
	(-7.30)	(-7.80)	(-6.83)
313.naics	-1602.8***	-130.5***	-59445.9***
	(-8.66)	(-9.12)	(-8.15)
314.naics	-948.4***	-77.49***	-34586.8***
	(-6.62)	(-7.05)	(-6.14)
315.naics	-1355.4***	-111.1***	-50301.7***
	(-8.31)	(-8.76)	(-7.69)
316.naics	-1894.3***	-151.6***	-71478.8***
	(-8.59)	(-8.81)	(-8.21)
321.naics	-158.6	-21.65*	-6680.3
	(-1.30)	(-2.21)	(-1.42)
322.naics	-1579.8***	-133.7***	-57758.7***
	(-8.53)	(-8.95)	(-7.99)
323.naics	-313.0**	-25.21**	-11178.7*
	(-2.74)	(-2.82)	(-2.50)
324.naics	-1667.0***	-133.5***	-62234.3***
	(-8.22)	(-8.46)	(-7.89)
325.naics	-410.2***	-50.49***	-8972.9*
	(-3.35)	(-4.73)	(-1.98)
326.naics	-528.9***	-57.50***	-19291.4***
	(-4.08)	(-5.31)	(-3.81)
327.naics	-137.6	-23.18*	-4025.5
	(-1.20)	(-2.43)	(-0.91)
331.naics	-1182.9***	-98.68***	-43435.1***
	(-7.12)	(-7.55)	(-6.72)
332.naics	70.47	-14.54	4175.4

	(0.66)	(-1.61)	(1.00)
333.naics	-259.3*	-39.79***	-4642.3
	(-2.19)	(-3.99)	(-1.04)
334.naics	-187.4	-81.74***	17827.2*
	(-1.19)	(-7.40)	(2.14)
335.naics	-1317.0***	-119.1***	-46839.9***
	(-8.25)	(-9.36)	(-7.46)
336.naics	-139.1	-39.45***	-929.1
	(-1.16)	(-3.94)	(-0.20)
337.naics	-478.1***	-39.47***	-17844.9***
	(-3.87)	(-4.13)	(-3.67)
339.naics	-149.5	-24.59**	-432.5
	(-1.29)	(-2.69)	(-0.09)
423.naics	1941.1***	210.4***	100467.6***
	(12.01)	(11.14)	(11.94)
424.naics	1255.7***	89.80***	59724.7***
	(10.21)	(9.00)	(10.88)
425.naics	-114.0	29.92***	-1849.4
	(-1.04)	(3.93)	(-0.45)
441.naics	1500.7***	94.62***	60411.0***
	(15.54)	(12.95)	(15.20)
442.naics	233.4*	38.85***	3711.9
	(2.30)	(5.11)	(0.91)
443.naics	76.01	18.32*	-2227.5
	(0.73)	(2.31)	(-0.52)
444.naics	805.3***	47.62***	21509.7***
	(8.65)	(6.19)	(5.78)
445.naics	2280.9***	102.0***	41519.4***
	(16.76)	(13.61)	(11.68)
446.naics	787.4***	69.85***	21231.8***
	(8.65)	(9.58)	(5.86)
447.naics	343.8***	68.12***	3148.0
	(3.31)	(9.03)	(0.73)
448.naics	1373.1***	139.2***	16020.4***

	(12.92)	(14.02)	(4.36)
451.naics	80.99	17.30*	-6120.1
	(0.76)	(2.13)	(-1.37)
452.naics	1518.9***	17.54*	29709.1***
	(13.26)	(2.07)	(8.37)
453.naics	499.8***	92.53***	4680.4
	(4.97)	(12.76)	(1.14)
454.naics	204.2*	28.78***	7634.7
	(1.98)	(3.68)	(1.93)
481.naics	-741.7***	-89.34***	-20073.1***
	(-5.99)	(-7.90)	(-4.04)
483.naics	-1291.7***	-119.0***	-41322.9***
	(-7.69)	(-8.55)	(-6.25)
484.naics	466.6***	53.76***	17805.3***
	(4.35)	(6.98)	(4.32)
485.naics	-470.8***	-41.41***	-19003.3***
	(-3.81)	(-4.21)	(-3.86)
486.naics	-1550.7***	-118.1***	-59590.7***
	(-6.89)	(-6.75)	(-6.80)
487.naics	-984.0***	-76.00***	-36331.3***
	(-6.41)	(-6.49)	(-6.07)
488.naics	207.7	13.57	7269.2
	(1.92)	(1.66)	(1.80)
492.naics	-444.7***	-61.27***	-16159.0***
	(-3.95)	(-6.08)	(-3.66)
493.naics	-333.8**	-55.50***	-11197.1*
	(-2.74)	(-5.35)	(-2.40)
511.naics	270.6**	-5.896	28110.1***
	(2.73)	(-0.67)	(6.45)
512.naics	-576.3***	-43.18***	-22524.1***
	(-4.64)	(-4.64)	(-4.56)
515.naics	-434.2***	-49.24***	-9025.6*
	(-3.61)	(-4.78)	(-2.08)
516.naics	-1600.8***	-127.8***	-56020.0***

	(-6.49)	(-6.65)	(-5.86)
517.naics	716.9***	19.36*	45378.0***
	(7.29)	(2.41)	(9.39)
518.naics	-575.2***	-53.72***	-10511.7*
	(-4.65)	(-5.41)	(-2.33)
519.naics	-1066.0***	-85.63***	-38240.3***
	(-6.85)	(-7.14)	(-6.36)
521.naics	-4018.0***	-315.9***	-156446.2***
	(-6.16)	(-5.44)	(-6.12)
522.naics	2196.7***	176.4***	114276.0***
	(12.29)	(16.51)	(11.71)
523.naics	138.6	59.96***	43346.1***
	(1.34)	(7.88)	(7.89)
524.naics	1671.1***	151.3***	98535.3***
	(11.39)	(15.50)	(11.35)
525.naics	-1785.1***	-140.3***	-65107.0***
	(-9.42)	(-9.77)	(-8.75)
531.naics	1429.5***	346.1***	54414.6***
	(12.72)	(16.44)	(11.74)
532.naics	276.6**	34.08***	8862.8*
	(2.84)	(4.42)	(2.33)
533.naics	-1736.5***	-139.9***	-63231.3***
	(-9.33)	(-9.78)	(-8.69)
541.naics	5610.8***	911.7***	318459.3***
	(13.61)	(12.91)	(11.99)
551.naics	1701.1***	-6.333	145022.5***
	(8.76)	(-0.74)	(9.82)
561.naics	6556.5***	417.2***	183295.0***
	(13.07)	(16.31)	(12.73)
562.naics	-146.9	-22.17*	-3541.4
	(-1.29)	(-2.33)	(-0.82)
611.naics	1536.5***	52.46***	44065.0***
	(12.22)	(7.14)	(9.71)
621.naics	4823.6***	537.3***	252203.7***

	(16.80)	(14.85)	(16.19)
622.naics	3070.7***	-34.44***	151389.0***
	(11.97)	(-3.47)	(11.62)
623.naics	2117.7***	22.21**	50976.2***
	(18.22)	(2.72)	(13.67)
624.naics	1085.3***	83.72***	18717.2***
	(11.30)	(11.43)	(5.12)
711.naics	-33.35	1.574	4382.1
	(-0.32)	(0.19)	(1.11)
712.naics	-669.8***	-59.67***	-25430.4***
	(-4.91)	(-5.43)	(-4.76)
713.naics	1455.6***	36.44***	25086.5***
	(8.04)	(4.69)	(5.26)
721.naics	1852.0***	30.27***	40375.3***
	(9.65)	(3.69)	(7.79)
722.naics	8077.3***	411.1***	107363.8***
	(17.07)	(17.67)	(17.80)
811.naics	703.8***	176.0***	19956.8***
	(7.78)	(17.17)	(5.55)
812.naics	899.8***	176.5***	13031.5***
	(9.94)	(16.04)	(3.50)
813.naics	2056.1***	205.4***	44191.3***
	(17.15)	(21.02)	(12.18)
1.fipscty	0	0	0
	(.)	(.)	(.)
3.fipscty	-1561.0***	-122.3***	-52802.0***
	(-16.67)	(-14.35)	(-13.76)
5.fipscty	-279.8***	-14.15***	-10511.8***
	(-5.05)	(-3.46)	(-4.10)
7.fipscty	-1395.6***	-110.9***	-45164.2***
	(-15.71)	(-13.97)	(-12.47)
9.fipscty	1098.4***	97.05***	45076.8***
	(11.97)	(17.09)	(9.56)
11.fipscty	6696.3***	610.5***	269196.2***

	(18.46)	(14.42)	(17.89)
13.fipscty	-1699.4***	-139.4***	-56324.6***
	(-16.51)	(-14.53)	(-13.75)
15.fipscty	-688.2***	-31.64***	-24245.1***
	(-10.66)	(-7.32)	(-8.40)
17.fipscty	-787.4***	-45.87***	-27806.3***
	(-11.43)	(-9.16)	(-9.21)
19.fipscty	-630.7***	-32.52***	-23521.4***
	(-10.08)	(-7.58)	(-8.24)
21.fipscty	365.6***	60.45***	16532.2***
	(5.90)	(14.79)	(6.18)
23.fipscty	-1007.5***	-76.29***	-34241.0***
	(-13.72)	(-11.97)	(-10.83)
27.fipscty	-1484.1***	-117.4***	-48560.9***
	(-16.22)	(-14.37)	(-13.18)
29.fipscty	-1781.4***	-152.0***	-57685.1***
	(-16.87)	(-15.24)	(-13.90)
31.fipscty	3860.8***	223.3***	166234.3***
	(18.02)	(18.20)	(16.66)
33.fipscty	258.6***	18.80***	7824.3**
	(4.83)	(6.50)	(3.19)
35.fipscty	-1063.4***	-69.22***	-36224.1***
	(-14.20)	(-11.91)	(-11.27)
37.fipscty	-1494.8***	-123.0***	-47691.8***
	(-15.44)	(-13.94)	(-12.42)
39.fipscty	-1160.6***	-89.56***	-37861.4***
	(-13.75)	(-12.30)	(-11.01)
41.fipscty	-1499.4***	-126.6***	-47606.0***
	(-14.96)	(-13.87)	(-12.22)
43.fipscty	-1837.4***	-159.4***	-58565.0***
	(-14.84)	(-13.62)	(-12.40)
45.fipscty	-1569.6***	-127.4***	-52012.9***
	(-16.46)	(-14.65)	(-13.52)
47.fipscty	-1707.4***	-140.3***	-55272.9***

	(-16.57)	(-14.55)	(-13.54)
49.fipscty	-1484.5***	-118.3***	-48319.0***
	(-16.27)	(-14.13)	(-13.03)
51.fipscty	-1317.6***	-100.9***	-44498.5***
	(-15.34)	(-13.38)	(-12.53)
53.fipscty	-720.4***	-41.28***	-25856.7***
	(-11.13)	(-8.77)	(-8.81)
55.fipscty	-967.4***	-65.50***	-33694.5***
	(-13.27)	(-11.12)	(-10.63)
57.fipscty	5123.8***	318.9***	217499.6***
	(17.77)	(15.65)	(15.56)
59.fipscty	-1451.3***	-115.2***	-46947.2***
	(-15.40)	(-13.59)	(-12.44)
61.fipscty	-554.2***	-25.30***	-18291.6***
	(-8.88)	(-6.28)	(-6.65)
63.fipscty	-1196.4***	-90.33***	-40631.6***
	(-14.64)	(-12.63)	(-11.87)
65.fipscty	-1520.4***	-124.5***	-50372.6***
	(-16.26)	(-14.61)	(-13.42)
67.fipscty	-1862.6***	-161.0***	-60541.1***
	(-16.00)	(-13.92)	(-13.15)
69.fipscty	-116.6*	11.36***	-5524.8*
	(-2.11)	(3.56)	(-2.09)
71.fipscty	1222.7***	129.9***	40041.5***
	(13.65)	(17.40)	(12.03)
73.fipscty	149.6**	19.52***	5485.9
	(2.62)	(5.92)	(1.91)
75.fipscty	-1225.0***	-94.03***	-40340.2***
	(-14.46)	(-12.77)	(-11.59)
77.fipscty	-2074.1***	-172.8***	-70357.8***
	(-15.48)	(-13.25)	(-13.56)
79.fipscty	-1538.8***	-122.0***	-51600.0***
	(-16.89)	(-14.78)	(-13.83)
81.fipscty	68.24	27.78***	2826.4

	(1.26)	(10.40)	(1.14)
83.fipscty	6.165	17.74***	-2084.0
	(0.11)	(5.88)	(-0.79)
85.fipscty	-381.9***	-9.419**	-12881.2***
	(-6.69)	(-2.85)	(-4.91)
86.fipscty	9055.9***	803.3***	346544.7***
	(12.98)	(10.20)	(11.94)
87.fipscty	-683.1***	-32.02***	-24331.2***
	(-9.74)	(-6.45)	(-7.79)
89.fipscty	-933.4***	-62.61***	-31167.7***
	(-12.29)	(-10.51)	(-9.66)
91.fipscty	-216.6***	-6.532	-7733.9**
	(-3.95)	(-1.92)	(-3.06)
93.fipscty	-1286.8***	-97.55***	-43042.4***
	(-15.11)	(-13.17)	(-12.22)
95.fipscty	6523.2***	321.1***	249131.9***
	(17.80)	(16.39)	(17.34)
97.fipscty	-341.2***	-13.38***	-14847.6***
	(-5.70)	(-3.45)	(-5.27)
99.fipscty	4565.7***	438.3***	191071.4***
	(17.69)	(14.19)	(16.39)
101.fipscty	31.04	35.16***	-2451.6
	(0.57)	(11.50)	(-0.93)
103.fipscty	3463.1***	264.4***	130062.1***
	(18.55)	(16.31)	(16.72)
105.fipscty	1092.6***	71.20***	38236.7***
	(14.57)	(20.61)	(11.94)
107.fipscty	-931.9***	-65.02***	-31446.2***
	(-12.29)	(-10.60)	(-9.75)
109.fipscty	-445.7***	-10.06**	-14941.5***
	(-7.46)	(-3.00)	(-5.44)
111.fipscty	-324.8***	-6.040	-12450.2***
	(-5.36)	(-1.67)	(-4.47)
113.fipscty	-807.1***	-47.63***	-27185.3***

	(-11.58)	(-9.18)	(-8.96)
115.fipscty	658.8***	93.25***	24225.0***
	(10.81)	(15.74)	(9.28)
117.fipscty	965.9***	92.95***	37909.4***
	(13.16)	(17.05)	(12.52)
119.fipscty	-1199.3***	-90.52***	-40702.4***
	(-15.30)	(-13.23)	(-12.11)
121.fipscty	-1329.4***	-100.6***	-45004.3***
	(-15.87)	(-13.78)	(-12.92)
123.fipscty	-1342.3***	-104.3***	-44740.2***
	(-15.10)	(-13.35)	(-12.32)
125.fipscty	-1886.6***	-158.0***	-61692.2***
	(-17.06)	(-15.11)	(-14.01)
127.fipscty	691.1***	85.94***	17743.7***
	(10.63)	(17.79)	(6.84)
129.fipscty	-1352.6***	-112.1***	-43584.7***
	(-14.83)	(-13.82)	(-11.95)
131.fipscty	-978.2***	-67.35***	-33112.0***
	(-12.69)	(-11.05)	(-10.05)
133.fipscty	-1465.9***	-115.5***	-47943.2***
	(-16.03)	(-14.13)	(-12.97)
2004.year	0	0	0
	(.)	(.)	(.)
2005.year	11.76	3.842	2832.0
	(0.16)	(0.56)	(1.03)
2006.year	67.34	6.510	6181.8*
	(0.91)	(0.94)	(2.16)
2007.year	53.26	8.955	7470.6*
	(0.73)	(1.28)	(2.55)
2008.year	5.049	4.001	6135.0*
	(0.07)	(0.65)	(2.25)
2009.year	-76.91	0.489	3554.4
	(-1.16)	(0.08)	(1.33)
2010.year	-102.3	0.310	3797.1

	(-1.55)	(0.05)	(1.42)
2011.year	-86.61	-0.00583	5284.4
	(-1.30)	(-0.00)	(1.94)
2012.year	-51.11	2.988	7193.0*
	(-0.74)	(0.48)	(2.54)
volusia	0	0	0
	(.)	(.)	(.)
_cons	410.1***	27.08**	5441.4
	(3.79)	(3.13)	(1.26)
N	39972	39972	39972

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.8: Difference-in-difference estimates

VITA

DAVID VITT

2011	B.A., Economics University of Florida
2013	M.A., Economics Florida International University
2016	Ph.D., Economics Florida International University
2011-2016	Graduate Teaching Assistant Florida International University
2016	Joins SUNY Farmingdale faculty in September