Essays in Industrial Organization

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Mark Ponder

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Amil Petrin, Advisor

June, 2022

Acknowledgments

I am very grateful to my advisor, Amil Petrin, and my committee members, and Thomas J. Holmes, Joel Waldfogel and Alessandra Fogli, for their advice and support over the past six years. Their guidance has been invaluable in teaching me how to select, develop and present research projects. Additionally, my research has benefited from conversations with members of the University of Minnesota Applied Microeconomics workshop and fellow graduate students in the Department of Economics. My research has further benefited from discussions with several economists at the Federal Reserve Board of Minneapolis and, in particular, Fabrizio Perri. I owe special thanks to Veronica Postal who co-authored the final chapter of my dissertation. Further, I acknowledge the financial support of the Graduate School of the University of Minnesota. Finally, I would like to thank the Department of Economics staff, especially Wendy Williamson and Catherine Bach.

Dedication

To my son Luca, the greatest motivator and greatest impediment to completing my dissertation. To my wife Veronica, without whose support this would not have been possible. And to my parents, Darryl and Juli, for their care and guidance over the years.

Abstract

This dissertation is comprised of three essays, each dealing with topics in empirical Industrial Organization and Applied Microeconomics. The second chapter was co-authored with Amil Petrin and Boyoung Seo and the third chapter was co-authored with Veronica Postal.

In the first chapter, I develop a dynamic model of the oil pipeline industry to estimate the impact of direct price regulation on investment. Since the shale boom began in 2010, crude oil production in the United States has surged over 100% leading to a dramatic increase in demand for pipeline transportation. However, the profitability of investing in oil pipelines is constrained as transportation rates are set subject to a price cap. In this chapter, I examine the impact of direct price regulation on pipeline investment in response to the shale boom. I develop a theoretical model of the pipeline industry, where firms make production and investment decisions while being subject to a dynamically changing price ceiling. I estimate the model using detailed operational data derived from regulatory filings and compare welfare under three separate regulatory environments: price cap regulation, cost-of-service regulation, and price deregulation. I find that price cap regulation was superior to the alternative mechanisms considered, as it increased market entry by 15% and incentivized firms to operate 17% more efficiently. I find evidence suggesting that prices were allowed to increase too quickly. While this led to an increased rate of entry into new markets it came at the expense of higher prices in existing markets. This ultimately resulted in a transfer in consumer surplus from existing customers to new customers and a slight decrease in total relative to what could have been achieved under a fixed price ceiling.

In the second chapter, we propose a novel approach to estimating supply and demand in a discrete choice setting. The standard Berry, Levinsohn, and Pakes (1995) (BLP) approach to estimation of demand and supply parameters assumes that the product characteristic un-

observed to the researcher but observed by consumers and producers is conditionally mean independent of all characteristics observed by the researcher. We extend this framework to allow all product characteristics to be endogenous, so the unobserved characteristic can be correlated with the other observed characteristics. We derive moment conditions based on the assumption that firms - when choosing product characteristics - are maximizing expected profits given their beliefs at that time about preferences, costs, and competitors' actions with respect to the product characteristics they choose. Following Hansen and Singleton (1982), we assume that the "mistake" in the choice of the amount of the characteristic that is revealed once all products are on the market is conditionally mean independent of anything the firm knows when it chooses its product characteristics. We develop an approximation to the optimal instruments and we also show how to use the standard BLP instruments. Using the original BLP automobile data we find all parameters to be of the correct sign and to be much more precisely estimated. Our estimates imply observed and unobserved product characteristics are highly positively correlated, biasing demand elasticities upward significantly, as our average estimated price elasticities double in absolute value and average markups fall by 50%.

In the third chapter, we estimate the benefit households derived from the introduction of light rail transit in Minneapolis. The primary goal of this chapter is to decompose this benefit into two components: the direct effect from improved access to public transportation and the indirect-effect from the endogenous change in local amenities. The literature has predominantly relied on two methods to estimate the impact of public transportation: difference-in-differences models and hedonic pricing models. Difference-in-difference models yield convincing treatment effect estimates but do not readily provide a decomposition of the direct and indirect effect. Hedonic pricing models can provide such a decomposition but have historically relied on parsimonious specifications that do not control for omitted variable bias. Recently, researchers have proposed refining the hedonic pricing approach by incorporating predictive modeling, where the researcher trains a predictive model on a control group using a high-dimensional dataset and then uses this model to predict what prices would have been in the "but-for" world for the treatment group. The difference between actual and predicted prices provides a valid estimate of the average treatment effect.

However, if important sources of heterogeneity are excluded from the model then this approach will still suffer from omitted variable bias. We propose augmenting the estimation of the predictive model with instrumental variables allowing us to control for the selection bias induced by unobserved heterogeneity. We find close agreement between our predictive model and the difference-in-differences approach, estimating an increase in house prices of 10.4-11.3%. Using the predictive model, we estimate that prices increased by 5.5% due to improved access to public transportation and 5.8% due to improved access to amenities.

Contents

List of Figures

List of Tables

Chapter 1

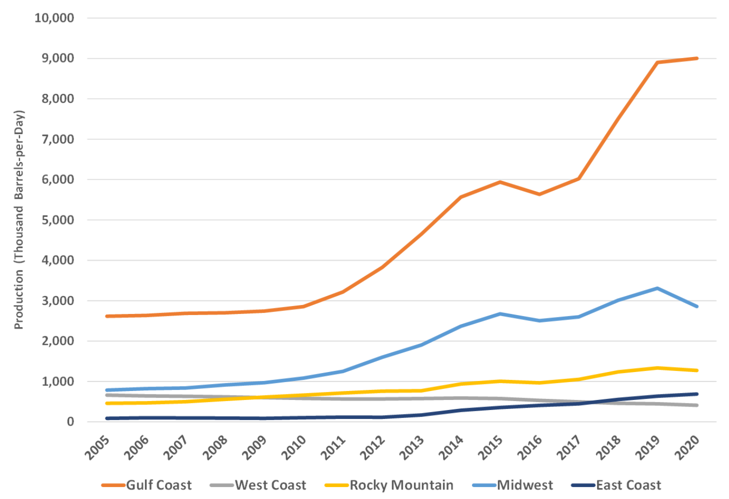
The Impact of Price Regulation on U.S. Pipeline Investment During the Shale Revolution

1.1 Introduction

The United States oil pipeline industry has experienced significant change over the past two decades due to an unprecedented demand shock and increased regulatory scrutiny. Innovations in hydraulic fracturing, horizontal drilling, and seismic imaging led to a boom in oil production with domestic supply increasing by over 100%, from 5 million barrels per day (bpd) in 2000 to over 11 million bpd by 2020. As shown in Figure (1.1), the impact of the shale boom was felt across the United States as previously marginally productive fields greatly increased their supply. However, the surge in production was largely accommodated using the existing processing infrastructure, as refining capacity only expanded 6% over the same period. This required new means of transportation to connect new wells to the processing infrastructure, generating a large increase in demand for additional crude oil pipeline construction.

The profitability of constructing new pipelines is constrained by the use of price caps to regulate transportation prices in the industry. Oil pipelines primarily generate revenue

Figure 1.1: Growth in Oil and NGL Production by Region

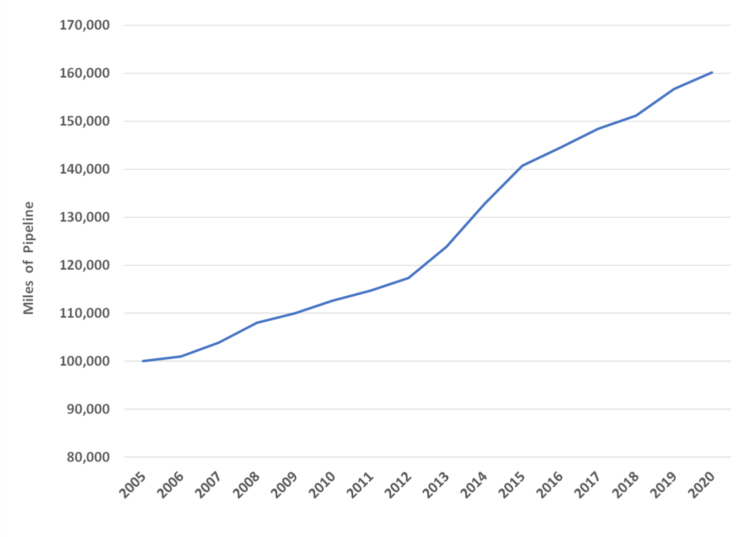


Note: Production figures include crude oil and natural gas liquids (NGLs). Data are from the EIA and include production of crude oil and plant condensate. For region definitions, see Figure (A.1).

through the provision of transportation services, charging a fixed price to transport a barrel of oil. These prices are set subject to a price cap that is determined by the Federal Energy Regulatory Commission (FERC). The initial price cap depends on the pipeline's average total cost while the evolution of the price cap depends on movements in the producer price index plus the difference between the mean industry average total cost and the change in the producer price index. The price cap mechanism was implemented to prevent pipelines from generating excessive rents, but this created a trade-off as it also disincentivizes firms from undertaking potentially welfare improving investment. In this paper, I examine the extent to which the use of price cap regulation impacted the incentive for pipelines to invest in response to the shale boom, and how their investment decisions would have changed under alternate forms of regulation.

Price cap regulation can impact investment decisions through two channels: the initial price ceiling level and how the ceiling trends over time. In principle, the regulator can set initial price levels to compensate firms for sunk investment costs in order to incentivize construction. However, when making entry decisions, firms also have to anticipate whether

Figure 1.2: Growth in Crude Oil and Natural Gas Liquids Pipelines



Note: Data are from the PHMSA and include mileage for crude oil and highly volatile liquids pipelines.

the price ceiling will rise fast enough to compensate them for exogenous increases in unit cost. Since 2000, firms saw operational costs grow by roughly 65%.[1] The exact cause of this increase is difficult to determine, however, the regulatory record shows that it was in part due to expanded environmental and safety requirements which increased operational and capital costs for pipelines. If pipelines anticipated that costs would grow faster than the price cap, they could be further deterred from undertaking additional capital expansion.

To determine the impact of price cap regulation on investment, I develop a model of pipeline production and investment where pipelines are subject to price regulation. Pipelines supply oil transportation services to shippers, which include oil producers, refineries, and industrial manufacturers. Pipelines are modeled as local monopolist that choose the optimal timing of market entry to maximize their expected profit and face a sunk cost of entry. After entering a market, pipelines can further invest to lower their unit cost and to expand their system. Pipelines must set prices subject to a price cap, which limits the expected return to pipeline construction. The price cap evolves over time, so pipelines must predict how the price cap will change in response to changing industry costs and demand. Pipelines can then exit if

[1] See Appendix Figures (A.2).

the price cap falls below the cost of production.

Pipeline decisions impact welfare in three ways. First, to the extent that the price cap is above the marginal cost of production, pipelines generate dead-weight-loss and lower total welfare. Second, firms are able to invest to lower their unit-cost of production and lower production costs are potentially eventually passed on to shippers in the form of lower prices. Finally, pipelines may expand their system in response to higher expected returns, potentially increasing welfare.

To determine the relative importance of these various effects, I estimate the model using a rich dataset gathered from regulatory filings made to FERC. I begin by estimating a demand curve for pipeline transportation, regressing data on the total quantity transported (in barrel-miles) on the transportation price and additional covariates. These covariates include the number of routes a pipeline provides, market-specific fixed-effects, and an indicator variable for the shale boom. I estimate that the shale boom increased demand for pipeline transportation by roughly 90% on average.

Next, I estimate the variable cost of production in a two-step process. First, I estimate an industry production function using data on physical units of output - total barrel-miles transported - and detailed data on variable and capital costs. Data on variable costs include wages, materials and supplies, operating fuel, and outside services. My capital series is construct following Olley and Pakes (1996), using capital expenditures on land, equipment, and structures. The production function estimates recover a measure of firm-level productivity and how it evolved over time. Average productivity is estimated to decline by 50% over the sample period, suggesting that increased safety regulation led to higher costs.

Following Dhyne et al. (2020), I then maintain cost-minimization to recover firm-level marginal costs. These marginal costs are then regressed on observed output levels and pipeline routes to recover a marginal cost function in a manner analogous to Dhyne et al. (2020). Integrating the marginal cost function yields the variable cost of production. The marginal cost of transportation is estimated to be U-shaped and is relatively flat over the relevant range of output.

With estimates of the variable cost function and demand, I use the model to recover the industry fixed cost structure by matching predicted investment, entry, and exit decisions to their empirical counterpart. I estimate large sunk costs of market entry of roughly \$1.3 billion, with 95% of sunk costs falling between \$700 million and \$1.9 billion. The fixed cost of system expansion is similarly large but has a higher variance. Together, these large costs suggest that pipelines need reasonably large expected returns in order to either build new systems or expand an existing system.

I use the estimated model to determine how welfare would have changed under three alternative regulatory scenarios: cost-of-service, price deregulation, and a non-adjusting price cap. I find that the price cap led to a significant increase in returns for oil pipelines relative to the cost-of-service mechanism, leading to 15% increase in market participation after the shale boom. System expansion is marginally impacted, increasing by only 0.5%, as both forms of regulation encourage pipelines to operate on a larger scale. Average pipeline productivity is estimated to be 17% higher under the price cap. However, welfare gains from entry and productivity investment are somewhat offset by the higher prices charged to existing customers, who see their consumer surplus decline by 15%.

Price deregulation leads to the highest level of market participation among regulatory scenarios I consider and pipelines undertake roughly 18% more system expansion. However, this is more than offset by an 8% decline in average productivity and much higher prices for customers so that welfare declines by roughly 1.4%.

The price cap yielded better results than a traditional cost-of-service regime and price deregulation, however welfare could have been further improved by not allowing the index to adjust dynamically. Under the fixed price cap, entry would have been 6% lower after the shale boom. However, the lower prices that existing customers would have faced leads to an increase in consumer surplus of 5%. The ultimate impact is that welfare would have been 2.4% higher given a fixed price cap.

1.1.0.0.1 Related Literature and Contribution A few papers have analyzed oil pipeline investment after the shale boom. Covert and Kellog (2017) studies the impact of rail transportation on pipeline investment in the Bakken and finds that railroads can provide an important alternative to pipelines due to their flexibility. McRae (2017) studies how the expansion of pipeline capacity in the Permian impacted oil price differentials. In this paper, I use a new dataset to study entry and investment decisions by oil pipelines across the United Sates over the same time period. Detailed data on pipeline operations allows me to take a comprehensive look at how the industry was impacted by increased regulatory scrutiny and significant changes in demand. Additionally, this paper focuses on how price regulation changed investment decisions.

I build on the literature studying the impact of incentive-based regulatory mechanisms. Several papers have used regressions techniques, exploiting variation in firm regulatory environments, to estimate the impact of price caps, including Ai and Sappington (2002), Ai et al. (2004), Majumdar (2016), and Sappington (2003). Domah and Pollitt (2001) and Bottasso and Conti (2009) use a structural approach where they estimate cost functions. This paper provides one of the few dynamic structural models which have been used to determine the efficacy of incentive-based regulation. Pint (1992) also develops a structural model to compare price cap vs cost-of-service regulation, however the author does not estimate the model parameters. By using a structural model, I am able to explore different regulatory environments and decompose the impact of price cap regulation through different channels. This decomposition is important, especially when the industry of interest experiences significant changes after the regulation is introduced. For instance, several paper haves found that productivity falls after the implementation of price cap regulation, for instance Jenkins (2004). I find a similar results with unit-costs increasing by over 50% over the past two decades. However, the structural model implies that most of this increase was exogenous and that the price cap actually served to make firms more efficient relative to cost-of-service regulation.

Additionally, rather than focus on the trade-off between achieving the optimal price structure and incentivizing productivity gains, I allow firms to make optimal investment and market participation decisions. This adds an additional margin, of particular importance for the oil pipeline industry after 2008, through which the price cap regulation can impact market outcomes. Evans and Guthrie (2012) also study the impact of price cap regulation on investment using a continuous time model but do not consider entry or exit. Complementing the theoretical literature on price cap regulation, I find that productivity was substantially improved compared to traditional cost-of-service regulation. However, I also find that price cap regulation was important for encouraging firm entry into markets that were previously not served. These markets can provide significant gains in consumer surplus, and so constitute an important margin for regulators to consider when implementing a price cap mechanism.

My paper proceeds as follows. In Section (1.2), I give a high level look at the oil pipeline industry and discuss my primary confounding event, the shale boom. Next, in Section (1.3) I provide an overview of price cap regulation and specifically how it is implemented in the oil pipeline industry. Section (3.5) discusses the various datasets that I use during my analysis, with a particular emphasis on the FERC Form 6, the primary oil pipeline regulatory filing. Sections (1.5) and (1.6) provides an overview of my model of an oil pipeline production and investment. Section (1.7) provides an in-depth discussion of my empirical strategy. Readers interested in the empirical results can skip directly to Sections (3.6) and (1.9). Section (3.6) provides the estimates for the model primitives, including demand and cost functions, as well as the distribution of entry, exit, and investment costs. Section (1.9) shows the evolution of markups and productivity over the past two decades in the industry and provides my main counterfactuals, namely how investment and entry would have changed as we alter the price cap.

1.2 Industry Overview

In this section I provide a high-level overview of the oil pipeline industry and how it was impacted by the shale boom and increased governmental oversight. I provide a more detailed description of the industry in my online appendix. Readers not interested in industry details can go directly to Section 1.7.

Oil pipelines primarily generate revenue through the transportation of oil, charging prices that vary based on the total distance the oil is shipped. Oil pipelines are limited in their ability to provide dedicated transportation and must reasonably accommodate all shippers that that demand transportation at the posted price. Broadly speaking, pipelines can be classified by whether they transport crude oil or refined petroleum product. Shippers on crude pipelines are generally oil producers and refineries while shippers on refined product pipelines including terminaling companies, airports, and large industrial customers. As of 2020, there was roughly 225,000 miles of oil pipeline installed in the United States. Of this, nearly 70% was dedicated to transporting crude oil and highly volatile liquids (HVL)[2], with the rest dedicated to transporting refined petroleum product.[3]

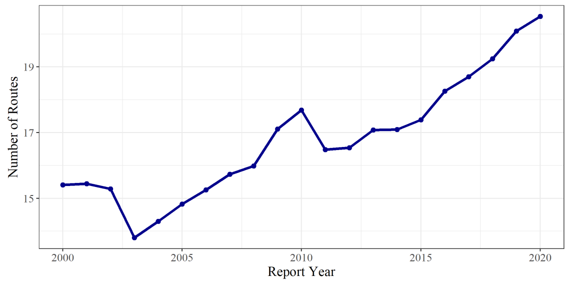
1.2.0.0.1 Shale Boom The development of horizontal drilling, hydraulic fracturing, and seismic imagine made previously high-cost oil fields economically viable to drill. Starting around 2010, low interest rates and a high price-per-barrel for crude oil led upstream oil companies to invest heavily in shale exploration and production. The results were striking as oil production rose over 100% in less than a decade. This massive increase in oil supply led to a similar increase in demand for pipeline transportation. The total crude/HVL pipeline footprint increased from roughly 100,000 to 160,000 miles by 2020, while existing pipelines delivering product into Northeast were reversed and started moving petroleum from the Marcellus/Utica to processing plants along the Gulf Coast. The increase in pipeline capacity was driven in part by the expansion of existing systems and by the construction of new pipeline systems. Panel (a) of Figure (1.3) shows the increase in average number of routes offered by pipelines from 2000 to 2020, increasing from 15 to 20. Panel (b) shows the introduction of new pipeline systems. Following a similar pattern, there was an average of 5 entrants per year prior to 2011, but this increased to over 15 in the subsequent decade.

1.2.0.0.2 Increased Safety Regulation The pipeline industry saw an increase in environmental and safety regulation over this time period, starting with the passing of the Pipeline Safety Improvement Act of 2002 and the Pipeline Inspection, Protection, Enforcement and Safety Act of 2006. The Pipeline and Hazardous Materials Safety Administration (PHMSA) imposed new integrity management regulations that required pipelines to invest in physical modifications to their pipelines and repairs, as well as increased operational

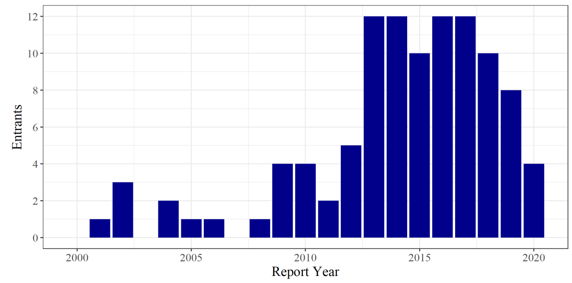
[2]Highly volatile liquids include propane, butane, and other condensates.

[3]See PHMSA “Annual Report Mileage for Hazardous Liquid or Carbon Dioxide Systems"

Figure 1.3: Dynamics of Firm Entry and Route Expansion



(a) Average No. Routes



(b) Enterants per Year

costs associated with more frequent inspections. While PHMSA does not collect quantitative data on the cost of this regulation, they did provide anecdotal evidence in the 2010 price cap index review noting that pipelines have reported compliance costs between \$2.0 - \$2.5 billion. The Deepwater Horizon Spill and the Kalamazoo River oil spill put further scrutiny on the pipeline industry, leading to additional regulation under the Pipeline Safety, Regulatory Control, and Job Creation Act of 2011 and reviews of pipeline safety by the National Transportation Safety Board and the Government Accounting Office. These reviews recommended further measures to ensure the integrity of pipeline systems, culminating in the Protecting our Infrastructure of Pipelines and Enhancing Safety Act of 2016. Each of these acts, along with new rules adopted by PHMSA, appear to have contributed to increased in costs for oil pipelines over the past two decades.[4] The price cap potentially limited the ability of pipelines to raise prices in response to these costs. I discuss the mechanics of price regulation in the oil pipeline industry next.

1.3 Price Cap Regulation

Many regulated industries have historically operated under a cost-of-service mechanism. Under this scheme, the regulator compensates firms for all costs accrued during operations and provides them with a pre-determined return on their unit-cost. However, this mechanism distorts the firm's incentive to minimize their costs because they are compensated for all

[4]Pipeline also face increased regulatory uncertainty as concerns over pipeline spills, climate change, and social justice have increase. The Keystone XL pipeline incurred over \$1.5 billion of development costs before TC Energy was forced to abandoned the project.

incurred costs. Littlechild (1983) proposed a different method of regulating British Telecom after its privatization in the 1980s based on a dynamically changing price cap, called price cap regulation. Firms set prices subject to a ceiling and become the residual claimant on any reduction in their cost base, providing the proper incentives to operate efficiently. Since then, price cap regulation has been applied extensively to utilities around the world.[5] Most important to this paper, FERC adopted price caps as a form of price control when deregulating midstream oil services in 1996.

The initial level of each firm's price cap is determined using a standard cost-of-service filing, such that the ceiling is set equal to average total cost. Firms report their cost of production and the regulator sets the maximum price level based on the allowed rate of return. Firms can then improve their profit margin by reducing their unit-cost. The evolution of the price cap is based on the movement of a price index (FERC uses the producer price index) and average industry performance. This typically takes the form of $PPI + X$, where PPI is a measure of inflation and X is a term determined by the regulator. Firm-specific price cap then evolves according to

$$\bar{P}_{t+1} = \bar{P}_t \left(\frac{PPI_{t+1}}{PPI_t} + X \right)$$

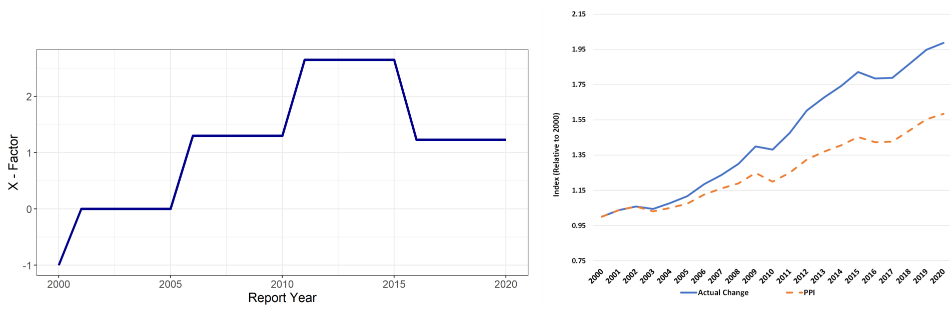
The term X is meant to reflect how industry costs and productivity are expected to change relative to PPI over a predetermined interval of time, called the review period. At the beginning of each review period the term X is reset in order to pass any cost savings on to consumers. Every 5 years, FERC sets this term equal to

$$X = \left(\frac{1}{N} \sum_i \mathrm{CG}_{i \in I} \right)^{\frac{1}{5}} - \left(\frac{PPI_5}{PPI_0} \right)^{\frac{1}{5}}$$

where

[5] Ofwat adopted price caps to regulate water and sewage services at its inception in 1989 and Ofgem adopted price caps for the downstream natural gas market after it formation in 2000. In 1989, the FCC adopted price caps to regulate interstate telecommunication services, and was followed by several states soon after.

Figure 1.4: Adjustment Factor



(a) Adjustment factor (b) Cumulative Change

Note: Panel (a) shows the adjustment factor established after each review period. Panel (b) shows the cumulative change in price caps since 2000.

$$\mathrm{CG}_{iT} = \left((1 - OR_i) \frac{\mathrm{AFC}_{iT}}{\mathrm{AFC}_{i0}} + OR_i \frac{\mathrm{AVC}_T}{\mathrm{AVC}_0} \right)$$

Here, OR_i is the operating ratio of firm i, defined as the ratio of operating expenses to operating revenue. AFC_i is the average fixed cost, defined as the average capital net of depreciation, and AVC_i is the average variable cost, defined as the the average operating and maintenance expense. The use of the operating ratio when calculating the cost growth index is meant to capture the relative importance of operating expenses for certain pipelines. Figure (1.4) shows the realized value of X for each review period, along with the cumulative change in price cap since 1999. With the exception of the prior review, the adjustment factor has increased steadily in each review period and price levels have been allowed to increase over 100%.

One potential explanation for the steady rise in the adjustment factor is that increases in market power can lead to an increase in unit operational cost. In the online appendix, I show that when marginal costs are constant the change in average total cost to a change in price is roughly equal to

$$\frac{\partial \mathrm{ATC}}{\partial p} \frac{P}{\mathrm{ATC}} = \frac{\mathrm{FC}}{\mathrm{TC}} \cdot |\epsilon_D|$$

When firms restrict output to increase price their average total cost increases as well. In the pipeline industry, capital expenses are on the order of 60% and I estimate the elasticity of demand to be roughly 1.2 to 1.5, so a 10% increase in price would translate to an 7.2% - 9.0% increase in average total cost. This has the potential to create a feedback loop, where increased prices contribute to increased unit costs, which in turn leads to a higher rate index and therefore prices. The extent to which this is an issue is an empirical question.

1.4 Data Sources

My primary data sources on pipeline operations come from regulatory filings made to FERC, including the Form 6 and responses to FERC Order 342.3. These sources provide information on prices, physical output, and costs, which I discuss them in the following section.

1.4.1 Form 6

The principal data source for this analysis is the FERC Form 6, a mandatory, quarterly filing for all interstate oil pipeline that have at least $500,000 in annual revenue.[6] Form 6 databases are provided annually by FERC for the years 2000 to 2020. These data include information on firm revenue by interstate and intrastate transportation. Data is provided on transportation quantities by product type, type of transportation, and region. Output is reported in both barrels and in barrel-miles, where the latter reflects the total distance the oil was transported. The Form 6 also contains detailed data on pipeline operating and capital costs.

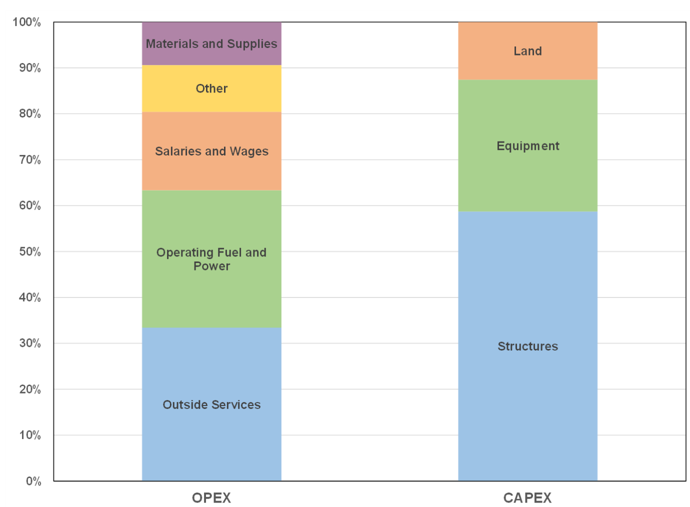
Operating expenses are broken out into two categories: general expenses and operating and maintenance expenses. I use operating and maintenance expenses (OPEX) as my measure of the variable inputs to production. Figure (1.5) shows the average share of each cost category in OPEX. The largest two shares include Outside Services and Operating Fuel and Power. Together, they account for 63% of the pipelines variable cost. A significant compo-

[6]Interstate pipelines are pipelines that transport product that has crossed state lines. Importantly, it is not the pipeline which needs to cross state lines. So pipelines that are entirely contained within a single state may still show up as interstate pipelines and therefore report to FERC. This increases the coverage of my dataset.

nent of outside services is the use of outside contractors so I include Outside Services in the labor expense. Salaries and Wages directly payed by the firm account for roughly 15% of pipeline variable costs. Materials and Supplies account for roughly 10% and the remainder is Other Expenses[7] To convert operating expenses into variable inputs, I deflate these costs using the input price index for the pipeline transportation industry (NAICS code 486210) provided by the BEA.

Capital expenditures (CAPEX) recorded separately for carrier property (capital that is used to directly transport petroleum) and non-carrier property (capital that is not used in the transportation of petroleum). I limit my data series to carrier property, as this is the capital stock most directly tied to production. These data are then further broken out into line items, including land, right of way, pipe, and machines and tools. I partition capital into three components: land, structures, and equipment. Figure (1.5) shows the average share of each component in CAPEX. The largest component is structures, which include line pipe and oil tanks, at almost 60%. The next largest component is equipment at roughly 25%. Finally, land accounts for roughly 15% of pipeline CAPEX.

Figure 1.5: Cost Category Components



I construct a capital index using the perpetual inventory method, analogously to Olley and

[7]Other expenses include oil losses and shortages, where companies incur the cost of spilled oil.

Pakes (1996). Capital at time t is equal to the depreciated stock at time t plus deflated investment. I use a separate depreciation rate for each CAPEX category. Following FERC, land is not depreciated and is deflated using the price index for nonresidential investment from the BEA. Pipeline companies are required to report line item depreciation rates, which I aggregate to generate a weighted average depreciation rate for each of the three components. Structures and equipment are deflated by their analogous nonresidential price indices, also provided by the BEA.[8]

Pipelines must report all major changes in operations when submitting the Form 6. This allows me to track major divestitures or changes in ownership that happen during the sample period for which the data need to be adjusted. A fairly common occurrence is that a pipeline will change change its legal status, generating a new FERC identifier. For instance, “Minnesota Pipe Line Company" became an LLC in 2006 and changed its name to “Minnesota Pipe Line Company, LLC". Observations such as these are combined into a single reporting unit for my analysis.[9]

1.4.2 Tariffs and Responses to Order 342.3

Pipeline tariffs provide system prices, terms of service, and a list of pipeline routes. Pipelines report the evolution of their price cap and their current rate in their annual response to Order 342.3. Both of these sources further help me identify which pipelines operate under the price cap mechanism and which pipelines operate under a different regulatory regime. For example, pipelines in the the Trans-Alaskan Pipeline System still use cost-of-service filings to set rates which Explorer Pipeline may charge market base rates. These pipelines are excluded from my analysis when necessary.

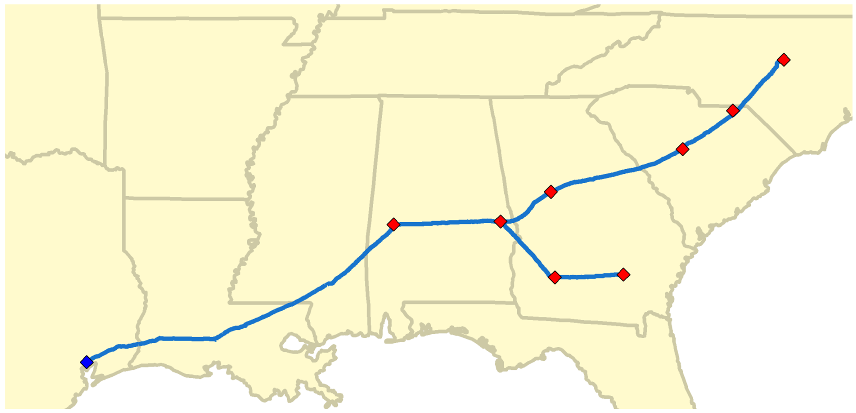
1.4.3 Unit of Analysis

The unit of analysis is an oil pipeline system. Figure (1.6) shows a typical pipeline system, Dixie Pipeline, which transports propane from Mont Belvieu, TX to locations in the south

[8]Several pipelines in my sample lease capacity on other pipeline systems. To convert to an equivalent capital stock, I take the rental expense and scale it by the firm's weighted average cost of capital and then deflate this measure. This impacts fewer than 5% of firms.

[9]See the online appendix for other sample adjustments and how costs were categorized.

Figure 1.6: Dixie Pipeline System



Note: The blue point identifies a receipt poin at a fractionator in Mont Belvieu, TX. The red points represent delivery terminals.

and southeast. Pipeline systems are may be held larger holding companies, however they generally operate independently. So while Dixie Pipeline is owned by Enterprise Products Partners, it shares a limited geographic footprint with the the company's other assets. As such, its operations are not generally impacted by the operations of the company's other pipeline systems. Given this, I make the simplifying assumption that all pipeline systems make independent production and investment decisions.

1.5 Demand

In this section, I describe demand for pipeline transportation and how I derive a firm-specific demand curve.

Several pipelines can service the same oil field or the same downstream market, but pipelines infrequently overlap in a given origin-destination pair, generating a degree of product differentiation. In principle, pipelines may compete with other modes of transportation such as railroads and trucks between within a given origin and destination. However, these other modes tend to have much higher prices for transporting oil than pipelines. For instance, a Congressional Research Service report estimated an average per-barrel cost of pipeline transportation of \$5 and an average cost of \$10 to \$15 for equivalent transportation by rail.[10] As

[10]See Frittelli et al. (2014).

such, I assume that each transportation demand can be approximated by an pipeline-specific demand curves.

I model transportation demand as shippers choosing among their various transportation alternatives. I assume that there is a continuum of oil shippers, each wanting to transport a fixed amount of output each period. These producers choose the lowest cost of transportation to deliver their product to a destination market and sell their output at a uniform price. Let i index the pipelines serving a basin, t the period, and j the oil producers. Following Eaton and Kortum (2002), let the cost of transportation be given by the pipeline's transportation rate divided by a Frechet distributed idiosyncratic shock, $\tau_i(j)$.[11] As such, the shipper solves

$$\min_i \left\{ \frac{P_i}{\tau_i(j)} \right\}$$

Integrating over shipper gives the following standard demand functional form

$$Q_{it} = M_t \frac{T_{it} p_i^{-\theta}}{\sum_k^N T_{kt} p_k^{-\theta}}$$

where Q_{it} represents demand for pipeline i and M_t is the basins potential level of production in period t. The distribution of cost shocks τ_{it} is governed by the parameters T_{it} and θ. T_{it} governs the absolute cost advantage of pipeline i in period t and θ determines the dispersion of the shocks. Taking logarithms, demand has the following functional form

$$\ln(Q_{it}) = \ln(M_t) + \ln(T_{it}) - \theta \ln(P_i) - \ln(\Phi_t)$$

where $\Phi_t = \sum_k^N T_{kt} p_k^{-\theta}$ is common to all pipelines serving an oil field. Note that $\gamma \Phi_t^{\frac{1}{\theta}}$ is the expected transportation cost for oil producers and it is through this term that potential

[11] These shocks represent exogenous factors that change the cost of transportation for a given mode for a specific shipper. For instance, an oil well might be situated in a geological formation not readily accessible by a pipeline.

competitive effects can arise. I abstract away from these competitive effects and assume that firms treat Φ_t as fixed in a given period. Collecting terms, this results in the following constant elasticity demand function for pipeline transportation for firm i

$$\ln\left(Q_{it}\right) = \lambda_{it} - \theta \ln\left(P_{it}\right) + \epsilon_{it}$$

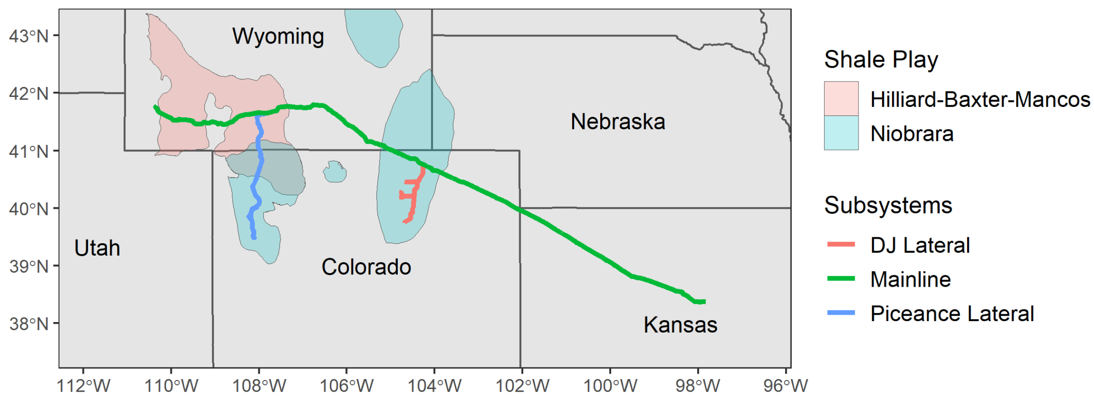
where $\lambda_{it} = \ln(M_t) + \ln(T_{it}) - \ln(\Phi_t)$. I assume that $\ln(T_{it}) = \delta_1 Prod_i + \beta \ln(N_{it})$ where $Prod_i$ indicates that a pipeline transports refined petroleum product instead of crude oil and N_{it} is the number of routes the pipeline serves. The term $Prod_i$ is accounts for the fact that product pipelines tend to have lower rates than crude pipelines. I include the number of routes in demand as a pipeline with more routes will be closer to more wells and therefore will more frequently be the lowest cost option for transportation. Consider the addition of the Denver-Julesburg Lateral to the Overland Pass Pipeline system (see Figure (1.7)). This expansion increased demand for the pipeline system by connecting the pipeline to a new region of the Niobrara shale in the Denver Basin. Adding this route is analogous to the introduction of a new product, as the addition gave customers in the Denver Basin access to the NGL market in Kansas. As such, increasing the number of routes shifts out demand and increases consumer surplus at a given price point.

I model the potential level of production as $\ln(M_t) = \ln(M) + \delta_2 \cdot Shale_t$. $Shale_t$ is an indicator variable for the years after 2010 which reflects the increase in cost-effective reserves after the shale boom. The term ϵ_{it} is a mean-zero i.i.d. shock to T_{it} that is unanticipated by the firm. However, they know the underlying distribution of these demand shocks. Combining the various components, the firm's demand curve is given by

$$\ln(Q_{it}) = \ln(\Phi_t) + \delta_1 Prod_i + \beta \ln(N_{it}) + \ln(M) + \delta_2 \cdot Shale_t - \theta \ln\left(P_{it}\right) + \epsilon_{it}$$

In my baseline specification, I assume that $\Phi_t = \Phi$. I turn now to the pipelines production and investment decisions.

Figure 1.7: Overland Pass Pipeline Subsystems



1.6 Supply

My model of pipeline investment and market participation builds off of the model proposed in Ryan (2012). In my model I abstract away from oligpolistic competition and instead incorporate price regulation.

1.6.1 Markets and Timing

There are N markets which are defined as regional origin-destination pairs and are assumed to be served by a single pipeline.[12] There is an infinite number of periods, with each period corresponding to a year. In each period, the sequence of events is as follows

1. Operating pipelines receive a random scrap value, a random fixed cost of expansion, and an i.i.d. productivity shock. Pipelines observe the evolution of their price cap and the state of the industry. The state of the industry includes the level of the producer price index, the average cost change for operating pipelines from the previous period, and the current term X. They then decide whether or not to exit.

2. Pipelines that continue to operate choose their level of investment in productivity and

[12] I assume further that pipelines do not operate in multiple markets.

system expansion. Investments mature in the subsequent period.

3. Operating pipelines choose a price level subject to the price cap and shippers demand transportation at the posted price. Pipelines incur both a variable and fixed cost of production.

4. Pipelines that are potential entrants receive a random sunk cost of entry and a productivity draw. They then decide whether to enter and, if so, how large of a system to build. Entrants do not produce in the period they enter.

In the following section, I discuss the firm's per-period profit function and my functional form assumptions. I then discuss the state-evolution of each pipeline and conclude with the Bellman equations for operating pipelines and potential entrants.

1.6.2 Per-Period Profits

The pipeline's per-period profits are determined by the revenue generated from providing transportation services less the costs associated with providing the service and any investment costs incurred during the period. Letting i index the pipeline system and t the period, the per-period profit in state $s_{it} = \{\bar{P}_{it}, N_{it}, \omega_{it}, \lambda_{it}\}$ is given by

$$\pi(P_{it}, \xi_{it}, \Delta_{N,it}, s_{it}) = P_{it} \cdot Q(P_{it}, N_{it}, \lambda_{it}) - c(Q_{it}, N_{it}, \omega_{it}) - i_{it} - \Gamma_N(\Delta_{N,it}, \gamma_0) - FC * N_{it}$$

where price must satisfy $P_{it} \leq \bar{P}_{it}$. The variable i_{it} represents the pipeline's level of productive investment and $\Delta_{N,it}$ the pipeline's investment in system expansion. The variable cost of production is $c(Q_{it}, N_{it}, \omega_{it})$, which depends on the level of production Q_{it}, the number of routes N_{it}, and the productivity of the pipeline ω_{it}. I allow variable costs to increase with the size of a pipeline as increasing the geographic footprint of a pipeline may come with additional costs.

Pipelines are subject to convex adjustment costs when changing their number of routes, given by $\Gamma_N(\Delta_{N,it}, \gamma_{0,it})$. Additionally, I assume that pipelines draw a random fixed cost in each period associate with expanding their system, given by $\gamma_{0,it}$. Pipelines incur a fixed cost FC each period that depends on the total number of routes that they provide.

Potential entrants draw a random sunk cost each period, given by κ_{it}, which they must pay in order to start producing. These costs can include the development costs and regulatory risk.[13] Operating pipelines draw a random scrap value ϕ_{it} each period that they receive if they exit. I summarize these costs in the following function, where a is the pipelines action

$$\Phi(a;\kappa_{it},\phi_{it}) = \begin{cases} -\kappa_{it}, & \text{if the firm is an entrant} \\ \phi_{it}, & \text{if the firm exits} \end{cases}$$

The per-period profits including entry and exit costs are given by

$$\tilde{\pi}(P_{it},\xi_{it},\Delta_{N,it},a_{it},s_{it}) = \pi(P_{it},\xi_{it},\Delta_{N,it},s_{it};\theta) + \Phi(a_{it};\kappa_{it},\phi_{it})$$

I next discuss my parameterization of the firm's variable and adjustment cost functions, as well as the distributional assumptions for the various fixed costs.

1.6.2.1 Cost Functions

Firm marginal costs are assumed to have the following functional form

$$\ln \mathrm{mc}_{it} = \gamma_1 + \gamma_2 \ln(\hat{Q}_{it}) + \gamma_3 \ln(\hat{Q}_{it})^2 + \gamma_4 \ln(N_{it}) + \gamma_4 \ln(N_{it})^2$$

where $\hat{Q}_{it} = \frac{Q_{it}}{e^{\omega_{it}}}$ is the productivity adjusted output. This functional form nests the commonly assumed case where marginal cost is constant in the quantity produced, i.e. $\gamma_2 = \gamma_3 = 0$. However, it also allows for the case of U-shaped marginal costs as well as monotonically increasing or decreasing costs.

I assume that the convex cost of adjusting routes is given by

[13] A prime example comes from the proposed Keystone XL Pipeline, which incurred \$1.5 billion in development costs before ultimately being canceled after its permit was revoked by the Biden Administration. See the TC Energy 2019 Annual Report.

$$\Gamma_N(\Delta_{N,it}, \gamma_{0,it}) = \gamma_{0,it} \cdot I(\Delta_{N,it} \neq 0) + \gamma_{\Delta 0}\Delta_{N,it} + \gamma_{\Delta 1}\Delta_{N,it}^2$$

I assume that the fixed cost of investing $\gamma_{0,it}$, the sunk cost of entry κ_{it}, and the scrap value upon exit ϕ_{it} are all normally distributed and independent of one another.

1.6.3 State Transition

The size of the pipeline system evolves endogenously, with

$$N_{it+1} = N_{it} + \Delta_{N,it} \tag{1.1}$$

An investments at time t does not mature until the following period. Productivity evolves endogenously but is subject to a stochastic shock. Following Ericson and Pakes (1994), pipelines invest an amount i_{it} to improve their distribution of productivity draws in the subsequent period. That is, the distribution of productivity at $t+1$ is given by $p(\omega_{it+1}|\omega_{it}, i_{it})$. The authors specify a productivity process where productivity can be decomposed as $\omega_{it+1} = \omega_{it} + \xi_{it}(\omega_{it}, i_{it}) + \eta_{it}$, where τ_{it} represents the deterministic change in productivity from investing i_{it} and η_{it} is a stochastic component. I make three modifications. First, I assume ω_{it} is an AR(1) process, absent investment. Second, I assume that $\xi_{it}(\omega_{it}, i_{it}) = \xi_{it}(i_{it})$. Finally, I assume that ξ_{it} is continuous instead of discrete. This yields the following process

$$\omega_{it+1} = \psi_0 + \psi_1\omega_{it} + \xi_{it}(i_{it}) + \eta_{it+1} \tag{1.2}$$

where $\eta_{it} \sim N(0, \sigma_\eta)$ is a productivity shock. I implicitly define $\xi(i_{it})$ by the equation $i_{it} = \gamma_{\xi,0}\xi_{it} + \gamma_{\xi,1}\xi_{it}^2$. The productivity shock η_{it} can represent decreased output due to a spill, unanticipated pipeline maintenance, or cyber-attacks. I assume that the firm's demand shifter λ evolves according to

$$\lambda_{it+1} = \lambda_{it} + \ln(N_{it+1}) - \ln(N_{it})$$

This assumes that firms do not anticipate the shale boom or the potential entry that may occur after a pipelines has entered the market. This assumption is made to simplify the problem for estimation. An alternative assumption would be that firms use a reduced-form rule that predicts the evolution of λ_{it} and that this rule would account for changes in production and changes in the competitive environment.

The price caps evolve according to a PPI + X rule, i.e.

$$\bar{P}_{it+1} = \bar{P}_{it}\left(\frac{PPI_{t+1}}{PPI_t} + X\right)$$

All that remains is to specify the price cap rule. During a review period, which I index to $\tau = 0$, the adjustment factor is set to match the difference in the growth in annualized average total cost less the change in the producer price index.[14] That is

$$X_{t+1} = \begin{cases} \bar{a}_t - E\left[\left(\frac{PPI_{t+5}}{PPI_t}\right)^{\frac{1}{5}}\right] & \text{if mod(t,5)} = 4 \\ X_t & \text{otherwise} \end{cases} \tag{1.3}$$

Firms keep track of an aggregate state variable, $\bar{a}$ which represents the annualized change in average total cost.[15] This would normally require firms to keep track of the joint distribution of productivity, network size, and price caps to determine the likely evolution of this state variable. As this would be intractable, I following Krusell and Smith (1998) and Gowrisankaran and Rysman (2012) and assume that the firms approximate the evolution of the aggregate state of the economy according to

$$\bar{a}_{t+1} = f(S_t) + u \tag{1.4}$$

[14]See my comment on the Kahn Methodology and the adjustments that are necessary to better reflect this process.

[15]In my counterfactuals this aggregate state can represent different measures of cost or productivity. For instance, this measure can represent the evolution of total factor productivity or of marginal costs.

where $S_t = \{PPI_t, X_t, \bar{a}_t\}$ is the aggregate state. Here, f represents a reduced form approximation of the transition dynamics that depends only on the aggregate state variables. In equilibrium, the firm's approximation must be consistent with the actual evolution of the aggregate state. I include the residual u to represent the fact that f is only an approximation.

1.6.4 Value Function

Pipelines seek to maximize the real discounted sum of cash flows net investment given their information at time 0. Firms discount the future at a rate $\beta = 0.87$, which I chose to coincide with the mean weighted average cost of capital (WACC), using quantities as weights.

$$V_0(s_{i0}) = \max_{\{N_{it},\xi_{it},P_{it}\}_{t\geq 0}} E\left[\sum_t \beta^t \tilde{\pi}(s_{it};\theta) \middle| I_0\right] \tag{1.5}$$

$$\text{s.t. } P_{it} \leq \bar{P}_{it}, \forall t \tag{1.6}$$

and subject to the transition laws in (1.1), (1.2), (1.3) and the perceived law of motion in (1.4). The value function for an incumbent incumbent is given by

$$V_\tau(s;\theta,\epsilon) = \max_{\{P,\xi,\Delta\}} \left\{\tilde{\pi}(s;\theta) + \int \max\left\{\phi, \max_{\Delta,\xi}\left[\beta \int E_\epsilon V_{\tau'}(s';\theta,\epsilon) dP(s'|s)\right]\right\} dF(\phi)\right\} \tag{1.7}$$

with $\epsilon = \{\phi, \kappa, \gamma\}$. Here, τ indexes the time that has elapsed since a review period, and evolves according to $\tau' = \tau + 1 \mod T$. I assume that the price cap is reset before the period $\tau = 0$. The corresponding value function for an entrant is given by

$$V_\tau^e(s;\theta,\epsilon) = \int \max\left\{0, \max_{\xi,\Delta}\left[-\Phi_N(\Delta_N;\gamma) + \beta \int E_\epsilon V_{\tau'}(s';\theta,\epsilon) dP(s'|s)\right] - \kappa\right\} dF(\kappa) \tag{1.8}$$

1.6.4.0.1 Equilibrium An equilibrium in this model is a set of policy functions and a law of motion for the aggregate states $\{PPI, X, \bar{a}\}_t$ such that, given the evolution of the aggregate state, the policy functions solve (1.7) for operating firms and (1.8) for potential entrants, the policy functions generate the law of motion (1.4), and the policy functions are consistent with evolution of X.

Now that I have described the theoretical model, I will explain how I map the model to the data.

1.7 Empirical Strategy

Estimation proceeds in two stages. In the first stage, I recover the determinants of firm profitability in each period, including the demand curve, the cost curve, and firm level productivity. In the second stage, I use a nested fixed point (NFXP) estimation routine where I guess a value for the parameters, solve the firm's dynamic programming problem, and minimize a Generalized Method of Moments (GMM) criterion. I describe the details of each stage in turn. The reader who is uninterested in the estimation details may proceed directly to Section (3.6).

1.7.1 Demand Estimates

Demand parameters are recovered using two-stage least squares. I estimate several versions of demand, with the most comprehensive having the following form

$$\ln(Q_{it}) = \beta_{i0} + \beta_1 \ln(N_{it}) + \beta_2 Prod_{it} + \beta_3 Shale_t - \alpha \ln(P_{it}) + \epsilon_{it} \tag{1.9}$$

where β_{i0} represents a market fixed effect that does not vary over time, $Prod_{it}$ classifies a pipeline as either carrying refined petroleum product or crude oil and controls for changes in the product type, and $Shale_t$ is an indicator variable for the years 2010 and onward. Here, ϵ_{it} is estimated as the residual of this regression. The market-specific intercept control for variations in the level of demand along specific routes. The shale indicator captures the change in demand for oil transportation that resulted from the shale boom. I allow all pipelines to be equally impacted by the shale boom as most oil producing regions have shale deposits. Finally, I control for the type of pipeline as product pipelines generally charge lower rates than crude oil pipelines.

The exogenous movement of the price cap helps trace out the demand curve. However, while the large majority of pipelines have prices regulated using the price cap there are still several that are regulated under different regulatory mechanisms. For instance, pipelines in the Trans-Alaskan Pipeline System (TAPS) are regulated using a cost-of-service scheme because their costs increase fast enough that they are unlikely to recover their cost of capital under the price cap mechanism. I exclude these pipelines when estimating the full model but include them when estimating demand as they still provide useful variation for recovering price elasticities. Including these pipelines can reintroduce an endogeneity problem as prices are now jointly determined with output. To account for this, I use several different instruments. First, I include lags of the firm state variables, including productivity, system size, and the producer price index. These variable directly impact firm costs and therefore provide valid instruments to identify the demand curve.

1.7.2 Supply Estimates

Two key inputs in my analysis are firm productivity and marginal costs. My approach is to recover productivity by estimating an industry production function and then treating the residual as pipeline level total factor productivity. Then, I follow Dhyne et al. (2020) and use cost minimization first-order conditions to recover marginal costs. I then regress the marginal cost on productivity-adjusted output and system routes to recover the cost function. I discuss each step in turn.

1.7.2.0.1 Production Function Estimation I assume that the production technology is translog during estimation. The translog production function is given by

$$q_{it} = \beta_0 + v_{it}\beta_l + k_{it}\beta_k + v_{it}^2\beta_{ll} + k_{it}\beta_{kk} + k_{it}v_{it}\beta_{lk} + \omega_{it} + u_{it}$$

where v_{it} and k_{it} are logged variable and fixed inputs, respectively, ω_{it} is productivity known to the firm but unobserved by the researcher, and u_{it} is an i.i.d. error that can be thought of as an unanticipated productivity shock or approximation error. Here, ω_{it} is the researchers problem as firms know their productivity when optimally choosing the level of variable input, v_{it}. As such, we need to control for the endogeneity of v_{it} and ω_{it} during estimation. I follow the literature and assume that ω_{it} follows a Markov process. Since the policy function for productivity investment depends on the pipeline's information set at time $t-1$, I assume that technology evolves according to the following Markov process

$$\omega_{it} = g(\omega_{it-1}, s_{it-1}, S_{it-1}) + \eta_{it}$$

where s_{it-1} is the pipeline's individual state variables and S_{it-1} are aggregate state variables.

Two issues have recently been highlighted in the literature regarding when estimating production functions. First, it is common to use deflated revenue in place of physical output when estimating production functions as output is generally not observed in accounting data. However, if firm's are heterogeneous in their markups then using deflated revenue will bias the parameter estimates. I observe output directly, so this issue does not impact my results. Second, the assumptions underlying the most popular method for production function estimation, the control function approach, often do not hold in imperfectly competitive or regulated markets. I account for this by using the estimation routine proposed in Ponder (2021) which remains valid even when these assumptions are violated. The author shows that the parameters of the production function can be recovered using semi-parameteric two-stage least squares if u_{it} is independent of a set of instruments. This avoids the need to use a control function and therefore does not rely on a monotonicity assumption.

1.7.2.0.2 Estimating Markups and Marginal Costs To estimate firm level markups, I follow the insight of Hall (1988) and De Loecker and Warzynski (2012) and use a cost minimization approach to approximates markups. In order to produce a given quantity $\bar{Q}_{it}$, they solve the following cost minimization problem

$$\min_{V_{it}} \; w_{it}V_{it} + r_{it}K_{it} \tag{1.10}$$

$$\text{s.t. } Q_{it} \geq \bar{Q}_{it} \tag{1.11}$$

where w_{it} is the price of the variable input. The cost minimization first order conditions require that

$$w_{it} = \lambda_{it} \frac{\partial Q_{it}}{\partial V_{it}}$$

where λ is the Lagrange multiplier on the production constraint. Because this describes the increase in costs associated with a unit increase in output, the multiplier is exactly equal to the pipeline's marginal cost, mc_{it}. Multiplying both sides by V_{itk} and dividing by $Q_{it}P_{it}$ yields

$$\frac{V_{itk}w_{itk}}{P_{it}Q_{it}} = \frac{\lambda_{it}}{P_{it}} \frac{\partial Q_{it}}{\partial V_{itk}} \frac{V_{itk}}{Q_{it}} = \frac{\theta_{itk}}{\mu_{it}}$$

The term μ_{it} is the markup. Importantly, this relationship holds for any flexible input that directly enters the production function. Operating expenses and revenue are reported in pipeline financial statements. To estimate markups, wI only need an estimate of the input elasticity θ_{it} which comes from the production function estimates. Dhyne et al. (2020) extend this procedure to the multi-product setting and show that observing quantity data allows us to consistently estimate marginal costs using

$$\mathrm{mc}_{it} = \frac{V_{itk}w_{itk}}{Q_{it}} \frac{1}{\theta_{itk}}$$

Given estimates of marginal costs we can determine whether markup changes were driven by changes in price, marginal cost, or both. Additionally, the marginal cost estimates allow us to test whether marginal costs are constant, as is commonly assumed in the regulatory literature, or depend on the level of output.

1.7.2.0.3 Cost Function

Given estimates of firm productivity, I generate a productivity adjusted output level, $\hat{q} = q_t - \hat{\omega}_t$. I then estimate the marginal cost function using the marginal costs recovered from the cost-minimization FOCs. This procedure follows that used in Dhyne et al. (2020). The authors directly estimate the variable cost function, controlling for firm productivity, capital, and input prices. I do not directly observe input prices, so I assume they are common across the industry and control for them using annual fixed effects. Additionally, I directly regress marginal costs on productivity adjusted quantities, rather than variable costs. The marginal cost function is given by

$$\ln \hat{\text{mc}} = \gamma_{1t} + \gamma_2 \ln(\hat{Q}) + \gamma_3 \ln(\hat{Q})^2 + \gamma_4 \ln(N) + \gamma_4 \ln(N)^2 + u_{it}$$

Here, u_{it} is assumed to be a residual that is due to either measurement error or misspecification error. In order to control for the potential endogeneity between u_{it} and the covariates, I use two-stage least squares to estimate $\{\gamma_k\}$. Specifically, I use the price cap index and the spot price of West Texas Intermediate in Cushing, Oklahoma as demand shifters. The price cap index is exogenous to firm decisions (by design) and moves around the price of transportation. The spot price of oil is largely driven by movements in global supply, which oil pipeline rates play a de minimis role in determining. As such, the spot price of oil is assumed to be exogenous and therefore a valid instrument. Further, the spot price of oil determines the level at which fields produce, making it a relevant instrument. Finally, I include individual firm dummies as instruments, imposing the u_{it} is mean zero for each pipeline. I estimate models with and without the quadratic terms.

1.7.2.1 Adjustment and Fixed Costs

Given the parameter estimates that determine per-period profits, I estimate the remaining model parameters using a nested-fixed point algorithm following Rust (1987). In the inner loop, I solve for the firm's value function and optimal policy functions given a guess of the parameter coefficients. Then, I form four residuals

$$\eta_{1it} = \omega_{it} - \psi_0 - \psi_1\omega_{it-1} - \xi_{it}(X_{it-1}) \tag{1.12}$$

$$\eta_{2it} = \Delta_{it} - \Delta(X_{it}) \tag{1.13}$$

$$\eta_{3it} = I\left[exit_{it} = 1\right] - \Phi\left(\frac{EV(X_{it}) - \mu_\phi}{\sigma_\phi}\right) \tag{1.14}$$

$$\eta_{4it} = I\left[enter_{it} = 1\right] - \Phi\left(\frac{EV^e(X_{it}) - \mu_\kappa}{\sigma_\kappa}\right) \tag{1.15}$$

The first residual represents is the stochastic shock to productivity. The second residual is the difference between observed investment and the models predication. I assume that this is specification error and that the residual is mean zero, conditional on a set of instruments. Finally, the third and fourth residuals represents the difference between the observed exit and entry of firms and the models predicted probabilities. As the sample size grows, the sample frequency of exit and the predicted probability of exit should converge. I then interact these residuals with a set of instruments, Z_{kit}. Valid instruments include the state variables of firms at $t-1$, as these should be correlated with their decision at time t but should be orthogonal to expectational errors at time t.

Letting Z be a block diagonal matrix of Z_k, $\eta = \{\eta_1, \eta_2, \eta_3, \eta_4\}$, and θ the vector of parameters, we can form the following residual

$$g(\theta) = Z'\eta(\theta)$$

such that in expectation

$$E[g(\theta)|Z] = 0$$

The GMM objective function is then

$$GMM(\theta) = g(\theta)'Wg(\theta)$$

and the outer loop searches over θ to minimize this quantity. Note that the size of the residuals can be quite different, which leads the objective function to place more weight on residual (1.13). To remedy this, I take a two-step approach. First, I use an initial guess of θ_0 and calculate the implied residual $\hat{\eta}_0$. I generate

$$g_{it}(\theta_0) = Z_{it} \cdot \hat{\eta}_0$$

and approximate the optimal weight matrix

$$\hat{W}_0 = \frac{1}{NT} \sum_{it} g_{it}(\theta_0) \cdot g_{it}(\theta_0)'$$

I estimate the model parameters using the nested-fixed point algorithm to generate the first set of consistent estimates. I recalculate the optimal weighted matrix using these estimates and then re-run the estimation routine.

1.8 Estimation Results

1.8.1 Demand

Table (1.1) I present several different specifications for demand. The simplest is presented in column (1), where I only use price as an independent variable and estimate the equation using OLS. The coefficient implies a relatively low elasticity of -1.35. Interestingly, the R^2 for this regression is 0.512, meaning that a constant term and price have a significant amount of explanatory power. Column (2) presents the same regression, but this time uses two-stage least squares. The estimated elasticity decreases marginally to -1.43. This modest change is likely due to the fact that prices are almost set exogenously. When firms price at the ceiling, then price changes are driven entirely by movements in PPI and this exogenous changes serves to trace out the demand curve.

Table 1.1: Demand Estimates

	Dependent variable: ln(Q)					
	OLS	*Instrumental Variable*				
	(1)	(2)	(3)	(4)	(5)	(6)
ln(P)	-1.346^{***}	-1.432^{***}	-1.440^{***}	-1.351^{***}	-1.488^{***}	-1.522^{***}
	(0.024)	(0.027)	(0.027)	(0.024)	(0.025)	(0.066)
Prod. Pipeline			-0.159^{***}	-0.471^{***}	-0.445^{***}	-0.239^{***}
			(0.058)	(0.054)	(0.053)	(0.059)
ln(N)				0.493^{***}	0.479^{***}	0.178^{***}
				(0.019)	(0.019)	(0.034)
Shale Boom					0.72^{***}	0.64^{***}
					(0.006)	(0.006)
Constant	21.116^{***}	22.079^{***}	22.237^{***}	20.537^{***}	21.637^{***}	25.324^{***}
	(0.270)	(0.300)	(0.305)	(0.284)	(0.285)	(0.731)
Market FE	No	No	No	No	No	Yes
Observations	3,005	3,005	3,005	3,005	3,005	3,005
R^2	0.512	0.510	0.511	0.601	0.624	0.877
Adjusted R^2	0.512	0.510	0.510	0.601	0.624	0.864
Residual Std. Error	1.590 (df = 3003)	1.593 (df = 3003)	1.592 (df = 3002)	1.438 (df = 3001)	1.396 (df = 3000)	0.839 (df = 2730)
F Statistic	$3{,}151.042^{***}$ (df = 1; 3003)					

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Pipelines move different types of product and these products might have different demand shifters. To account for this difference, column (3) adds a dummy variable to describe whether or not the pipeline ships crude oil or refined petroleum product. The estimates imply that demand for transporting refined petroleum product is lower by roughly 15% at any price point. Column (4) includes the log number of routes that a pipeline provides. Ideally, we would be able to estimate demand at the individual route level. However, data are not available at this level of granularity. Including the total number of routes provided captures the fact that demand shifts out as pipeline enter additional markets. Including the log number of routes increases the R^2 by almost 0.1, showing the importance of this variable in explaining transportation demand.

I consider two additional demand specifications. In column (5) I include a shifter for the Shale Boom, which I assume starts in 2010. The coefficient is statistically significant, and shows that demand increased by roughly 100% by 2020. However, this covariate explains little of the residual variation.[16] Finally, I include a pipeline fixed effect, reported in column (6). This increases the R^2 from 0.624 to 0.877, implying that market specific demand shifters are important component in describing demand.

Notably, the estimated demand elasticities all fall within a range of -1.52 to -1.35. The specification with the most covariates also has the most elastic demand. One potential problem with these estimates is that they place a limit on the maximum markup that a monopolist will charge. Specifically, the ratio of a monopolist's marginal revenue to demand is given by $\left(1 - \frac{1}{|\epsilon|}\right)^{-1}$. An estimated elasticity of -1.5 implies a ratio of 3, so a monopolist pricing at the profit maximizing price will never have a price to marginal cost ratio above 3. With the production function approach, the estimated ratio is 6 as of 2020, meaning that these demand estimates are not able to rationalize the markups in the second half of my sample. I turn now to the supply side estimates.

[16] One could allow this coefficient to have a time trend, reflecting the fact that production did not increase instantaneously. However, this leads to very similar results and does not improve the fit of the estimating equation.

1.8.2 Supply

I begin with the results from the first stage of estimation, including the production function estimation results and implied marginal cost function. Then, I discuss the results from the second stage of estimation, which includes the parameters associated with the exogenous part of productivity, the convex adjustment costs, and the distributional costs associated with entry, exit, and investment.

1.8.2.1 Production Function Estimates

I estimate the production function parameters using several different specifications. The benchmark estimates are recovered using OLS, ignoring the endogeneity between ω_{it} and the inputs. To the extent that the covariance between the variable inputs and productivity is low, this would tend to give a reasonable approximation to the truth and does not depend on the other modeling assumptions. For my first specification, I use a Cobb-Douglas production function and for the second I use a translog production function. Next, I assume that productivity follows a Markov process and use the semi-parametric estimator from Ponder (2021) to control for endogeneity. Cross-validation is used to determine the number of terms included in the polynomial approximation to the productivity process. As with the OLS estimates, I estimate both a Cobb-Douglas and a translog production function.

Before presenting the results, a quick comment on my measure of output is in order.[17] Researchers have often used deflated revenue when estimating production functions because financial data rarely provides information on physical output. However, I see several measures of pipeline output in my dataset, which I use to bring the estimation routine closer to economic theory. My preferred measure of output is barrel-miles, i.e. the number of barrels times the total distance each barrel traveled. However, these data are only reported on an annual basis, which limits the size of my dataset. Alternatively, I can use barrels or deflated operating revenue, which are reported quarterly. The disadvantage of using barrels is that it does not take into account the distance traveled. As such, a long-haul pipeline and a short-haul pipeline might have the same reported barrels in a year, but the long-haul pipeline has considerably higher input costs. This ultimately leads to estimates that imply

[17] I provide a detailed description of the various variables I use and how I construct each input time series in the appendix.

significant decreasing returns to scale. Operating revenue circumvents this issue, as the long-haul shipments yield a significantly higher revenue than short-haul movements. However, operating revenue potentially biases the estimates because an increase in revenue can come from increasing prices or increasing output. Using barrel-miles comes with the cost that my sample size is reduced to a quarter of the full sample. I present the results using annual data here, and then provide results using other measures of output in the appendix. As an additional robustness check, I estimate the production function assuming that there are errors in the measurement of capital. These estimates are also discussed in the appendix.

Table 1.8 shows the results of the various estimators. Column (1) shows the Cobb-Douglas estimates using OLS. The mean variable input elasticity of 0.87 is the largest of all my estimates, roughly 20% greater than my preferred estimates. Column (2) shows the translog estimates, again using OLS. None of the second-order terms are statistically significantly and a F-test fails to rejected the hypothesis that these terms are zero. The mean variable elasticity is then estimated to be roughly 0.83 and the mean capital elasticity 0.4, implying a returns-to-scale of roughly 1.23. Column (3) shows the results for the semi-parametric estimator using the Cobb-Douglas specification. The mean variable input elasticity decreases by roughly 20% while the capital elasticity remains relatively unchanged. Column (4) shows my preferred estimator. Production is assumed to be translog and I use the semi-parametric estimator. The second-order terms are now statistically significant, save for v_{it}^2, and the mean capital elasticity increases from 0.38 to 0.53. The implied returns-to-scale are comparable to those implied by the OLS estimates. The principal difference is that the variable elasticity is much lower and the capital elasticity is higher, which we would expect if variable inputs were correlated with unobserved productivity.

1.8.2.2 Cost Function

The results of the marginal cost regression are presented in Table (1.2). The dependent variable is the estimate log marginal cost of production for each pipeline, in each period. The independent variables include the productivity-adjusted measure of output, $\hat{q} = q_t - \omega_t$ and the number of routes. All coefficients are statistically significant at $\alpha = 0.1$. The model with quadratic terms has a slightly better fit, with an adjusted-R^2 of 0.867 compared to an

Figure 1.8: Production Function Estimates

	Dependent variable: q_t			
	OLS		Semi-Parametric	
	CD	Translog	CD	Translog
k_t	0.362***	0.348***	0.381***	0.242***
	(0.026)	(0.118)	(0.017)	(0.029)
k_t^2		0.007		0.031***
		(0.021)		(0.007)
v_t	0.873***	0.907***	0.728***	0.470 ***
	(0.028)	(0.102)	(0.017)	(0.070)
v_t^2		-0.018		0.024
		(0.024)		(0.015)
$v_t k_t$		-0.001		0.023***
		(0.041)		(0.004)
Constant	2.535***	2.471***		
	(0.081)	(0.185)		
Observations	2,863	2,863	2,863	2,863
OPEX Elast.	0.873	0.836	0.728	0.730
CAPEX Elast.	0.362	0.403	0.381	0.528

Note: *p<0.1; **p<0.05; ***p<0.01

adjusted-R^2 of 0.8 for the log-log specification.

A model with constant marginal costs would estimate the coefficient on output to be 0. The null hypothesis that marginal costs are constant is rejected at $\alpha = 0.01$. This implies that assuming a constant marginal cost may be missing an important dimension of the regulated firm's environment. Marginal costs are shown to U-shapes in that they increase significantly as production approaches zero and increase gradually as output increases. Increasing the number of routes has a positive impact on the marginal cost of production, reflecting the fact that increasing the size of a system requires additional pumping stations and adds complexity to the scheduling process. However, the quadratic term implies that, over the range of the data, adding additional routes increases marginal cost but at a decreasing rate.

1.8.2.3 Productivity, Adjustment Costs, and Distributional Parameters

Table (1.3) shows the results of the nonlinear GMM estimation. The first set of parameters determine the exogenous changes to productivity and productive investment. The average productivity level in my sample is roughly 3.0, meaning that the average pipeline would expected their efficiency to decline exogenously by 30% over the sample period. Shocks to productivity are quite large relative to the mean level reflecting the significant role of outside forces. Productivity investment costs are quite large so that most firms only invest to improve efficiency by 3% - 10% annually and a large number of firms make no investment at all.

Each additional route is estimated to cost roughly \$74 million and this cost increases as more routes are added. However, the main cost of construction appears to be the fixed cost of investment, which the model estimates to be \$1.1 billion. Interestingly, the mean fixed cost of expanding an existing system is comparable to the mean entry cost of creating a new system, which is \$1.3 billion.[18] The principal difference is that the variance of the investment fixed cost is almost three times are large. This likely reflects the fact that system expansions can be relatively minor or can be comparable to building an entirely new system. The mean scrap value of a pipeline system is estimated to be fairly small at

[18]For reference, the Oil and Gas Journal estimated pipeline construction costs of \$6.57 million per mile in 2014. Since 2010, the average length of new pipeline construction was 260 miles so the average construction cost was \$1.7 billion, slightly higher than what I estimate here.

Table 1.2: Marginal Cost Function Estimates

	Dependent variable:
	ln(mc)
$\ln(\hat{q})$	-0.296^{***}
	(0.097)
$\ln(\hat{q})^2$	0.037^{***}
	(0.007)
$\ln(N)$	0.134^{**}
	(0.065)
$\ln(N)^2$	-0.023^{*}
	(0.012)
Constant	10.001^{***}
	(0.366)
Observations	2,122
R^2	0.882
Adjusted R^2	0.867
Residual Std. Error	0.454 (df = 1893)
F Statistic	61.867^{***} (df = 228; 1893)
Note:	*p<0.1; **p<0.05; ***p<0.01

roughly $\$10,000,000$. However, unlike the other distributional parameters, the variance is substantially larger than the mean value. This likely accounts for the fact that pipelines exit under a variety of circumstances. For instance, existing can incur substantial costs related to pipeline abandonment or generate revenue from the selling of assets.

1.9 Oil Pipeline Industry Performance

Before turning to the results of my full model, I explore the estimated change in industry markups and productivity derived from the production function estimation and cost-minimization FOCs. price caps are in principal meant to ensure that prices cannot diverge substantially from the first-best (price equal to marginal cost) or second-best (price equal to average total cost) price level. However, if the adjustment factor X is set too high then firms may generate excessive rents. A simple test is to use the markups from the cost-minimization FOCs determine how markups evolved over the past two decades. An increase in the markup ratio implies that either prices have increased faster than costs or that cost have declines but these gains have not been passed on to customers. I find that both average price and the average marginal cost have increased over my sample period but that prices increased much faster. The price cap is also meant to encourage productivity gains and several papers in the regulatory literature have documented the evolution of productivity after the introduction of a price cap. While I find that firm productivity declined over the same period, this was likely due to exogenous factors. As I show in the final section, my model implies that the higher price cap led to considerable gains in both productivity and entry than we would have seen under a the traditional cost-of-service regulation.

1.9.1 Evolution of Markups

Using the estimated production function and the observed variable cost-revenue shares, I used cost-minimization FOCs to recover the average industry markup. The evolution of the markups since 2000 are shown in Figure (1.9). The rise in markups has been substantial since 2000, increasing from roughly 2 to 6. In terms of price-cost margins, this is an increase from 50% to 80%. There are two principal terms in the first-order conditions: the revenue-to-operating expense ratio and the input elasticity. We have already seen that the first term increased significantly over the past two decades. It is worth asking to what extent did the

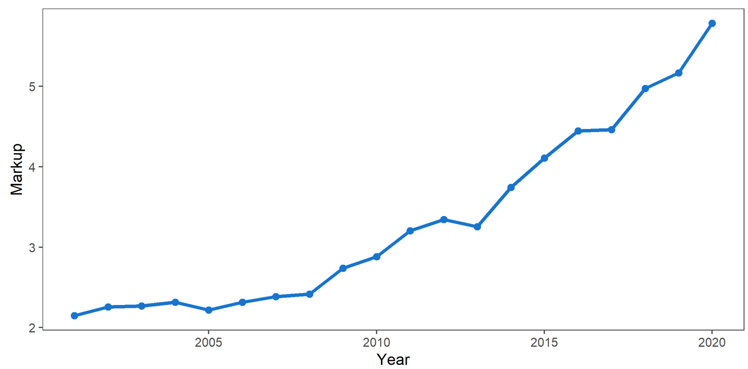
Table 1.3: Parameter Estimates

Parameter	Mean	SE
Exogenous Productivity		
ψ_0	0.183	0.006
ψ_1	0.931	0.005
σ_η	0.527	0.001
Productivity Investment		
$\gamma_{\xi 0}$	178	2.1
$\gamma_{\xi 1}$	10,889	256.9
System Investment		
$\gamma_{\Delta 0}$	74,035	2,172
$\gamma_{\Delta 1}$	985	49
Fixed Cost (FC)	17,269	2,091
Scrap Value		
μ_ϕ	9,995	1,660
σ_ϕ	89,670	16,155
Entry Cost		
μ_κ	1,292,410	975,756
σ_κ	312,315	160,852
Investment Fixed Cost		
μ_γ	1,102,962	152,708
σ_γ	861,524	25,286

Note: Units are in millions of dollars for productivity investment. Units are in thousands of dollars for system investment and all distributions.

second term change. Appendix Figure (A.3) plots the weighted average input elasticity over the same time period. The average level of the input elasticity has increased slightly over this period, meaning that most of the increase in markups has been driven by the first term. I do not report the decomposition of Figure (A.3) here, but each component has been stable as well. Firms appear to be producing roughly the same level of output using the same mixture of inputs, but have been given a higher price, in real terms, for each unit of output.

Figure 1.9: Evolution of Markups



Recently, the literature has noted that markup and marginal cost estimates can be very sensitive to which flexible input is used to estimate their level. In my baseline results, I use a deflate measure of all operating and maintenance expenses, similar to the approach taken in De Loecker and Warzynski (2012) and De Loecker et al. (2020). An alternative approach would be to estimate production function with a separate term for each cost category and estimate the change in markups using different measures of flexible inputs. Figure (1.10) shows the results of this analysis. Using "Materials and Supplies" shows an evolution in markups highly similar to my baseline approach. Markups start at comparable levels in 2000 and increase by roughly the same amount. However, markups increase more uniformly over time, as opposed to baseline results where markups are flat before 2007. Using "Operating Fuel" or "Labor" results in less of an increase in markups, largely due to the higher level that they start at. While the baseline estimates have an average markup of roughly 2 in 2000, using either "Operating Fuel" or "Labor" results in a markup of 3 in the same year. Ultimately, we can bound the increase in markups between 160% and 300%, which represents a substantial divergence from marginal cost pricing.

A rise in markups does not necessitate that firms are acting anti-competitively. It has long

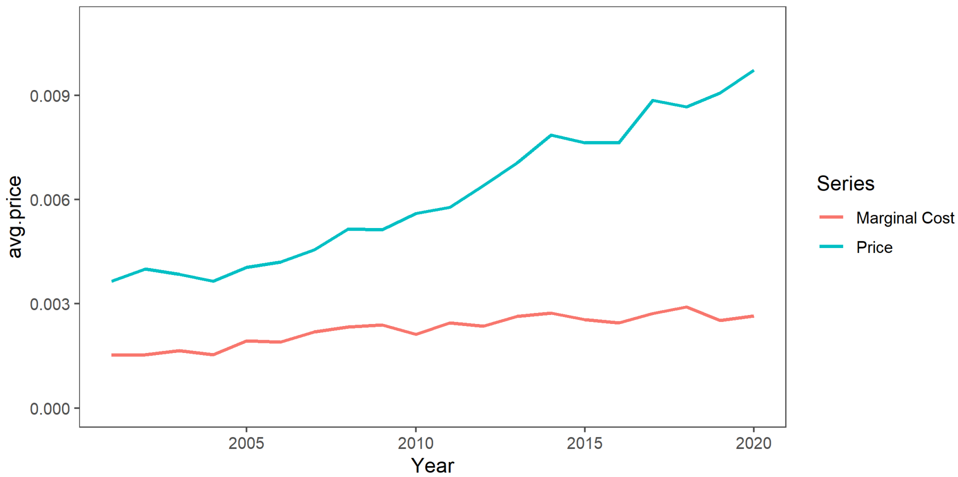
Figure 1.10: Markup by Flexible Input

been noted that high markups can be explained by low marginal costs, high prices, or a combination of the two. As noted in Dhyne et al. (2020), observing physical output allows us to separate out changes in price from changes in marginal cost. Figure (1.11) plots the average price and the average marginal cost over my sample period. Real prices have increased at a faster rate than real marginal costs since 2004, implying that decreasing marginal costs are not driving the change in estimated markups. Note that an increasing average marginal cost is not necessarily an indication of increasing inefficiency. The oil pipeline industry has seen substantial entry and investment since 2000. Many dynamic investment models[19] predict that low cost opportunities are chosen first and high cost investments are delayed. Therefore, new construction will generally be higher cost than the established firm, increasing the industry average cost across time.

Markups increased more than the cumulative change in the average price cap over this period. To understand why, I follow Melitz and Polanec (2015) and decompose the change in the weighted average markup into four components: the unweighted mean change in markups, the change in the covariance between market share and markups, the impact of firm entry, and the impact of firm exit. Figure (1.12) shows this decomposition. The first thing to note is that markups increased significantly due to firm entry. Entrants on average had a higher initial markup than existing firms and this difference contributed to roughly a third of the increase in markups. Firm exit had minimal impact on the change in markups, as firms that exited tended to have markups similar to surviving firms. Surviving firms saw the

[19] For instance, see Hopenhayn (1992).

Figure 1.11: Evolution of Average Price and Marginal Cost



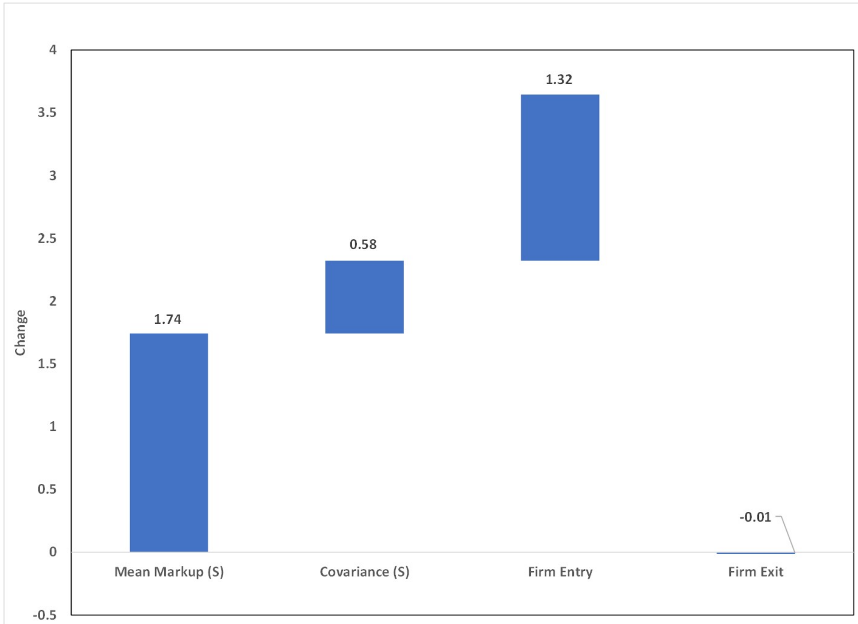
covariance between their market share and their markups increase by 0.58, accounting for 16% of the markup increase. Finally, the average markup ratio increased by 1.74, accounting for roughly 50% of the overall increase. As such, a significantly portion of the increase in markups is due to firm entry and reallocation of markups to firms with higher output. Of the overall increase, 50% is due to the average price level increasing across time and it is this share that could be attributable to changes in the price cap index.

1.9.2 Evolution of Productivity

Figure 1.13 displays the change in (demeaned) weighted average productivity over the sample period. The log change of roughly -0.45 between 2001 and 2020 is consistent across all specifications and estimation routines, and corresponds to a roughly 50% decrease in total factor productivity. Figure (1.14) performs the same Melitz and Polanec (2015) decomposition described in the section on markups. On average, surviving pipelines were considerably less productive at the end of the sample and that the covariance between size and productivity declined. However some of this decline was offset by firm entry, where new firms entered with a higher average level of productivity than the existing firms.

This decline in factor productivity is not necessarily caused by an ineffective price cap as there have been several changes to the regulatory environment since 2000. The decrease in

Figure 1.12: Decomposition of the Change in Markups



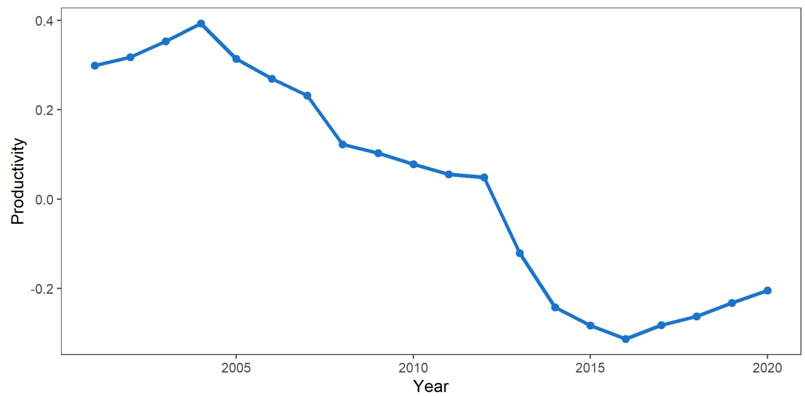
factor productivity might, in part, reflect a change in the quality of service through increased safety and environmental regulation. Pipeline repairs lead to a decrease in the likelihood of spillage and the associated loss of service, which is arguably good for both shippers and households (who would experience a significant externality if a pipeline ruptured).[20] Figure (A.5) in the appendix shows that total pipeline incidents have declined over time, after accounting for the number of active pipelines, reflecting an increase in the quality of transportation. Therefore, in order to disentangle the impact of exgoneous changes in productivity and changes in response to the price cap index, it is necessary to analyze the full theoretical model.

1.9.3 Impact of the price cap on Industry Dynamics

To assess the impact of the price cap regulation on industry performance, I rely on my theoretical model. The first counterfactual that I run assumes that FERC maintained cost-of-service regulation in the industry. In this experiment, firms are allowed a maximum 13% return on their cost-base in each period, where 13% was chosen to match the observed average weighted average cost of capital. I estimate the impact of the regulation change on two sets of firms: those that have operated continuously since 2000, which I call mature

[20]See the discussion in 18 CFR Part 342, "Five-Year Review of Oil Pipeline Pricing Index".

Figure 1.13: Change in Weighted Average Productivity

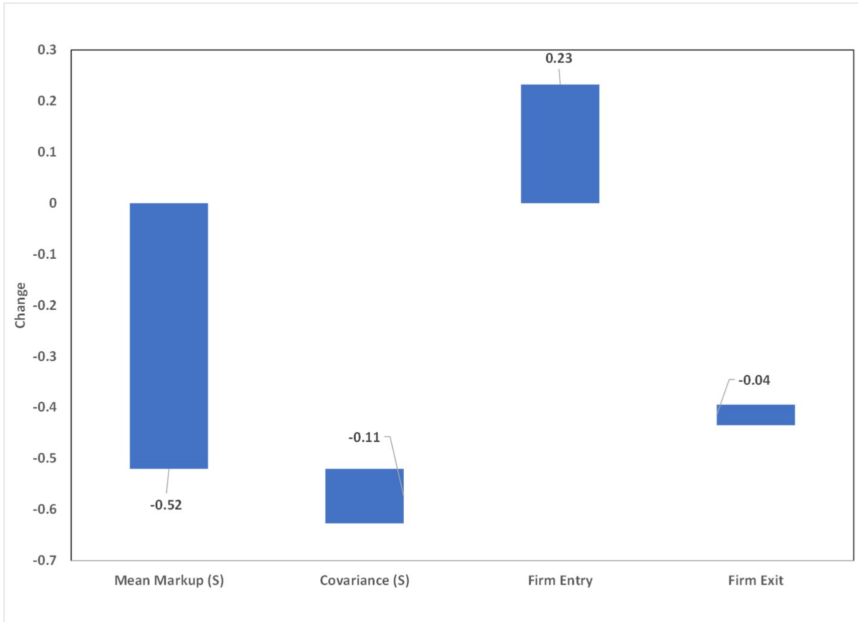


markets, and all firms. Table (1.15) presents the results. As predicted by theory, firm profits are greatly constrained leading to a significant decline in producer profits. Customers in mature markets would have seen a substantial increase in their consumer surplus due to the lower prices. Some of this welfare gain for existing customers was offset by a roughly 19% decline in average productivity and the lower producer profits, however welfare would have been roughly 15% higher in mature markets.

After accounting for the impact on firm entry, total welfare across all markets declines by 7%. Entry is roughly 15% lower, showing that fewer firms found it profitable to enter given the constraint on profits. Interestingly, there was little impact on consumer surplus, as the reduction in consumer surplus from entry was almost exactly offset by increases in consumer surplus in mature markets. Total welfare declined by 7.2% by 2020 driven largely by declines in producer surplus.

The previous results imply that entry plays an important role in determining total welfare. Deregulating prices would have yields the greatest incentive to undertake market expansion, so next counterfactual considers what would have happened if the price cap was removed. Table (1.16) shows the results of this experiment. I again consider the impact on mature markets and all markets. Absent a price cap, firms in mature markets would have seen their profits increase by 8.3% while consumer surplus would have decreased by 3.0%, ultimately resulting in a total welfare decline of 2.2%. There was considerable increase in system expan-

Figure 1.14: Change in Weighted Average Productivity



sion, as pipelines added an additional 3.6 routes. The welfare gains from system expansion would have been offset by a decline in productivity and higher prices. This decline would have been roughly 8%.

Price deregulation did lead to additional firms operating as of 2020. After accounting for the reduction in exit rates, I see 3% more firms operating under price deregulation. [21] However, these gains were offset by the higher prices that existing customers would have ended up paying. When considering all markets, total welfare would have decreased by 1.4% if the price cap was not put in place.

Taken together, the price cap appears to have performed better than either price deregulation or maintaining a cost-of-service mechanism, as it struck a balance between allowing high enough returns to stimulate entry but constraining profits sufficiently that the increase in dead-weight-loss in mature markets did not offset these gains.

[21] I fix the number of markets in the data to those that I observe. The shale boom led to an over-expansion in capacity meaning that it is unlikely additional markets would have been entered, event absent the price cap. As such, the difference in observed operating firms and predicted operating firms in this experiment is due to exit.

Figure 1.15: Welfare Impact of Cost-of-Service

Measure	Price-cap	Cost-of-Service	Percent Difference
Markets with a pipeline operating since 2000			
Producers Profit	25	6	-76.3%
Net Consumer Surplus	298	367	23.1%
Total Welfare	323	373	15.5%
Avg. Routes	19.0	19.1	0.3%
Avg. Productivity	2.55	2.38	-15.7%
All Markets			
Producers Profit	53	9	-83.9%
Net Consumer Surplus	630	626	-0.7%
Total Welfare	684	634	-7.2%
Operating Firms	177	151	-14.7%

Note: Profits and welfare measures are in billions of dollars.
Results are based on simulating 50 samples.

The observed increase in markups since 2000 indicates that there may have been room for improvement. The final counterfactual I consider is to only let price caps rise according to the change in PPI, which is equivalent to removing the adjustment factor X. I call this experiment the "fixed price cap", reflecting that X does not adjust. Table (1.17) shows the impact of fixing the price cap in perpetuity. In mature markets, firm profits would have declined significantly, by roughly 15%. Consumer surplus would have increased by roughly 5.1% by 2020 leading to an overall welfare gain of 3.5%. As expected, firm investment would have declined under a fixed price cap. This decline in investment happens gradually, as can be seen in Figure (1.18), and is a result of firms having lower expectations of future earnings.

Similar to the previous results, productivity would have increased. However, the gains are substantially less than we saw going from no price cap to the current mechanism. The reason for this is that the productivity investment policy function is non-monotonic. A small reduction of a high price cap increases the gain from investing in productivity. The lower price cap results in firms producing more, and since productivity gains lower the cost of infra-marginal production firms will see more of a benefit from making investments in

Figure 1.16: Welfare Impact of Deregulation

Measure	Price-cap	Price Deregulation	Percent Difference
Markets with a pipeline operating since 2000			
Producers Profit	25	27	8.3%
Net Consumer Surplus	298	289	-3.0%
Total Welfare	323	316	-2.2%
Avg. Routes	19.0	22.6	18.7%
Avg. Productivity	2.55	2.47	-8.2%
All Markets			
Producers Profit	53	56	4.8%
Net Consumer Surplus	630	618	-2.0%
Total Welfare	684	674	-1.4%
Operating Firms	177	182	2.8%

Note: Profits and welfare measures are in billions of dollars.
Results are based on simulating 50 samples.

productivity gains. However, as the price cap continues to decrease, firms eventually see the returns to productivity investment decrease as well. Prices being to fall below the cost of production and firms anticipate that they may have to exit from the market. This results in them decreasing their efforts at reducing costs. For precisely this reason, we see only modest gain in productivity under a fixed price cap.

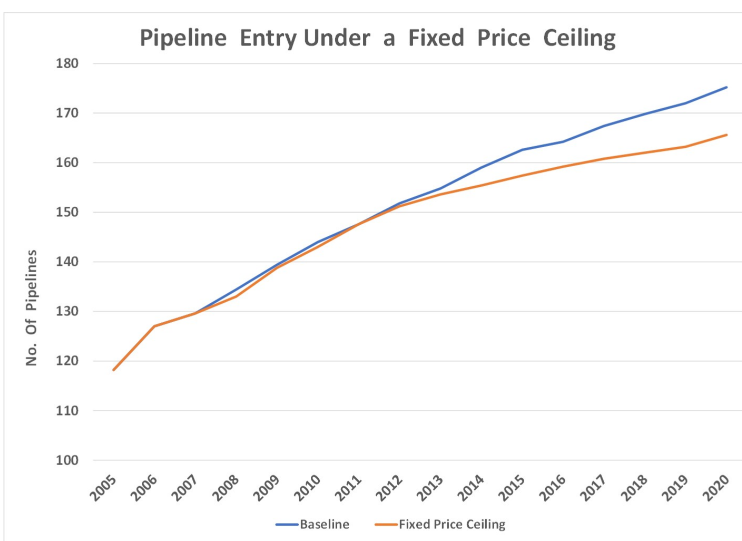
Firms enter new markets at a lower rate, and several firms exit markets that they otherwise would not have, under the fixed price cap. Figure (1.18) shows the difference in active firms under the fixed price cap relative to baseline entry. The decline in firm entry and increase in firm exit have a negative impact on total welfare. However, most of this entry and exit occurs in smaller markets. The larger markets are also the mature markets, which would have seen a large increase in consumer surplus under a fixed price cap. Therefore, the model estimates that welfare would actually have been higher under a fixed price cap by roughly 2.4%. By allowing the price cap to dynamically adjust, FERC essentially incentivized firms to enter new markets by reducing consumer surplus in established markets. Customers of firms in mature markets saw their consumer surplus decrease by 5.1% relative to a fixed-price cap at the same time the industry saw roughly 5% more entry.

Figure 1.17: Welfare Impact of a price cap without the Factor X

Measure	Price-Cap	PPI Price Ceiling	Percent Difference
Markets with a pipeline operating since 2000			
Producers Profit	25	21	-14.8%
Net Consumer Surplus	298	313	5.1%
Total Welfare	323	334	3.5%
Avg. Routes	19.0	18.0	-5.3%
Avg. Productivity	2.55	2.57	2.4%
All Markets			
Producers Profit	53	49	-8.7%
Net Consumer Surplus	630	652	3.4%
Total Welfare	684	700	2.4%
Operating Firms	177	167	-5.6%

Notes: Profits and welfare measures are in billions of dollars.
Results are based on simulating 50 samples.

Figure 1.18: Impact of a Fixed Price cap on Entry and Exit



1.10 Conclusion

In this paper, I examined the impact of price regulation on firm investment and welfare in the oil pipeline industry during the years surrounding the shale revolution. I find that the existing price cap regulation led to an increase in returns for firms over the past two decades relative to cost-of-service regulation, but that the welfare loss from higher prices was more than offset by increased firm productivity and investment. While price cap regulation led to welfare gains relative to either cost-of-service or price deregulation, there was still room for improvement. In fact, welfare could have been increased by an additional 2.4% had the adjustment factor been held fixed over the 20 year period. While this would have led to less investment in response to the shale boom, existing customers would have paid considerably lower prices, increasing their consumer surplus. I find that using a structural model is important when assessing the impact of different forms of regulation. Standard methods of estimating firm level productivity would have found a large decline in productivity after the adoption of the price cap, but I find this to be largely due to exogenous changes in firm costs. Instead, the full model predicts that firms actually did significantly decrease their unit-cost of production relative to cost-of-service regulation.

The current analysis uses a theoretical model to determine the impact of different regulatory regimes on investment and welfare. However, direct evidence of the regulatory impact is more difficult to find. One potential area for future research is to compare the experience of oil pipelines to that of natural gas pipelines, which operate under cost-of-service regulation. Despite natural gas production seeing an increase comparable to that of oil production, natural gas pipelines added less than a quarter of the total mileage that oil pipelines did. A similar analysis would need to be taken to determine if the difference in investment was due to higher sunk investment costs, lower transportation demand, or lower expected returns.

This analysis has abstracted away from competitive effects, both between pipelines and other forms of transportation. While I have argued that this is a reasonable approximation, recent papers have documented that this margin may be important in determining pipeline investment.[22] In 2010, rail transportation accounted for less than 1% of all crude

[22] For instance, see Covert and Kellog (2017).

movements but increased to over 7% by 2020.[23]. Future analyses can extend the current framework to account for the impact of competition, where it exists, and substitution to other forms of transportation. The current analysis also abstracts away from the impact of foreign production on domestic supply and transportation demand. The increased domestic production largely served to displace foreign imports so that existing processing capacity only expanded 6% in response to the shale revolution. Future analyses can study the impact of foreign oil supply on pipeline investment and vice versa. Finally, an important caveat to the welfare analysis in this paper is that it fails to account for the environmental impacts of oil pipelines, both in terms of oil spills and more broadly to facilitating the consumption of fossil fuels. These externalities can be important determinants of welfare and estimating their significance would constitute an important extension of the current work.

[23]"Crude Oil and Petroleum Products Transported in the United States by Mode", Department of Transportation

Chapter 2

Identification and Estimation of Discrete Choice Demand Models when Observed and Unobserved Characteristics are Correlated

2.1 Introduction

The identification of discrete choice demand models since Berry, Levinsohn, and Pakes (1995) (BLP) has relied on the assumption that the product characteristic unobserved to the researcher but observed to producers and consumers is conditionally mean independent of all observed product characteristics. Under this identification assumption any function of observed characteristics of all products in the market is a valid instrument for any product's price. Given the abundance of instruments - many of them likely to be very weak - BLP use the structure of their competitive setting to develop product-specific instruments for price that are likely to be highly correlated with that product's price. More recently Gandhi and Houde (2015) show how to extend this logic to develop even more powerful instruments.

Since the inception of its use this assumption has been criticized as being inconsistent with profit-maximizing behavior; it is not clear why firms would choose a level of the unobserved quality for a product independently of the choice of the products' observed characteristics. Empirically we see a high positive correlation among the observed attributes of products, suggesting unobserved product quality is likely to be positively correlated with observed characteristics. If firms do choose to put more unobserved-by-the-researcher quality on products that have more attractive observed characteristics, then instrumenting price with observed product characteristics will not break the positive correlation between price and unobserved quality that BLP are trying to address. Demand elasticities will then continue to be biased in a positive direction because higher prices mean consumers are getting higher unobserved quality, leading consumers to look less price sensitive than they actually are in reality.

In this paper we extend BLP to allow all product characteristics to be endogenous so the unobserved characteristic can be correlated with the other observed characteristics. Spence (1976) formalized the notion that firms' decisions about characteristics' choices are driven by their beliefs about consumer preferences for them and the costs of providing them by showing their first-order conditions for profit-maximization contain terms related to marginal and infra-marginal consumers and costs.[1] We use these first-order conditions for the optimal choice of price and observed and unobserved product characteristics to try to infer firms' beliefs about the distribution of consumers tastes and the structure of costs.

We estimate a model of BLP-type demand and supply under the assumption that firms choose characteristics first given some information set. They do so knowing that once all of the product characteristics and other demand and supply factors have been realized they will compete in prices in a Bertrand–Nash manner. Using the insight from Hansen and Singleton (1982), our identification is based on the assumption that firms' ex-post optimization mistakes are conditionally mean independent of anything the firm knows at the time the firm chooses its product characteristics. This will be true as long as firms do not condition on something that we do not observe that affects their profitability and their characteristics' choices (see Pakes et al. (2015)). An advantage of these setups is that we do not have to completely specify the firm's information set at the time it chooses characteristics; it may

[1] See also the more recent generalization by Veiga and Weyl (2014).

include other firms' product lagged or contemporaneous characteristics and demand/cost shocks, signals on all of these, or no information on them at all.

Our approach is complementary to the many papers previous to ours that have exploited Spence's insight that optimization can help with identification of model parameters, including Mazzeo (2002), Sweeting (2007), Crawford and Shum (2007), Lustig (2008), Gramlich (2009), Fan (2013), Eizenberg (2014), and Blonigen et al. (2013).[2] These papers are more general than our approach in the sense that they consider (e.g.) the use of optimization to help with identification of fixed costs, sunk costs, or identification in the face of restricted sets of characteristics from which firms can choose for product characteristics. However, all of these papers maintain some kind of independence between the level or change in the demand or supply shock and the observed product characteristics. Our identification assumption is straightforward to adopt to all of these settings and would allow researchers to sidestep imposing mean independence of observed and unobserved characteristics while at the same time estimating (e.g.) fixed or sunk costs.

The steps necessary to calculate the value of our objective function are identical to the steps in BLP's two-step GMM estimator except we replace the mean independence moments with our optimization moments. The BLP inversion allows us to – for any given parameter value – solve for the unobserved characteristics for every vehicle so we can treat them as another observed characteristic that the firm is choosing optimally. Using characteristics of competitors vehicles from prior years - which should be known to the firm at the time they made characteristic choices in those prior years - we develop an approximation to the optimal instruments implied by the model's structure. The standard BLP instruments are also valid instruments in our setting and we provide results using these instruments as well. The only other difference with the BLP estimation routine is we include these characteristics in the marginal cost function. Formulating our estimator in the GMM framework means our estimator can easily be supplemented with moments that may further help with identification, as in Petrin (2002) or Berry et al. (2004).

[2] See the review in Crawford (2012) for a complete list of all papers that use optimization in characteristics for identification.

The most striking difference between the BLP estimates and the optimization estimates is that the coefficient on price is much larger under optimization. The impact of this change is that relative to BLP on average elasticities double and estimated markups fall by 50%. We investigate whether a positive correlation between observed and unobserved characteristics is a possible explanation by constructing a "BLP instrumented price", that is, we regress price on the BLP instruments and construct predicted values. We find our unobserved qualities are significantly positively correlated with the instrumented prices with a correlation of approximately 0.5.

A second related difference in model fit relates to the fact that only 10% of U.S. households buy new cars in any given year so both fitted demand models need a way to explain why 90% of households choose the outside good. BLP fits 90% of households choosing the outside good by having consumers view the average unobserved quality of new cars as much worse than the outside good. In contrast, the optimization-fit has consumers strongly desiring new cars relative to the outside good but the significantly higher price elasticity causes 90% not to buy a new car.

Our estimates are almost always much more precisely estimated relative to the BLP-fit model. We also find some of the anomalies in the BLP point estimates are not present in the optimization-fit point estimates. The BLP point estimates imply consumers dislike fuel efficiency but in our setup they strongly and significantly prefer fuel efficiency. The BLP point estimates also imply it cost less to build a bigger and more fuel efficient vehicle while we find the opposite.

The differences we report here between the optimization-fit model and the BLP-fit model have also been found in European automobile data (see Miravete et al. (2015)). They adopt our approach to estimating demand and supply to look at competition in the Spanish automobile market. They report that using the optimization moments on average estimated price elasticities double and estimated markups fall by 50% relative to when they use the BLP moments. Anomalous demand and supply point estimates under the BLP-fit are not present under optimization-fit, the standard errors are much smaller, and their unobserved quality term is positively correlated with observed characteristics.

In Section (2.2), we specify demand and supply system. In Section (2.3) we discuss our identifying restrictions and give conditions under which we are locally identified. Section (2.4) discusses estimation and our choice of instruments. Certain commonly used data generating processes can complicate estimation using our approach, which we discuss in Section (2.5). In Section (2.6), we provide Monte Carlo simulation results for the proposed estimator which demonstrate its properties. Lastly, we apply our approach to the same automobile data BLP used in (2.7).

2.2 Demand and Supply

In this section we provide the demand and supply framework for our estimator.

2.2.1 Demand

Each product is defined as a vector of K observed characteristics and price $(X_j, p_j) \in \mathbb{R}^{K+1}$ and an unobserved (to the econometrician) characteristic ξ_j which is observed by both consumers and producers. Product $j = 0$ is the option of not buying a new vehicle and it is standard to normalize its characteristics and price to zero ($X_0 = p_0 = \xi_0 = 0$).

A consumer i is indexed by $(D_i,\, v_i,\, \varepsilon_i)$, where D_i is a vector of their demographic characteristics, v_i is vector of their K idiosyncratic taste draws $(v_{ik})_{k=1}^{K}$ drawn from a known distribution, one for each of the K characteristics, and ε_i is the vector of their product-specific “tastes" $(\varepsilon_{ij})_{j=1}^{J}$ which are assumed to be independent and identically distributed extreme value across consumers and products. The demand model parameters are given as $\theta^D = (\theta_l, \theta_{nl})$. Utility that consumer i derives from good j is given as

$$u_{ij}\left(\theta^D\right) \quad = \quad \delta(X_j, p_j, \xi_j; \theta^D) + \mu(X_j, p_j, D_i, v_i; \theta^D) + \epsilon_{ij}$$

The term $\delta_j = \delta(X_j, p_j, \xi_j; \theta^D)$ is a product specific utility component and is common to all consumers. The term $\mu(X_j, p_j, D_i, v_i; \theta^D)$ captures the individual specific taste for the characteristics of good j.[3]

[3] It is commonly assumed that these functions are linear in the characteristics so that $\delta_j = X_j\beta - \alpha p_j + \xi_j$, and consumer i's taste for characteristic X_k is then given by $\beta_{ik} = \beta_k + \Pi_k D_i + \Sigma_k v_i$. See Nevo (2004).

Consumer chooses the one and only one product j which yields the highest utility:

$$u_{ij}\left(\theta^{D}\right) \geq u_{ij'}\left(\theta^{D}\right), \quad \forall j'.$$

Define s_{ij} as individual i's probability of purchasing good j prior to the realization of ε and is given by

$$s_{ij}\left(p, X, \xi, D_i, v_i;\ \theta^{D}\right) \quad = \quad \frac{\exp\left(\delta(X_j, p_j, \xi_j; \theta^D) + \mu(X_j, p_j, D_i, v_i; \theta^D)\right)}{\sum_{j' \in J} \exp\left(\delta(X_{j'}, p_{j'}, \xi_{j'}; \theta^D) + \mu(X_{j'}, p_{j'}, D_i, v_i; \theta^D)\right)}.$$

Letting $F(D, v)$ denote the distribution of consumer characteristics the market share for good j is given as the integral over these consumers:

$$s_j\left(p, X, \xi;\ \theta^{D}\right) \quad = \quad \int s_{ij}\left(p, X, \xi, D, v;\ \theta^{D}\right)\ dF(D, v).$$

We turn now to our model of supply.

2.2.2 Supply

Define $Z_j = (X_j, W_j, \omega_j)$ and $Z = (Z_j)_{j \in J}$, where W_j are cost shifters, including X_j itself or the log of it and ω_j is the unobserved cost shock for good j. We follow the literature and specify firm profits as

$$\Pi_f = \sum_{j' \in J_f} \left(p_{j'} - mc_{j'}\left(W_j, \omega_j; \theta\right)\right)\ s_{j'}\left(p, Z, \xi; \theta\right)$$

Here f indexes firms and J_f is the set of products produced by firm f. As in many papers in this literature we assume firms compete with each other every period in two stages. In the first-stage, firms choose their product characteristics to maximize their expected profit. In the second stage, all uncertainty is resolved and firms compete Nash-Bertrand in prices.

Specifically, in the second stage, firms have the following J pricing first-order conditions:

$$s_j + \sum_{j' \in J_f} \left(p_{j'} - mc_{j'}\left(W_j, \omega_j; \theta\right)\right)\ \frac{\partial s_{j'}}{\partial p_j} \quad = \quad 0, \quad \forall j \in J_f. \tag{2.1}$$

where firms are indexed by f and J_f is the set of goods that firm f produces.

In the first stage, firms know that they will compete in prices in a Bertrand-Nash manner given the chosen characteristics of all products in the market and that prices are set according to the first-order conditions (2.1). With this knowledge, firms choose characteristics to maximize expected profits given their information set, denoted I_f for firm f. This information set may differ across firms, and it may include other firms' product characteristics, own- and other-firm cost shifters, some signals on the variables, or no information at all on them.

Redefine K be the number of characteristics including ξ and let $\theta = (\alpha, \beta, \sigma, \gamma)$ include the cost parameters. In the first step, firm f chooses vectors $X_f = (X_j)_{j \in J_f}$ and $\xi_f = (\xi_j)_{j \in J_f}$ to solve:

$$\max_{X_f,\ \xi_f} \qquad E\left[\Pi_f \mid I_f\right]$$

with prices determined after characteristics are set in a Bertrand-Nash manner. Given (Z, ξ), the realized value of the first-order condition for characteristic k of product j is given by $\nu_{jk}(\theta)$ and written as

$$\frac{\partial \Pi_f}{\partial X_{jk}} = \quad \textstyle\sum_{j' \in J_f} \left[\left(p_{j'} - mc_{j'}\left(W_j, \omega_j; \theta\right)\right) \frac{d\ s_{j'}(p,Z,\xi;\theta)}{d\ X_{jk}} + \right. \tag{2.2}$$

$$\left. s_{j'}\left(p, Z, \xi; \theta\right) \ \frac{\partial\left(p_{j'} - mc_{j'}(W_j, \omega_j; \theta)\right)}{\partial X_{jk}} \right] \tag{2.3}$$

for $k < K$. If $k = K$, the above first-order condition is taken with respect to ξ_j. Firms anticipate the change in equilibrium prices that will occur in the second step if they change their product characteristics and this shows up in the first-order condition in the derivative of shares with respect to characteristics X_j (and ξ_j):

$$\frac{d\ s_{j'}\left(p, Z, \xi; \theta\right)}{d\ X_{jk}} \quad = \quad \frac{\partial s_{j'}}{\partial X_{jk}} + \sum_{j'' \in J} \frac{\partial s_{j'}}{\partial p_{j''}} \frac{\partial p_{j''}}{\partial X_{jk}}.$$

The first-order condition illustrates that multi-product firms internalize the externality of changing X_j on the profits of its other products $j' \in J_f$. The term in (2.2) represents the change in profits attributable to *marginal* consumers while the second term in (2.3) captures those attributable to *infra marginal* consumers. Rational expectations requires that

$$E\left[\frac{\partial \Pi_f}{\partial X_{jk}}\bigg|\ I_f\right] = 0,\ \forall k, j \in J_f,\ \forall f \tag{2.4}$$

In equilibrium, the optimal level of X_f chosen by the firm maximizes expected profits given what the firm knows at the time the characteristics are chosen. Sometimes firms will provide too little of a characteristic and sometimes it will provide too much, but on average these "mistakes" average out.

2.3 Identifying Restrictions

We turn now to the issue of identification. Readers interested in the empirical results can skip this section and go directly to Section (2.6). Our identification is based on the K first-order conditions in (2.3) coupled with the J first-order conditions with respect to p in (2.1). Given θ and the data, marginal costs can be recovered from (2.1):

$$mc\left(p, Z, \xi; \theta\right) = p - \Delta^{-1}\left(p, Z, \xi; \theta\right) s$$

where

$$\Delta_{ij} = \begin{cases} -\frac{\partial s_j}{\partial p'_j}, & \text{if } j, j' \in J_f \\ 0 & \text{otherwise} \end{cases}$$

Then the realized value of the firm's first-order conditions, evaluated at θ, are given by

$$\nu_{jk}\left(\theta\right) = \sum_{j' \in J_f}\Bigg[\left(p_{j'} - mc_{j'}\left(p, Z, \xi; \theta\right)\right)\frac{d\ s_{j'}\left(p, Z, \xi; \theta\right)}{d\ X_{jk}} + \\ s_{j'}\left(p, Z, \xi; \theta\right)\ \frac{\partial\left(p_{j'} - mc_{j'}\left(p, Z, \xi; \theta\right)\right)}{\partial X_{jk}}\Bigg]$$

Profit maximization and rational expectations imply the following moment conditions

$$E\left[\nu_{jk}\left(\theta\right)\ \mid I_f\right] = 0 \quad \forall k, j \in J_f,\ \forall f, \quad \text{if } \theta = \theta_0 \tag{2.5}$$

The residual ν_{jk} may include expectational errors that arise due to asymmetric information across competing firms on each others' costs and product characteristics or it may be incomplete information on the outcomes own-firm payoff-relevant variables (like realized cost shocks). ν_{jk} may also include model approximation error or measurement error in the data. As Pakes et al. (2015) note, if there is something known to the firm but not seen by the researcher and if it affects the firm's profits and thus its decisions, the mean of these selected observations will not generally be zero.

We use the insight from Hansen and Singleton (1982) that (2.5) implies that any function of the arguments of I_f are possible instruments that can be used to identify the model parameters. We propose such instruments in the subsequent section. Before doing so, we discuss the local identification of the model.

2.3.0.0.1 Local Identification Let $\nu_k\left(\theta\right)$ be the vector of J residuals $\nu_{jk}\left(\theta\right)$. We will show that it is possible to write this ν_k as an affine function of θ_l:

$$\nu_k\left(\theta\right) = \nu_{kc}(\theta_{nl}, X, W, s) + \nu_{k\theta}(\theta_{nl}, X, W, s)\theta_l = 0, \tag{2.6}$$

for some new functions ν_{kc} and $\nu_{k\theta}$, neither of which have as arguments θ_l.

The linearized residuals help clarify when the model is locally identified. Local identification is achieved when there are sufficient excluded instruments Z that are correlated with the derivative of the residuals with respect to the model parameters, $\frac{\partial \nu_k}{\partial \theta'}$, and $Z' \frac{\partial \nu_k}{\partial \theta'}$ is full rank. From equation (2.6), we have

$$\frac{\partial \nu_k}{\partial \theta_l'} = \nu_{k\theta}(\theta_{nl}, X, W, s)$$

and

$$\frac{\partial \nu_k}{\partial \theta_{nl}'} = \frac{\partial \nu_{kc}(\theta_{nl}, X, W, s)}{\partial \theta_{nl}'} + \frac{\partial \nu_{k\theta}(\theta_{nl}, X, W, s)}{\partial \theta_{nl}'}\theta_l$$

When $\nu_{kc} = 0$, then $E[Z'\nu_k] = 0$ implies $E[Z'\nu_{k\theta}(\theta_{nl}, X, W, s)]\theta_l = 0$. This then implies that the components of $E\left[Z'\frac{\partial \nu_k}{\partial \theta_l'}\right]$ are collinear and that the rank condition is violated. Below, we briefly discuss a class of DGPs where the model parameters cannot be consistently estimated from the FOCs alone, as they violate this rank condition. This class of models includes demand systems derived from generalized extreme value idiosyncratic preferences and therefore will be of interest to many practitioners.

In the case of a single-characteristic single-product monopolist with utility given by $u_i = X(\beta + \sigma v_i) - \alpha p + \xi$ and marginal costs given by $\ln(\text{mc}_j) = X\gamma + \omega$, we can easily derive expressions for ν_{kc} and $\nu_{k\theta}$. The residual is given by

$$\begin{aligned}\nu_j &= \frac{\beta}{\alpha}s_j + \frac{\sum_i \sigma v_i s_{ji}(1 - s_{ji})}{\sum_i \alpha s_{ji}(1 - s_{ji})}s_j - \frac{\text{mc}_j s_j}{X_j}\gamma \\ &= \underbrace{\frac{\sum_i \sigma v_i s_{ji}(1 - s_{ji})}{\sum_i \alpha s_{ji}(1 - s_{ji})}s_j}_{\nu_c} + \underbrace{\left[\frac{s_j}{\alpha}, -\frac{\text{mc}_j s_j}{X_j}\right]'}_{\nu_\theta}[\beta, \gamma]\end{aligned}$$

where s_{ji} are choice probabilities for consumer i. It is clear that (a properly scaled) ν_c will be non-zero at the true parameter values, and so the model will be locally identified. A similar expression can be derived for the more general case of multi-product oligopolists with an arbitrary number of product characteristics. Because the functions are more involved, we relegate them to Appendix Section (B.3).

To show that (2.5) holds, we start by noting that the J residuals for any characteristic k can be written in matrix notation as

$$\nu_k = \left(\frac{\partial(p-\mathrm{mc})}{\partial X_k'} \circ T\right) s + \left(\frac{ds}{dX_k'} \circ T\right)(p-\mathrm{mc}),\ k=1,...,K, \tag{2.7}$$

where $\circ$ denotes element-wise multiplication and T, the ownership matrix, is a block diagonal matrix that identifies products owned by the same firm, so if $i \in J_f$ and $j \in J_f$ for some f, then $t_{ij}=1$; otherwise, $t_{ij}=0$. Conditional on a given value of θ_{nl}, several of the terms in ν_k are held constant, including $s(\delta(\theta_{nl}))$ and $(p-\mathrm{mc})$. $\frac{\partial s_i}{\partial p_j'} = \frac{\alpha}{y_i} s_{ij}(1-s_{ij})$, and with s_{ij} also not a function θ_l, conditional on θ_{nl}, implying the matrix $\frac{\partial s}{\partial p'}$ is also constant. Since $\frac{ds}{dX_k} = \frac{\partial s}{\partial X_k'} + \frac{\partial s}{\partial p'}\frac{\partial p}{\partial X_k'}$, (2.7) is comprised of three terms that vary with θ_l: $\frac{\partial \mathrm{mc}}{\partial X_k'}$, $\frac{\partial s}{\partial X_k'}$ and $\frac{\partial p}{\partial X_{jk}'}$. None of these terms multiply one another, so if each is established to be affine in θ_l then the entire expression will be as well. Given our specification of utility and marginal cost, it is clear that the first two terms are affine in θ_l, so ν_k will have the required form if $\frac{\partial p}{\partial X_{jk}'}$ is affine as well, which Proposition 2 proves below.

We begin with Proposition 1, which establishes that equation (2.7) may be written as (2.6) if the matrix $\frac{\partial p}{\partial X_{jk}'}$ can be written as an affine function of θ_l.

Proposition 1. *Assume that mean utility is linear in β and the derivative of marginal costs with respect to X_{jk} is affine in γ_k, conditional on observables. If $\frac{\partial p}{\partial X_{jk}'}$ can be written as an affine function of θ_l, conditional on observed market shares and the nonlinear parameters, then ν_k can be written as an additive function of two non-linear functions of (θ_{nl}, X, s, p), one of which is linear in θ_l.*

2.3.0.0.2 Proof See appendix.

Turning to Proposition 2, we now establish that $\frac{\partial p}{\partial X_{jk}'}$ is affine in θ_l. By the Implicit Function Theorem, the derivative of prices with respect to characteristics has the following form

$$\frac{\partial p}{\partial X_{jk}'} = -\left(\frac{\partial R}{\partial p'}\right)^{-1} \frac{\partial R}{\partial X_{jk}} \tag{2.8}$$

where R is a vector of residual equations defined by the pricing FOCs. Proposition 2 shows that $\frac{\partial R}{\partial p}$ does not depend on θ_l and that $\frac{\partial R}{\partial X_{jk}}$ is affine in θ_l, thereby establishing that $\frac{\partial p}{\partial X'_{jk}}$ is affine in θ_l.

Proposition 2. *Assume that mean utility is linear in β and the derivative of marginal costs with respect to X_{jk} is affine in γ_k, conditional on observables. Then the derivative of price with respect to each product characteristic can be written as the sum of two nonlinear functions of θ_{nl}, one of which does not depend on θ_l and the other that is a linear function of θ_l.*

2.3.0.0.3 Proof See appendix.

We not turn now to the estimator.

2.4 Estimation

There are three sections to this estimation section. Section 4.1 describes how we implement the GMM objective function. Section 4.2 then describes our approximation to the optimal instruments. Section 4.3 show hows to concentrate out all of the "linear" parameters during estimation to reduce the dimensionality of the nonlinear search. Readers only interested in these last details can skip directly there.

2.4.1 The estimator

Estimation follows two-stage GMM. For an initial guess at θ_0 we calculate the approximation to the optimal instruments described in the following section. Given those instruments we calculate the optimal weighting matrix and the first stage estimates.[4] At the first stage estimates we recalculate the optimal instruments and the efficient weighting matrix and the re-estimate to get the two-step GMM estimates.

In each stage, estimation has the following steps. Let $\nu(\hat{\theta})$ be a vector created by eval-

[4] As in BLP we use importance sampling to minimize simulation error. We draw importance samples at an initial estimate θ_1, and then evaluate instruments H and optimal weighting matrix Ω at θ_1 for GMM estimation. Once the first step estimates are converged at θ_2, we re-draw importance samples, re-derive instruments, and re-evaluate optimal weighting matrix at θ_2. Then, we repeat the search over θ.

uating $\nu_{jk}(\theta)$ at a guess $\hat{\theta}$. Given a set of instruments Z and a weight matrix W, the empirical moment conditions may be written as[5]

$$g(\theta) = \frac{1}{N} Z' \nu(\hat{\theta})$$

and the GMM objective function is given by

$$GMM(\theta) = g(\theta)' Z W Z' g(\theta)$$

Each stage of estimation then reduces to the following steps:

1. Fix θ_{nl} equal to initial guess θ_{nl}^1
2. Solve for $\delta(\theta_{nl}^1)$ that matches shares
3. Use the first-order conditions for prices to recover marginal costs.
4. Evaluate the GMM objective function.
5. Repeat from Step 1 until the objective function is minimized.

There is a pathological case that needs to be dealt with, where $\alpha \to \infty$ and $\gamma \to 0$ can set the moment conditions exactly equal to zero. We view this case as un-economic as it implies that costs are independent of the characteristics and consumers are infinitely price sensitive. There are several potential solutions to eliminating this case. First, the researcher may re-write the moments conditions to rule out this case. Given our model specification, each term in ν_{kc} is multiplied by $\frac{\sigma_k}{\alpha}$. Pulling this common term out, we can write ν_{kc} as $\nu_{kc}(\theta_{nl}, X, W, s) = \frac{\sigma_k}{\alpha} \tilde{\nu}_{kc}(\theta_{nl}, X, W, s)$, for some function $\tilde{\nu}_{kc}(\theta_{nl}, X, W, s)$. Importantly, the function $\tilde{\nu}_{kc}(\theta_{nl}, X, W, s)$ is always non-zero and varies non-linearly with θ_{nl}. Also,

$$\begin{aligned}\frac{\alpha}{\sigma_k}\nu_k &= \frac{\alpha}{\sigma_k}\left(\nu_{kc}(\theta_{nl}, X, W, s) + \nu_{k\theta}(\theta_{nl}, X, W, s)\tilde{\theta}_l\right) \\ &= \tilde{\nu}_{kc}(\theta_{nl}, X, W, s) + \nu_{k\theta}(\theta_{nl}, X, W, s)\tilde{\theta}_l\end{aligned}$$

[5] We suppress the dependence of the moment conditions on the data for notational simplicity.

is zero whenever ν_k is zero, but not when $\alpha \to \infty$. It is clear that we can recover the true θ_l given estimates of $\tilde{\theta}_l$ and θ_{nl}. [6]

Alternatively, the researcher may "compactify" the parameter space, choosing an upper bound on the absolute value of the parameters such that the true parameter values lie in the compact set. In practice, the researcher would set an upper-bound on the α (say 300) and search over local minima that satisfy the bound. This is a similar procedure as is proposed in Newey and McFadden (1994) to deal with a likelihood function that becomes unbounded as the variance parameter approaches 0 (see page 2,136). Assuming that only the true θ_0 satisfies in the first-order conditions in the compact set would then lead to consistency of our proposed estimator.

2.4.2 Instruments

We have more than K unknown parameters but only K first-order condition conditions. As such, we need instruments that are orthogonal to ν_{jk} in order to estimate the model parameters. Chamberlain (1987) shows that the efficient set of instruments are the expected value of the derivatives of the error term with respect to the parameters evaluated at the true parameter θ_0.

In our context this optimal instrument H is a $JK \times |\theta|$ matrix

$$H = E\left[\nu\left(\theta_0\right)\nu\left(\theta_0\right)' \mid I\right]^{-1} E\left[\frac{\partial\nu\left(\theta_0\right)'}{\partial\theta} \,\middle|\, I\right]'. \tag{2.9}$$

[6] We note the similarity between this problem and estimating Euler Equations with CES utility. The Euler Equation is given by

$$E\left[R_{t+1}\beta_0(1-\alpha_0)c_{t+1}^{-\alpha_0}|I_t\right] = (1-\alpha_0)c_t^{-\alpha_0}$$

which can be set exactly to zero by taking $\alpha_0 \to \infty$ or setting $\alpha_0 = 1$. However, transforming the equation to an equivalent form

$$E\left[R_{t+1}\beta_0\left(\frac{c_{t+1}}{c_t}\right)^{-\alpha_0}\middle| I_t\right] - 1 = 0$$

rules out these pathological cases during estimation.

Letting $E\left[\nu(\theta_0)\nu(\theta_0)' \mid I\right] = I_{JK}$ for now, H_{jkl}, the element $(jk,\, l)$ of the derivative, is given as

$$H_{jkl} = E\left[\frac{\partial \nu_{jk}(\theta_0)}{\partial \theta_l} \;\middle|\; I_f\right] \quad \forall k, l, j \in J_f,\ \forall f. \tag{2.10}$$

where the derivative of ν_{jk} with respect to θ_l is given as:

$$\begin{aligned}
\frac{\partial \nu_{jk}(\theta_0)}{\partial \theta_l} &= \sum_{j' \in J_f} \left[\frac{\partial\left(p_{j'} - mc_{j'}\right)}{\partial \theta_l} \frac{d\ s_{j'}}{d\ X_{jk}} + \left(p_{j'} - mc_{j'}\right) \frac{d^2\ s_{j'}}{d\ \theta_l\ d\ X_{jk}} \right. \\
&\qquad \left. + \frac{d\ s_{j'}}{d\ \theta_l} \frac{\partial\left(p_{j'} - mc_{j'}\right)}{\partial X_{jk}} + s_{j'} \frac{\partial^2\ \left(p_{j'} - mc_{j'}\right)}{\partial \theta_l\ \partial X_{jk}} \right] \\
\text{where} \quad \frac{d\ s_j}{d\ \theta_l} &= \frac{\partial s_j}{\partial \theta_l} + \sum_{j' \in J} \frac{\partial s_j}{\partial X_{j'}} \frac{\partial X_{j'}}{\partial \theta_l} + \sum_{j' \in J} \frac{\partial s_j}{\partial p_{j'}} \frac{\partial p_{j'}}{\partial \theta_l} + \sum_{j' \in J} \sum_{j'' \in J} \frac{\partial s_j}{\partial p_{j'}} \frac{\partial p_{j'}}{\partial X_{j''}} \frac{\partial X_{j''}}{\partial \theta_l}
\end{aligned}$$

for $k < K$. If $k = K$, $d\ X$ or ∂X is substituted to $d\ \xi$ or $\partial \xi$. We follow Fan (2013) and recover $\frac{\partial p_{j''}}{\partial X_{jk}}$ using the Implicit Function Theorem. In principle we are exactly identified, i.e. the total number of instruments is equal to the total number of model parameters. These instruments place larger weights on the first-order conditions which are most responsive to changes in the parameters contained in θ.

There are four significant challenges to calculating the optimal instruments. We do not know the true value of parameters θ_0 and we do not know the information set I_f of any firm. Even if we knew I_f we would have to specify the distribution of the remaining unknown random variables conditional on the information set to be able to integrate over it. Finally, $\frac{\partial X}{\partial \theta}$ and $\frac{\partial p}{\partial \theta}$ are complicated unknown equilibrium objects.

We follow Berry et al. (1999) and choose an informed guess θ_g and then approximate the optimal instrument H_{jkl} by using the value of the derivative itself $\frac{\partial \nu_{jk}}{\partial \theta_l}$ calculated under different assumptions about what is known to the firm at time when the characteristics' decisions are made. We set the terms $\frac{\partial X}{\partial \theta}$ and $\frac{\partial p}{\partial \theta}$ to zero because of the difficulties of estimating them so $\frac{d\ s_j}{d\ \theta_l} = \frac{\partial s_j}{\partial \theta_l}$ in our estimation routine.[7] Let $X_t = \left(X_{jt}\right)_{j \in J_t}$ be a vector of characteristics of all products available in year t, and define ξ_t, W_t, p_t, and ω_t similarly. Let

[7] Leaving out these terms as well as letting $E\left[\nu(\theta_0)\nu(\theta_0)' \mid I\right] = I_{JK}$ is not a consistency issue but instead an efficiency issue. In our monte carlos it is possible to calculate these terms we compare the monte carlo results with and without them to check on the importance of these terms for precision.

$X_{f,t} = (X_{jt})_{j \in J_{ft}}$ be the set of firm f's products, and $X_{-f,t} = (X_{jt})_{j \notin J_{ft}}$ be the set of firm f's competitors' products.

In the benchmark setup we assume firms' information sets contain their own contemporaneous costs shocks and their competitors' characteristics from the previous year:

$$\{X_{-f,\ t-1},\ \xi_{-f,\ t-1},\ W_{-f,\ t-1},\ \omega_{-f,\ t-1},\ \omega_{f,\ t}\} \quad \subset \quad I_{f,t}^{\text{lagged}}$$

When calculating the derivative for a product characteristic for firm f we use observed and unobserved characteristics of products of the firms competing against f from the previous year. At those characteristics and firm f's current observed and unobserved characteristics we solve for the Bertrand-Nash equilibrium prices. Then, we evaluate the derivative at those prices for firm f.

Given the lagged information set, lagged version of BLP-instruments are valid instruments for ν_{jk} – own product characteristics $X_{j,t}$, other product characteristics within own firm $X_{j' \neq j,f,t}$, and competitor's product characteristics $X_{-f,\ t-1}$ from the previous year. Moreover, BLP type instruments can be applied to price – own price, other prices within own firm, and competitor's prices from the previous year are valid instruments in our setup.

In the second case we assume

$$\{X_{-f,\ t},\ \xi_{-f,\ t},\ p_{-f,\ t},\ W_{-f,\ t},\ \omega_t\} \quad \subset \quad I_{f,t}^{\text{contemporaneous}}$$

so firm f knows its competitors' contemporaneous choices of characteristics and costs at the time of decision. This information set implies that the conditional expectation of the FOCs are taken with respect to approximation or measurement error. In this case, the derivative is evaluated at realized values of X_t, ξ_t, W_t, p_t, and ω_t. In addition, given the contemporaneous information set, BLP-type instruments evaluated at the realized values are valid instruments for ν_{jk}.

2.4.3 Concentrating Out Linear Parameters

BLP reduce the dimensionality of their parameter search by "concentrating out" parameters $\theta_l = (\beta, \gamma)$, which enter their moment conditions linearly. An implication of (2.6) is that

θ_l enters the moment conditions linearly, conditional on θ_{nl} and the data. As such, we can concentrate out θ_l, thereby allowing us to restrict our search to the space of the non-linear parameters, denoted $\theta_{nl} = (\alpha, \sigma)$. This reduces the dimensionality of the nonlinear search by $\#\beta \times \#\gamma$ parameters and leads to a more robust estimation algorithm. For each guess of θ_{nl}, the value of θ_l that minimizes the GMM objective function for a given guess of θ_{nl} is given by

$$\theta_l(\theta_{nl}, X, W, s) = -\left(\nu_\theta' Z W^{-1} Z' \nu_\theta\right)^{-1} \nu_\theta' Z W^{-1} Z' \nu_c$$

where ν_c comes from stacking the K ν_{kc} functions and ν_θ comes from stacking the $\nu_{k\theta}$ functions. In general, the terms ν_{kc} and $\nu_{k\theta}$ will have complex functional forms. We provide expressions for ν_{kc} and $\nu_{k\theta}$ in Appendix Section (B.3) given our specification of marginal costs and demand.

2.5 Identification Challenges with Logit and Nested-Logit Demand

In this section, we show that local identification can fail when idiosyncratic shocks are characterized by logit or nested-logit demand. This result requires a specific functional form for marginal costs, namely log-linearity in the unobserved characteristic, and so is not a general feature of the estimation routine. However, due to the common usage of GEV demand and log-linear marginal costs in the literature it is an important case to note. We demonstrate the lack of identification for the case of a multi-product monopolist under two specifications for demand: logit and nested-logit. We derive expressions for the FOCs and show how this linearity hampers identification. The logic of this section extends directly to multiple firms and GEV demand.

2.5.1 Multi-Product Monopolist

Consider a monopolist that produces J products. Market shares are a generic function of the vector of observed characteristics X, price p, and unobserved characteristics ξ, i.e. $s_j = s_j(X, p, \xi)$. For simplicity, assume that each product has a single observed characteristic. In equilibrium, markups must satisfy the Nash-Bertrand first-order conditions $(p - \text{mc}) =$

$-\left(\frac{\partial s}{\partial p'}\right)^{-1} s$. Plugging the expression for the market share total derivative and the mark-up into the characteristic FOCs gives

$$\begin{aligned}
\nu &= E\left[-\left(\frac{\partial s'}{\partial X} + \frac{\partial p'}{\partial X}\frac{\partial s'}{\partial p}\right)\left(\frac{\partial s}{\partial p'}\right)^{-1} s + \frac{\partial p'}{\partial X}s - \frac{\partial \mathrm{mc}'}{\partial X}s \middle| I_m\right] \\
&= E\left[-\frac{\partial s'}{\partial X}\left(\frac{\partial s}{\partial p'}\right)^{-1} s - \frac{\partial \mathrm{mc}'}{\partial X}s \middle| I_m\right]
\end{aligned}$$

where I_m denotes the firm's information set in market m. Note that the monopolist does not need to anticipate the change in prices due to a change in product characteristics, greatly simplifying the analysis. Using this expression, we show how identification fails given

2.5.1.0.1 Logit Demand Assuming logit demand, we have the following market share derivatives

$$\frac{\partial s}{\partial p} = -\alpha\left(s \circ I - ss'\right)$$

$$\frac{\partial s}{\partial X} = \beta\left(s \circ I - ss'\right)$$

Plugging these terms into the FOCs gives the following expression for ν_j, where j indexes the elements of ν,

$$\nu_j = \left(\frac{\beta}{\alpha} - \frac{\mathrm{mc}_j}{X_j}\gamma\right)s_j$$

It is clear that $\beta = 0$ and $\gamma = 0$ sets the FOCs exactly equal to zero, so the model is not globally identified. The model is not locally identified either as multiplying β and γ by a common term leaves the FOCs unchanged. We can see that these moments violate the rank condition, as the moment condition $E[Z'\nu_j]$ implies

$$E\left[Z'\frac{\iota}{\alpha}\right] = E\left[Z'\frac{\mathrm{mc}_j}{X_j}\right]\frac{\gamma}{\beta}$$

which means that $E[Z'\frac{\partial \nu_j}{\partial \beta}]$ is an exact linear combination of $E[Z'\frac{\partial \nu_j}{\partial \gamma}]$.

Contrast this to the case of a monopolist discussed in Section (2.3). The FOC for product j is equivalent to

$$\nu_j = \left(\frac{\beta}{\sigma} + \frac{\sum_i \eta_i s_{ji}(1 - s_{ji})}{\sum_i s_{ji}(1 - s_{ji})} - \frac{\mathrm{mc}_j}{X_j} \frac{\gamma}{\alpha} \sigma \right) s_j$$

Note that both σ and α enter the model non-linearly through s_{ij} and mc_j. Now, any rescaling of β and γ around the true parameters will violate the moment conditions.

The lack of identification given logit demand is not unique within the class of GEV demand systems. We turn next to nested logit demand and demonstrate that the model fails to be identified following a similar argument.

2.5.1.0.2 Nested Logit The market share of good j is given by

$$s_j = s_{j|g} s_g$$

where $s_{j|g}$ is product js share in product group g and s_g is the market share of group g. That is, the share of good j is equal to the conditional market share of j in group g, multiplied by the group share g. The derivatives of market shares with respect to characteristics of product j are

$$\frac{\partial s_k}{\partial p_j} = \begin{cases} -\frac{\alpha}{1-\sigma} s_j + \frac{\alpha}{1-\sigma} s_j s_{j|g} - \alpha(1 - s_g) s_{j|g} s_j & \text{if } j = k \\ \frac{\alpha}{1-\sigma} s_{j|g} s_k - \alpha(1 - s_g) s_{j|g} s_k, & \text{if } j, k \in g \\ \alpha s_g s_{j|g} s_k & \text{otherwise} \end{cases}$$

$$\frac{\partial s_k}{\partial X_j} = \begin{cases} \frac{\beta}{1-\sigma} s_j - \frac{\beta}{1-\sigma} s_j s_{j|g} + \beta(1-s_g) s_{j|g} s_j & \text{if } j = k \\ -\frac{\beta}{1-\sigma} s_{j|g} s_k + \beta(1-s_g) s_{j|g} s_k, & \text{if } j, k \in g \\ -\beta s_g s_{j|g} s_k & \text{otherwise} \end{cases}$$

Note that the derivative of market shares with respect to X_j is linear in β, conditional on observed market shares. Therefore, we can write $\frac{\partial s}{\partial X} = \beta H(s, \sigma)$, where H is a matrix that depends only on observed shares and the nonlinear parameter σ. The FOCs are given by

$$\nu_j = \left(-\beta H_j(s, \sigma) \left(\frac{\partial s}{\partial p'} \right)^{-1} s - \gamma \frac{\mathrm{mc}_j}{X_j} s_j \right)$$

As with logit demand, we cannot separately identify β and γ. These results suggest that additional structure is necessary to estimate the model parameters.

2.5.2 Additional Restrictions

When idiosyncratic preferences are assumed to follow a generalized extreme value distribution, identification will either require augmenting the first-order conditions with additional moments or using a different functional form where the effects of β_k and γ_k can be separated.

2.5.2.0.1 Marginal Costs Moments One potential solution is to augment the FOCs with moments from the marginal cost equation. The marginal cost equation can be used to identify the cost side parameters and the FOCs can be used to identify the ratio of β and γ, from which we can recover the demand side parameters. Another potential solution is to make marginal cost nonlinear in ξ_j. If (2.13) held for a subset of product characteristics, or if the researcher had access to instruments Z_j such that

$$E[\omega_j \left(\theta_0^D, \gamma_0 \right) | Z_j] = 0, \ \forall j$$

then these moments could be included to improve efficiency. However, if these conditions are incorrectly specified then including them will lead to inconsistent results. Because the model is identified from the FOCs alone, these moments are a testable implication of the model. The researcher could then include the marginal cost moments for efficiency gains if the orthogonality condition cannot be rejected using just the first-order condition moments.

2.5.2.0.2 Marginal Costs with Nonlinear ξ A key feature of the non-identification result is that β and γ enter the moments linearly, conditional on observed market shares. This is in part due to the fact that ξ_j does not appear directly in the first-order condition moments. Consider an alternative case where ξ_j enters log-marginal cost nonlinearly

$$\ln(\mathrm{mc}_j) = \ln(X_j)'\gamma_x + \gamma_\xi \ln(\xi_j) + \omega_j$$

For a single-product monopolist facing logit demand, the first-order conditions are given by

$$\nu_j(s_j, X_j; \theta) = \begin{bmatrix} 0 & \frac{1}{\alpha} s_j & -\frac{\mathrm{mc}_j}{X_j} s_j & 0 \\ s_j & 0 & 0 & -\alpha \cdot \frac{\mathrm{mc}_j}{\xi_j} s_j \end{bmatrix} [1, \beta, \gamma_x, \gamma_\xi]'$$

which would be sufficient to separately identify the parameters β and γ. The parameters β enter the first-order conditions directly, as well as through their impact on $\xi = \delta - X\beta$. Therefore, the mean utility parameters no longer enter the moments linearly. However, care must be taken when using a marginal cost function that is nonlinear in ξ_j. The above specification rules out cases where $\xi_j < 0$, and would therefore place a material restriction on the estimated results. An alternative specification would be to use

$$\ln(\mathrm{mc}_j) = \ln(X_j)'\gamma_x + \gamma_\xi \operatorname{arcsinh}(\xi_j) + \omega_j$$

where $\operatorname{arcsinh}(x) = \ln\left(x + \sqrt{x^2 + 1}\right)$. This function has approximately the same curvature as the natural logarithm for values of x greater than 1, but has the added benefit that it is defined for negative numbers as well.

2.6 Monte Carlo Simulation

We investigate the properties of our estimator with a simple Monte Carlo that closely resembles our empirical analysis. Readers not interested in these details can skip directly to the BLP application. We consider single-product oligopolists who choose a single, observed characteristic and an unobserved quality level to maximize expected profits. Production is subject to a marginal cost shock, the distribution of which is known to all firms in the industry when they choose their optimal bundle of characteristics. They also knows the profit maximizing equilibrium price for any given demand-cost-characteristic tuple, so they can calculate expected profits for any chosen level of product characteristic. The specifics follow.

2.6.1 Demand

Demand takes a random coefficients, discrete-choice functional form. Let β_{i0} denote the base-level of utility consumer i derives from purchasing the good. β_i is the taste for the single good characteristic X. Both ε_i and ε_{i0} are distributed i.i.d. extreme value. Consumer i purchases the good if u_i is greater than or equal to $u_{i0} = \varepsilon_{i0}$ where

$$\begin{aligned} u_{ij} &= X_j\beta_i - \alpha_i p_j + \xi_j + \varepsilon_{ji} \\ \alpha_i &= \frac{\alpha}{y_i} \\ \ln(y_i) &\sim N(\mu_y, \sigma_y), \\ \beta_i &\sim N(\beta, \sigma_X) \end{aligned}$$

with β defined as the mean taste for X and σ_X characterizing the heterogeneity in taste, both of which need to be estimates. Following the empirical set-up of BLP, the price elasticity depends on the distribution of income, assumed to follow a known log-normal distribution with mean μ_y and standard deviation σ_y. These parameters are known to the researcher. The coefficient α, which governs the degree of price sensitivity, is unknown and needs to be estimated.

We write the demand parameters together as $\theta^D = (\alpha, \beta, \sigma_X)$. Individual choice probabilities have the standard logit form

$$s_{ij}\left(p, X, \xi; \theta^{D}\right) = \frac{\exp\left(X_j\beta_i - \alpha_i p_j + \xi_j\right)}{1 + \sum_m \exp\left(X_m\beta_i - \alpha_i p_m + \xi_m\right)}$$

and market shares come from integrating over the distribution of consumers $G(i)$,

$$s_j\left(p, X, \xi; \theta^{D}\right) = \int_i s_{ij}\left(p_j, X_j, \xi_j; \theta^{D}\right) dG(i).$$

As is standard in the literature, the utility from the outside good is normalized to 0.

2.6.2 Supply

Firms play a two-stage game. In the first stage, they optimally choose their level of X and ξ to maximize expected profits, where the expectation is taken over a known distribution of cost shocks. In the second stage Marginal cost is given by the following quadratic in the log characteristic and the unobserved quality

$$\ln\left(\mathrm{mc}_j\right) = \gamma_0 + \gamma_x \ln\left(X_j\right) + \gamma_{x^2} \ln\left(X_j\right)^2 + \gamma_\xi \xi_j + \gamma_{\xi^2}\xi_j^2 + \gamma_{x\xi} \ln\left(X_j\right)\xi + \omega_j$$

with cost shock ω_j.[8] The cost shock ω_j is split into two parts: ω_{1j} which is known in stage 1 and ω_{2j} which is realized in stage 2. All parameters together are denoted $\theta = (\alpha, \beta, \sigma_X, \gamma)$.

Let $Z = (X, \xi, \omega)$. Profits for firm j are given as

$$\Pi\left(Z; \theta\right) \quad = \quad \left(p_j - mc\left(X_j, \omega_j;\ \theta\right)\right)\ s_j\left(p, X, \xi;\ \theta\right)\ .$$

The timing is as follows. The vector of cost shocks ω_1 is realized. The oligopolist knows the demand parameters and the distribution of the stage 2 cost shocks $F\left(\omega_2\right)$ but she does not see the realized shock ω_2 before the characteristic choice is made. She solves for X

$$\begin{aligned}\max_{X_j, \xi_j} \quad & E\left[\Pi\left(Z; \theta\right) \middle| \omega_1, F(\omega_2)\right] \\ = \quad & \int \left(p_j - mc\left(X_j, \xi_j, \omega_j;\ \theta\right)\right)\ s_j\left(p, X, \xi;\ \theta\right)\ dF\left(\omega_2\right),\end{aligned}$$

[8] When characteristics are optimally chosen, a linear index in demand with linear marginal costs will lead to either a corner solution or a continuum of solutions. We use a quadratic form in X and ξ for log-marginal cost to ensure a unique, interior solution.

knowing p_j will be set to maximize profits once the characteristic is set.

Let $\hat{Z} = (X^*, \xi^*, \omega)$ with X_j^* and ξ_j^* are the optimal amounts of X_j and ξ_j chosen before ω_2 is realized. For ease of notation, let $\tilde{X}_j = (X_j, \xi_j)$ and $\Delta\left(\tilde{X}^*, \omega; \theta\right) = p_j\left(\tilde{X}^*, \omega; \theta\right) - mc\left(\tilde{X}_j^*, \omega;\ \theta\right)$. Then X_j^* satisfies

$$
\begin{aligned}
E\left[\left.\frac{\partial \Pi\left(\hat{Z};\theta\right)}{\partial X_j}\right| \omega_1, F(\omega_2)\right] &= \int \left(\Delta\left(\tilde{X}^*, \omega; \theta\right) \frac{d\ s_j\left(p_j\left(\tilde{X}^*, \omega; \theta\right), \tilde{X}_j^*;\ \theta\right)}{d\ X_j}\right) dF\left(\omega_2\right) + \\
&\quad \int \left(\frac{\partial \Delta\left(\tilde{X}^*, \omega; \theta\right)}{\partial X_j} s_j\left(p_j\left(\tilde{X}^*, \omega; \theta\right), \tilde{X}_j^*;\ \theta\right)\right) dF\left(\omega_2\right) \\
&= 0
\end{aligned}
$$

where $p_j\left(\tilde{X}^*, \omega; \theta\right)$ is the optimal price given $\tilde{X}^*$ and ω, maximizing the expected profits where expectation is taken over $F\left(\omega_2\right)$. The firms also solve an analogous equation for the unobserved characteristic ξ_j. Letting $Z^* = (X^*, \xi^*, \omega^*)$, where ω^* corresponds to the realized cost shocks, the value of the derivative is given as

$$
\begin{aligned}
\nu_j(Z^*;\theta) &= \frac{\partial \Pi\left(Z^*;\theta\right)}{\partial X} \\
&= \Delta\left(\tilde{X}^*, \omega; \theta\right) \frac{d\ s_j\left(p_j\left(\tilde{X}^*, \omega^*; \theta\right), \tilde{X}^*;\ \theta\right)}{d\ X_j} + \\
&\quad \frac{\partial \Delta\left(\tilde{X}^*, \omega; \theta\right)}{\partial X_j} s_j\left(p_j\left(\tilde{X}^*, \omega^*; \theta\right), \tilde{X}^*;\ \theta\right).
\end{aligned} \tag{2.11}
$$

This derivative will sometimes be positive and sometimes be negative depending upon whether “too much" or “too little" of X_j and ξ_j was chosen prior to the realized demand shock. By the way the data are constructed on average these “mistakes" will average out:

$$E\left[\nu_j(Z^*;\theta) \middle| \omega_1, F(\omega_2)\right] = 0,$$

and this moment, along with the marginal cost equation, is our source of identification.

When evaluating $\nu_j(Z^*;\theta)$ in Equation (2.11), we mimic the standard empirical settings where econometricians cannot observe ω_2 and ξ. At each θ we invert demand to find the realized demand shock $\xi\left(X^*, p^*;\ \theta^D\right)$. Then, we apply the FOC with respect to price in (2.1) to find $mc\left(p^*, s\left(p^*, X^*, \xi\left(X^*, p^*;\ \theta^D\right)\right);\ \theta^D\right)$. Then, $\nu_j(Z^*;\theta)$ in Equation (2.11) is evaluated at the realized price, characteristic, recovered demand and cost shocks and marginal cost at each θ.

The optimal instruments are given by

$$H = E\left[\frac{\partial \nu\left(Z^{*};\theta\right)}{\partial \theta}, \frac{\partial \omega_{2}\left(Z^{*};\theta\right)}{\partial \theta}\Big| \omega_{1}, F(\omega_{2})\right] T(Z^{*})$$

where $T(Z^{*})$ serves to normalize the error matrix and is equal to the expected inverse second-moment matrix of the residuals. As we are interested in the properties of our estimator in standard empirical settings where the researcher could not compute the optimal instruments we mimic our proposed empirical approach to approximating the optimal instruments by evaluating the derivative at the realized values Z^{*}. The derivative is approximated at some initial value $\hat{\theta}$, which we take to be a random draw centered at the true parameter value.

Note that the second stage objects are correlated with the random shocks to marginal cost and therefore cannot be used when constructing the optimal instruments. This will in general be true when violations of the first-order conditions are due to optimization error, rather than approximation error. We follow suggestions in Berry et al. (1999) and Gandhi and Houde (2015) and approximate these second stage objects using variation from the first-stage variables. Specifically, we non-parametrically regress prices on a flexible functional form in X_j and X_{-j} to approximate the pricing function. Let P_x be a potentially high-order polynomial in X_j and X_{-j}. We estimate $\hat{p}$ as

$$\hat{p} = \exp(P_x(P_x' P_x)^{-1} P_x' \ln(p))$$

and approximate the market shares as

$$\hat{s}_j = \frac{1}{N}\sum_i \frac{\exp(X_j\hat{\beta}_i + \hat{\xi}_j - \hat{\alpha}_i\hat{p}_j)}{1 + \sum_k \exp(X_k\hat{\beta}_i + \hat{\xi}_k - \hat{\alpha}_i\hat{p}_k)}$$

where $\hat{\xi}_j$ is a consistent estimate of ξ_j using the parameter estimates. The approximate optimal instruments can then be constructed using $\hat{p}$ and $\hat{s}$ in place of their counterparts.

The derivative of the residual is then evaluated at the observed level of X and the approximate values of $\hat{\xi}$, $\hat{p}$, and $\hat{s}$.

The moment condition we use is given by

$$\begin{aligned} G_l\left(Z^*;\theta\right) &\equiv E\left[\hat{H}_l \cdot \left[\nu(Z^*;\theta), \omega_2(Z^*;\theta)\right]\right] \quad \forall l \\ &= 0 \end{aligned}$$

By applying two step GMM with moment condition $G_l\left(\theta\right)$, we estimated the parameters.[9]

We simulate $M = 1,500$ markets with two firms for each of $N = 100$ times. Table 2.1 shows summary statistics for our data generating process, which was calibrated to closely resemble the automobile data used in in Berry et al. (1995), and that we use in our empirical exercise. We chose a value of 80 for α, leading to price elasticities similar to our empirical estimates. The parameters governing income were chosen to generate a dispersion in prices similar to those in the BLP automobile data. Each market has only two firms so the market shares are notably larger, however the share of consumers choosing the outside good is approximately the same as the BLP data. Finally, by choosing $1,500$ markets, we have a similar number of observations as the BLP automobile data.

Table 2.2 shows the estimated results under two assumptions. First, we assume that we have a consistent estimate of the model parameters, which we generate by taking the true values and adding noise to them. Second, we consider the case where the optimal instruments are constructed at the true parameter values. This allows us to compare the loss in efficiency from using an approximation. In both cases, the parameter estimates are close to the population values. The price coefficient is estimated accurately and precisely, which is encouraging as price elasticities are a common object of interest. The mean parameters for X are estimated with the most error and the largest RMSE, relative to the true population value. When the true parameter values are used to construct the optimal instruments, we see the RMSE decrease by a factor of 3 for most parameters. This shows that the approximation does introduce noise into our estimates and this primarily impacts the mean cost and utility

[9] We provide a linearized version of these moment conditions in the appendix.

coefficients. However, in most cases the additional noise has little impact on inference, and the estimates are not statistically significantly different from the true parameters values. We turn now to our empirical application

2.7 Application to BLP Data

2.7.1 Empirical Framework

We use the exact same data used in BLP. There are twenty new U.S. automobile markets - one for each year from 1971 to 1990 - for a total of 2217 observations on prices, quantities, and characteristics of different vehicle models. We assume the firms set the same $K = 5$ characteristics as those that enter into the BLP utility function, including the ratio of horsepower to weight, interior space (length times width), miles per dollar, whether air conditioning is standard (a proxy for luxury), and the unobserved quality. The five cost shifters (W) are the unobserved quality, the log of ratio of horsepower to weight, the log of interior space, air conditioning, and the log of miles per gallon.[10] In a market with J products there are J observations on the K realized first-order conditions. The outside good quality ξ_0 is normalized to zero and we do not separately estimate the mean utility for new vehicles (i.e. constant term) instead letting it remain in the unobserved quality so in our setup $\beta \in R^{K-1}$. A $\xi_j > 0$ implies that new car on average is preferred to not purchasing a new good. Parameter $\theta = (\beta, \sigma, \alpha, \gamma)$ consists also of $\sigma \in \mathbb{R}^K$, $\alpha \in \mathbb{R}$, and $\gamma \in \mathbb{R}^K$ for a total of 3K=15 parameters to be identified.

Following the base specification in BLP, we assume that utility is given by

$$u_{ij}(\theta) \quad = \quad \alpha \, ln\left(y_i - p_j\right) + \delta_j + \sum_{k=1}^{K} \sigma_k v_{ik} X_{jk} + \varepsilon_{ij}$$

where $\delta_j = X_j' \beta + \xi_j$. Income draws y_i follow the same log-normal distribution estimated

[10] Air conditioning is an indicator variable which raises the issue of differentiability. We estimate the model both with and without the air conditioning first-order condition as we remain overidentified even when we do not use this condition. At the cost of complicating the estimator by having to combine moment equalities with moment inequalities we could add an inequality related to air conditioning or any other indicator-type characteristic.

in BLP and v_{ik} are normally distributed. Additionally, we assume that marginal costs are independent of the output level and are comprised of two terms: one term that is log-linear in the product characteristics and a second term which is unobserved by the econometrician. Specifically,

$$\ln(mc_j) = \gamma_\xi \xi_j + W_j' \gamma + \omega_j$$

The only difference between our marginal cost specification and that of BLP is that the level of unobserved quality, ξ_j, impacts marginal costs.

BLP assume that all X are exogenous and so any function of them can serve as instruments for any vehicle j. Using the firm pricing first-order conditions Pakes (1994) provides motivation for using the following as instruments for good j (which we call the BLP instruments): own product characteristic X_{jk}, $\forall k$, the sum of characteristic across own-firm products $\sum_{j' \neq j, j' \in J_f} X_{j'}$, and the sum of all characteristics across competing firms, $\sum_{j' \notin J_f} X_{j'}$. These instruments approximate the equilibrium pricing function where markup of a product depends on other products' characteristics. These instruments remain valid in our approach when $\{X_{jk}\} \in I_f$, and so can provide the basis for estimation.

To estimate the model parameters BLP impose that unobserved quality, ξ_j, is orthogonal to observed product characteristics. Formally, let $X = (X_j)_{j \in J}$ denote all of the characteristics observed to consumers, producers, and the researcher. Additionally, they assume that unobserved cost shocks are orthogonal to observed characteristics. BLP then use the following identifying restrictions,

$$E\left[Z_{jl}^D \xi_j\left(\theta_0^D\right) \mid X\right] = 0 \quad \forall j, l. \tag{2.12}$$

$$E\left[Z_{jl}^S \omega_j\left(\theta_0^D, \gamma_0\right) \mid W\right] = 0 \quad \forall j, l. \tag{2.13}$$

where Z_{jl}^D and Z_{jl}^S are functions of X and W, respectively. These conditions rule out correlation between observed and unobserved product characteristics. Ignoring this correlation can result in demand estimates that too inelastic when price is positively correlated with unobserved product quality (see e.g. Trajtenberg (1989))). Therefore, we replace these moment conditions with our moment conditions

$$E\left[H_{jkl}\nu_{jk}(\theta)\ \mid I_f\right],\ \forall\ k,l \tag{2.14}$$

where H_{jkl} is the approximation to the optimal instruments. We estimate the model under two different choices for I_f. The "lagged" information set, I^{lagged}, includes only last years observed and unobserved characteristics to construct $\nu_{jk}(\theta)$. In this case when firm f chooses her characteristics she does so using the configuration of competitors' last years products and characteristics to forecast her best guesses at profit maximizing characteristics' choices. In doing so she calculates the Bertrand-Nash prices that would be realized given her choices of observed and unobserved characteristics and the realized characteristics of her competitors products in the previous year. On the other hand, "contemporaneous" information set, I^{cont}, uses contemporaneous characteristics to construct $\nu_{jk}(\theta)$. We approximate $E\left[\nu(\theta)\nu(\theta)'|I\right] = I_{JK}$.[11] We transform the instrument $\hat{H}_{jkl}$ into a block diagonal matrix so that we have $K * 15 = 75$ instruments as a benchmark specification.

In addition, we estimate a version of the model where we augment the first-order condition moments with simple OLS moments on the marginal cost equation,

$$\begin{aligned} G_{2k}(X,\xi(X,\theta);\theta) &\equiv E\left[W_k(X,\xi(X,\theta))\,\omega(X,\xi(X,\theta);\theta)\right] \quad \forall k=1,...,K \\ &= 0. \end{aligned}$$

2.7.2 Results

Table 2.3 shows the demand and supply estimates. The first column restates the original BLP results and columns two and three labeled with FOC (first-order condition) are estimated using the optimization conditions given the information set $I_{f,t}^{\text{lagged}}$. That is, we

[11]This simplification does not affect consistency, only the efficiency. Another approximation of $E\left[\nu(\theta)\nu(\theta)'|I\right]$ can be done by a block diagonal matrix where a block is a K by K variance-covariance matrix of $\nu_{jk}(\theta)|_{k=1,\dots,K}$ for each firm f and year t.

estimate the parameters under the assumption that firms only know last year's characteristics of their competitors' cars when choosing their characteristics.[12] Column two uses the full set of instruments (Full IV) of which there are 43, one for each parameter-FOC pair after dropping one instrument due to high correlation. It is well known that while additional instruments always improve standard errors, if many of them are weak bias can be introduced into the estimates.[13] For this reason we also use a subset of these instruments that we think are likely to be the most informative (Partial IV). For each characteristic X_k (except ξ) we use only the derivatives with respect to $(\alpha, \beta_k, \gamma_k, \sigma_k)$. For ξ we use the derivatives with respect to $(\alpha, \gamma_k, \sigma_k)$ giving us a total of 19 instruments. In the robustness section we estimate parameters based on the assumption that firms know their competitors' contemporaneous characteristics when making choices.

The most striking difference between the BLP estimates in column one and optimization estimates in columns two and three is that the coefficient on price is much larger in the latter cases; consumers are significantly more price sensitive when optimization conditions are used for identification. Table 2.4 investigates the impact of this difference on estimated elasticities and markups. On average elasticities increase by 31% in absolute value in response to the increase in price sensitivity. This causes estimated markups to fall by on average around 22%.

One potential explanation for these changes across identification conditions is that observed and unobserved characteristics are positively correlated because firms put more unobserved quality into cars with high observed quality to the researcher. In this case the instrumented price in the BLP setup will be positively correlated with unobserved quality and this may be leading to an upward bias in the price coefficient. Table 2.5 explores whether ξ is positively correlated with X by regressing estimated ξ's on all of the BLP demand instruments. Consistent with the price coefficient changes, the BLP instruments explain 50% of the variation in ξ across vehicles and except for miles per dollar – which is negatively correlated with ξ – all other characteristics are positively correlated with ξ. The negative correlation between miles per dollar and ξ might be the reason that the coefficient in the BLP setup of miles per dollar is negative, that is, why people appear not to like fuel efficiency. In reality they like

[12] This makes us to drop the first year observations, resulting in the total number of models 2,125.

[13] For example see Bekker (1994), Newey and Smith (2004), or Hansen et al. (2008).

fuel efficiency but it is negatively correlated with other unobserved features of the vehicle that consumers' value.

The last step is to check whether the BLP instrumented price is positively correlated with ξ. We construct the instrumented price by regressing price on the BLP instruments to get a predicted price for each vehicle. Table 2.6 reports the estimates of the regression of these instrumented prices on an intercept and ξ. The coefficient is significant and positive and the correlation between the instrumented price and ξ is approximately 0.15. Thus the hypothesis that the price coefficient is biased up under the assumption of mean independence because observed and unobserved product characteristics are positively correlated is consistent with all of our findings from the model estimated with the optimization conditions. Similarly, Table 2.7 reports the results of the regression of the cost shocks on the BLP instruments for the cost function. Observed cost characteristics explain almost half of the movement in the unobserved cost shock, implying that exogeneity between W and ω does not hold. This would naturally further bias the estimated coefficients, although the direction is ambiguous.

A second related difference is in how the two demand models fit the data. Both models exactly match market shares of products using the BLP inversion. Only 10% of U.S. households buy new cars in any given year so both fitted demand models need a way to explain why 90% of households choose the outside good. The way they do so is quite different and the difference can be found in the final row of Table 2.3, which reports the average utility of purchasing a new vehicle net of price, $u_{ij} - \alpha \ln(y_i - p_j)$, from each model's fit: BLP predicts it at -4 while our approach predicts it at 7.5. BLP fits 90% of households not buying by having consumers derive strong negative utility from the act of buying a car relative to the outside option of no new car (excluding negative utility from price). In contrast, the optimization-fit has consumers strongly desiring new cars relative to the outside good but the significantly higher price elasticity causes 90% not to buy a new car.

Another difference is that some of the anomalies in the BLP point estimates are not present in the optimization-fit point estimates. The BLP point estimates imply consumers dislike fuel efficiency but in our setup they strongly and significantly like fuel efficiency. They also find costs are *decreasing* as interior space and fuel efficiency increases. We find costs increasing in all of the characteristics, including the unobserved characteristic ξ which enters our cost function but does not enter the BLP cost function.

Table 2.3 also shows that our estimates are almost always much more precisely estimated relative to the BLP-fit model whether we use the full or partial set of instruments. With the full set of instruments our standard errors are on average a fourth of the standard errors from BLP. The data is exactly the same data so the optimization moments appear to contain more information on both the distribution of consumer preferences and on the cost parameters.

Before turning to the rest of our results we note that the differences we report here between the optimization-fit model and the BLP-fit model have also been found in European automobile data (see Miravete et al. (2015)). They adopt our approach to estimating demand and supply to look at competition in the European automobile market. Using the optimization moments estimated price elasticities double on average and estimated markups fall relative to when they use the BLP moments, They find some anomalous demand and supply point estimates under the BLP-fit that are not present under optimization-fit. Under the optimization-fit their standard errors are much smaller and their unobserved quality term is positively correlated with observed characteristics.

2.7.3 Robustness

We explore the robustness of our results by estimating the model given different assumptions on the information sets of firms, the moments used during estimation, and firm optimizing behavior. First, we estimate the model assuming that the characteristics of all products at time t are known to firms at time t. We call this the contemporaneous information set. Then we estimate the model assuming that marginal cost shocks are orthogonal to the observed characteristics. Finally, we estimate a model where firms have dynamic first-order conditions, where their FOCs depend on the future stream of profit after a change in a product's characteristics.

Contemporaneous Information Set

Column 4 uses the "contemporaneous" information set, which assumes that $X_t \subset I_f$. While most of the coefficients in Column 4 are similar to Column 2 some are different although none significantly. The point estimate of the price coefficient decreases 5% but again the difference is not significant. Standard errors for some parameters go down relative to the "lagged" information set assumption.

Marginal Cost Moments

Column 5 of Table 2.3 includes marginal cost moments during estimation, using the additional restriction that the cost shock is orthogonal to the observed characteristics, i.e. $E[\omega_j|W] = 0$. The estimated coefficients are broadly consistent, however there are a couple key differences. First, the standard errors tend to be smaller than in the baseline model, likely due to the additional restrictions. Second, it is worth noting that the coefficient on price is roughly half that of the unrestricted model, but still larger than the estimates found in BLP. The lower estimates may be due to misspecification, as the marginal cost shock is conceivable correlated with observed characteristics.

Dynamic Optimal Decision

Some product characteristics may not be update every period, implying that firms maximize the sum of the future stream of profits at the time of decision. We allow for the "dynamic" optimization FOCs, by approximating the sum of the future stream of profits at the time of changing the characteristics to $E\left[\Pi_f \mid I_{f,t}\right] = E\left[\sum_{\tau=t}^{T-1} \pi_{f,\tau} \mid I_{f,t}\right] \approx E\left[\pi_{f,t} \mid I_{f,t}\right]$. T refers to the year when at least one characteristic is updated. In the estimation we restrict the observations to new models or existing car models where at least one characteristic is changed more than 10%. The last column in Table 2.3 reports the "dynamic" results. Although this reduces the number of observations to approximately half, the standard errors do not increase much as most of the variation in moments originate from the restricted observations. The estimated parameter coefficients are consistent with the base specification in both direction and magnitude.

2.8 Conclusions

Traditional identification since BLP in discrete choice demand model has been to assume no correlation between observed and unobserved characteristics. The major concern of this identification assumption is that it may lead to biased price elasticities if observed and unobserved characteristics are correlated with one other. We avoid this mean independence assumption and infer the distribution of consumer tastes in demand and supply estimation by exploiting optimal choices of product characteristics and prices by firms. We allow firms' information sets at the time they choose characteristics to potentially include competitors' product characteristics, demand, and cost shocks, signals on all of these, or no information at all on them. Following Hansen and Singleton (1982), our identification is based on the assumption that firms are correct in their choices on average even though firms may wish they had made different decisions ex-post.

Using the same automobile data from BLP, we find elasticities double and markups fall by 50%. We also find significantly more precise estimates given the same exact data and some of the slightly puzzling parameter estimates of BLP go away as all of our parameter estimates are of the correct sign.

Tables

Table 2.1: Data Generating Process Summary Statistics

	X	ξ	Price	Shares
Mean	1.41	1.82	4.98	0.01
Std Dev.	0.16	0.18	2.12	0.01
Max	2.62	2.42	45.88	0.08
Min	1.10	1.34	1.48	0.00
J	3000			

Table 2.2: Monte Carlo Simulation - Duopoly, J=3000

Parameter	Truth	IV at non-truth		IV at truth	
		Mean	RMSE	Mean	RMSE
α	80	79.39	1.50	79.82	0.37
β	2	1.84	0.33	1.92	0.11
γ_c	1.25	1.25	0.06	1.25	0.02
γ_X	-0.15	-0.14	0.16	-0.15	0.04
γ_{X^2}	1	0.94	0.10	0.97	0.04
γ_ξ	-0.25	-0.24	0.05	-0.25	0.02
γ_{ξ^2}	0.12	0.11	0.02	0.11	0.01
$\gamma_{X\xi}$	0.10	0.10	0.08	0.10	0.03
σ	0.50	0.50	0.01	0.51	0.01

Table 2.3: Estimated Parameters of the Demand and Supply

Parameter	Characteristic	BLP	FOC				
			Full IV	Part IV	Contemp.	MC moms	Dynamic
Term on price α	$ln(y-p)$	43.501	144.131	155.661	137.032	88.477	162.919
		(6.427)	(34.054)	(81.672)	(17.737)	(3.003)	(48.834)
Means (β's)	Constant	-7.061					
		(0.941)					
	HP/weight	2.883	1.168	1.527	0.886	5.169	0.937
		(2.019)	(0.192)	(0.884)	(0.158)	(0.418)	(0.229)
	Size	3.460	0.108	0.316	0.618	3.693	0.133
		(0.610)	(0.036)	(0.422)	(0.062)	(0.236)	(0.034)
	Air	1.521	1.722	0.975	1.324	0.706	1.319
		(0.891)	(0.331)	(0.812)	(0.140)	(0.092)	(0.276)
	MP$	-0.122	2.442	2.401	2.644	1.493	1.678
		(0.320)	(0.414)	(1.633)	(0.295)	(0.078)	(0.345)
Std. Dev. (σ's)	Constant	3.612	3.190	3.093	2.276	2.799	3.330
		(1.485)	(1.093)	(6.304)	(0.471)	(0.238)	(1.886)
	HP/weight	4.628	3.007	2.818	2.963	3.820	2.986
		(1.885)	(0.587)	(2.351)	(0.389)	(0.521)	(0.662)
	Size	2.056	0.934	0.919	0.371	0.621	0.641
		(0.585)	(0.150)	(0.747)	(0.169)	(0.375)	(0.128)
	Air	1.818	1.773	1.607	1.286	1.803	2.009
		(1.695)	(0.257)	(1.247)	(0.150)	(0.118)	(0.394)
	MP$	1.050	0.859	1.612	0.771	1.007	0.846
		(0.272)	(0.286)	(0.984)	(0.150)	(0.088)	(0.356)

Continued on next page

Table 2.3: Estimated Parameters of the Demand and Supply

Parameter	Characteristic	BLP	FOC				
			Full IV	Part IV	Contemp.	MC moms	Dynamic
Cost params. (γ's)	Constant	0.952					
		(0.194)					
	Mean charac. (ξ)		0.122	0.121	0.134	0.120	0.113
			(0.028)	(0.056)	(0.017)	(0.004)	(0.033)
	HPweight	0.477	0.059	0.067	0.053	0.414	0.049
		(0.056)	(0.015)	(0.041)	(0.011)	(0.033)	(0.016)
	Size	-0.046	0.030	0.054	0.118	0.767	0.026
		(0.081)	(0.006)	(0.066)	(0.013)	(0.039)	(0.007)
	Air	0.619	0.226	0.135	0.188	0.120	0.170
		(0.038)	(0.042)	(0.078)	(0.024)	(0.014)	(0.042)
	MPG	-0.415	0.551	0.538	0.636	0.925	0.423
		(0.055)	(0.081)	(0.386)	(0.075)	(0.014)	(0.086)
J		2,217	2,125	2,125	2,217	2,125	902
Num. IVs		15	43	19	72	78	59
Median net util.	$u_{ij} - \alpha \ln(y_i - p_j)$	-0.915	5.432	7.289	3.520	4.211	4.856

Table 2.4: Implied Elasticities and Markups

	Elasticities			Markups ($)		
	BLP	FOC		BLP	FOC	
		Full IV	Part IV		Full IV	Part IV
Lexus LS400	-3.027	-4.836	-5.008	9,214.54	5,754.35	5,553.02
Lincoln Towncar	-3.030	-5.708	-5.973	8,310.82	4,633.86	4,435.15
Nissan Maxima	-4.124	-7.867	-8.155	3,385.84	1,780.48	1,716.92
Ford Taurus	-3.952	-8.205	-8.966	2,679.14	1,363.66	1,244.59
Chevy Cavalier	-5.899	-10.284	-11.668	1,327.75	755.42	654.93
Nissan Sentra	-6.304	-10.751	-12.420	909.79	533.02	459.87
Mean	-4.087	-7.796	-8.482	4,051.87	2,393.99	2,280.53
Median	-3.975	-8.219	-8.789	2,751.77	1,397.99	1,324.34
Std. Deviation	1.120	2.130	2.577	3,905.32	2,821.63	2,712.61

“Full IV" uses the full set of instruments while “Part IV" uses a subset of instruments.

Table 2.5: $E\left[\xi_j \mid X\right] \neq 0$

ξ	Full IV		Part IV	
Constant	0.898	-3.102	1.700	-1.686
	(0.594)	(1.122)	(0.673)	(1.241)
HP/weight	6.693	6.123	7.506	5.919
	(0.593)	(0.613)	(0.672)	(0.678)
Size	5.463	4.239	5.607	3.888
	(0.294)	(0.335)	(0.334)	(0.371)
Air	1.397	0.863	2.503	1.693
	(0.135)	(0.136)	(0.154)	(0.150)
MP$	-2.808	-3.952	-3.122	-4.844
	(0.0996)	(0.143)	(0.113)	(0.158)
Other BLP instruments	No	Yes	No	Yes
R-squared	0.621	0.736	0.624	0.751

“Full IV" uses the full set of instruments while “Part IV" uses a subset of instruments.

Table 2.6: Correlation Between Instrumented price ($\hat{p}(IV_X)$) and ξ

$\hat{p}(IV_X)$	Full IV	Part IV
Constant	7.734	7.048
	(0.206)	(0.202)
ξ	0.778	0.773
	(0.031)	(0.026)
R-squared	0.220	0.281

$\hat{p}(IV_X)$ is predicted price on BLP demand instruments.

"Full IV" uses the full set of instruments while "Part IV" uses a subset of instruments.

Table 2.7: $E\left[\omega_j \mid W\right] \neq 0$

ω	Full IV		Part IV	
Constant	0.090	0.993	-0.165	0.655
	(0.136)	(0.164)	(0.148)	(0.184)
ln(HP/weight)	0.326	0.048	0.294	0.031
	(0.035)	(0.033)	(0.038)	(0.036)
ln(Size)	-0.800	-0.418	-0.871	-0.467
	(0.064)	(0.060)	(0.069)	(0.067)
Air	0.341	0.155	0.299	0.134
	(0.017)	(0.0160)	(0.019)	(0.018)
ln(MPG)	0.049	-0.136	0.123	0.025
	(0.043)	(0.0441)	(0.047)	(0.049)
Other BLP instruments	No	Yes	No	Yes
R-squared	0.314	0.578	0.283	0.533

"Full IV" uses the full set of instruments while "Part IV" uses a subset of instruments.

Chapter 3

Accessibility or Amenities? Estimating the Value of Light Rail Transit

3.1 Introduction

The costs associated with a mass transit system are readily apparent, from the large initial capital investment to the sustained public funding necessary for its operation. However, the value the public derives from these systems is less tangible making it difficult to weigh the transit system's benefit against its costs. For this reason, substantial work has been done in the economics literature to value the public's willingness-to-pay for public transportation.[1]However, less work has been done to quantify the channels through which this value is derived. he public not only directly benefits from improved access to public transportation, it also indirectly benefits from the endogenous response of local amenities to the introduction of a transit system. The aim of this paper is to quantify the total value of a new transit system and to decompose this amount into a direct and indirect valuation, with an application to the introduction of the METRO Blue Line in Minneapolis, Minnesota.

[1]A number of studies are available on the subject, spanning dozens of cities. These include Atlanta (Cervero, 1994; Ihlanfeldt, 2003; Immergluck, 2009), Buffalo (Hess and Almeida, 2007), Charlotte (Billings, 2011), Chicago (McDonald and Osuji, 1995; McMillen and McDonald, 2004), Dallas (Clower et al., 2002; Nelson et al., 2015), Hampton Roads (Wagner et al., 2017), Huston (Pan, 2013), Miami (Gatzlaff and Smith, 1993), Los Angeles (Cervero and Duncan, 2002), Philadelphia (Kilpatrick et al., 2007), Phoenix (Seo et al.,

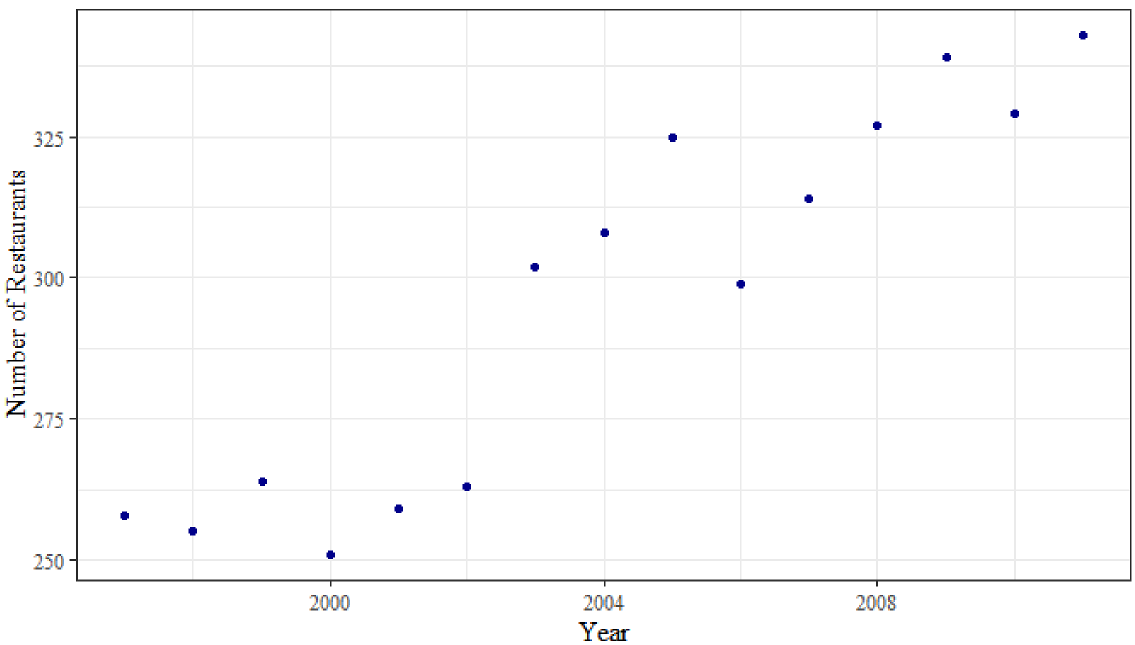
Formerly known as Hiawatha Line, the construction of the METRO Blue Line was first proposed by the Minnesota Department of Transportation in 1985, but it was not until January 2001 that construction began. The Blue Line started operations between a subset of 12 stations in June 2004, and full service started in November 2004. It connects downtown Minneapolis with its southern suburbs, counting 18 stations and spanning a total of 12 miles. Two studies have examined the impact of the Blue Line in Minneapolis on housing properties.[2] Goetz et al. (2010) published a comprehensive study focusing on the Blue Line's impact on property prices, housing investment and land use. They find modest price premiums (in the order of 3.8-4.0%) for single family homes located within a half mile of a station in South Minneapolis, with the net effect varying non-linearly as a function of distance. Pilgram and West (2018) use repeat sales to establish the effect of opening the Blue Line within a difference-in-difference setup. They find that single-family homes located within half a mile from a station in South Minneapolis experience a positive price premium (2.5-4%), but that the premium is diminishing over time, potentially as a result of the Great Recession.

Even before construction was completed, neighborhoods surrounding Blue Line stations started seeing an uptick in the number of business being opened. For example, Figure 3.1 illustrates the number of restaurants, art, and entertainment establishments within one mile of a Blue Line station between 1997 and 2011. A significant increase in the number of these establishments is evident starting around 2003. This is in line with the findings of Berry and Waldfogel (2010) who find that restaurants and other businesses with high variable cost increase in number and diversity as market size increases. The large increase in local amenities in response to the introduction of the Blue Line likely had a sustained impact on house prices, and prior studies of the Blue Line have not attempted to quantify the value of this response.

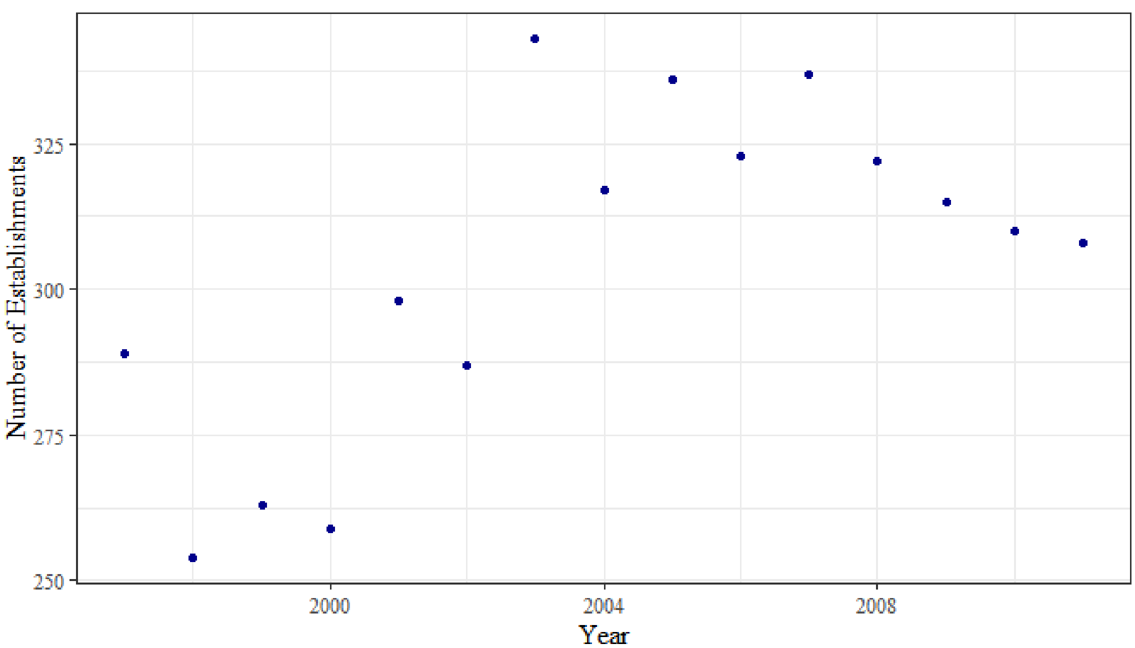
2014), Portland (Dueker and Bianco, 1999), Sacramento (Rewers, 2010), Santa Clara County (Weinberger, 2001), San Diego (Duncan, 2008), Washington County (Knaap et al., 2001), Washington DC (Damm et al., 1980; Grass, 1992; Cervero, 1994). Outside the United States, there are studies focusing on Amsterdam (Debrezion et al., 2011), Beijing (Zheng et al., 2016), Bogotá (Tsivanidis, 2018), Haifa (Portnov et al., 2009), London (Gibbons and Machin, 2005), Manchester (Forrest et al., 1996), Ottawa (Hewitt and Hewitt, 2012), Seoul (Bae et al., 2003), Shangai (Pan and Zhang, 2008), Toronto (Dewees, 1976), (Bajic, 1983), among others.

[2] Two additional studies examine on the effect of the Blue Line on other outcomes: Ko and Cao (2013) focus on industrial and commercial properties values, and Hurst and West (2014) investigate the effect of the Blue Line on land-use changes.

Figure 3.1: Establishments within 1 mile of a Blue Line Station



(a) Restaurants



(b) Entertainment

A few studies (Bowes and Ihlanfeldt, 2001; Zheng et al., 2016) have attempted to disentangle the direct and indirect effects of mass transit investments, but they typically assume that the level of amenities is independent of unobserved characteristics that impact housing prices, conditional on observed characteristics. However, this assumption fails in the presence of preference externalities. If residents tend to cluster based on shared unobserved preferences, each neighborhood will see a different composition of establishments entering the local market in response to the introduction of a new transit system. The mix of new establishments will naturally be correlated with unobserved preferences and will therefore confound estimation. This paper adapts recent techniques from the machine learning literature to decompose the benefits of introducing a light rail system into direct and indirect effects. Furthermore, we estimate how both the direct and indirect effects can vary heterogeneously across different types of neighborhoods. Our approach allows us to identify relevant features given a large selection of covariates in a data-driven manner, while simultaneously incorporating instrumental variables to control for endogeneity.

The closest paper to ours in terms of methodology is Ho (2016), which uses gradient boosting techniques to estimate the effects of air pollution on house prices. Ho (2016) follows Varian (2014) to identify which properties are unaffected by air pollution from a first-stage estimation and uses these observations as a control group. A second-stage estimation is then performed based on these observations and property prices are predicted for the treatment group. The difference between the predicted and realized prices is the estimated effect of air pollution, which is then regressed onto observed covariates to model how the estimated effect varies heterogeneously. We argue that this approach only allows us to recover the direct effect of adding public transportation and that instrumental variables are necessary to recover the indirect effect. To this end, we incorporate the insight of Athey et al. (2019) into the Boosted Smooth Tree framework of Fonseca et al. (2018) to estimate a predictive model with causal interpretation.

The results of our estimation routine using Boosted Smooth Trees show that the price of properties located within a half mile of a light rail station increased by around 11.3%. This total effect is in line with that estimated via DiD (10.4% in our preferred specification), and somewhat higher than the overall effect estimated by Goetz et al. (2010) and Pilgram and West (2018). This might be due at least in part to a different geographic focus, since we examined all neighborhoods along the path of the Blue Line, while other studies on the impact on this topic focus exclusively on neighborhoods in Southern Minneapolis. The Boosted Smooth Trees estimation procedure also allows us to directly calculate the estimated spillover due to changes in amenities, quantifiable at 5.8%, while the direct impact of access to the light rail itself is estimated to increase local housing prices by 5.5%. Thus over 51% of the overall appreciation in housing prices after the introduction of the Blue Line is attributable to an increase in the number of new amenities around light rail stations. The only comparable result in the literature is from Zheng et al. (2016) who found that the increase in neighborhood restaurant activities due to the introduction of a new subway station in Beijing captures 20 to 40% of the overall appreciation in home values. The discrepancy might be explained by the fact that we control for a far greater variety of businesses than Zheng et al. (2016), who focus exclusively on restaurants.

The rest of the paper will proceed as follows: Section 3.2 introduces the theoretical framework for our estimation, while Section 3.3 discusses the different possible approaches to estimating treatment effects, such as difference-in-differences and machine learning methods. Section 3.5 summarizes our data sources for housing values, neighborhood amenities and demographics, while Section 3.6 presents our results under the different estimated approaches, and Section 3.7 concludes.

3.2 Theoretical Framework

The starting point for our analysis is the hedonic model introduced by Rosen (1974) who proposes generating a pricing surface based on a vector of characteristics of the good of interest, in our application housing. Under certain regularity assumptions, the derivative of the pricing surface with respect to a given set of characteristics represents the consumers marginal willingness-to-pay for said characteristics.

Consider z, a vector describing the characteristics of a good (in our case, residential housing). The good has a market price which arises as an equilibrium object from the endogenous sorting of buyers and sellers, where the buyers' indifference curve and sellers' offer curve are tangent, conditional on the housing characteristics. Housing prices can thus be written in terms of the vector of housing characteristics z, $p(z)$. Let x represent the consumption bundle of all other goods. Then a consumer with utility function U solves the following utility maximization problem:

$$\max_{\{z_j\}} U(x, z_1, z_2, ..., z_J) \quad \text{subject to} \quad y = x + p(z)$$

where income y is measured in units of x. For each housing characteristic j, the consumer's first-order conditions are given by:

$$\frac{\partial U(y - p(z), z_1, z_2, ..., z_J)}{\partial z_i} = -U_x \frac{\partial p(z)}{\partial z_i} + U_{z_i} = 0$$

So that the consumer marginal willingness-to-pay for characteristic z_j can be written as:

$$\frac{\partial p(z)}{\partial z_i} = \frac{U_{z_i}}{U_x}$$

The pricing surface is an equilibrium object and therefore can change over time, so it will not necessarily be the case that $p_{t-1}(z) = p_t(z) = \bar{p}(z)$. When using a before and after approach (such as difference-in-differences), the estimated effect is a combination of the marginal effect and the equilibrium response. This complicates the interpretation of the parameter estimates.[3] Dealing with a shifting pricing surface is beyond the scope of this paper, so we assume that $p_t(z) = \bar{p}(z)$.

[3]See Taylor (2003) and Palmquist (2005) for a review of the empirical literature and various difficulties that arise when using a hedonic approach.

3.3 Estimating Treatment Effects

In this section we briefly describe the most popular methods for estimating the impact of public transit on house prices. We begin with a discussion of the panel data and difference-in-differences approaches that are commonly employed in the literature. We then turn to the method of Ho (2016) which applies machine learning to the hedonic framework. We demonstrate that this approach can be augmented with instrumental variables in order to decompose the total effect into a direct and indirect effect. Finally, we end on a detailed description of our estimator.

3.3.1 Panel Data

Early studies on the impact of mass transit projects focused on the estimation of hedonic pricing surface using a panel of housing sales. Typically, the (log) price is regressed on a set of covariates which include the distance to the nearest transit stop, d. The MWTP for access to transit in this setting is given by:

$$\frac{\partial p_{it}}{\partial d_{it}} = \beta_d p_{it}$$

where i and t index homes and time, respectively. Because house prices are observed, consistent estimation of the MWTP is equivalent to consistently estimating β_d.

The simplest approach to estimating the impact of the introduction of light rail transit is to use a time-varying cross sectional regression of home sales and estimate the coefficient with respect to distance to public transit.[4] This approach is valid under the assumption that the covariates included control for all channels through which unobserved preferences impact housing prices. Therefore, a simple regression of the distance to public transit, amenities, and exogenous covariates would provide valid estimates for each channel. Then the impact of public transit on amenities could be estimated to generate the desired decomposition.[5] This method is easy to implement and the assumptions for valid causal identification are

[4] This approach is perhaps the most popular in the literature. See Bajic (1983), Gatzlaff and Smith (1993), Forrest et al. (1996) Bowes and Ihlanfeldt (2001), Cervero and Duncan (2002), Bae et al. (2003), Duncan (2008), Immergluck (2009), Portnov et al. (2009), Rewers (2010), Weinberger (2001), Debrezion et al. (2011) among others.

[5] See for example Zheng et al. (2016).

apparent. Further, it is easy to extend this approach to allow for nonlinear effects and interactions across covariates, allowing the researcher to specify a model with rich heterogeneous effects. Finally, if there are still concerns about endogeneity, implementing a cross-sectional approach with instruments is straightforward.

Controlling for the impact of unobserved preferences on house prices requires knowing which exogenous covariates to condition on, which the researcher will not know a priori. The general strategy is then to include a large number of demographic and individual characteristics to avoid any omitted variable bias. However, the inclusion of irrelevant regressors or collinear regressors leads to higher variance estimates. Additionally, adding interactions and higher-order terms can quickly lead to a situation where $K \gg N$. For instance, the number of terms in a fully saturated model grows exponentially in K and can therefore dominate N even for a modest number of covariates. To avoid this, the researcher needs to determine which terms to include a priori, without guidance from the data. Machine learning techniques such as LASSO are effective at generating a parsimonious specification but tend to lead to overly sparse models. Additionally, they have a harder time to adapt to the local nature of the data generating process.

3.3.2 Difference-in-Differences

Recent papers have employed a difference-in-differences empirical strategy to estimate treatment effects.[6] This strategy entails defining treatment and control groups using concentric circles around each transit station, treating the inner circle as the treatment group and the outer circle as the control group. Typically, these papers assign houses that are within a certain radius (usually 1 km or 0.5 miles) of a new station to a treatment group and use houses located further from the station as a control group. The pre-treatment period can be defined in several ways: before the system is announced, before construction begins, or before the system opens. Based on these definitions, a simple difference-in-differences estimator is implemented to produce an estimate of the total effect of public transportation on housing prices.

The causal impact of the transit system may be recovered by looking at the difference in house prices between treatment and control group before and after the introduction of the light rail system, assuming that this difference would be constant absent any treatment. This strategy has several strengths. It is easy to implement and gives valid causal estimates if the underlying assumptions hold. The average treatment effect may be consistently estimated with few assumptions about functional form and the researcher is not required to control for all determinants of house prices, since it mitigates some of the endogeneity concerns arising from omitted variable bias in traditional cross sectional hedonic models. Additionally, this method can be extended to allow for the estimation of continuous pricing surfaces, such as in Diamond and McQuade (2019).

A limitation of this type of analysis is that the estimated average treatment effect is a combination of the direct effect from access to public transit and the indirect effect of amenity changes. To decompose these effects, we need a consistent estimate of the impact of transit on amenities and the impact of amenities on housing prices. Estimating the impact of the Blue Line on amenities is straightforward, but the existence of preference externalities and other confounding factors can once again make the estimates of amenities on house prices inconsistent. Because we consider a wide selection of amenities, a simple before and after approach will not identify each individual effects. Our proposed solution is to explicitly model all channels that affect housing prices and find relevant instruments to obtain causal identification.

3.3.3 Machine Learning

Varian (2014) proposes a method for estimating treatment effects given a well defined treatment and control group, and a predictive model. Let $\mathcal{C}$ denote the subset of houses in the control group and $\mathcal{T}$ the set of houses in the treatment group. Assume that the price of a house in period t is given by

[6] Among others, these include Baum-Snow and Kahn (2000), Gibbons and Machin (2005), Goetz et al. (2010), Billings (2011), Wagner et al. (2017), Pilgram and West (2018).

$$p_{it} = f(X_{it}) + \epsilon_{it}$$

for both the treatment and control groups. The price of a house in period $t+1$ is given by

$$p_{it+1} = f(X_{it+1}) + g(\tilde{X}_{it+1}) + \epsilon_{it+1}$$

for the treated group, where $\tilde{X}_{it+1} \subset X_{it+1}$ and the function $g(\cdot)$ captures the direct effect of public transportation on house prices. The researcher first train a model on the control group, estimating $E_{\mathcal{C}}[p_{it}|X_{it}] = f(X_{it}) + E_{\mathcal{C}}[\epsilon_{it}|X_{it}]$. Under the assumption that $E_{\mathcal{C}}[\epsilon_{it}|X_{it}] = E_{\mathcal{T}}[\epsilon_{it}|X_{it}]$, one can use the model to predict the price in period $t+1$ of the treatment group to get

$$E_{\mathcal{T}}[p_{it+1} - E_{\mathcal{C}}[p_{it+1}|X_{it+1}]|X_{it+1}] = g(\tilde{X}_{it+1})$$

generating an estimate of the direct treatment effect.[7] Following Bajari and Benkard (2005), it is then possible to take the residual $r_{it} = p_{it+1} - E_{\mathcal{C}}[p_{it+1}|X_{it+1}]$ and regress it on the covariates X_{it+1} to uncover heterogeneous treatment effect resulting from the treatment.

Given a well defined pre-treatment and post-treatment period, this method also provides a check on the definition of the control group. Before treatment occurs, it must be that:

$$E_{\mathcal{C}}[p_{it}|X_{it}] = E_{\mathcal{T}}[p_{it}|X_{it}]$$

The residual from predicting outcomes in the pre-treatment treatment group using the control group should have mean zero, but will in general not be mean zero in the post-treatment period. If estimates using the control group are not mean zero, then the control group is not

[7]This strategy is similar to the synthetic control method (Abadie and Gardeazabal, 2003; Abadie et al., 2010), which uses a weighted combination of observations in the control group in order to approximate the desired attributes in the treatment group in order to estimate a pricing function analogous to $f(X_{it})$.

sufficiently similar to the treatment group to provide reliable estimates. Of course, this does not necessarily guarantee that the control group is valid, especially if the control group is contaminated by the treatment. In this case, we would expect the control group to perform well at predicting the treatment group because the control group should have been included in the treatment group to begin with.

Unfortunately, following this approach does not allow us to explicitly recover the indirect treatment effect as well. The indirect effect is given by:

$$IE = E_{\mathcal{T}}\left[f(X_{it+1}) - f(X_{it})\right] - E_{\mathcal{C}}\left[f(X_{it+1}) - f(X_{it})\right]$$

However, we can only estimate the term:

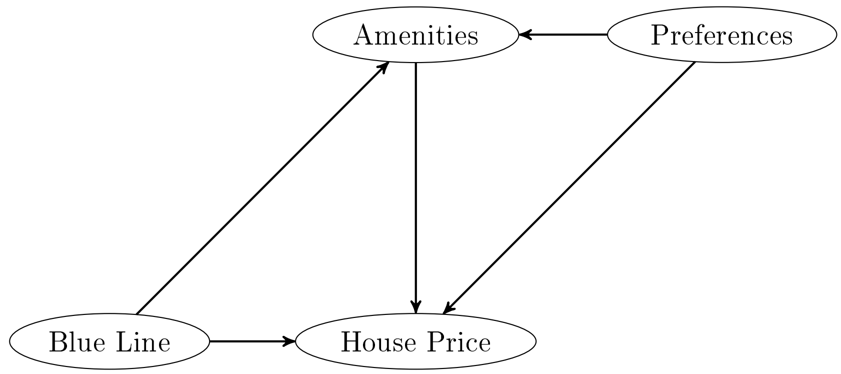
$$\begin{aligned} Biased\ IE = & \underbrace{E_{\mathcal{T}}\left[f(X_{it+1}) - f(X_{it})\right] - E_{\mathcal{C}}\left[f(X_{it+1}) - f(X_{it})\right]}_{\text{Indirect Effect}} + \\ & \underbrace{E_{\mathcal{T}}\left[\epsilon_{it+1} - \epsilon_{it}\right] - E_{\mathcal{C}}\left[\epsilon_{it+1} - \epsilon_{it}\right]}_{\text{Selection Bias}} \end{aligned} \tag{3.1}$$

While the distribution of ϵ_{it} conditional on X_{it} might be the same for each group, differences in the group compositions can result in selection bias. We propose augmenting the approach of Varian (2014) with instrumental variables, so that the predictive model has a causal interpretation. This eliminates the selection bias an allows us to consistently recover indirect effects. Before introducing our proposed estimator, we discuss the issue of endogeneity in our framework and our proposed instruments.

3.4 Instrumental Variables

Consider the causal relationship depicted in Figure 3.2. The introduction of the Blue Line has a direct impact on house prices, due to consumers valuing access to public transportation. However, it also causes an increase in local amenities, providing an indirect channel through which it again impacts home values. Public transportation increases the catchment area for local businesses, making it more likely that businesses will locate near stops.

Figure 3.2: Blue Line DAG



Nearby houses benefit from this increase in amenities, further increasing their value. However, this channel is confounded by the presence of unobserved preferences which impact both the level of amenities available in a given neighborhood as well as house prices. For instance, more expensive restaurants might locate in wealthier neighborhoods or bars and more movie theaters might open up in neighborhoods that are relatively younger. This is especially problematic in our setting where we do not observe household income or the age demographics of the household.

To account for this endogeneity, we instrument for the level of local amenities using the level of amenities in all neighborhoods excluding the location of interest. A firm's entry decision depends not only on the observed and unobserved characteristics of the neighborhood, but also on aggregate trends in demand and supply. For instance, a general rise in income will lead to more restaurants entering all markets and reflects shifts in demand that are uncorrelated with local unobservables. Similarly, citywide changes in the cost structure of firms will impact entry decisions, but will be orthogonal to local unobservables. Firm entry outside of the neighborhood will therefore be correlated with local entry but will be orthogonal to the unobservable error. This logic is similar to that of the instruments used in Fan (2013).

If aggregate trends in demand and supply shifters also impact individual house prices then these instruments would be invalidated. This would be the case if house prices increase dues to increases in wages or increase or asset prices. However, we include several covariates that control for aggregate trends in house prices, such as the Case-Shiller index for Minneapolis.

The identifying assumption is that conditional on the observed city-wide covariates, our instruments are orthogonal to local unobservables.

3.4.1 Boosted Smooth Trees

In this section we provide a detailed description of how we implement our preferred estimator. Readers interested in the empirical application may go directly to Section 3.5.

3.4.1.1 Gradient Boosting

A common approach in the machine learning literature to estimate a predictive model is to use gradient boosting, as proposed in Friedman (2001). This approach builds up an estimate of $F(X_{it}) = f(X_{it}) + E[\epsilon_{it}|X_{it}]$ by using functional gradient descent to iteratively improve the performance of regression function. The main goal is to solve

$$\hat{F} = \text{argmin}_F E_{p,x} \left[L(p, F(X)) \right]$$

where F is the function to be approximated, p is the dependent variable, and L is a loss function. Solving for F directly is infeasible, but we can use functional gradient descent to update an approximation in step m as:

$$F_m(X) = F_{m-1}(X) - \gamma_m \bigtriangledown_{F_{m-1}} L(p, F_{m-1}(X))$$

At step $m-1$, the researcher calculates $r = - \bigtriangledown_{F_{m-1}} L(p, F_{m-1}(X))$ from the data and the current approximation $F_{m-1}(X)$. They then approximate the function r with a weak learner h_m. The weak learners are chosen so that they have high bias and low variance, meaning that an individual h_m does a poor job approximating a given function, but an ensemble of weak learners can provide an arbitrarily close approximation. The gradient r is regressed on h_m, and the estimator is updated according to:

$$F_m(X) = F_{m-1}(X) + \gamma_m \hat{h}_m(X)$$

The step size γ_m is estimated by regressing $\hat{h}_m(X)$ on r. It is common to take the loss function to be the quadratic loss $L(p_i, F(X_i)) = (p_i - F(X_i))^2$ and h_m to be decision trees with depths ranging from 2 to 6 (Hastie et al., 2009).

3.4.1.2 Decision Trees

Decision trees are commonly used because they are flexible and can adapt well to the local structure of the function. Decision trees approximate the target function using a piecewise constant function over a partition of the observations. Formally, let J be the set of parent nodes and T be the set of terminal nodes. Then the decision tree can be written as:

$$h_m(x_i) = \sum_{k \in T} \beta_k B_{J_k}(x_i; \theta_k)$$

where:

$$B_{J_k}(x_i; \theta_k) = \prod_{j \in J} I(x_{s_j}; c_j)^{\frac{n_{kj}(1+n_{kj})}{2}} \left(1 - I(x_{s_j}; c_j)\right)^{(1-n_{kj})(1+n_{kj})}$$

and

$$I(x_{s_j}; c_j) = \begin{cases} 1 & \text{if } x_{s_j} \leq c_j \\ 0 & \text{otherwise} \end{cases}$$

and

$$n_{kj} = \begin{cases} -1 & \text{if the path of leaf k does not include the parent node j} \\ 0 & \text{if the path of leaf k includes the right-hand child of parent node j} \\ 1 & \text{if the path of leaf k includes the left-hand child of parent node j} \end{cases}$$

Note that $\sum_{k \in T} B_{J_k}(x_i; \theta_k) = 1$ and each observation x_i is mapped uniquely to some region of space.

3.4.1.3 Boosted Smooth Trees

Decision trees use local averaging, leading to function approximations that are step functions. As such, the approximation's derivative is zero almost everywhere. Because MWTP is based on the derivative of the hedonic pricing function, we prefer an approximation that is smooth. Fonseca et al. (2018) propose replacing the indicator $I(x_{s_j}; c_j)$ with a sigmoid function:

$$L(x_{s_j,i}; \gamma_j, c_j) = \frac{1}{1 + e^{-\gamma_j (x_{s_j,i} - c_j)}}$$

so that every point has a positive probability of being assigned to any terminal leaf. As the term γ_j increases, the model converges to a standard decision tree. Moderate values of γ_j smooth the estimates and the authors show that this allows for better estimation of the derivatives.

Unfortunately, this specification is far more computationally demanding than using a regression tree. The main issue is that the gradient boosting algorithm does not require us to actually construct the matrix $\{B_{J_k}(x_i; \theta_k)\}_k$ and regress it on r for each potential split. However, this step is unavoidable when using $L(\cdot)$ because testing a new split requires recalculating the choice probabilities for every leaf. This makes the Fonseca et al. (2018) algorithm, BooST, impractical for very large datasets. In Appendix C.2, we propose two refinements to the BooST algorithm to remove the runtime's quadratic dependence on the number of observations and to test all potential splits with a single pass through the data. This allows us to efficiently scale the algorithm to problems with several hundred covariates and have it run in a couple minutes, rather than a few days.

3.4.1.4 Incorporating Instruments

An advantage of the linear regression formulation is that it is straightforward to incorporate instruments in the estimation routines. Assume that we have access to a set of instruments Z, such that local estimation equation holds:

$$E[Z'(p - F(X))|X] = 0$$

Then we can introduce the following loss function:

$$L(p, F) = (p - F(X))' P_Z (p - F(X))$$

and apply the gradient boosting algorithm with smooth trees. At each step m, we fit the residual:

$$r = -\gamma_m \bigtriangledown_{F_{m-1}} L(p, F_{m-1}(X))$$

with a weak learner $h_m(x)$ that is a smooth tree, using Z as a matrix of instruments. By construction, the residual is orthogonal to the matrix of instruments at each step of the estimation routine, resulting in a final estimator that satisfies the local moment condition for all values of X. We do not currently have a proof of consistency, but provide Monte Carlos in Appendix C.2 to justify this approach. Further, we note the similarity between this approach and that of Athey et al. (2019), which uses decision trees rather than smooth trees, but provides some theoretical guarantees of consistency.

This algorithm provides several advantages. First, it provides a smooth pricing surface for which derivatives can be easily calculated. Second, it allows us to choose relevant regressors in a data driven manner, akin to the standard gradient boosting algorithm. Finally, it allows us to instrument for amenities values, and therefore approximate the indirect effect of public transportation on house values. We turn next to a discussion of the instruments we use during estimation.

3.5 Data

3.5.1 Housing Data

This analysis quantifies the effect of the construction of the Blue Line by examining its impact on the sale price of residential properties. The sale records for each property were collected from the City of Minneapolis Tax Assessor Office, along with basic property characteristics, such as the year of construction, the square footage, the number of stories, the number of bedrooms and bathrooms. An identifier number (PID) unique to each property allowed us to merge this information with Hennepin County records in order to geocode the location of each property. Geocoding allowed us to determine the distance of each property from the closest Blue Line station, as well as other transit options and nearby amenities. The analysis focuses on sales occurring between 2002 and 2006, the two years before and after the introduction of the Blue Line in 2004. This yields a total of 38,930 individual transactions, after excluding foreclosures and other non-market sales.

Table 3.1: Summary Statistics of Housing Data

	Mean	St. Dev.	Min	p25	p50	p75	Max	N
Sale Price	224,801	108,139	11,000	158,500	200,988	263,500	779,737	38,930
Distance to BL	2.046	1.301	0.046	0.904	1.888	2.942	5.236	38,930
Year Built	1,938	32.105	1,900	1,913	1,926	1,955	2,006	38,922
Sq. Feet	2,020	808.906	224	1,514	1,978	2,450	4,996	38,930
# of Stories	1.457	0.463	1.000	1.000	1.200	2.000	5.000	38,561
# of Baths	1.764	0.775	0.000	1.000	2.000	2.000	6.000	38,598

3.5.2 Transportation Data

Information on the public transit system in Minneapolis was obtained from the Minnesota Geospatial Commons, which yielded a dataset containing the location of over 5,518 transit stops within the City of Minneapolis across 147 separate transit routes. The closest transit stop for each transit line was identified for each residential property in the sample. In order to reduce the dimensionality of the data, the closest stop along the major transit axes between Minneapolis and its suburbs was also identified (see Appendix C.1 for details.)

3.5.3 Neighborhood Amenities

A list of amenities within 0.5 miles of each property was compiled using ReferenceUSA data on local businesses, updated for each year between 2002 and 2006. We were thus able to track new businesses openings, existing businesses changing locations and businesses closing down within the City of Minneapolis over this time period. NAICS codes were used to categorize of each business, in order to calculate the density of each type of amenity around each individual property. This exercise yielded 28 amenity categories, such as "Full-Service Restaurants" or "Museums, Historical Sites, and Similar Institutions", to be used in later analysis (see Appendix C.1 for details.) To accomplish this, we calculated the number of businesses within a 0.5 miles radius of each property for each category. Further information on the quality of the amenities in the neighborhood of each individual property was scraped from Yelp, in particular the average rating of shopping outlets and restaurants, as well as information on the distance to the closest educational institution (childcare centers, elementary schools, high schools and colleges) to each property.

3.5.4 Demographic Data

Demographic information for each Census Tract was downloaded from Social Explorer for the 1990 and 2000 Decennial Census, and the 2008 - 2012 American Community Survey (ACS). The key variables of interest include the demographic make up of each neighborhood (% white residents, % black residents, % female residents), educational attainment (% college graduates, % high school graduates), economic variables (median household income, % living in poverty, % receiving public assistance, % unemployed), the share of owner occupied units and of vacant units, information about means of transportation to work (% commuting by car, % commuting by public transit) and the average commute length.

3.6 Results

3.6.1 Difference-in-Difference

The standard approach to a problem such as this is using a difference-in-differences framework where outcomes of properties located within a certain radius from the closest Blue Line station are compared to those of properties located beyond this radius. Thus, properties located within a 0.5 miles radius from the closest Blue Line stop have been assigned to the treatment group, while properties located between 0.5 and 1 miles of the closest Blue Line stop were assigned to the control group. Estimation results for this technique are reported in Table 3.2. The first specification reports results for a DiD routine with no controls, the second specification adds year and month fixed effects, and the last specification controls for housing characteristics, such as the year of construction, square footage, number of bed-

Table 3.2: Difference in Difference Regression Results, Log Sale Price

VARIABLES	(1) Log Sale Price	(2) Log Sale Price	(3) Log Sale Price
Treatment	0.0100	0.0120	0.0361***
	(0.0164)	(0.0154)	(0.0139)
Treatment * Post	0.126***	0.123***	0.104***
	(0.0193)	(0.0187)	(0.0168)
Year Built			0.00260***
			(0.000122)
Sq. Feet			0.000123***
			(7.59e-06)
# of Stories			-0.119***
			(0.0104)
# of Baths			0.161***
			(0.00729)
Post	0.145***		
	(0.0130)		
Constant	12.09***	12.07***	6.673***
	(0.0109)	(0.0219)	(0.239)
Observations	10,541	10,541	10,295
R-squared	0.061	0.068	0.248
Month Fixed Effects	No	Yes	Yes
Year Fixed Effects	No	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

rooms, bathrooms and stories.

All specifications display strongly significant coefficients for the interaction terms capturing the DiD effect. The impact of the Blue Line in treatment neighborhoods is estimated to increase housing prices between 10.4 and 12.6%. These results should however be interpreted with some caution for the reasons discussed in Section 3.3.

3.6.2 Boosted Smooth Trees

Table 3.3 reports the estimation results using the Boosted Smooth Trees estimation routine with our proposed instruments. The Pre-Treatment column shows that the algorithm trained on the control group is able to correctly predict sale prices in the treatment group, with the prediction residuals for sale prices in the treatment group clustering around zero. After the introduction of the Blue Line, Post-Treatment prices in the treatment group increase by 5.5% as a direct effect of the Blue Line on property prices. This algorithm also allows us to directly approximate the spillover. To do this, we hold the level of amenities fixed at their pre-Blue Line levels and predict what housing prices would have been after it was introduced and compare these results to the predicted values post-introduction. This give us an approximation of $E[f(X_{it}^t) - f(X_{it}^c)]$, and thus the indirect effect. The change in amenities are predicted to increase the sale prices of properties located in the treatment group by a further 5.8%, implying that the total effect of the Blue Line on property prices is around 11.3%, remarkably close to the DiD prediction reported in Table 3.2. Following a similar procedure without instruments found a spillover of 1.3%, meaning that results that do not account for endogeneity would be downwardly biased and would tend to overstate the direct effect of the Blue Line relative to its indirect effect. With instruments, we find that the indirect effect accounts for over 51% of the total effect and is therefore an important channel through which public transportation impacts housing prices and consumer welfare.

These effects are not homogeneous and depend on where houses are located along the Blue Line. Figure 3.3 plots the direct treatment effect averaged across groups of houses along the path of the Blue Line. The direct treatment effect is lowest for the suburbs in Southern

Figure 3.3: Predicted Change in Housing Prices, Direct Effect

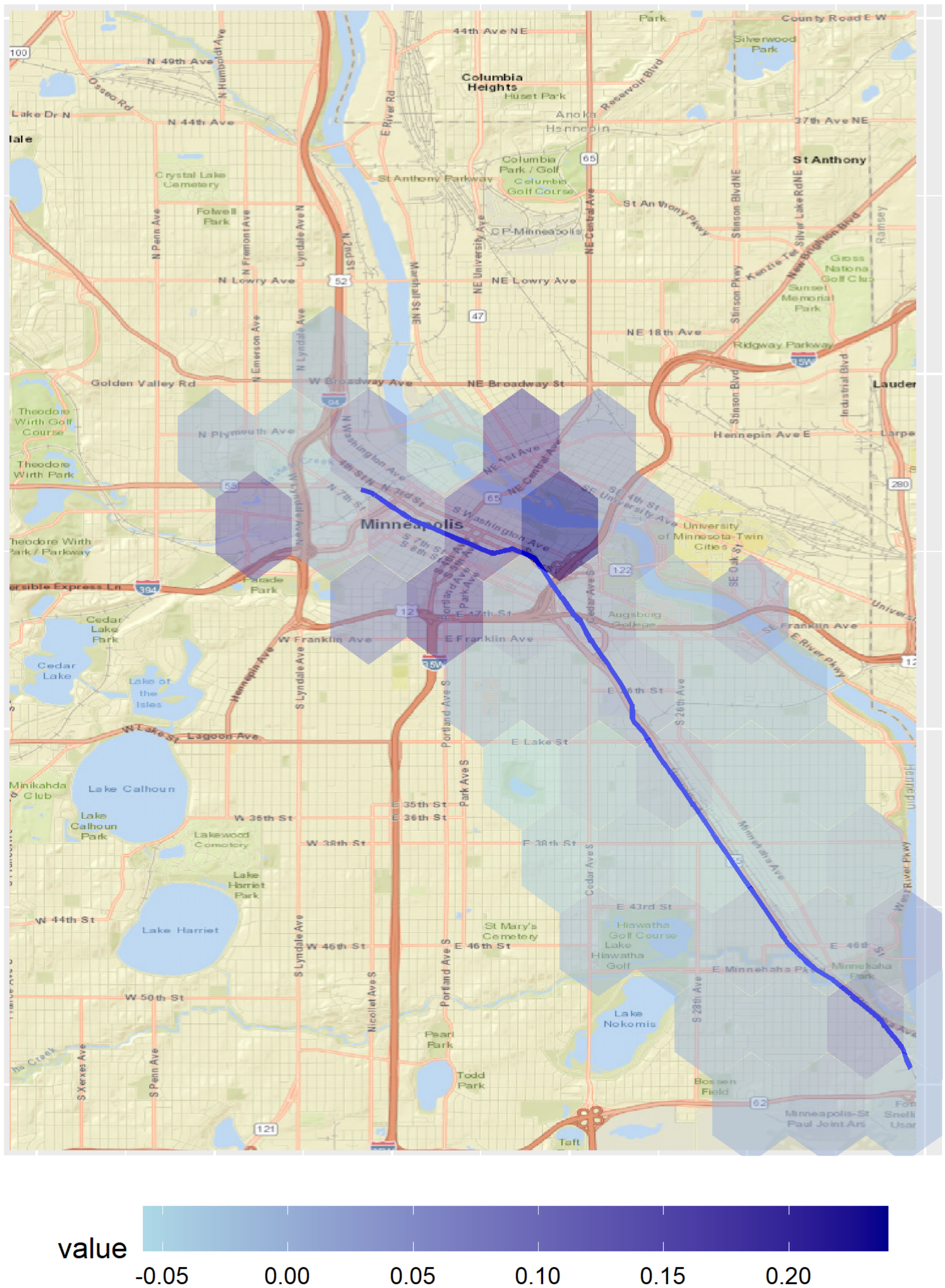


Figure 3.4: Predicted Change in Housing Prices, Indirect Effect

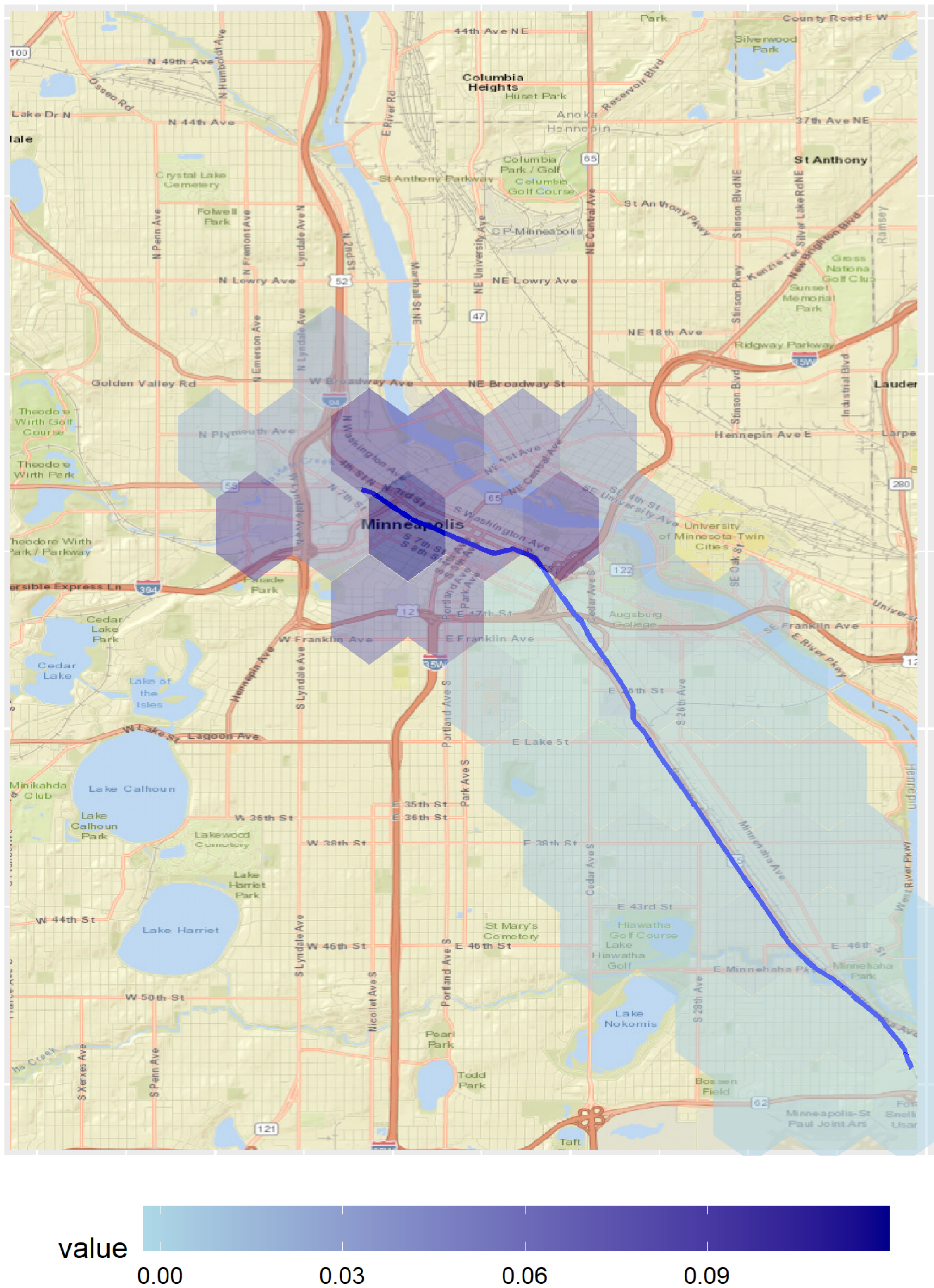


Table 3.3: Boosted Smooth Trees

Predicted Residual:	Pre-Treatment	Post-Treatment	Spillover
Mean	0.0009	0.0546	0.0584
Std Dev.	(0.0166)	(0.0198)	(0.0255)

Minneapolis and certain parts of the downtown area, meaning that houses located in these neighborhoods did not see much of a pricing effect after the Blue Line was introduced as a result of better access to public transportation. This could be due to the fact that the downtown area is already served by several bus lines and the fact that the southern suburbs have a higher rate of drivers, so there is less need for public transportation. Houses just outside of the city-center benefited the most. This includes houses in gentrifying neighborhoods such as East Phillips and Corcoran. These neighborhoods benefited from having additional direct transportation to downtown, while also seeing a significant boom in local businesses. The indirect treatment effect (Figure 3.4) is instead highest in the downtown area, which saw the largest increase in the entry of new amenities around the introduction of the Blue Line, while the the suburbs in Southern Minneapolis were relatively unaffected by this channel.

3.7 Conclusion

This paper applies recent advances in machine learning methods to investigate the impact that the construction of the METRO Blue Line had on housing prices and neighborhood amenities in Minneapolis. While many studies exist on the impact of mass transit on the urban environment, these studies generally do not decompose the overall impact of the introduction of a new mass transit system into direct and indirect effects. We apply a Boosted Smooth Tree learning algorithm to predict the direct and indirect effect of the introduction of the Blue Line. Our methodological contribution is a scalable algorithm for smooth tree boosting and a framework to incorporate instruments within this technique to control for endogeneity.

Our results show that that the price of properties located within 0.5 miles of a light rail station increased by around 11.3% compared to houses located further away. This can be thought of as the total impact of the Blue Line on local housing prices, encompassing both the direct benefit of improved access to public transit and the indirect benefit of an increase in the number neighborhood amenities. The direct impact of access to the light rail itself is estimated to increase local housing prices by 5.5%, while the spillover effect due to changes in amenities is quantifiable at 5.8%. Thus, just over half of the overall appreciation in housing prices following the introduction of the Blue Line is not due to residents MWTP for public transit but is rather a spillover effect attributable to an increase in the number of amenities around light rail stations.

Bibliography

Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American statistical Association 105*(490), 493–505.

Abadie, A. and J. Gardeazabal (2003). The economic costs of conflict: A case study of the basque country. *American economic review 93*(1), 113–132.

Ai, C., S. Martinez, and D. Sappington (2004). Incentive regulation and telecommunications service quality. *Journal of Regulatory Economics 26*(3), 263–285.

Ai, C. and D. Sappington (2002). The impact of state incentive regulation on the u.s. telecommunications industry. *Journal of Regulatory Economics 22*, 133–160.

Athey, S., J. Tibshirani, S. Wager, et al. (2019). Generalized random forests. *The Annals of Statistics 47*(2), 1148–1178.

Bae, C.-H. C., M.-J. Jun, and H. Park (2003). The impact of seoul's subway line 5 on residential property values. *Transport policy 10*(2), 85–94.

Bajari, P. and C. L. Benkard (2005). Demand estimation with heterogeneous consumers and unobserved product characteristics: A hedonic approach. *Journal of political economy 113*(6), 1239–1276.

Bajic, V. (1983). The effects of a new subway line on housing prices in metropolitan toronto. *Urban studies 20*(2), 147–158.

Baum-Snow, N. and M. E. Kahn (2000). The effects of new public projects to expand urban rail transit. *Journal of Public Economics 77*(2), 241–263.

Bekker, P. A. (1994). Alternative Approximations to the Distributions of Instrumental Variable Estimators. *Econometrica 62*(3), 657–681.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile Prices in Market Equilibrium. *Econometrica 63*(4), 841–890.

Berry, S., J. Levinsohn, and A. Pakes (1999). Voluntary export restraints on automobiles: Evaluating a trade policy. *American Economic Review*.

Berry, S., J. Levinsohn, and A. Pakes (2004). Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market. *Journal of Political Economy 112*(1), 68–105.

Berry, S. and J. Waldfogel (2010). Product quality and market size. *The Journal of Industrial Economics 58*(1), 1–31.

Billings, S. B. (2011). Estimating the value of a new transit option. *Regional Science and Urban Economics 41*(6), 525–536.

Blonigen, B. A., C. R. Knittel, and A. Soderbery (2013). Keeping it Fresh: Strategic Product Redesigns and Welfare.

Bottasso, A. and M. Conti (2009). Price cap regulation and the ratchet effect: a generalized index approach. *Journal of Productivity Analysis 32*(3), 191–201.

Bowes, D. R. and K. R. Ihlanfeldt (2001). Identifying the impacts of rail transit stations on residential property values. *Journal of Urban Economics 50*(1), 1–25.

Brumm, J. and S. Scheidegger (2017). Using adaptive sparse grids to solve high-dimensional dynamic models. *Econometrica 85*(5), 1575–1612.

Cervero, R. (1994). Rail transit and joint development: Land market impacts in washington, dc and atlanta. *Journal of the American Planning Association 60*(1), 83–94.

Cervero, R. and M. Duncan (2002). Land value impacts of rail transit services in los angeles county. *Report prepared for National Association of Realtors Urban Land Institute*.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics 34*(3), 305–334.

Clower, T. L., B. L. Weinstein, et al. (2002). The impact of dallas (texas) area rapid transit light rail stations on taxable property valuations. *Australasian Journal of Regional Studies, The 8*(3), 389.

Collard-Wexler, A. and J. De Loecker (2016). Production function estimation with measurement error in inputs. *National Bureau of Economic Research* (w22437).

Covert, T. and R. Kellog (2017). Crude by rail, option value, and pipeline investment. *NBER Working Paper Series* (23855).

Crawford, G. S. (2012). Endogenous product choice: A progress report. *International Journal of Industrial Organization 30*(3), 315–320.

Crawford, G. S. and M. Shum (2007). Monopoly Quality Degradation and Regulation in Cable Television. *The Journal of Law and Economics 50*(1), 181–219.

Damm, D., S. R. Lerman, E. Lerner-Lam, and J. Young (1980). Response of urban real estate values in anticipation of the washington metro. *Journal of Transport Economics and Policy*, 315–336.

De Loecker, J., J. Eeckhout, and G. Unger (2020). The rise of market power and the macroeonomic implications. *Quarterly Journal of Economics 135*(2), 561–644.

De Loecker, J. and F. Warzynski (2012). Markups and firm-level export status. *American Economic Review 102*(6), 2437–71.

Debrezion, G., E. Pels, and P. Rietveld (2011). The impact of rail transport on real estate prices: an empirical analysis of the dutch housing market. *Urban Studies 48*(5), 997–1015.

Dewees, D. N. (1976). The effect of a subway on residential property values in toronto. *Journal of Urban Economics 3*(4), 357–369.

Dhyne, E., A. Petrin, V. Smeets, and F. Warzynski (2020). Theory for extending single-product production function estimation to multi-product settings. *Working Paper*.

Dhyne, E., A. Petrin, and F. Warzynski (2020). Deregulation and investment spillovers in multi-product production settings. *Working Paper*.

Diamond, R. and T. McQuade (2019). Who wants affordable housing in their backyard? an equilibrium analysis of low-income property development. *Journal of Political Economy 127*(3), 1063–1117.

Domah, P. and M. Pollitt (2001). The restructuring and privatisation of electricity distribution and supply businesses in england and wales: A social cost–benefit analysis. *Fiscal Studies 22*(1), 107–146.

Dueker, K. J. and M. J. Bianco (1999). Light-rail-transit impacts in portland: The first ten years. *Transportation Research Record 1685*(1), 171–180.

Duncan, M. (2008). Comparing rail transit capitalization benefits for single-family and condominium units in san diego, california. *Transportation Research Record 2067*(1), 120–130.

Eaton, J. and S. Kortum (2002). Technology, geography, and trade. *Econometrica 70*(5), 1741–1779.

Eizenberg, A. (2014). Upstream Innovation and Product Variety in the United States Home PC Market. *The Review of Economic Studies 81*, 1003–1045.

Ericson, R. and A. Pakes (1994). Markov-perfect industry dynamics: A framework for empirical work. *The Review of Economic Studies 62*(1), 53–82.

Evans, L. and G. Guthrie (2012). Price-cap regulation and the scale and timing of investment. *The RAND Journal of Economics 43*(3), 537–561.

Fan, Y. (2013). Ownership consolidation and product characteristics: A study of the us daily newspaper market. *American Economic Review 103*(5), 1598–1628.

Fonseca, Y., M. Medeiros, G. Vasconcelos, and A. Veiga (2018). Boost: Boosting smooth trees for partial effect estimation in nonlinear regressions. *arXiv preprint arXiv:1808.03698*.

Forrest, D., J. Glen, and R. Ward (1996). The impact of a light rail system on the structure of house prices: a hedonic longitudinal study. *Journal of Transport Economics and Policy*, 15–29.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.

Frittelli, J., A. Andrews, P. Parfomak, R. Pirog, J. Ramseur, and M. Ratner (2014). U.s. rail transportation of crude oil: Background and issues for congress. *Congressional Research Service*.

Gandhi, A. and J.-F. Houde (2015). Measuring substitution patterns in differentiated products industries,. *University of Wisconsin Working Paper*.

Gatzlaff, D. H. and M. T. Smith (1993). The impact of the miami metrorail on the value of residences near station locations. *Land Economics*, 54–66.

Gibbons, S. and S. Machin (2005). Valuing rail access using transport innovations. *Journal of urban Economics 57*(1), 148–169.

Goetz, E. G., K. Ko, A. Hagar, H. Ton, and J. Matson (2010). The hiawatha line: impacts on land use and residential housing value.

Gowrisankaran, G. and M. Rysman (2012). Dynamics of consumer demand for new durable goods. *Journal of Political Economy 120*(6), 1173–1219.

Gramlich, J. (2009). *Gas Prices, Fuel Efficiency, and Endogenous Product Choice in the U.S. Automobile Industry.* Ph. D. thesis, Georgetown University.

Grass, R. G. (1992). The estimation of residential property values around transit station sites in washington, dc. *Journal of Economics and Finance 16*(2), 139–146.

Hall, R. (1988). The relation between price and marginal cost in u.s. industry. *Journal of Political Economy 96*(5), 921–947.

Hansen, C., J. Hausman, and W. Newey (2008). Estimation With Many Instrumental Variables. *Journal of Business & Economic Statistics 26*(4), 398–422.

Hansen, L. P. and K. J. Singleton (1982). Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models. *Econometrica 50*(5), 1269–1286.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction.* Springer Science & Business Media.

Hess, D. B. and T. M. Almeida (2007). Impact of proximity to light rail rapid transit on station-area property values in buffalo, new york. *Urban studies 44*(5-6), 1041–1068.

Hewitt, C. M. and W. Hewitt (2012). The effect of proximity to urban rail on housing prices in ottawa. *Journal of Public Transportation 15*(4), 3.

Ho, J. (2016). Machine learning for causal inference: An application to air quality impacts on house prices.

Hopenhayn, H. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica 60*(5), 1127–1150.

Hurst, N. B. and S. E. West (2014). Public transit and urban redevelopment: The effect of light rail transit on land use in minneapolis, minnesota. *Regional Science and Urban Economics 46*, 57–72.

Ihlanfeldt, K. R. (2003). Rail transit and neighborhood crime: the case of atlanta, georgia. *Southern Economic Journal*, 273–294.

Immergluck, D. (2009). Large redevelopment initiatives, housing values and gentrification: the case of the atlanta beltline. *Urban Studies 46*(8), 1723–1745.

Jenkins, C. (2004). Efficiency properties of price cap regulation. *Unpublished Thesis*.

Kilpatrick, J., R. Throupe, J. Carruthers, and A. Krause (2007). The impact of transit corridors on residential property values. *Journal of Real Estate Research 29*(3), 303–320.

Knaap, G. J., C. Ding, and L. D. Hopkins (2001). Do plans matter? the effects of light rail plans on land values in station areas. *Journal of Planning Education and Research 21*(1), 32–39.

Ko, K. and X. J. Cao (2013). The impact of hiawatha light rail on commercial and industrial property values in minneapolis. *Journal of Public Transportation 16*(1), 3.

Krusell, P. and A. Smith (1998). Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy 106*(5), 867–896.

Littlechild, S. (1983). Regulation of british telecommunications' profitability. *Department of Industry, HMSO*.

Lustig, J. (2008). *The Welfare Effects of Adverse Selection in Privatized Medicare.* Ph. D. thesis, UC Berkeley.

Majumdar, S. (2016). Pricing regulations and network technology investments: A retrospective evaluation. *Annals of Public and Cooperative Economics 87*(1), 23–53.

Mazzeo, M. J. (2002). Product choice and oligopoly market structure. *RAND Journal of Economics 33*(2), 221–242.

McDonald, J. F. and C. I. Osuji (1995). The effect of anticipated transportation improvement on residential land values. *Regional science and urban economics 25*(3), 261–278.

McMillen, D. P. and J. McDonald (2004). Reaction of house prices to a new rapid transit line: Chicago's midway line, 1983–1999. *Real Estate Economics 32*(3), 463–486.

McRae, S. (2017). Crude oil price differentials and pipeline infrastructure. *NBER Working Paper Series* (24170).

Melitz, M. and S. Polanec (2015). Dynamic olley-pakes productivity decomposition with entry and exit. *The RAND Journal of Economics 46*(2), 362–375.

Miravete, E. J., M. J. Moral, and J. Thurk (2015). Innovation, Emissions Policy, and Competitive Advantage in the Diffusion of European Diesel Automobiles. *Working Paper*.

Murarasu, A., J. Weidendorfer, G. Buse, D. Butnaru, and D. Pflüger (2011). Compact data structure and scalable algorithms for the sparse grid technique. *ACM SIGPLAN Notices 46*(8), 25–34.

Nelson, A. C., D. Eskic, S. Hamidi, S. J. Petheram, R. Ewing, and J. H. Liu (2015). Office rent premiums with respect to light rail transit stations: Case study of dallas, texas, with implications for planning of transit-oriented development. *Transportation Research Record 2500*(1), 110–115.

Nevo, A. (2004). A practitioner's guide to estimation of random-coefficients logit models of demand,. *Journal of Economics and Management Strategy 9*(4), 513–548.

Newey, W. K. and R. J. Smith (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica 72*(1), 219–255.

Olley, S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica 64*(6), 1263–1297.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2015). Moment Inequalities and Their Application. *Econometrica 83*(1), 315–334.

Palmquist, R. B. (2005). Property value models. *Handbook of environmental economics 2*, 763–819.

Pan, H. and M. Zhang (2008). Rail transit impacts on land use: Evidence from shanghai, china. *Transportation Research Record 2048*(1), 16–25.

Pan, Q. (2013). The impacts of an urban light rail system on residential property values: a case study of the houston metrorail transit line. *Transportation Planning and Technology 36*(2), 145–169.

Petrin, A. (2002). Quantifying the Benefits of New Products: The Case of the Minivan. *Journal of Political Economy* (110), 705–729.

Pilgram, C. A. and S. E. West (2018). Fading premiums: The effect of light rail on residential property values in minneapolis, minnesota. *Regional Science and Urban Economics 69*, 1–10.

Pint, E. (1992). Price-cap versus rate-of-return regulation in a stochastic-cost model. *The RAND Journal of Economics 23*(4), 564–578.

Ponder, M. (2021). Production functions without control functions: Estimating input elasticities when monotonicity does not hold. *Working Paper*.

Portnov, B., B. Genkin, and B. Barzilay (2009). Investigating the effect of train proximity on apartment prices: Haifa, israel as a case study. *Journal of Real Estate Research 31*(4), 371–395.

Rewers, J. M. (2010). *Identifying the impacts of light rail station location on residential property values in the city of Sacramento*. Ph. D. thesis.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy 82*(1), 34–55.

Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica 55*(5), 999–1033.

Ryan, S. (2012). The costs of environmental regulation in a concentrated industry. *Econometrica 80*(3), 1019–1061.

Sappington, D. (2003). The effects of incentive regulation on retail telephone service quality in the united states. *Review of Network Economics 2*(4), 355–375.

Seo, K., A. Golub, and M. Kuby (2014). Combined impacts of highways and light rail transit on residential property values: a spatial hedonic price model for phoenix, arizona. *Journal of Transport Geography 41*, 53–62.

Spence, M. (1976). Product Selection, Fixed Costs, and Monopolistic Competition. *The Review of Economic Studies 43*(2), 217–235.

Sweeting, A. (2007). Dynamic Product Repositioning in Differentiated Product Markets: The Case of Format Switching in the Commercial Radio Industry.

Taylor, L. O. (2003). The hedonic method. In *A primer on nonmarket valuation*, pp. 331–393. Springer.

Trajtenberg, M. (1989). The Welfare Analysis of Product Innovations, with an Application to Computed Tomography Scanners. *Journal of Political Economy 94*, 444–479.

Tsivanidis, N. (2018). The aggregate and distributional effects of urban transit infrastructure: Evidence from bogotá's transmilenio. *Job Market Paper*.

Varian, H. R. (2014, May). Big data: New tricks for econometrics. *Journal of Economic Perspectives 28*(2), 3–28.

Veiga, A. and E. G. Weyl (2014). Product Design in Selection Markets.

Wagner, G. A., T. Komarek, and J. Martin (2017). Is the light rail "tide" lifting property values? evidence from hampton roads, va. *Regional Science and Urban Economics 65*, 25–37.

Weinberger, R. R. (2001). Light rail proximity: Benefit or detriment in the case of santa clara county, california? *Transportation Research Record 1747*(1), 104–113.

Zhang, A. (2020). Estimation of high-dimensional dynamic games: Fintech lenders and bank branch closures. *Working Paper*.

Zheng, S., Y. Xu, X. Zhang, and R. Wang (2016). Transit development, consumer amenities and home values: Evidence from beijing's subway neighborhoods. *Journal of Housing Economics 33*, 22–33.

Appendix A

Appendix to Chapter 1

A.1 Additional Figures

Figure A.1: Petroleum Administration for Defense Districts (PADD)

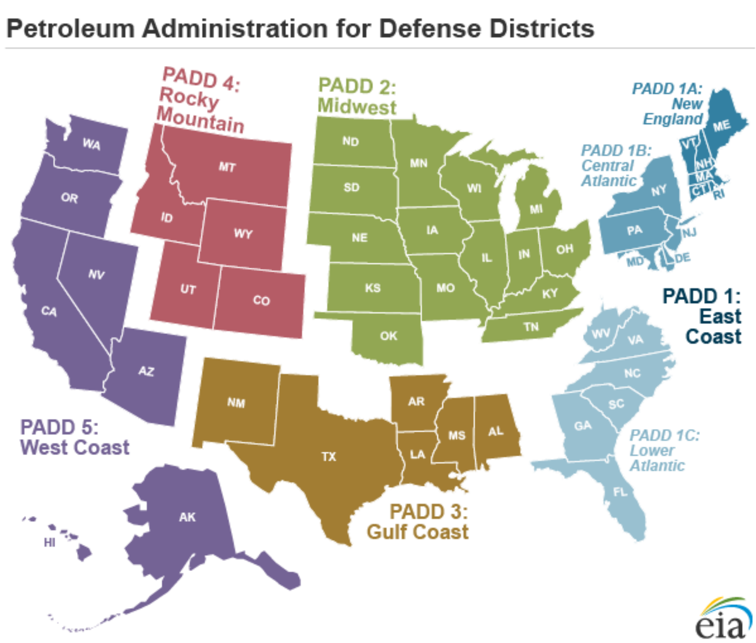


Figure A.2: Average Unit Operating Cost

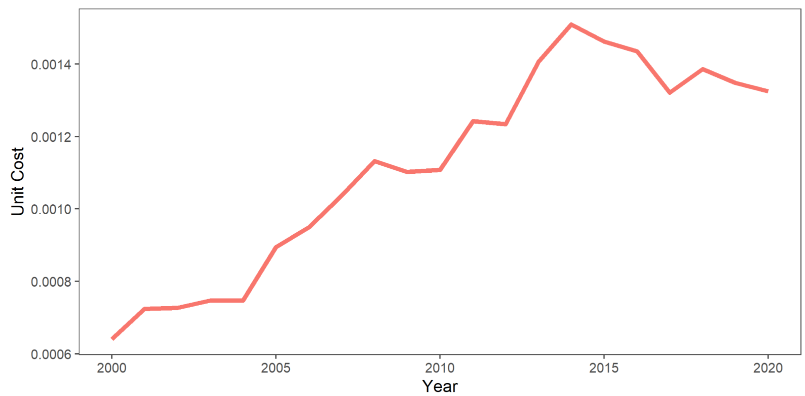


Figure A.3: Variable Input Elasticity

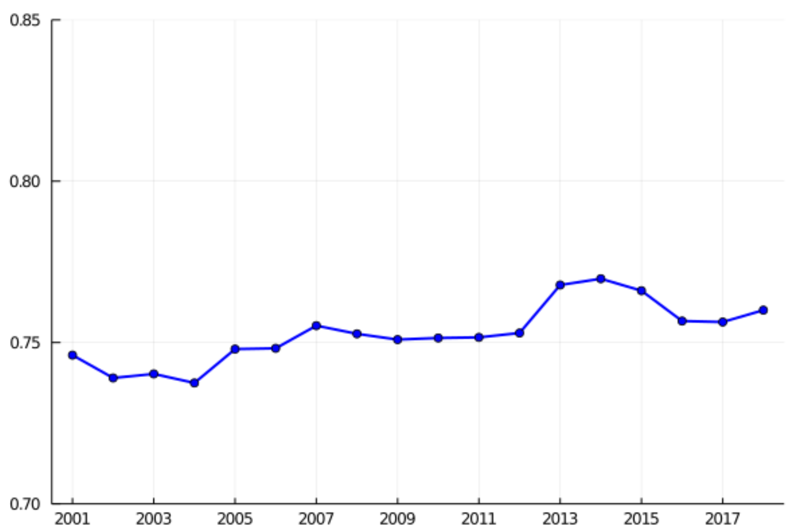


Figure A.4: Weighted Average Cost of Capital

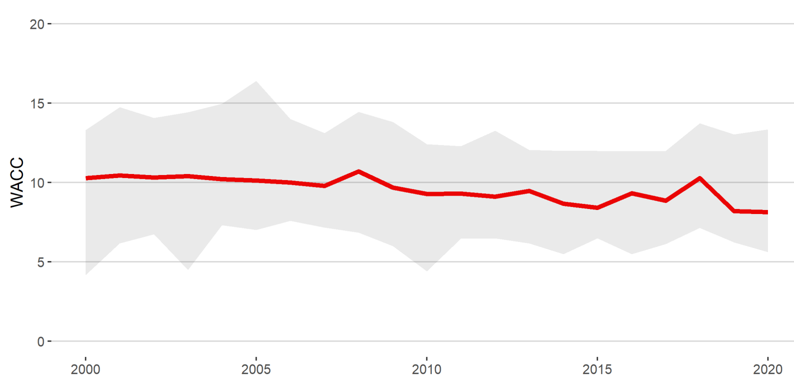
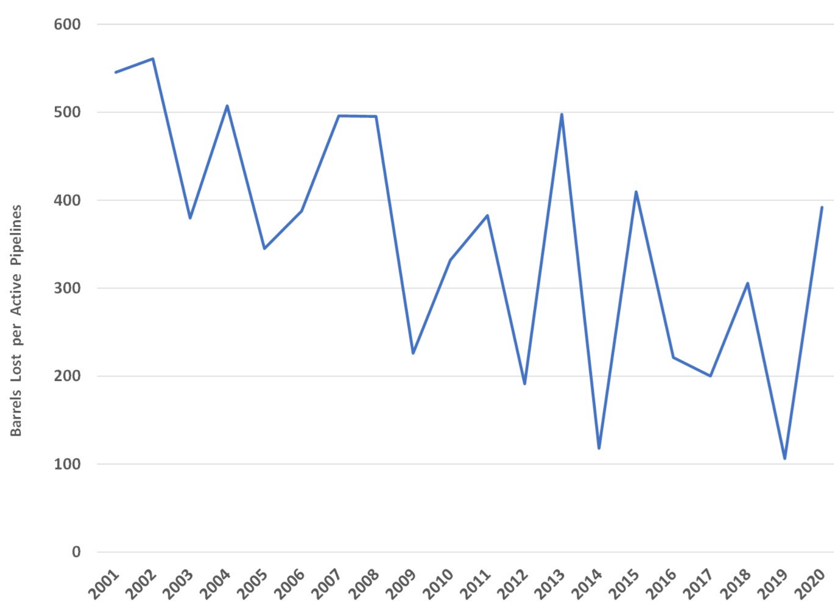


Figure A.5: Barrels Lost per Active Pipelines



A.2 Robustness

A.2.1 Error in Capital

We can extend this method to handle measurement error in the capital stock in a manner similar to that of Collard-Wexler and De Loecker (2016). They assume that the measurement error is uncorrelated with the true capital stock and that we see

$$k_{jt} = k^*_{jt} + u_{jt}$$

If we strengthen this to statistically independent, then we can follow an approach analogous to before, this time instrumenting for capital with investment, lags of capital, or other macro-indicators that are uncorrelated with the error term. Note that this relies on the production function being linear in parameters. In this way, $E[k_{jt}|Z]$ and $E[k^*_{jt}|Z]$ span the same space (when including a constant), so we consistently recover the parameter β_k. Additional restrictions must be placed on u_{jt} to consistently estimate a translog production function. The most natural would be to assume that u_{jt} is mean zero. Then following the previous logic

$$\begin{aligned}
E[k^2_{jt}|Z] &= E[k^{*2}_{jt}|Z] + E[k_{jt}u_{ij}|Z] + E[u^2_{jt}|Z] \\
&= E[k^{*2}_{jt}|Z] + 2E[k^*_{jt}|Z]E[u_{ij}] + E[u^2_{jt}|Z] \\
&= E[k^{*2}_{jt}|Z] + E[u^2_{jt}]
\end{aligned}$$

Note that the variation in $E[k^2_{jt}|Z]$ is driven entirely by the term $E[k^{*2}_{jt}|Z]$, meaning that after conditioning on a constant we will get consistent estimates for all model parameters.

This result is interesting in that we can consistently estimate the parameters of the Cobb-Douglas and translog production functions under some natural assumptions on the measurement error. However, these assumptions are still strong. For instance, the measurement error in the capital stock arises from how we calculate the proxy. Therefore, it is unlikely that the measurement error will truly be independent of the true stock. Further, even if the measurement error is independent of the capital stock, there is little reason to believe it will be mean zero. In most regression models, the mean zero assumption is innocuous because we can always add a constant to control for a non-zero mean. Here, however, this assumption is used to separate out the variation of $E[k^2_{jt}|Z]$ and $E[k_{jt}|Z]$, and is therefore necessary for identification.

While this approach appears to be more general than what I use in the text, it has one serious drawback. There is no way to separate out u_{it} from k^*_{jt}. This means that all of the elasticity estimates will also be measured with error. As such, I do not use the results as a starting point in the analysis. However, Table A.1 presents the results for a translog assuming that capital has been measured with error. I no longer use k_t as an instrument for itself, but instead using k_{t-1}. Additionally, I add in the price cap index as an instrument. Column (1) reproduces the baseline estimates from the text, while column (2) presents the results of this alternative procedure. Both input elasticities are estimated to be higher on average in the baseline specification, and the returns-to-scale is also larger. Most significantly, the estimated elasticities are all positive. In the baseline model (and for the other methods that I used), a few pipelines were estimated to have negative input elasticities. This problem disappeared when instrumenting for capital.

Table A.1: Error-in-Capital Estimates

	Baseline (1)	Error-in-Capital (2)
β_v	0.470	0.752
	(0.070)	(0.147)
β_k	0.242	0.415
	(0.029)	(0.07)
β_{v^2}	0.024	0.045
	(0.015)	(0.011)
β_{k^2}	0.031	0.023
	(0.007)	(0.011)
β_{vk}	0.023	-0.064
	(0.004)	(0.029)
Avg. Capex Elast.	0.730	0.642
Avg. Opex Elast.	0.528	0.492
Local RTS	1.258	1.134
Observations	2,863	2,863

A.2.2 Different Measures of Output

As mentioned in the main text, barrel-miles is not the only potential dependent variable. We can instead use either barrels or deflated revenue. Using barrels runs the risk of using outputs with a different “quality", by which I mean two barrels of oil traveling different distances have a different inherent value. Deflated revenue is consistently used in the literature when measures of physical output are not available. To check the robustness of my results and to have a point of comparison with the literature, I estimate the model using each measure separately. Because I have quarterly data available for revenue and barrels, I use this for estimation. Unfortunately, I do not have the line item data for capital at the quarterly basis, so I use Net Carrier Property as a proxy. Estimation using annual data generates similar results.

Column (2) shows the results using barrels rather than barrel-miles. The most striking difference is the implied returns to scale when using barrels. Rather than being increasing returns to scale, the estimated input elasticities imply a decreasing returns to scale technology. This makes intuitive sense in that larger pipelines tend not to produce more barrels but instead transport barrels over a greater distance. Therefore, we see that increases in capital tend to lead to marginal changes in output, measured in barrels. This has the effect of making the capital elasticity very nearly zero. The variable input responds more readily to changes in throughput, but it is still significantly attenuated. This demonstrates the importance of using quality adjusted output during estimation.

Column (3) shows the results using deflated revenue. The model predicts that the input elasticity is declining in capital at low levels and then increasing in capital at higher levels. The variable elasticity is strongly decreasing in the level of capital. Combined, these results imply that the standard deviation of the input elasticities are over twice as large. Additionally, 19% of the observations have a negative capital elasticity. The production function is estimated to have constant returns to scale. The lower estimated returns to scale makes sense as firms with higher throughput tend to charge lower prices. So doubling output will increase revenue by less than a factor of two.

Table A.2: Estimates for Alternative Dependent Variables

	Barrel-Miles (1)	Barrels (2)	Deflated Revenue (3)
β_v	0.470	0.176	1.673
	(0.070)	(0.404)	(0.134)
β_k	0.242	-0.233	-0.724
	(0.029)	(0.157)	(0.059)
β_{v^2}	0.024	0.016	0.061
	(0.015)	(0.03)	(0.007)
β_{k^2}	0.014	0.032	0.137
	(0.013)	(0.03)	(0.007)
β_{vk}	0.023	0.048	-0.233
	(0.004)	(0.063)	(0.023)
Avg. Opex Elast.	0.730	0.407	0.761
Avg. Capex Elast.	0.528	0.074	0.242
Local RTS	1.258	0.481	1.003
Observations	2,863	7,404	6,999

Appendix B

Appendix to Chapter 2

B.1 Proof of Proposition 1

We start by observing that linearity of β implies that the derivative of market shares can be written as an affine function of β, conditional on θ_{nl} and data. The derivative of market shares with respect to the product characteristics are given by

$$\begin{aligned}
\frac{ds_j(\delta, X, p, \theta_{nl})}{dX_k'} &= \frac{\partial s_j}{\partial \delta'} \frac{\partial \delta}{\partial X_k'} + \frac{\partial s_j(\delta, X, p, \theta_{nl})}{\partial X_k'} \\
&= \frac{\partial s_j}{\partial \delta'}(I \circ \beta_k) + \frac{\partial s_j(\delta, X, p, \theta_{nl})}{\partial X_k'} \\
&= g_1(\delta, X, p, \theta_{nl}) + g_2(\delta, X, p, \theta_{nl})\beta_k
\end{aligned}$$

for functions g_1 and g_2 that do not depend on θ_l, conditional on δ. Additionally,

$$\begin{aligned}
\frac{ds_j(\delta, X, p, \theta_{nl})}{dp'} &= \frac{\partial s_j(\delta, X, p, \theta_{nl})}{\partial p'} \\
&= g_p(\delta, X, p, \theta_{nl})
\end{aligned} \tag{B.1}$$

We use the total derivative to emphasize that X_k impacts shares both directly and through its impact on δ. Similar expressions are straightforward to derive for $\frac{\partial^2 s_j}{\partial p \partial p'}$ and $\frac{\partial^2 s_j}{\partial p \partial X_k'}$.
We focus on each term ν_k in (2.7) in turn. Market shares, $s_{j'}$, are set equal to their observed values during estimation and so can be treated as data. Markups are inferred from the data and θ_{nl} by $(p - \mathrm{mc}) = \Delta^{-1}s$, where

$$\Delta_{ij} = \begin{cases} -\frac{\partial s_j}{\partial p_j'}, & \text{if } j, j' \in J_f \\ 0 & \text{otherwise} \end{cases}$$

Because the derivative of $s_{j'}$ with respect to price does not directly depend on θ_l, Δ and markups are a function of θ_{nl} and observed data alone. By assumption, the derivative of marginal cost with respective to X_k is affine in θ_l. We have assumed that $\frac{\partial p}{\partial X'_{jk}}$ is linear in θ_l, so all that remains is are the terms $\frac{ds_{j'}}{dX_{jk}}$. Again, we can write these out as

$$\frac{ds_j}{dX_{jk}} = \frac{\partial s_j}{\partial X_{jk}} + \sum_{j' \in J} \frac{\partial s_j}{\partial p_{j'}} \frac{\partial p_{j'}}{\partial X_{jk}}$$

so that the total derivative is given as

$$\frac{ds_j}{dX_{jk}} = g_1(\delta, X, p, \theta_{nl}) + g_2(\delta, X, p, \theta_{nl})\beta_k + g_p(\delta, X, p, \theta_{nl})'\tilde{H}_{jk}$$

where, by assumption, $\frac{\partial p}{\partial X'_{jk}} = h_1(\delta, X, p, \theta_{nl}) + h_2(\delta, X, p, \theta_{nl})\theta_l$ for some matrices h_1 and h_2. The parameter β_k is an element of θ_l, and so this reduces to

$$\frac{ds_j}{dX_{jk}} = \tilde{d}_1(\delta, X, p, \theta_{nl}) + \tilde{d}_2(\delta, X, p, \theta_{nl})\theta_l$$

for some matrices $\tilde{d}_1$ and $\tilde{d}_2$. This establishes that all of the terms in ν_{jk} are nonlinear functions of θ_{nl} and either do not depend on θ_l or are linear in θ_l. Plugging these terms back into the residual equation yields

$$\nu_{jk} = \sum_{j' \in J_f} \left(\left(h_1(\delta, X, p, \theta_{nl}) + h_2(\delta, X, p, \theta_{nl})\theta_l - \frac{\partial \mathrm{mc}_j}{\partial X_{jk}} \right) s_{j'} + \left(\tilde{d}_1 + \tilde{d}_2\theta_l \right) (p_{j'} - \mathrm{mc}_{j'}) \right) \tag{B.2}$$

Pulling out the components of θ_l, we end up with an expression of the form

$$\nu_k = \nu_{kc}(\delta, X, p, \theta_{nl}) + \nu_{k\theta}(\delta, X, p, \theta_{nl})\theta_l$$

which completes the proof.

B.2 Proof of Proposition 2

Let R be the residual equation for the first-order condition with respect to price. That is,

$$R = s + \left(T \circ \frac{\partial s}{\partial p'} \right) (p - \mathrm{mc})$$

so that the derivative of price with respect to characteristic X_{jk} is given by (2.8). We show $\frac{\partial R}{\partial p'}$ does not depend on θ_l and that $\frac{\partial R}{\partial X'_{jk}}$ is affine in θ_l. The derivative of the residual equation with respect to price

$$\frac{\partial R}{\partial p'} = \frac{\partial s}{\partial p'} + \sum_j \frac{\partial}{\partial p'}(p_j - \mathrm{mc}_j)\left(T_j \circ \frac{\partial s_j}{\partial p'}\right) \tag{B.3}$$

$$= \frac{\partial s}{\partial p'} + \sum_j \left(e_j\left(T_j \circ \frac{\partial s_j}{\partial p'}\right) + (p_j - \mathrm{mc}_j)\left(T_j \circ \frac{\partial^2 s_j}{\partial p \partial p'}\right)\right) \tag{B.4}$$

where e_j is the jth column of the identity matrix. We previously noted that $\frac{\partial s}{\partial p'}$ does not depend on θ_l by equation (B.1) and that $(p_j - \mathrm{mc}_j)$ does not depend on θ_l as $(p - \mathrm{mc}) = \Delta^{-1}s$, where Δ is defined as before. The second derivative $\frac{\partial^2 s_j}{\partial p \partial p'}$ does not depend directly on θ_l, so $\frac{\partial R}{\partial p'}$ is a nonlinear function of θ_{nl} and the data, and independent of θ_l. Taken together, this means that θ_l does not appear in $\frac{\partial R}{\partial p'}$.

Turning to $\frac{\partial R}{\partial X'_k}$, we can write this derivative as

$$\begin{aligned}\frac{\partial R}{\partial X'_k} &= \frac{\partial s}{\partial X'_k} + \sum_j \frac{\partial}{\partial X'_k}(p_j - \mathrm{mc}_j)\left(T_j \circ \frac{\partial s_j}{\partial p'}\right)\\ &= \frac{\partial s}{\partial X'_k} + \sum_j \left(-\frac{\partial \mathrm{mc}_j}{\partial X'_k}\left(T_j \circ \frac{\partial s_j}{\partial p'}\right) + (p_j - \mathrm{mc}_j)\left(T_j \circ \frac{\partial^2 s_j}{\partial p X'_k}\right)\right)\end{aligned}$$

By assumption, $\frac{\partial \mathrm{mc}_j}{\partial X'_k}$ is affine in θ_l. We have previously shown that the terms $\frac{\partial s}{\partial X'_k}$ and $\frac{\partial^2 s_j}{\partial p X'_k}$ are affine in θ_l and that the values of $\frac{\partial s_j}{\partial p'}$ and $(p_j - \mathrm{mc}_j)$ do not depend on θ_l. All terms that are affine in θ_l multiply terms that are fixed in θ_l. As such, the resulting expression will be affine in θ_l and that establishes the result.

B.3 Application to Mixed Logit Models

Assume the random coefficients are normally distributed and let s_i denote the vector of choice probabilities. Then the derivative of market shares with respect to characteristic k is given by

$$\begin{aligned}\frac{ds(\delta, X, p, \theta_{nl})}{dX'_k} &= \frac{1}{ns}\sum_{i=1}^{ns}(\beta_k + \sigma_k \eta_{ik})\left(s_i \circ I - s_i s'_i\right)\\ &= \frac{1}{ns}\sum_{i=1}^{ns}\sigma_k \eta_{ik}\left(s_i \circ I - s_i s'_i\right) + \frac{1}{ns}\sum_{i=1}^{ns}\left(s_i \circ I - s_i s'_i\right)\beta_k\\ &= g_1(\delta, X, p, \theta_{nl}) + g_2(\delta, X, p, \theta_{nl})\beta_k\end{aligned}$$

with an analogous expression for $\frac{\partial s}{\partial p'}$.

We now provide the expressions for the residuals when demand has a mixed logit specification. Let P denote the matrix of choice probabilities, which is a $J \times n_s$ matrix. Next, define the following $n_s \times 1$ weight vectors $w_\alpha = \left\{-\frac{\alpha}{y_i}\frac{1}{n_s}\right\}_{i=1}^{n_s}$, $w_{\alpha,2} = \left\{\left(\frac{\alpha}{y_i}\right)^2 \frac{1}{n_s}\right\}_{i=1}^{n_s}$, $w_\sigma = \left\{\sigma_k \eta_{ik} \frac{1}{n_s}\right\}_{i=1}^{n_s}$, and $w_{\alpha,\sigma} = \left\{-\left(\frac{\alpha}{y_i}\right)\sigma_k \eta_{ik}\frac{1}{n_s}\right\}_{i=1}^{n_s}$. We denote the $J \times n_s$ weighted matrix of choice probabilities by $P_{w_j} = P \circ w_j'$, for each w_j above. Finally, define the $J \times n_s$ matrix $\zeta = P \circ [T(P \circ (p - \text{mc}))]$. Recall that the residuals are given by

$$\nu_k = \left(\frac{\partial(p-\text{mc})}{\partial X_k'} \circ T\right) s + \left(\frac{ds}{dX_k'} \circ T\right)(p - \text{mc}),\ \ k = 1, ..., K,$$

The first term is given by

$$\begin{aligned}\left(\frac{\partial(p-\text{mc})}{\partial X_k'} \circ T\right) s = & \underbrace{-\left(T \circ G^{-1} H_{k,c}\right) s}_{\nu_{k,c,1}} \\ & \underbrace{-\left(T \circ G^{-1} H_{k,\beta}\right) s\, \beta_k}_{\nu_{k,\beta,1}} \\ & \underbrace{-\left(T \circ \left(\frac{\text{mc}}{X_k} \circ I\right)\right) s\, \gamma_k}_{\nu_{k,\gamma,1}}\end{aligned}$$

and the second term is given by

$$\begin{aligned}\left(\frac{ds}{dX_k'} \circ T\right)(p-\text{mc}) = & \underbrace{\left(T \circ \left(P w_\sigma \circ I - P P_{w_\sigma}' - \frac{\partial s}{\partial p'} G^{-1} H_{k,c}\right)\right)(p-\text{mc})}_{\nu_{k,c,2}} + \\ & \underbrace{\left(T \circ \left(s \circ I - \frac{1}{n_s} P P' - \frac{\partial s}{\partial p'} G^{-1} H_{k,\beta}\right)\right)(p-\text{mc})\, \beta_k}_{\nu_{k,\beta,2}} - \\ & \underbrace{\left(T \circ \frac{\partial s}{\partial p'} G^{-1} H_{k,\gamma}\right)(p-\text{mc})\, \gamma_k}_{\nu_{k,\gamma,2}}\end{aligned}$$

where $G = \frac{\partial R'}{\partial p}$ and $H_k = \frac{\partial R}{\partial X_k'} = H_{k,c} + H_{k,\beta}\beta_k + H_{k,\gamma}\gamma_k$, with their terms are defined below. We then the expressions for $\nu_{k,\theta_l} = (\nu_{k,\beta}, \nu_{k,\gamma})$ and $\nu_{k,c}$ are given by

$$\begin{aligned}\nu_{k,\beta} &= \nu_{k,\beta,1} + \nu_{k,\beta,2} \\ \nu_{k,\gamma} &= \nu_{k,\gamma,1} + \nu_{k,\gamma,2} \\ \nu_{k,c} &= \nu_{k,c,1} + \nu_{k,c,2}\end{aligned}$$

Finally, after some tedious algebra, the expressions for G and H_k can be shown to be given by

$$
\begin{aligned}
G = &\left((Pw_\alpha \circ I) - PP'_{w_\alpha}\right) + \left((Pw_{\alpha,2} \circ I) - PP'_{w_{\alpha,2}}\right) \circ (p - \mathrm{mc}) + (\zeta w_{\alpha,2}) \circ I + \\
&\left(PP'_{w_{\alpha,2}}\right) \circ T \circ (p - \mathrm{mc})' - 2\zeta P'_{w_{\alpha,2}} + \frac{\partial s}{\partial p'} \circ T,
\end{aligned}
$$

and

$$
\begin{aligned}
H_{k\mathrm{c}} = &\left((Pw_\sigma \circ I) - PP'_{w_\sigma}\right) + \left((Pw_{\alpha,\sigma} \circ I) - PP'_{w_{\alpha,\sigma}}\right) \circ (p - \mathrm{mc}) + (\zeta w_{\alpha,\sigma}) \circ I + \\
&\left(PP'_{w_{\alpha,\sigma}}\right) \circ T \circ (p - \mathrm{mc})' - 2\zeta P'_{w_{\alpha,\sigma}} \\
H_{k\beta} = &\left((s \circ I) - \frac{1}{n_s} PP'\right) + \left((Pw_\alpha \circ I) - PP'_{w_\alpha}\right) \circ (p - \mathrm{mc}) + \zeta w_\alpha \circ I + \\
&\left(PP'_{w_\alpha}\right) \circ T \circ (p - \mathrm{mc})' - 2\zeta P'_{w_\alpha} \\
H_{k\gamma} = &- \frac{\partial s}{\partial p'} \circ T \circ \frac{\mathrm{mc}}{X_k},
\end{aligned}
$$

Appendix C

Appendix to Chapter 3

C.1 Data Appendix

C.1.1 Transit Data

The full set of transit data includes information on the location of 5,518 transit stops across 147 transit routes. The closest stop to each of the downtown and local routes (route numbers 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 14, 17, 18, 19, 21, 22, 23, 25, 27, 30, 32, 39, 46, 53, 59, 133, 134, 135, 141, 146, 156) was calculated for each property in Minneapolis.

In order to reduce the dimensionality of the data, the closest stop along each of the major axes connecting Minneapolis to its suburbs was calculated as illustrated in Table C.1, based on information obtained from the Twin Cities Metro Transit.

Table C.1: Summarized Transit Variables

Variable	Direction	Route Number
Closest 200 Route	Roseville, White Bear Lake	250, 252, 261, 263, 264, 270, 272, 288.
Closest 300 Route	Woodbury	353, 355, 365, 375.
Closest 400 Route	Eagan	436, 446, 452, 460, 464, 465, 467, 470, 472, 475, 476, 477, 478, 479, 490, 491.
Closest 500 Route	Edina, Bloomington	515, 535, 552, 553, 554, 558, 578, 579, 587, 588, 589, 597.
Closest 600 Route	Minnetonka, Eden Prairie	600, 612, 643, 645, 652, 663, 664, 667, 668, 670, 671, 672, 673, 677, 679, 690, 695, 697, 698, 699.
Closest 700 Route	Plymouth, Maple Grove, Brooklin Park	721, 724, 742, 747, 755, 756, 758, 760, 761, 762, 763, 764, 765, 766, 767, 768.
Closest 800 Route	Coon Rapids	824, 825, 850, 852, 854, 865, 887, 888.
Closest UofM Route	University of Minnesota	111, 113, 114, 115, 118, 121, 122, 129.
Closest Route to St Paul	St Paul	61, 67, 74, 94.

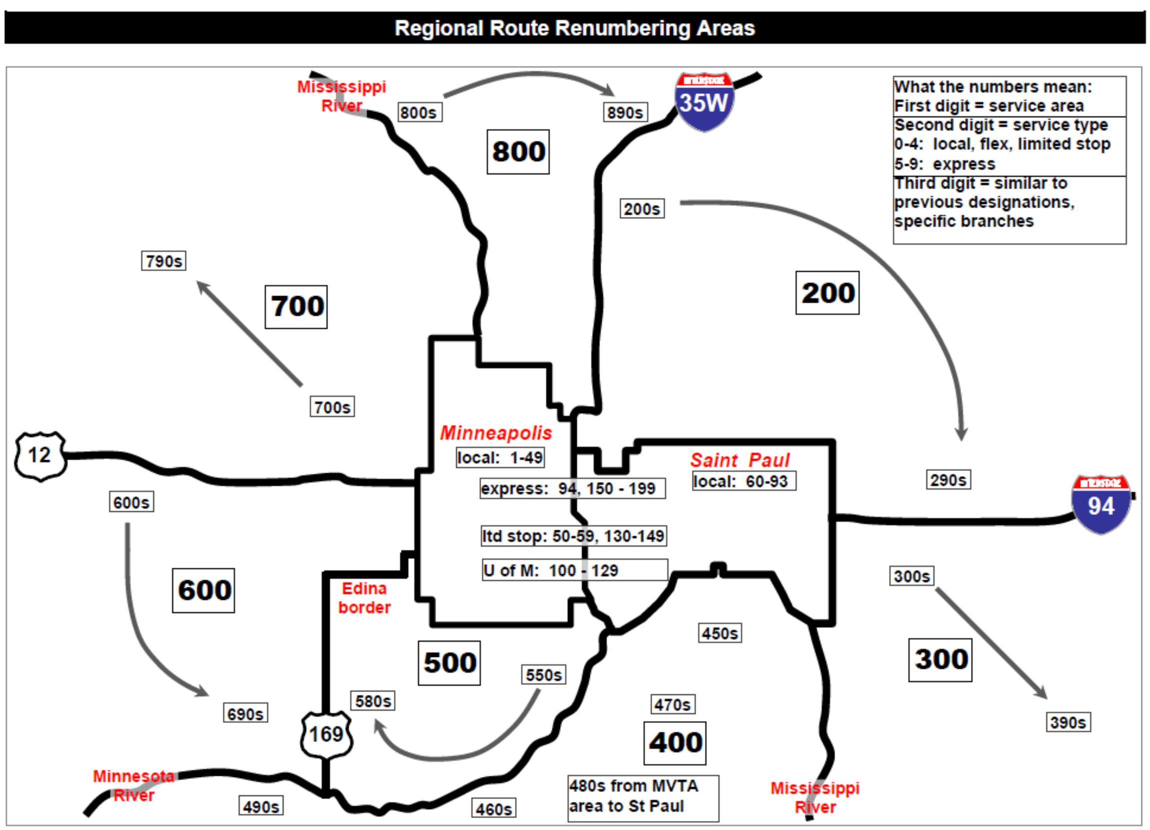
C.1.2 Amenities Data

Table C.2 reports the NAICS codes and the average number of observations per year for the amenity categories included in our analysis obtained from ReferenceUSA.

C.1.3 Population Changes by Neighborhood

C.1.4 Results Using XGBoost

To recover the direct effect of the Blue Line, we estimated the conditional mean of the pricing surface using gradient boosting with decision trees. Following our difference-in-differences specification, we separated the sample into treatment and control groups using concentric circles with a radius of 0.5 miles and 1 mile respectively, training the model on the control group. We used a 10% hold-out sample and cross-validated the model by searching for the set of parameters that minimized the out-of-sample mean-squared error. The parameters we searched over included the maximum tree depth, the number of iterations, the rate of



Source: Twin Cities Metro Transit

Table C.2: NAICS Codes of Neighborhood Amenities Variables

NAICS Category	Avg. Obs per Year	Description
22, 562	1,948	Utilities and Waste Management and Remediation Services
442	4,838	Furniture and Home Furnishings Stores
443	5,662	Electronics and Appliance Stores
444	7,271	Building Material and Garden Equipment and Supplies Dealers
445	13,065	Food and Beverage Stores
446	6,713	Health and Personal Care Stores
447	2,830	Gasoline Stations
448	10,931	Clothing and Clothing Accessories Stores
451	7,061	Sporting Goods, Hobby, Musical Instrument, and Book Stores
452, 453	18,007	General Merchandise Stores, Miscellaneous Store Retailers
481	532	Air Transportation
541	162,156	Professional, Scientific, and Technical Services
55	888	Management of Companies and Enterprises
61	17,909	Educational Services
621	130,663	Ambulatory Health Care Services
622	1,496	Hospitals
623	4,231	Nursing and Residential Care Facilities
711	6,040	Performing Arts, Spectator Sports, and Related Industries
712	2,077	Museums, Historical Sites, and Similar Institutions
713	4,982	Amusement, Gambling, and Recreation Industries
721	2,899	Accommodation
722310, 722320	2,003	Food Service Contractors, Caterers
722410, 722515	6,019	Drinking Places (Alcoholic Beverages), Snack and Nonalcoholic Beverage Bars
722511	26,799	Full-Service Restaurants
722513, 722514	847	Limited-Service Restaurants, Cafeterias, Grill Buffets, and Buffets
812	27,770	Personal and Laundry Services
813	34,529	Religious, Grantmaking, Civic, Professional, and Similar Organizations
92	12,029	Public Administration

convergence, and the ℓ_2 regularization weight.

An advantage of using this machine learning approach is that we can select relevant covariates in a data driven way without imposing our model be sparse. This allows for localized, non-linear interactions across a high number of covariates. Of the 198 covariates included in our analysis, 132 were estimated to have a non-zero effect. Table C.4 shows the top 20 covariates, ranked by their contribution to the reduction in mean-squared error. Unsur-

Table C.3: Population Change in Census Tracts Adjacent to the Blue Line, 2000-2010

Census Tract	Neighborhood	Pop. 2000	Pop. 2010	Growth Rate
5901	Elliot Park	3,060	3,166	0.03
11998	Minnehaha	4,058	3,980	- 0.02
104400	Downtown West	1,499	2,097	0.40
104800	Cedar Riverside	7,551	8,094	0.07
105400	Elliot Park	3,416	3,527	0.03
106000	Ventura Village	3,462	3,339	- 0.04
106200	Seward	3,356	3,499	0.04
107400	Longfellow	1,713	1,726	0.01
107500	Longfellow/Seward	2,019	1,988	- 0.02
108600	Corcoran/Powderhorn Park	3,087	2,880	- 0.07
108700	Corcoran/Standish	3,550	3,274	- 0.08
108800	Howe/Longfellow	3,813	3,786	- 0.01
110200	Standish	3,518	3,522	0.00
110400	Hiawata/Howe	2,929	2,733	- 0.07
110500	Hiawata/Howe	4,438	4,694	0.06
111100	Ericsson	3,149	3,192	0.01
125900	East Phillips	4,147	4,269	0.03
126100	Downtown East/West	3,210	4,938	0.54
126200	North Loop	1,515	4,291	1.83

Source: US Cenus Bureau.

prisingly, the total area of a property is very predictive of the house price, as well as the number of bathrooms and the age of the house. The Case-Shiller Index was also highly predictive, showing the sensitivity of individual house prices to aggregate trends. One measure of amenities, the number of restaurants within a half-mile, was also strongly predictive of house prices. Several transit routes were significant as well, including the minimum distance to bus stops along routes 12 and 4. These routes connect downtown Minneapolis with its wealthier suburbs, and so it is unsurprising that they are important determinants of house prices. Finally, several demographic characteristics were significant, including the percent of college graduates in a census block and the average commute time.

Table C.5 reports the estimated direct effect using gradient boosting with regression trees. The Pre-Treatment column shows that the algorithm trained on the control group, that is, property sales occurring between 0.5 and 1 miles of a Blue Line station, is able to correctly predict sale prices in the treatment group before the introduction of the Blue Line. While mean residual is positive, it is not significantly different from zero. After the introduction of the Blue Line, Post-Treatment prices in the treatment group increase by 7.1% more than predicted, even though the algorithm accounts for the introduction of new amenities and for demographic shifts in treatment neighborhoods (see Section 3.5 for a list of control variables). Thus the XGBoost estimation routine implies that the direct effect of the Blue Line is an increase in sale prices of 7.1% for properties located within 0.5 miles of a station.

Comparing these results with those from the DiD regression we can obtain an approximate measure of the implied spillover effect arising from the introduction of the Blue Line. Our preferred (and most conservative) specification for the DiD results predicts prices will increase by 10.4% in treatment neighborhoods. This increase can be thought of as the total effect arising from the introduction of the Blue Line, compounding both the direct effect of access to light rail transit itself and the effect of the amenities changing because of increased accessibility in treatment neighborhoods. The difference between these two estimates (3.3%)

Table C.4: Top 20 covariates by importance

Feature	Gain	Cover	Frequency
Building Area	0.120	0.081	0.072
Ground Floor Area	0.107	0.040	0.036
No. of Bathrooms	0.068	0.022	0.012
Case-Shiller Index	0.060	0.041	0.035
No. of Restaurants (within 0.5 miles)	0.040	0.002	0.004
Minimum Distance Route 12	0.039	0.002	0.003
Second Floor Area	0.037	0.035	0.031
Minimum Distance Route 4	0.035	0.013	0.010
Age of House	0.035	0.047	0.038
Percent College	0.022	0.009	0.006
Minimum Distance Route 32	0.022	0.009	0.005
No. of Bedrooms	0.020	0.016	0.013
Minimum Distance Route 46	0.020	0.020	0.013
Housing Stock	0.016	0.006	0.004
Minimum Distance Route 27	0.015	0.022	0.014
Average Commute	0.014	0.001	0.003
Minimum Distance Route 21	0.012	0.031	0.018
Minimum Distance Route 22	0.011	0.046	0.026
Percent High School	0.010	0.004	0.006
Finished Basement	0.010	0.022	0.015

Note: 132 of 198 covariates had nonzero gain.

Table C.5: XGBoost: Treatment Effect

Predicted Residual:	Pre-Treatment	Post-Treatment	Implied Spillover
Mean	0.005	0.071	0.033
Std Dev.	(0.01)	(0.01)	–

Note: The implied spillover is calculated as the difference between the post-treatment prediction of the direct impact of the Blue Line (0.071) and the overall treatment effect calculated via DiD in specification (3) of Table 3.2 (0.104).

Table C.6: XGBoost: Heterogeneous Effects

Predicted Residual:	Intercept	Distance to BL	% White	% Driving	Median Income
Mean	0.24	0.09	0.33	-0.60	-0.012
Std Dev.	(0.07)	(0.08)	(0.07)	(0.09)	(0.006)

can be interpreted as the implied spillover effect, that is, the impact that amenities changing as a result of the introduction of the Blue Line have on sale prices.

Heterogeneous effects for different types of neighborhoods can be obtained by regressing the residuals from the Post-Treatment predictions (Table C.5) on neighborhood attributes. This type of regression allows us to explore how the direct effect of the Blue Line varies with respect to neighborhood characteristics, although it does not allow us to do the same for the indirect effect. Table C.6 reports the results of such a regression on tract-level characteristics captured by the 2000 Census. Neighborhoods that before the introduction of the Blue Line had a higher share of white residents saw a significant increase in their home values after the transit line was introduced. An increase in the share of white residents by 10 percentage points translates to a 3.3% increase in house prices. Wealthier neighborhoods and neighborhoods where a greater share of residents commute by car saw less of a benefit from the Blue Lines introduction. This is unsurprising as these neighborhoods can be expected to have a lower MWTP for access to public transportation. Interestingly, properties located further from the Blue Line also tend to see an appreciation in sale prices in the Post-Treatment period, although this effect is not statistically significant.

C.2 Efficient Boosted Smooth Trees

There are two main difficulties to using smooth trees for gradient boosting. First, for each branch split we test, we need to re-regress the residual on the matrix of leaf node probabilities. The time complexity of this regression is $O(C^2N)$, where C is the number of leaves and N the number of observations. If we test each observation as a splitting point, then the total time complexity is given by $O(C^2N^2)$. So smooth trees increase quadratically in the depth of the trees and the number of observations. Second, when using instruments, we need to form the product of the leaf probabilities with the instruments. The time complexity of this step is $O(KN)$, where K is the number of instruments. Repeating this multiplication N times yields an asymptotic rate of $O(KN^2)$. The purpose of this appendix is to proposed an algorithm that cuts these rates by a factor of N.

The key idea is to transform the problem so that we can update the gain by changing a single covariate at a time, eliminating the factor C^2. This is done in a manner analogous to updating a Kalman filter, where we use the bordering method and a PR-recalculated matrix inverse to perform the regression. We then use a sigmoid function that closely approximates the logit sigmoid which has the added property of being multiplicatively separable in its inputs. This allows us to efficiently calculate the instrument moments for any split in the data.

Let P_{t-1} be the matrix of choice probabilities as of step $t-1$. Note that each column of P_{t-1} represents a leaf of the smooth tree, with each row of $P_{t-1,j}$ being the probability that X_i ends up in leaf j. We want to test whether a branch is added to leaf j, such that:

$$P_t = [P_{t-1,-j}, P_{t-1,j}L(X_i), P_{t-1,j}(1 - L(X_i))]$$

Let P_z be the projection matrix for the instruments Z. Define the new regressors:

$$\tilde{y} = P_z y$$
$$\tilde{P}_t = P_z P_t$$
$$\tilde{P}_{t-1} = P_z P_{t-1}$$

The residual from a ridge regression is given by:

$$R(y, X, \lambda) = y'(I - (1-\lambda)X(X'X + \lambda I)^{-1}X')y$$

We accept this addition if it maximizes the gain:

$$\begin{aligned} G(\tilde{y}, \tilde{P}_t, \tilde{P}_{t-1}, \lambda) &= \frac{1}{(1-\lambda)}\Big(R(\tilde{y}, \tilde{P}_{t-1}, \lambda) - R(\tilde{y}, \tilde{P}_t, \lambda)\Big) \\ &= \tilde{y}'\tilde{P}_t(\tilde{P}_t'\tilde{P}_t + \lambda I)^{-1}\tilde{P}_t'\tilde{y} - \tilde{y}'\tilde{P}_{t-1}(\tilde{P}_{t-1}'\tilde{P}_{t-1} + \lambda I)^{-1}\tilde{P}_{t-1}'\tilde{y} \end{aligned}$$

Redefine P_t as

$$P_t = [P_{t-1}, P_{t-1,j}L(X_i, c_t)]$$

and $\tilde{P}_t$ is constructed as before. This $\tilde{P}_t$ gives identical coefficients and residuals as the previous one, but only involves a single new regressor, rather than two. As we update $L(X_i, c_t)$, the term P_{t-1} stays fixed. For ease of notation, let $B = P_{t-1}$ and $A = P_{t-1,j}L(X_i, c_t)$. Then the term $(\tilde{P}_t'\tilde{P}_t + \lambda I)^{-1}$ can be written as the inverse of a symmetric block matrix:

$$(\tilde{P}_t'\tilde{P}_t + \lambda I)^{-1} = \begin{bmatrix} B'B + \lambda I & B'A \\ A'B & A'A + \lambda \end{bmatrix}^{-1}$$

Here, A is a $N \times 1$ vector, so we can re-write this inverse using the bordering method. This states that the inverse of a a bordered matrix is given by

$$\begin{bmatrix} Q & \delta \\ \delta' & Z \end{bmatrix}^{-1} = \begin{bmatrix} Q^{-1} + \frac{Q^{-1}\delta\delta' Q^{-1}}{\mu} & -\frac{Q^{-1}\delta}{\mu} \\ -\frac{\delta' Q^{-1}}{\mu} & \frac{1}{\mu} \end{bmatrix}$$

where

$$\mu = Z - \delta' Q^{-1}\delta$$

Therefore, our inverse becomes:

$$\begin{bmatrix} (B'B + \lambda I)^{-1} + \frac{(B'B+\lambda I)^{-1}B'AA'B(B'B+\lambda I)^{-1}}{\mu} & -\frac{(B'B+\lambda I)^{-1}B'A}{\mu} \\ -\frac{A'B(B'B+\lambda I)^{-1}}{\mu} & \frac{1}{\mu} \end{bmatrix}$$

with

$$\mu = A'A + \lambda - A'B(B'B + \lambda I)^{-1}B'A$$

The residual can be expressed as:

$$\tilde{y}'\tilde{P}_t(\tilde{P}_t'\tilde{P}_t + \lambda I)^{-1}\tilde{P}_t'\tilde{y} =$$

$$y'B(B'B + \lambda I)^{-1}B'y + \frac{1}{\mu}y'B(B'B + \lambda I)^{-1}B'AA'B(B'B + \lambda I)^{-1}B'y$$

$$-\frac{2}{\mu}y'AA'B(B'B + \lambda I)^{-1}B'y$$

$$\frac{1}{\mu}y'AA'y$$

Note that:

$$\tilde{y}'\tilde{P}_{t-1}(\tilde{P}'_{t-1}\tilde{P}_{t-1}+\lambda I)^{-1}\tilde{P}'_{t-1}\tilde{y} = y'B(B'B+\lambda I)^{-1}B'y$$

so these terms cancel. This leaves:

$$G(\tilde{y},\tilde{P}_t,\tilde{P}_{t-1},\lambda) = \frac{1}{\mu}\left(y'B(B'B+\lambda I)^{-1}B'A - y'A\right)^2$$

which, conditional on A, can be calculated with three dot products ($A'B$, $y'B(B'B+\lambda I)^{-1}B'A$ and $A'A$) and one low-dimensional matrix multiplication that scales with the depth of the trees.

The principal question then is how quickly can we construct the matrix $P'_{t-1}A$ or $Z'A$. The semi-naive approach would be to update A in each step and calculate these products. This approach is semi-naive because this is necessary for most sigmoid functions, and is what greatly reduces the computational efficiency of smooth trees. An alternative procedure would be to use the following sigmoid:

$$L(X_i,c_j) = \begin{cases} 1-\frac{1}{2}\frac{2^{c_j}}{2^{X_i}} & \text{if } c_j < X_i \\ \frac{1}{2}\frac{2^{X_i}}{2^{c_j}} & \text{if } c_j \geq X_i \end{cases}$$

Assume that the kth regressor, X_k, is sorted from smallest to largest, and that Z_k is sorted based on the ordering of X_k.[1] The goal is to test all elements of X_k to find the split that maximizes the gain. At iteration 1, we have:

$$\begin{aligned} A_1 &= Z'(P_{t-1,j}L(X_k,X_1)) && \text{(C.1)} \\ &= \sum_i Z_i P_{ij}^{t-1}\left(1-\frac{1}{2}\frac{2^{X_{1k}}}{2^{X_{ik}}}\right) && \text{(C.2)} \end{aligned}$$

This can be broken into two parts:

[1]This only needs to be done once at the start of the algorithm for all regressors.

$$\bar{Z}_1 = \sum_i Z_i P_{ij}^{t-1}$$

and

$$\tilde{Z}_1 = \frac{1}{2} \sum_i Z_i P_{ij}^{t-1} \frac{2^{X_{1k}}}{2^{X_{ik}}}$$

This defines:

$$Z_1^r = \bar{Z}_1 - \tilde{Z}_1$$

and

$$Z_1^l = 0$$

so that $A_1 = Z_1^l + Z_1^r$. The key idea is that going from $X_{m-1,k}$ to X_{mk} only involves updating according to:

$$\bar{Z}_m = \bar{Z}_{m-1} - Z_{m-1} P_{m-1,j}^{t-1}$$

$$\tilde{Z}_m = \left(\tilde{Z}_{m-1} - \frac{1}{2} Z_{t-1} P_{1j}^{t-1} \right) \frac{2^{X_{m,k}}}{2^{X_{m-1,k}}}$$

$$Z_m^l = \left(Z_{m-1}^l + \frac{1}{2} Z_{m-1} P_{m-1,j}^{t-1} \right) \frac{2^{X_{m-1,k}}}{2^{X_{m,k}}}$$

and

$$A_m = \bar{Z}_m - \tilde{Z}_m + Z_m^l$$

This allows us to calculate all potential A_m with a single pass through the data, reducing the time complexity of calculating A by a factor of N, from $O(N^2K)$ to $O(NK)$.

C.2.1 Monte Carlos

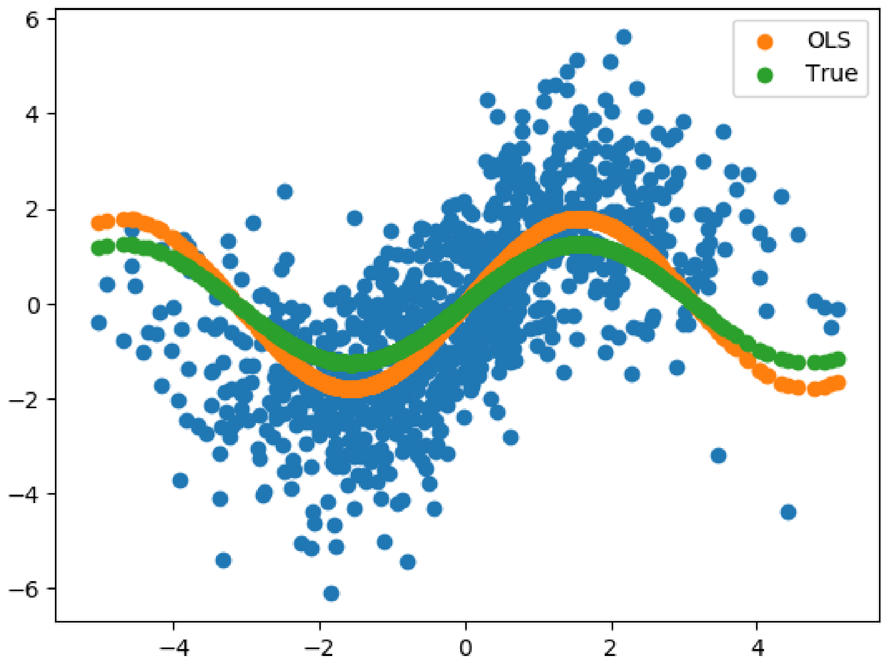
Since we do not currently have a proof for consistency of the Boosted Smooth Trees estimation routine, we provide a small Monte Carlo simulation to demonstrate the efficacy of our method. We used $N = 1,000$ observations with $Z_i, \epsilon_i \sim N(0, 0.5^2)$ and $X_i = Z_i + \frac{1}{2}\epsilon_i$ in order to simulate the effects of endogeneity. The dependent regressor is determined by the following nonlinear relationship:

$$y_i = 1.25 \sin(X_i) + \epsilon_i$$

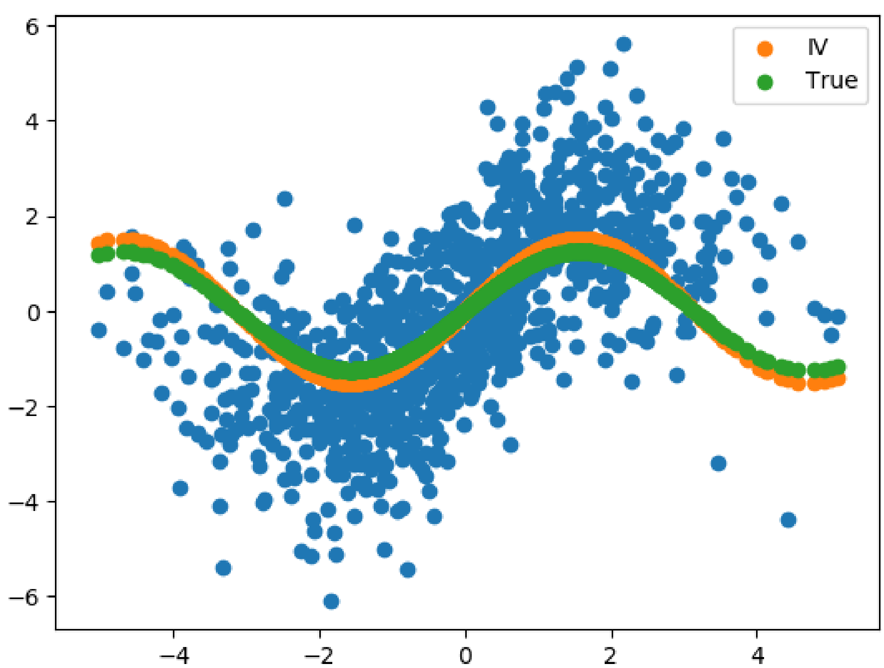
We used a learning rate of $\gamma = 0.05$ and a minimum of 50 observations per node. Finally, we trained with $M = 350$ iterations and cross-validated to determine the optimal λ. Figure C.1 plots the data and the parameter OLS and IV estimates. OLS tends to over-predict due to the positive correlation between the residual and the regressor. The instrumental variable estimates improve this somewhat but tend to still over-predict.

Figure C.2 shows the non-parametric estimates using Boosted Smooth Trees. There is close agreement in the range of -3.5 to 3.5, and divergence beyond that point. This is largely due to the lack of observations in the tails of the distribution. Using the above values, we repeated this exercise 100 times and calculate the RMSE for each sample. The mean RMSE was 0.09 and the standard deviation was 0.018.

Figure C.1: Parametric Estimation



(a) OLS



(b) IV

Figure C.2: Endogenous Regressor