



# Essays on Industrial Organization

## Citation

Wollmann, Thomas G. 2015. Essays on Industrial Organization. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:17467215>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

# Essays on Industrial Organization

A dissertation presented

by

Thomas G. Wollmann

to

The Department of Business Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Business Economics

Harvard University

Cambridge, Massachusetts

April 2015

© 2015 Thomas G. Wollmann

All rights reserved.

*Dissertation Advisor:*  
**Professor Ariel Pakes**

*Author:*  
**Thomas G. Wollmann**

## **Essays on Industrial Organization**

### **Abstract**

This dissertation comprises three essays on industrial organization. The first essay studies how product-level entry and exit decisions impact business and public policy analysis. It provides an empirical model that incorporates these decisions and then estimates it in the context of the commercial vehicle segment of the US automotive industry. Finally, it demonstrates the importance of accounting for product-level changes using the \$85 billion decision to rescue two US automakers in 2009. The second essay studies how two period strategies perform relative to Markov perfect strategies in discrete dynamic games. In particular, it considers a simple entry/exit game and shows that agents sacrifice very little in terms of expected discounted payoffs when they employ these simpler strategies. It also shows this result is robust to varying the underlying market characteristics. The third essay estimates the causal impact of research expenditures on scientific output. Unexpected college football outcomes provide exogenous variation to university funds, and in turn, research expenditures in the subsequent year. Using this variation, it estimates the dollar elasticity of scholarly articles, new patent applications, and the citations that accrue to each.

# Contents

Abstract . . . . .	iii
Acknowledgments . . . . .	viii
<b>Introduction</b>	<b>1</b>
<b>1 Trucks without Bailouts: Equilibrium Product Characteristics for Commercial Vehicles</b>	<b>3</b>
1.1 Introduction . . . . .	3
1.2 Market Setting and Data . . . . .	9
1.3 Model . . . . .	18
1.4 Estimation . . . . .	24
1.5 Descriptive Evidence . . . . .	35
1.6 Results . . . . .	38
1.7 The Impact of the “Bailout” . . . . .	46
1.8 Conclusions . . . . .	54
<b>2 Two Period Strategies for Discrete Dynamic Games</b>	<b>56</b>
2.1 Introduction . . . . .	56
2.2 Model . . . . .	58
2.2.1 Markov perfect strategy equilibrium . . . . .	59
2.2.2 Two period strategy equilibrium . . . . .	61
2.3 Base Case Performance . . . . .	62
2.3.1 Parameters . . . . .	62
2.3.2 Equilibrium Parameters . . . . .	63
2.3.3 Loss . . . . .	63
2.4 Varying Market Parameters . . . . .	64
2.5 Extensions . . . . .	68
2.6 Conclusion . . . . .	70

<b>3</b>	<b>The Impact of Money on Science: Evidence from Unexpected NCAA Football Outcomes</b>	<b>73</b>
3.1	Introduction . . . . .	73
3.2	US College and University Research . . . . .	80
3.2.1	How Research is Funded . . . . .	81
3.3	Data . . . . .	86
3.3.1	Sources . . . . .	86
3.3.2	Summary Statistics . . . . .	90
3.4	Empirical Model . . . . .	91
3.4.1	Overview . . . . .	91
3.4.2	Specification . . . . .	93
3.5	Results . . . . .	95
3.5.1	From Football to Money . . . . .	95
3.5.2	From Money to Scholarly Articles . . . . .	96
3.5.3	From Money to Scholarly Article Citations . . . . .	99
3.5.4	From Money to Patents . . . . .	101
3.5.5	From Money to Patent Citations . . . . .	102
3.5.6	Exogeneity Check . . . . .	105
3.5.7	Discussion of Timing . . . . .	107
3.6	Conclusion . . . . .	110
	<b>References</b>	<b>113</b>
	<b>Appendix A Appendix to Chapter 1</b>	<b>118</b>
A.1	Constructing empirical distribution of buyers . . . . .	118
A.2	Calculating market size. . . . .	119
A.3	Computing and selecting equilibrium . . . . .	119
	<b>Appendix B Appendix to Chapter 3</b>	<b>121</b>
B.1	Allocating patent filings to universities . . . . .	121
B.1.1	Disaggregating state educational systems . . . . .	124

## List of Tables

1.1	Summary statistics . . . . .	14
1.2	Mean characteristics conditional on buyer type and purchase . . . . .	16
1.3	Descriptive statistics by brand as of 2012 . . . . .	17
1.4	Ownership changes . . . . .	18
1.5	Demand estimates . . . . .	39
1.6	Marginal cost estimates . . . . .	42
1.7	Sunk cost estimates . . . . .	43
1.8	2009 GM and Chrysler Product Offerings . . . . .	50
1.9	Effects of model-level entry on products for liquidation . . . . .	51
1.10	Effects of model-level entry on purchases for liquidation . . . . .	52
3.1	Football and university finances . . . . .	84
3.2	The impact of football on non-federal research expenditures . . . . .	95
3.3	The impact of money on scholarly articles . . . . .	97
3.4	The impact of money on scholarly articles' citations . . . . .	100
3.5	The impact of money on patent applications . . . . .	103
3.6	The impact of money on patents' citations . . . . .	104
3.7	Exogeneity check for instrument . . . . .	107
3.8	First stage timing . . . . .	111
3.9	Second stage timing: scholarly articles . . . . .	112
3.10	Second stage timing: patents . . . . .	112

## List of Figures

1.1	Timing of the game . . . . .	19
1.2	Evidence from the construction industry . . . . .	36
1.3	Evidence from the freight industry . . . . .	37
1.4	Evidence from length deregulation . . . . .	38
1.5	Profit margin comparison . . . . .	41
1.6	Plot of log of marginal costs on GWR . . . . .	42
1.7	Comparisons across product space . . . . .	44
1.8	Spillovers in setup . . . . .	45
1.9	Earned vs expected discounted profits . . . . .	46
1.10	Increased markups induce entry . . . . .	53
1.11	Impact to consumers' probability of not purchasing . . . . .	54
2.1	Impact of discount rate on mean loss . . . . .	65
2.2	Impact of discount rate on maximum loss . . . . .	66
2.3	Impact of persistence on mean loss . . . . .	67
2.4	Impact of persistence on maximum loss . . . . .	68
2.5	Impact of scrap value distribution parameter on mean loss . . . . .	69
2.6	Impact of scrap value distribution parameter on maximum loss . . . . .	70
2.7	Impact of scrap value distribution parameter on mean loss . . . . .	71
2.8	Impact of scrap value distribution parameter on maximum loss . . . . .	72
3.1	Football and scholarly articles . . . . .	77
3.2	Football and scholarly article citations . . . . .	78
3.3	Football and new patent applications . . . . .	79
3.4	Football and patent citations . . . . .	80
3.5	Timing of football, research, and publishing . . . . .	92
3.6	Impact of football on federal vs. non-federal funding . . . . .	106
B.1	Patent applications filed over time . . . . .	122



## Acknowledgments

I am deeply appreciative of the instruction, guidance, and encouragement I received from my thesis committee: Ariel Pakes, Juan Alcacer, Greg Lewis, and Dennis Yao. Each had confidence in me when I lacked it in myself, overlooked my frequent mistakes, gently but clearly criticized my work when it veered off-track, invested time and energy in my learning, and instilled in me the notion that economic research is serious—as well as exciting—business.

I am grateful for the encouragement and advice of Thom Covert and Mike Sinkinson, the contributions of Richard Sweeney and Haris Tabakovic towards my development and our shared works, and engaging discussions with Oliver Hart, Sam Kruger, Jamie Lee, Robin Lee, Julie Mortimer, Mark Shepard, Eli Tamer, and countless other faculty and students.

I also thank Brenda Piquet, John Korn, Karla Cohen, Dianne Le, and Gia Petrakis for their superb administration of the graduate program at Harvard as well as Ann Richards for her help during my teaching fellowship and placement. Their competence, combined with the generosity of this institution, ensured that the only limits to my progress were my own ideas and effort. The estimation exercises in this dissertation relied heavily on the computational resources of the business school's Research Computing Services, where Troy Adair, John Galvin, Lauren Ross-Miller, and many others deserve considerable credit.

Last, I owe a debt to my close friends and family. In particular, I thank Michael Depalma, Stephanie Landry, and my parents for their support along the way.

Dedicated to my Mom, who at various times during the writing of this thesis has helped with everything from gathering data to editing copy.

# Introduction

In differentiated product markets, the entry and exit of individual product models—rather than of firms—often serve as the main equilibrating force. Market structure changes that lead to high prices also tend to encourage entry, partially offsetting the policy's effect on prices and purchases. Thus, accurately predicting changes from a merger or bankruptcy should incorporate this behavior. The first essay in the dissertation develops a model of equilibrium product characteristics in oligopoly and shows how to estimate the sunk costs of offering them. It then applies these methods to study the impact of the \$85 billion bailout of General Motors and Chrysler in the context of the commercial vehicle segment of the US automotive market. In the case where the troubled firms are liquidated rather than rescued, estimated sunk costs are low enough to induce product entry by rivals, and that this has a dramatic effect on prices and purchases. For example, allowing for model-level entry and exit moderates markup increases by over two-thirds for the most affected products. It also moderates the drop in total output by about one-half.

Markov perfect equilibrium strategies place a large burden on agents in all but the simplest dynamic games. The second essay in the dissertation, written jointly with Richard Sweeney, introduces the concept of two period strategies, which restrict agents to form cutoff rules based on multiples of the current period's profits. These strategies require keeping track of only one parameter, as opposed to the billions or more parameters required in a full forward-looking solution to the dynamic game. We simulate equilibrium of a simple entry/exit game under the set of market primitives considered in Pakes, Ostrovsky, and Berry (2007). Relative to a Markov perfect strategy, the loss from employing a two

period strategy is on average quite small. Loss increases with the discount factor and the number of potential entrants, but decreases with the mean of the selloff value and entry fee distributions.

Scientific discovery drives economic growth, but the high cost of research makes funding a limiting factor. Little is known about the causal impact of money on science, despite its importance for determining the socially-optimal level of R&D. The third essay in the dissertation, written jointly with Haris Tabakovic, estimates the dollar elasticity of research output at American universities by using unexpected NCAA football outcomes to exogenously shift research budgets across schools and time. It first demonstrates these outcomes are strong predictors of non-federal research expenditures, but not of federal expenditures, which lends support for the instrument. It finds that the dollar elasticity of scholarly publications and the citations that accrue to them are 0.27 and 0.53, respectively. It also finds that the dollar elasticity of new patent applications and the citations that accrue to them are 1.72 and 3.12, respectively. Each outcome contrasted sharply with the OLS estimates, which are significant but near zero and would lead policymakers to underinvest in research.

# Chapter 1

## Trucks without Bailouts: Equilibrium Product Characteristics for Commercial Vehicles

### 1.1 Introduction

The response of firms to heterogeneous consumer preferences has improved considerably since Henry Ford famously remarked, “You can have the Model T in any color, so long as it’s black.”<sup>1</sup> In many industries, a second pattern emerged alongside the surge in differentiated production. Industries would come to be organized around a single, relatively stable set of firms but a rapidly evolving set of product offerings. The US automotive industry, for example, has witnessed virtually no firm-level entry or exit for thirty years despite wide variation in the number and nature of products over time. Outside of autos there are many other examples, ranging from aircraft to bicycles to cat food. In each case, it is the entry and exit of individual product models—not the Schumpeter [1942] creation and destruction of firms—that serve to drive these markets into equilibrium. Yet despite rich evidence on

---

<sup>1</sup>There are variations of this famous saying and this seems to be the most common. Ford and Crowther [1922] recalled it slightly differently in his autobiography, but without a change in meaning.

the importance of accounting for entry and differentiation separately,<sup>2</sup> little is known about their combined impact. This paper examines how model-level entry and exit impacts policy analysis in differentiated product markets.

The theoretical motivation is simple. Entry and exit tend to work in the opposite direction of the price mechanism, so failing to account for them can overstate the impact to prices and purchases of a change in market structure. To illustrate, consider the acquisition or bankruptcy of a firm in oligopoly. Ignoring entry, this exit increases the market power of surviving firms, who raise markups and earn windfall profits. However, the prospect of high profits lures entrants. Even if startup costs are large enough to prohibit firm-level entry, the cost of adding or repositioning products for incumbents may be small enough to permit model-level entry. If the latter is true, firms will set lower prices that reflect a more crowded product space where consumers can more easily substitute. Thus, counterfactuals depend on accurately determining where and to what degree model-level entry will occur.

I provide methods for estimating the parameters governing model-level entry and exit decisions as well as computing counterfactuals that account for this behavior. Using these methods, I study the impact of the \$85B government bailout of the US automotive industry in late 2008 and early 2009. During this period, General Motors (“GM”) and Chrysler were headed for default.<sup>3</sup> Whether to provide federal assistance was hotly contested and even became a Presidential campaign topic in 2008<sup>4</sup> and 2012.<sup>5</sup> For tractability, I narrow the scope of my assessment to the commercial vehicle segment of the auto industry, and then construct an original dataset of all product offerings between 1987 and 2012. Taken together, the panel data and methods allow me to ask, “What would have happened to output and prices had the government not rescued the automakers?” There are two obvious policy alternatives. The first is liquidation, an effective removal of the GM and Chrysler brands and products

---

<sup>2</sup>For early examples of the former see Borenstein [1989] and Bresnahan and Reiss [1991] and of the latter see Berry et al. [1995] and Nevo [2001].

<sup>3</sup>Isidore, Chris. “Bush announces auto rescue.” *CNNMoney* 19 Dec 2008.

<sup>4</sup>Romney, Mitt. “Let Detroit Go Bankrupt.” *New York Times* 18 Nov 2008.

<sup>5</sup>Rosche, Jedd. “Obama quotes Romney’s auto bailout op-ed in UAW speech.” *Politico* 28 Feb 2012.

from the marketplace. The second is an acquisition by an existing, rival firm. To illustrate the range of outcomes, my counterfactuals consider a sale to Ford, which overlapped most heavily with GM and Chrysler in product space, and to Paccar, which overlapped the least.

The commercial vehicle market is representative of the auto industry overall and an ideal place to study this phenomenon. No firms have entered or exited in three decades but product offerings have changed frequently.<sup>6</sup> Ownership is concentrated among about ten firms and is especially concentrated in sub-segments of the market, even though most firms have produced most product variants at some time. Furthermore, the highly modular production of commercial vehicles allows manufacturers to quickly swap parts and introduce new models, often within months of changing demand or competitive conditions. This creates a tight link between the incentives to adjust products and the actual adjustment decisions. Finally, their physical weight, tariff structure, and legal standing isolate commercial vehicles from the passenger vehicle market and foreign automotive markets.

I find that in the event of liquidation, profit increases are high enough relative to sunk costs to induce product entry among surviving firms, and that this has a dramatic impact on the market. Relative to the bailout, markups for the most affected products rise by over 70% when entry is ignored but only 18% when it is accounted for. At the same time, the probability of not purchasing a vehicle rises nearly 50% for the most affected buyers when entry is ignored, but only 14% when it is accounted for. The median impacted buyer and product, on the other hand, see virtually zero change. This reflects the heavy overlap of GM and Chrysler products and strong preference heterogeneity in the buyers. For this reason, the impact on total output is much more muted: it falls 7.6% when entry is ignored and 3.2% when it is accounted for.

Acquisition is another alternative. If entry is ignored, an acquisition by Ford, which

---

<sup>6</sup>In the passenger segment, the most recent entrants are Hyundai and Kia, who began selling in 1986 (Kia began by re-badging its exports and did not legally incorporate a standalone US brand until 1992). This ignores Tesla, who at the time of writing is still very small (but growing). In the commercial segment, the most recent were Hino and UD in 1984 and 1985, respectively, although Hyundai exported a small number of vehicles between 2000 and 2001 under the Bering badge before being absorbed into Daimler.

overlaps heavily with GM and Chrysler in product space, looks qualitatively the same as a liquidation. In sharp contrast, an acquisition by Paccar, which does not overlap with GM or Chrysler, closely resembles the bailout. If model-level entry and exit are accounted for, however, all three counterfactual policies are essentially symmetric. Relative to the bailout, markups on the most affected products rise between 14% and 18% while total output falls between 2% and 3%. The policy choices—including the identity of the acquiring firm—appear to matter a lot when model-level entry and exit are ignored, but in fact matter little when this important equilibrating mechanism is accounted for.

These results require two methodological contributions. The first is a model of equilibrium product offerings that handles multi-product firms selling multi-attribute goods but remains tractable in applied settings. In this model, buyers vary in their preferences over characteristics, e.g. urban buyers on congested roads prefer short and maneuverable vehicles while long-distance freight haulers prefer large and rugged ones. Sellers face sunk costs to add and remove products as well as marginal costs of production. In each year, firms choose which vehicles to offer in the first stage and, conditional on those offerings, choose what prices to charge in the second stage. First stage product entry and exit decisions, therefore, weigh sunk costs against changes in second stage profits. Simultaneous first period choices form a Nash equilibrium in product space, while second period choices form a Bertrand-Nash equilibrium in prices. Sunk costs can induce forward-looking behavior among the producers, but the dynamic programming problem in this market—and most richly-defined differentiated product markets—places an unreasonably large computational burden on firms. It requires taking expectations over billions or more of states and rules out learning by repeated play, but raises the question of what managers actually do in these situations. Interviews with commercial vehicle product designers and division managers suggest the use of hurdle rates to best approximate this process. This capital budgeting rule greatly simplifies the entry and exit decisions that firms make. Later in the paper, I provide evidence that the discounted profits firms actually earn are quite close to what is implied by hurdle rates.



The second is a method of identifying sunk costs in industries with rich time variation but only one geographic market. The first step in the process entails recovering the primitives governing demand and using the Bertrand-Nash pricing condition to recover marginal costs. This yields an estimate of the profits that firms would earn under any hypothetical set of product offerings. The second step constructs inequalities based on the fact that firms must find the observed product offerings—i.e. equilibrium choices—preferable to all other alternative sets of product offerings. The model is identified by combining these inequalities with exogenous shifts in the composition of buyers over time, which firms respond to by changing product offerings. My data reveal that, for example, the size of the construction industry is a strong predictor of vehicle offerings tailored to the preferences of those buyers. This response is true for other industries as well as for a law change that shifts demand. These factors provide variation that can be combined with revealed preference and the necessary conditions for a Nash equilibrium to identify sunk costs.

Relying on only necessary equilibrium conditions is important since the presence of multiple equilibria rules out a one-to-one map between the parameters and the outcomes. For example, for a guess of the parameters I may only be able to say that Firm A will enter if Firm B does not and vice versa (but cannot say which is more likely). The econometrician can rarely take a stand on which equilibria is played, so calculating a likelihood is impossible here. Nonetheless, this still provides information. For instance, I can penalize instances where both Firm A and B enter or where neither does. Inequalities have been used in this way before (Tamer [2003], Pakes et al. [2015]), but combining them with exogenous demand shocks to identify parameters over a time series has not. Several plausibly exogenous ownership changes, which are another helpful feature of the market, also provide exogenous variation in product entry and exit incentives.

This paper contributes to a growing empirical literature on firm positioning. Prior work has largely relied on cross-sectional variation provided by multiple geographic markets (e.g. Mazzeo [2002], Seim [2006], Fan [2013], Draganska et al. [2009]). These papers cleanly demonstrate the inherent tradeoffs in differentiation in many important settings; however,

they require that industries comprise many local, isolated markets. This has tended to restrict their application to either goods that are costly to transport (e.g. ice cream, which melts and spoils) or service-related products (e.g. hotels, retail, or news production). In contrast, commercial vehicle models—like many consumer and industrial goods—are produced in at most one or two locations but distributed nationwide. This rules out geographic variation on the supply side. Thus, identifying fixed or sunk costs necessarily requires looking over a time series. Recent work on microprocessors do precisely that, although it assumes that the possible set of product offerings evolves exogenously over time (Nosko [2010], Eizenberg [2014]). Such an approach may prove difficult to extend past industries that are not guided by plausibly exogenous technological variation (e.g. Moore’s Law). I instead rely on exogenous shifts in demand. This rests on only observably heterogeneous buyer preferences, which have already been shown critical to understanding purchasing and pricing patterns in differentiated product markets (Petrip [2002], Berry et al. [2004]).

Recent work on dynamic games has also estimated sunk costs of entry, exit, and repositioning. Recovering the primitives of these games was made easier with a two-step approach introduced by Hotz et al. [1994], which in effect transferred the burden of solving these problems from the computation to the data (Pesendorfer and Schmidt-Dengler [2008], Aguirregabiria and Mira (2007), Pakes et al. [2007], Bajari et al. [2007]). Agents in these models fully internalize the future impact of their decisions. Applying these methods, however, has required a smaller action space than I consider here (e.g. Blonigen et al. [2013], who notably study passenger vehicle refresh and scrapping decisions), often with large numbers of geographic markets (e.g. Collard-Wexler [2013], Ryan [2012] and Sweeting [2013]). Again, in most differentiated product markets, the number of potential product types is large while the number of geographic markets rarely exceeds one. Morales et al. [2011] provide an alternative that is robust to large choice sets, although limited to the case of monopolistic competition, so firms face only a single agent problem. They show that Euler equation perturbations yield inequalities that can flexibly identify fixed and sunk costs (although these would be ruled out if there was strategic interactions).

## 1.2 Market Setting and Data

### Overview

The commercial vehicle segment of the US automotive industry accounts for about 10% of total US automotive sales (Wards [1986-2013]), which themselves account for 4% of gross domestic product.<sup>7</sup> The segment comprises any on-road vehicle rated for over 10,000 lbs gross vehicle weight (defined below) and sold domestically.<sup>8</sup> In terms of use and users, their scope is quite broad. For example, the market includes inner-city delivery vans, landscaping flatbeds, dump trucks, and highway tractor-trailers. In terms of capabilities, the high end of the segment has carried loads in excess of 250,000 lbs, such as oil rigs and turbine engines, and as cumbersome as an Airbus A320 fuselage, which was the case when the “Miracle of the Hudson” was hauled from New York City to the Smithsonian museum in Washington DC.<sup>9</sup>

### Data

Three main sources of data are used in this paper. First, I compile a panel of all commercial vehicle models sold in the US from 1986 to 2012 from annual issues of *The Truck Blue Book*. Each observation includes the brand, model year, model name and number, and a host of product characteristics. These include price as well as detailed specifications related to the load capacity, cab, chassis, powertrain and drivetrain. Each guide contains data for the prior ten years, easing compilation, but unfortunately does not contain retail price data

---

<sup>7</sup>Dept. of Commerce *Auto Industry Snapshot*. <http://selectusa.commerce.gov/industry-snapshots/automotive-industry-united-states>. Accessed 12/20/13.

<sup>8</sup>The “on-road” distinction here, or in US Dept. of Transportation terms “on-highway,” merely excludes irrelevant vehicles like one-seat “terminal tractors” and fifty-foot-tall mining trucks. I exclude motorhomes, buses, and “step” vans from the dataset because these are by-and-large manufactured by different firms and bought by different consumers than the vehicles in this dataset. Finally, I do not include low-entry cab-forward models. Although these are made by Mack/Volvo, and Peterbilt, who appear in this dataset, there is no variation in who produces them or their characteristics (with the exception of the short-lived Sterling Condor). Their exclusion should also not affect the demand system, since they are used almost exclusively in urban garbage collection.

<sup>9</sup>The “Miracle on the Hudson” refers to US Airways commercial flight 1549, which Capt. Chesley Sullenberger emergency landed on the Hudson River following a bird strike.

for the most recent model-year vehicles, so I proceed with MSRP. I convert all prices to 2005-equivalent dollars via the Consumer Price Index.

I merge the product characteristic data to unit sales data from the R.L. Polk & Co. *New Vehicle Registration Database*. The Polk data covers US vehicle sales Class 3 and above and is compiled from state motor vehicle registration records. Observations are broken down by brand, model name and/or number, as well as gross vehicle weight rating class, fuel-type, and body description. All sales figures are compiled on a calendar year basis. In rare cases, a model was identified in the quantity data by a model number and in the product characteristics data by a model name, or vice versa, although I was able to resolve these conflicts using the *Official Commercial Truck Guide* published by the American Truck Division of the National Automotive Dealers Association (NADA). In two cases I resolved the issue by calling a dealer.

The third main data source is microdata on commercial vehicle purchases available through the US Census. This microdata was collected up to and including 2002 under a program known as the Truck Inventory Use Survey (TIUS) and later the Vehicle Inventory and Use Survey (VIUS). Every five years, the Census mailed approximately 130,000 owners of trucks and vans and asked them various questions about their vehicle, the use of their vehicle, and about the owners themselves (relating to their vehicle use). The response rate was approximately 80% and relatively stable over time. I observe the industry and state that the buyer operates in, whether the vehicle was acquired new or used, and the characteristics of the vehicle the buyer owns.

Three additional pieces of information complete the dataset. First, the US Census *County Business Patterns* contributes the number of US firms by industry, state and year, while the US Department of Transportation ("DOT") *Highway Statistics* contributes urban and non-urban road mileage by state and year. Together, these provide an empirical distribution of buyer types that serve as the basis for taking simulation draws in the demand system and that determine changes in size of the total potential market for commercial vehicles. Last, the Bureau of Labor Statistics contributes product worker wages at the state and year

level, which I match to product models based on their respective factory locations.

## **Product Characteristics**

Commercial vehicles are conveniently summarized by a short list of product characteristics. The first and most important is gross vehicle weight rating (“GWR”), defined as the maximum load that may legally rest on the axles. GWR is the main means by which both the automotive industry and the Department of Transportation characterize vehicles. The threshold of 10,000 lbs GWR provides a natural separation between the passenger and commercial segments. Below this threshold are cars, minivans, station wagons, nearly all pickup trucks, SUVs, and cargo vans; above it are what would typically be thought of as work trucks. In terms of use, the Census microdata show a very obvious distinction. More than 95% of vehicles below 10,000 lbs are designed for “personal use,” while less than 5% above this threshold are. In terms of production, the physical design is also distinct. Vehicles below this cutoff usually feature unibody design, meaning the exterior skin provides primary support to the load. Vehicles above this cutoff feature “body-on-frame” design, meaning the load is supported by a ladder frame, onto which assemblers attach the axles, cab, power unit, and controls.<sup>10</sup> That is, the driver sits inside the load-bearing structure in a passenger vehicle but sits on top of it in a commercial one. Body-on-frame design allows vehicle assemblers to quickly modify the characteristics of a vehicle, often in as little as a few months.

GWR determines the possible uses of a vehicle. Since carrying loads in excess of it is illegal and unsafe, and since it increases price, buyers purchase vehicles with the minimum GWR that safely covers their needs. Other characteristics, like the transmission and engine, relate quite closely (even though they may not map exactly).<sup>11</sup> The tight relationship

---

<sup>10</sup>The technical term is monocoque construction, meaning the exterior skin supports the load. Variants of this term, like semi-monocoque and unitary construction, more precisely describe modern passenger vehicles, but all are distinct from body-on-frame.

<sup>11</sup>Engines deserve discussion. Although each model typically carries a large number of engine options, the base option is usually the same across manufacturers for a given GWR and cab type. Product guides even include charts that relate one engine type to another across manufacturers, suggesting close comparability.

between GWR and all other load-related attributes is not surprising, since firms want to minimize the cost of all components used in production, conditional on each being able to safely transport loads of a given GWR. The only exception to the usefulness of GWR as a load capacity measure is at the top end, where vehicles are more likely “pulling” rather than “carrying” loads. I adjust GWR to the gross combination weight rating (GCWR), which is the appropriate measure for these vehicles, although the measures are so correlated that in practice this adjustment only affects ten models in the panel.

The industry’s desire for compatible parts also means that while GWR is technically a continuous choice variable, it takes on only discrete values in practice. For example, over thirty vehicles in the data have a GWR of exactly 52,000 lbs. The small group of models that did not match exactly to any group were always close to some larger group, so the group’s GWR was substituted in for these models. After this adjustment, GWR takes up to 22 values.

The second characteristic is the style of cab, the portion of the vehicle that encloses the passengers, controls, and engine. Cabs come in three distinct varieties. The most popular cab type is the “conventional,” which is distinguished by its relatively long hood and placement of engine well ahead of the occupants. The conventional cab places the axle ahead of the driver, making for a smooth ride and spacious interior, but its long hood limits the amount of maneuverability in tight spaces. The cab-over-engine (“cabover”) type features a flat front and places the occupants directly over the front axle and engine. This makes for exceptional visibility and turning but an uncomfortable and less safe driving experience.<sup>12</sup> Although not particularly popular overall, they are common in congested city environments. A compromise between the two is the compact-front-end, or what is commonly called the “van.” Vans push steering and controls as forward as possible, making for a short but slanted hood. They feature some benefit of both the cabover and conventional design, but

---

Additionally, the market for large trucks (where engine quality matters most) is dominated by Cummins and Caterpillar, which are independent of any manufacturers in our data.

<sup>12</sup>Cab-over-engine vehicles place no distance between the driver and the vehicle in front of it, which creates the safety issue.

are limited in engine size.<sup>13</sup> A final characteristic specific to heavier conventional vehicles is a long-hood design. The long-conventional design provides maximal comfort and minimal noise to drivers making long or difficult hauls. The characteristic is sufficiently important to be listed separately in product guides (e.g. *The Truck Blue Book*).

Table 1.1 summarizes the product data. The top panel provides a summary of the data by year, while the bottom panel aggregates the data into equal nine-year periods. The most striking aspect of the data is that despite rich year-to-year variation in the number and type of products offered as well as units sold, the market exhibits few strong trends (with the exception of heavy cabover and compact-front-end vehicles, discussed at length later on). For example, although quantity swings considerably between years, the bottom panel shows that the size of the commercial vehicle market has grown little in the past 26 years. Relatedly, there is little movement in price over time (after a CPI adjustment to 2005 constant dollars). This suggests (but clearly does not strictly imply) production costs have not fallen much over time, which would not be surprising considering that modular production limits the extent to which large capital investments can automate the manufacturing process. GWR tells a similar story. Together these facts suggest stable preferences and slow long-term growth but large compositional changes.

Individual product types also move quite independently from one another over time. Although the number of models tends to peak during periods of high economic activity, it is clear this is driven almost completely by medium weight vehicles. The next section shows that their sensitivity to the construction industry—rather than the economy overall—drives this relationship. Other vehicle types do not exhibit this relationship. Nonetheless, model-level entry and exit is still quite frequent. In the case of heavy cabover vehicles, something else is clearly at work. In fact, this variation across product types and time is at the core of the identification of sunk costs.

---

<sup>13</sup>For this reason, compact-front-end vehicles take only the four lowest GWR values in the data.

**Table 1.1: Summary statistics**

	All Years		9-YEAR GROUPS		
	Min	Max	1986-1994	1995-2003	2004-2012
<u>Count of Offerings</u>					
Conventional, light-medium	8	10	8.0	7.9	8.3
Conventional, medium	21	35	21.3	24.8	27.1
Conventional, heavy	8	14	8.3	10.8	9.6
Conventional, long option	6	12	8.3	9.7	10.8
Cab-over-engine, light-medium	7	13	12.0	11.8	9.8
Cab-over-engine, medium	3	18	12.7	15.0	9.2
Cab-over-engine, heavy	0	8	6.0	1.8	0.0
Compact-front-end, all GWR	5	9	5.0	5.7	7.2
All vehicles	70	97	81.7	87.3	82.0
<u>Market Outcomes</u>					
Price, CPI-adjusted	\$65,958	\$71,334	\$64,510	\$68,012	\$67,644
Quantity, 000s of units	193	658	467	502	494
<u>Buyer Composition</u>					
Freight-related	27.1%	42.6%	30.9%	35.1%	34.7%
Construction-related	27.2%	50.6%	37.6%	41.3%	38.8%
Bus. & Pers. Service-related	13.3%	44.4%	31.6%	24.4%	26.5%

## Buyer Attributes

Buyers differ by industry, driving environment, and legal climate. Industry matters because it determines the load size and driving distance, which in turn influence preferences over GWR and cab size. According to US Census microdata, over 94% of buyers belong to either the for-hire transportation (“freight”) industry, one of three construction-related industries, or the business and personal service industry. Driving environments vary by whether surrounding areas are urban or not, which I measure by dividing urban road mileage by total road mileage. The third buyer attribute relates to a series of vehicle length laws. Initially states regulated the entire vehicle length, from the front of the power unit to the rear of the load being carried. A series of federal legislative and judicial decisions, beginning with the Surface Transportation Assistance Act, mandated that only the length of the load being carried be regulated. This allowed the power unit, i.e. truck or tractor, to be as long as



the driver desired. Rather than analyze each complicated decision, I take a simple count of the regulatory actions. In Section V, I discuss the law change in more depth and, in section VI, show the measure has good explanatory power for demand.

Table 1.2 highlights the tight link between buyer attributes and product characteristics. The left-most column in each row reports the average product characteristic conditional on purchase by any buyer, while the columns to the right report the average product characteristic conditional on purchase by a particular group of buyers. The first row shows how GWR varies by industry type. Delivery and service firms are purchasing most light commercial vehicles, construction-related firms are purchasing most medium weight units, and freight firms are purchasing the majority of heavy units. It should not surprise readers that, for example, the florist's van handles much lighter loads than the builder's dump truck. The degree of heterogeneity may be of some surprise, although these purchase patterns are not atypical (and are certainly inline with the Petrin [2002] findings that relate family size to minivan purchase). The second row shows that shorter cab types—with better visibility and agility—are preferred by urban buyers. The third row shows that compact-front-end cabs are preferred choice by local delivery firms. In most areas, the noisy and bumpy cabover is an unnecessary choice, although the extra visibility and tighter turning of compact front end vehicles is a help to these drivers.

## **Firms**

This paper focuses on automotive assembly firms. It treats upstream and downstream operations as either independent or completely determined by assembly operations. There are several reasons for this. First, the assemblers are few in number but large in size and serve as the central party to contracts between the disaggregated parts suppliers and geographically diverse dealerships. Unlike the carmakers, who typically build major components in-house or sign exclusive contracts with third-parties, commercial vehicle assemblers incorporate components from a host of suppliers. Whereas carmakers typically build many major components in-house, for example engines and axles, commercial builders

**Table 1.2:** Mean characteristics conditional on buyer type and purchase

		EXPECTED [ CHARACTERISTIC   BUYER TYPE, PURCHASE ]					
	All Industries	Bus. & Pers. Service	Contractor	General Construction	Heavy Building	General Freight	Specialty Freight
GWR	31,818 (16,612)	12,700 (5,263)	21,495 (6,191)	31,494 (4,193)	43,462 (5,216)	51,616 (1,465)	54,277 (1,205)
	All Buyers	+1.5 $\sigma$ Road Density					
Cabover	8%	56%					
	All Industries	Bus. & Pers. Service					
Compact-front-end	14%	46%					

rarely do. Cummins and Caterpillar, for example, account for a majority of heavy-duty truck engines but are not active in the market themselves. Also unlike the passenger vehicle segment, commercial dealers often carry competing brands.<sup>14</sup> One exception is that firms with both commercial and passenger vehicle operations occasionally leverage their passenger vehicle dealerships to sell commercial units (which I account for in estimation).<sup>15</sup> Second, assembly firms map directly to the “brand” that identifies the vehicle. Third, extending the analysis along the value chain is simply beyond the scope of the current data and methods.

US commercial vehicle production is also separate from its foreign market counterparts. The catalyst for this separation is a 25% import tariff on trucks imposed in 1963 and in effect today. The duty is part of Proclamation 3654 and ubiquitously referred to in the auto

<sup>14</sup>For example, the closest commercial vehicle dealership to Cambridge, Massachusetts (as of 12/1/13) sells both Ford and International vehicles. The second closest sells Mack, Western Star, Isuzu, Volvo, Isuzu, and Peterbilt vehicles, all rivals.

<sup>15</sup>In particular, I allow the sunk costs of offering some characteristics to vary by firm. See section VII for details.

industry as the “Chicken Tax” due to the fact that President Johnson aimed it primarily at stemming poultry imports. It applies to all “truck” imports to the United States. Although what is meant by “truck” is hotly contested, all vehicles in this paper are covered. Together with the heavy weight and high shipping costs of commercial vehicles, imports and exports are below 3% of this market.

As of 2012, nine independent parent companies offered fourteen brands of commercial vehicles. General Motors (GM), Ford, and Chrysler, which owns Dodge, are American firms with large passenger segment operations throughout the panel and collectively referred to as the “Big Three.”<sup>16</sup> Volvo and Daimler are European firms, while Hino and Isuzu are Japanese firms. International, also known as Navistar, and Paccar are American firms without passenger segment operations. Table 1.3 reports 2012 market share. The largest brand is Ford while the largest parent firm is Freightliner, which includes the Mitsubishi-Fuso (hereafter “Fuso”) and Western Star brands.

**Table 1.3:** *Descriptive statistics by brand as of 2012*

Brand	Parent Company (for subsidiary brands)	Average GWR	Average Price	Market Share
International	-	39,974	85,130	16.67%
Ford	-	21,555	39,680	12.50%
Freightliner	Daimler	39,994	85,529	12.50%
Peterbilt	Paccar	50,457	118,524	11.11%
GM	-	17,193	33,545	6.94%
Isuzu	-	18,554	40,192	6.94%
Kenworth	Paccar	49,848	118,498	6.94%
Dodge	-	13,139	28,389	5.56%
Fuso	Daimler	20,454	45,729	4.17%
Hino	-	22,044	47,464	4.17%
UD	Volvo	19,228	43,686	4.17%
Volvo	-	47,128	107,163	4.17%
Mack	Volvo	48,647	110,053	2.78%
Western Star	Daimler	52,000	131,854	1.39%

Notes. --- Brands ordered by market share.

<sup>16</sup>GM sells under the Chevrolet and GMC badges and although there is a distinction between these in the passenger segment, there is none in the commercial segment over my panel. I follow *The Truck Blue Book* by combining all GM models.

Some brands have changed owners, although the reasons behind these changes are plausibly exogenous to the US commercial vehicle market. European and Japanese commercial segment assemblers, as well as American passenger segment manufacturers, own subsidiaries in the domestic commercial vehicle market, although US commercial segment sales always comprise a small portion of total sales. Table 1.4 presents seven such changes over the panel. To illustrate with the first entry, Germany-based Daimler purchased Chrysler in 1998. The former owned the Mercedes brand while the latter owned the Dodge, Plymouth, Jeep, and Chrysler brands. At the time of the merger, only 1.6% of Chrysler’s sales were in the commercial vehicle market. It is then unlikely that this acquisition, as well as the others, were driven by concerns related to the assembly operations included in this panel.<sup>17</sup>

**Table 1.4:** *Ownership changes*

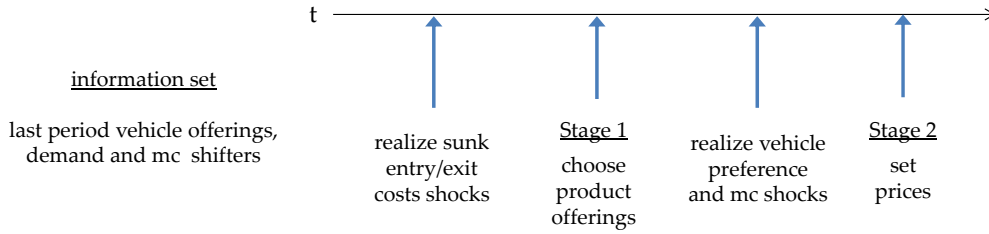
Parent	Action	Target	Year	Target's US CV Sales / Target's Total Sales
Daimler	acquisition	Chrysler	1998	1.6%
Daimler	acquisition	Hyundia Truck	2001	3.0%
Volvo	acquisition	Mack (Renault)	2001	24.1%
Daimler	spinoff	Hyundia Truck	2001	2.5%
Daimler	acquisition	Mitsu. Fuso	2004	4.2%
Daimler	spinoff	Chrysler	2006	1.3%
Volvo	acquisition	Nissan Diesel	2006	6.1%

### 1.3 Model

This section presents a two stage model that captures how firms endogenously adjust the set of products they offer to changing market conditions. In the first stage firms choose product offerings. In the second stage, firms set prices and consumers make purchase decisions. Figure 1.1 describes this timing.

---

<sup>17</sup>The one exception is the acquisition of Renault by Volvo. Renault at the time held a controlling stake in Mack, which accounted for about one-quarter of their combined size. Still, both are based in Europe and mentions of the merger in the annual report tended to focus on the European market.



**Figure 1.1:** *Timing of the game*

Firms solve the problem by working backwards from the second stage: calculate the equilibrium profits that will likely accrue to them under any possible set of product offerings and then choose the products that maximize those profits. For this reason, I also begin with the second stage decisions.

## Demand

Each buyer,  $r$ , decides whether to purchase a vehicle  $j$  from among  $J$  choices or the outside good so as to maximize utility. In the event of purchase, they derive utility from the vehicle based on an interaction between their attributes and the vehicle's characteristics. They also derive disutility from price. The total utility from product  $j$  is given by the following:

$$U_{r,j} = x_j (\beta_x + \beta_x^o z_r^o + \beta_x^u z_r^u) - p_j \beta_p + \zeta_j + \epsilon_{r,j} \quad (1.1)$$

$x_j$  denotes the vector of vehicle characteristics, excluding price. These include a constant, the gross weight rating, and dummies for the cab types and options, as well as an interaction between the gross weight rating and cab-over-engine ("cabover").  $z_r^o$  and  $z_r^u$  denote these buyer attributes, which can be observed or unobserved by the econometrician.  $\beta_x$  denotes the mean taste for each product characteristic, while  $\beta_x^o$  and  $\beta_x^u$  are coefficients on the interaction of buyer attributes and product characteristics.  $\beta_p$  denotes the consumers distaste for price,  $p_j$ . For convenience, let  $\beta$  denote the vector of taste parameters,  $(\beta_x, \beta_x^o, \beta_x^u, \beta_p)$ , and  $z$  denote the vector of both unobservable and observable consumer attributes,  $(z^o, z^u)$ .  $\zeta_j$  denotes a

product-specific preference shock while  $\epsilon_{r,j}$  denotes a preference shock specific to the choice and buyer. Buyers can also consume the outside good, whose mean utility is normalized to zero so that buyers receive only  $\epsilon_{r,0}$ . The distribution of  $\xi$  is only restricted to be i.i.d. across products.  $\epsilon$  is distributed extreme value and is i.i.d. across products and buyers.

This specification yields the familiar logit choice probabilities for each consumer. After integrating out over the total number of simulated consumers,  $ns$ , I arrive at the market share for any product  $j$ :

$$s_j = \frac{1}{ns} \sum_r s_{r,j} = \frac{1}{ns} \sum_r \left( \frac{e^{x_j(\beta_x + \beta_x^0 z_r^0 + \beta_x^u z_r^u) - p_j \beta_p + \xi_j}}{1 + \sum_k e^{x_k(\beta_x + \beta_x^0 z_r^0 + \beta_x^u z_r^u) - p_k \beta_p + \xi_k}} \right) \quad (1.2)$$

Adding a time  $t$  subscript,<sup>18</sup> we have that the product of market share and market size,  $M_t$ , yields total unit sales,  $q_{j,t}$ .

This setup assumes that static, unit demand closely approximates the actual purchasing decisions made and that buyers are price takers. In practice, many buyers do in fact own “fleets” of vehicles, although in most cases they are purchasing only one or two vehicles at a time.

## Pricing

The second stage decision from the firm’s perspective is to set prices. Firms,  $f$ , offering a set of products  $J_{f,t}$ , choose prices to maximize profits, given by:

$$\Pi_{f,t} = \sum_{j \in J(f)} [p_{j,t} - mc_{j,t}] s(x_{j,t}, x_{-j,t}, p_t, z_t; \beta, \xi_t, mc_{j,t}) M_t \quad (1.3)$$

where  $mc_{j,t}$  denotes the marginal costs of producing  $j$  at  $t$ .

This requires a first order condition of the profit function, rearranging terms, and taking other firms’ prices as fixed to arrive at Nash equilibrium prices given by:

$$p_{j,t}^* = mc_{j,t} - \frac{s_{j,t}}{\beta_p} \left[ s_{j,t} - \frac{1}{ns} \sum_r \sum_{k \in J(f)} s_{r,j,t} s_{r,k,t} \right]^{-1} \quad (1.4)$$

---

<sup>18</sup>The applied setting below considers annual decisions, so I use  $t$  and “year” interchangeably.

Marginal costs are a parametric function of observable product characteristics, wages, time, parameter  $\gamma$ , and an unobserved factor specific to the product and time. That is,  $mc_{j,t} = mc(x_{j,t}, w_{j,t}, t; \gamma, \omega_{j,t})$ . For now, I will remain agnostic as to the functional form; later on I will show that since the demand parameters are recovered without information from the supply side, I can be somewhat non-parametric with respect to the exact relationship of marginal costs and its determinants.

## Product Offerings

In the first period, firms choose product offerings, i.e. make model-level entry and exit decisions, with the understanding that their actions and their rivals' actions will impact the second stage. They do not know  $\zeta_t$  and  $\omega_t$  but do know the distribution of the disturbances,  $(F_\zeta, F_\omega)$ , so they form an expectation over them to compute the hypothetical expected profits from any set of production offerings. Then the expected “variable” profits are

$$\pi(J_{f,t}, J_{-f,t}, z_t, w_t, t, p^*; \beta, \gamma, F_\zeta, F_\omega) \equiv \int_{\zeta', \omega'} \Pi_f(J_t, z_t, w_t, t, p^*; \beta, \gamma, \zeta', \omega') dF_{\zeta'} dF_{\omega'} \quad (1.5)$$

The principal decision firms face in the first stage is to weigh the added profits of introducing or continuing to offer existing product models against the sunk costs of doing so. To proceed, I need to take a stand on the nature of sunk costs,<sup>19</sup> given below.

*Sunk Entry/Exit Costs Assumption (I).*

$$SC_{f,j,t} = x_j \tilde{\theta}_{f,x(j),t} \times [\{j \in J_{f,t}, j \notin J_{f,t-1}\} + \{j \notin J_{f,t}, j \in J_{f,t-1}\} \times \frac{1}{\lambda}]$$

The first braces term is an indicator function for products offered this year but not last year, whereas the second braces term is an indicator function for products not offered this year but offered last year. There are two features of sunk costs. First, they are linear in the observable product characteristics, although at this point can freely vary by product space,

---

<sup>19</sup>With sufficiently much data, one could imagine semi-parametric or even non-parametric identification of sunk costs along the lines of Matzkin [1992]. This would eliminate the need for this assumption, but requires more data than is available in this setting.

time, and firm. Second, the sunk cost of adding a model is a multiplicative scaling by  $\lambda$  of the sunk cost of retiring one.

Sunk costs can induce forward-looking behavior, but in differentiated product markets, the dynamic solution requires storage of and an expectation over billions or more of states. This computational burden is orders of magnitude too hard for firms in practice. A common rebuttal to this observation is that although agents do not appear to explicitly optimize, repeated play can nonetheless converge to equilibrium strategies<sup>20</sup>—although learning is also out of the question here. When the product space is rich, fifty years of annual—or even daily decisions—would not provide anywhere close to meaningful convergence over the state space, even if only a recurrent class of states are considered.

This raises an important question as to what managers actually do. Survey data suggests managers cut computational corners.<sup>21</sup> Solutions to these problems suggested by practitioners or research staff in private sector firms often suggest the same. For example, Jeff Alden, group manager of Manufacturing Systems Research at General Motors, and Robert Smith, Professor of Engineering, write in *Operations Research* that "by far the most common planning procedure found in practice is to approximate the solution" [Alden and Smith, 1992].<sup>22</sup> To figure out as accurately as possible what is done in this setting, I interviewed engineers, designers, and veteran managers. A common thread ran through these interviews, the clearest of which was given by the former head of General Motors Commercial Division, who said:

"Each year we look at demand, what we offer, and what the competition is going to offer. We consider changing the lineup like adding a vehicle... We know who

---

<sup>20</sup>Fudenberg and Levine [1998] explore how agents can learn through repeated play. Fershtman and Pakes [2012] provide an empirical model where simple updating converges to equilibrium strategies, and Doraszelski et al. [2014] estimate a related model in the context of electricity markets.

<sup>21</sup>Graham and Harvey [2001] found that CFOs are 2-3 times more likely to use contingent-free methods like the payback ratios and hurdle rates. With respect to adjusting for risk, Summers [1987] showed that 94% of surveyed Fortune 500 firms use the same discount rate across all projects, and that 23% used a discount rate above 19%.

<sup>22</sup>This article provides a simplified single-agent problem and shows how "rolling horizon procedures" approximate the solution.



the customers would be, what we can charge, and the production costs—so we have the added margin. The margin over the investment gives a return on capital, and *we'll build it when it crosses some threshold* (emphasis added)."

Hurdle rates are interpreted here in the following way.

*Capital Budgeting Assumption (II).*

$$\begin{aligned} \{j \in J_{f,t}, j \notin J_{f,t-1}\} &\Leftrightarrow \mathcal{E} \left[ \frac{\pi(J_{f,t}; J_{-f,t}) - \pi(J_{f,t} \setminus j, J_{-f,t})}{x_j \bar{\theta}_{f,t,x}} \middle| \mathcal{I}_{f,t} \right] \geq \text{HurdleRate} \\ \{j \in J_{f,t}, j \in J_{f,t-1}\} &\Leftrightarrow \mathcal{E} \left[ \frac{\pi(J_{f,t}; J_{-f,t}) - \pi(J_{f,t} \setminus j, J_{-f,t})}{x_j \bar{\theta}_{f,t,x} \times \frac{1}{\lambda}} \middle| \mathcal{I}_{f,t} \right] \geq \text{HurdleRate} \end{aligned}$$

The first inequality refers to products not offered by  $f$  at  $t - 1$ , while the second refers to products that are.

Hurdle rates are a straightforward rule-of-thumb. To illustrate, consider some product  $j$  that is not offered by  $f$  at  $t - 1$ . Introducing it at  $t$  would increase expected second-stage profits by \$20 and require a \$100 sunk costs. This action yields a 20% expected (static) return. Firms with a 19% hurdle rate would accept. Firms a 21% hurdle rate would not.

They also may capture a large share of what would be earned under more complex, fully dynamic strategies. In seeking to explain why “most firms do not make explicit use of real option techniques” and “projects are taken based on whether or not IRRs exceed arbitrarily high discount rates,” McDonald [2000] shows that at least in single-agent settings, hurdle rates do quite well. The reason is that because as option value increases, for example due to an increase in volatility, deviations from the optimal strategy are less costly. These hurdle rates are larger than the firms discount rate since these need to capture option value. Second, recall that the profits of adding products in any sub-segment of the market, or the market overall, are not predictably growing or shrinking over time. This presumably mitigates the impact of deterrence, since firms are not looking to move early and foreclose that action from rivals in the future. The second chapter of this dissertation provides simulations that support the use of hurdle rates in entry/exit games by comparing two-period strategies to multi-period dynamic strategies. Finally, I provide calculations showing that what firms actually earned is quite close to and centered around what they expected to earn.

## 1.4 Estimation

### Demand

The estimation of demand closely resembles Petrin [2002]. It minimizes a generalized method of moments objective function based on two sets of moments.<sup>23</sup> The first are constructed as follows. Firms do not know  $\zeta_t$  when they choose product characteristics, so  $\mathbb{E} [\tilde{\zeta}_{j,t} | x_{j,t}] = 0$ . On the other hand, firms do know  $\zeta_t$  when they choose prices. Approximations to the optimal instruments are constructed from variables that shift either marginal costs or markups. Wages are a valid instrument in the former case. I match vehicle models to the areas in which they were assembled and proxy for the factory wage rate with a Bureau of Labor Statistics estimate of the production wage in that area. Production locations are unlikely to relate to current commercial vehicle market conditions since the decision to launch a new facility precedes production by several years and, once launched, the mix of products are rarely re-allocated across factories. The competitive conditions provides a valid instrument in the latter case. The markup on  $j$  produced by  $f$  is decreasing in the number of competing products that are close in characteristic space but increasing in the proportion of these owned by  $f$ . The timing of choices, again, guarantees that  $\mathbb{E} [\tilde{\zeta}_{j,t} | x_{-f,t}] = 0$  and  $\mathbb{E} [\tilde{\zeta}_{-j,f,t} | x_{j,t}] = 0$ .<sup>24</sup> Observed shares are calculated by simply dividing the units sold by market size, the calculation of which is given in the Appendix.

The second set matches first and second moments from the microdata to model-predicted analogs of these moments. Specifically, I choose three subsets of moments to match for each buyer-product relationship: the mean and variance of the buyer attribute conditional on the product characteristic and the probability of purchase conditional on that buyer type. Formally these moments are

---

<sup>23</sup>In practice, I exploited a technique to speed up the parameter search given in Varadhan and Roland [2008]. I then checked the values using the standard contracting mapping proposed in Berry et al. [1995].

<sup>24</sup>Armstrong (2013) makes a strong case for including cost shifters, like production wages, in demand estimation. He shows that as the number of products grows large, markups converge to a constant. This leaves a model with only “BLP instruments,” i.e. those based on competing and own product characteristics, unidentified in the limit.

$$\begin{aligned}
mm_1 &= \mathbb{E}(z_r | r \text{ buys } j) \\
mm_2 &= \mathbb{E} \left[ (z_r)^2 - \mathbb{E}(z_r | r \text{ buys } j) \mid r \text{ buys } j \right] \\
mm_3 &= \mathbb{E}(z_r | r \text{ buys any } j)
\end{aligned}$$

The micro-moments include buyer-product relationships between the following: each industry type and the GWR; each industry type and a constant; the delivery industry and the compact-front-end cab type; the general freight industry and the cabover; the general freight industry interacted with length laws and the cabover; the urban measure and the cabover.

## Marginal Costs

With demand estimates in hand, only the Nash pricing condition is needed to back out marginal costs. This merely requires rearranging the pricing equation given in the previous section:

$$\hat{m}c_{j,t} = p_{j,t} + \frac{s_{j,t}}{\hat{\beta}_p} \left[ s_{j,t} - \frac{1}{ns} \sum_r \sum_{k \in I(f)} \hat{s}_{r,j,t} \hat{s}_{r,k,t} \right]^{-1} \quad (1.6)$$

Notice that  $\hat{m}c_{j,t} = mc(p_{j,t}, s_{j,t}, \hat{\beta}_p, s_{r,j,t})$ . That is, estimated marginal costs are a function of quantities available to the econometrician: prices and shares, which are observed in the data, as well as the price coefficient and the individual purchase probabilities, which are recovered from the demand system.

Much of the prior work has used this equation with an explicit functional form for marginal costs to add supply-side moments to the demand estimation. This is particularly helpful in pinning down the price coefficient, which can often be difficult to instrument for outside data. In contrast, I estimate demand without these assumptions and then confirm the implied price-cost margins are in line with what we see in audited financial data and that the elasticities are sensible. Presumably, rich microdata and an observable marginal cost shifter for marginal costs (wages) help a great deal here. The payoff lies in the fact that we can now analyze rather than assume the marginal cost shape with respect to the

right-hand side variables. With enough data, one could be non-parametric here or provide a formal shape test. In practice, I provide graphical evidence suggesting that the log of marginal costs is linear in the continuous product characteristic. That is, I have that:

$$\ln(mc_{j,t}) = [x_{j,t}, w_{j,t}, t]\gamma + \omega_{j,t} \quad (1.7)$$

Rearranging terms and solving for the additively separable error term provides:

$$\omega_{j,t} = \exp \left( p_{j,t} - \frac{s_{j,t}}{\hat{\beta}_p} \left[ s_{j,t} - \frac{1}{ns} \sum_r \sum_{k \in J_f} \hat{s}_{r,j,t} \hat{s}_{r,k,t} \right]^{-1} \right) - [x_{j,t}, w_{j,t}, t]\hat{\gamma} \quad (1.8)$$

$\gamma$  is estimated via ordinary least squares or a weighted least squares, since  $\omega_{j,t}$  is not known when the firms choose which products to offer. That is, we have  $\mathbb{E} [\omega_{j,t} | x_{j,t}, w_{j,t}, t] = 0$ . The last step is to plug back in for  $\hat{\beta}$  and  $\hat{\gamma}$  and recover an empirical distribution of  $\xi$  and  $\omega$ , which provide  $\hat{F}_\xi$  and  $\hat{F}_\omega$ . Together, these provide unbiased (but potentially measured with error) estimates of the second-stage payoffs firms would expect from offering any alternate set of products,  $\pi(J_{f,t}, J_{-f,t}, z, w; \hat{\beta}, \hat{\gamma}, \hat{F}_\xi, \hat{F}_\omega)$ .

## Sunk Costs

### *Setup.*

The estimation of fixed costs follows the logic of revealed preference. Firms were free to offer any alternative set of products to the ones that appear in the data (i.e. those that were chosen in equilibrium) but did not because these alternatives were less profitable. These alternatives provide intuitive upper and lower bounds on the parameters of interest: fixed costs could not be too low—or else firms would offer more products than appear in the data—and could not be too high—or else firms would offer less products than appear in the data. As a consequence of simultaneous moves, the necessary conditions for the Nash equilibrium provide that firms take rivals' decisions as fixed. Any unilateral deviation should be less profitable in expectation for the firm than the chosen product offerings.

To arrive at inequalities that are linear in observed and estimated quantities, the parameters of interest, and a set of disturbances, some algebra is necessary. Rewrite the

negative of the initial sunk cost term multiplied by the hurdle rate,  $-\tilde{\theta}_{f,x,t} \times \text{HurdleRate}$ , as  $\theta_{f,x,t}$ . If the hurdle rate is unaffected by the policy change, then these terms need not be separately identified.<sup>25</sup> Henceforth, refer to  $\theta_{f,x,t}$  as the “sunk cost” for convenience, rather than the term given in Assumptions I and II. Write these sunk costs as a mean and a deviation away from this mean which is firm, time, and product space specific. That is,  $\theta_{f,t,x}^k = \theta^k + v_{f,t,x}^k$ , where  $k$  refers to either the constant, gross weight rating, or dummy variable for cab-over-engine, compact-front-end, or long-option. Also re-write the expected profits based on the true parameters as the expected profits based on the econometrician’s estimate plus an error, so that  $\Delta\pi(\cdot; \beta, \gamma, F_{\xi}, F_{\omega}) = \Delta\pi(\cdot; \hat{\beta}, \hat{\gamma}, \hat{F}_{\xi}, \hat{F}_{\omega}) + v^{\pi} = \Delta\hat{\pi}(\cdot) + v^{\pi}$ . Last, notice that any comparison of the equilibrium offerings, denoted  $J_{f,t}$ , to an alternative set,  $J'_{f,t}$ , provides an inequality. Three such deviations are particularly helpful: not offering a product that is in  $J_{f,t}$ , offering a product that was in  $J'_{f,t}$ , and substituting one model for another. These yield

$$\mathcal{E} \left[ \Delta\hat{\pi}(J_{f,t}, J_{f,t} \setminus j, J_{-f,t}) + v_{J_{f,t}, J_{f,t} \setminus j; J_{-f,t}}^{\pi} + x_j (\theta + v_{f,t,x_j}) \left( \{j \notin J_{f,t-1}\} + \frac{1}{\lambda} \{j \in J_{f,t-1}\} \right) \middle| \mathcal{I}_{f,t} \right] \geq 0 \quad (1.9)$$

$$\mathcal{E} \left[ \Delta\hat{\pi}(J_{f,t}, J_{f,t} \cup j', J_{-f,t}) + v_{J_{f,t}, J_{f,t} \cup j'; J_{-f,t}}^{\pi} - x_{j'} (\theta + v_{f,t,x_{j'}}) \left( \{j' \notin J_{f,t-1}\} + \frac{1}{\lambda} \{j' \in J_{f,t-1}\} \right) \middle| \mathcal{I}_{f,t} \right] \geq 0 \quad (1.10)$$

$$\mathcal{E} \left[ \Delta\hat{\pi}(J_{f,t}, J_{f,t} \setminus j \cup j', J_{-f,t}) + v_{J_{f,t}, J_{f,t} \setminus j \cup j'; J_{-f,t}}^{\pi} + x_j (\theta + v_{f,t,x_j}) \left( \{j \notin J_{f,t-1}\} + \frac{1}{\lambda} \{j \in J_{f,t-1}\} \right) - x_{j'} (\theta + v_{f,t,x_{j'}}) \left( \{j' \notin J_{f,t-1}\} + \frac{1}{\lambda} \{j' \in J_{f,t-1}\} \right) \middle| \mathcal{I}_{f,t} \right] \geq 0 \quad (1.11)$$

It is easy to see how the equations above can provide bounds on the sunk cost parameters. The inequalities directly reflect the tradeoffs between changes in profits and changes in sunk costs in any product offering decision. Notice here that if the econometrician assumed all disturbance terms are zero, then it is sufficient to merely solve the system of linear inequalities that bound  $(\theta, \lambda)$ . This assumes that the econometrician has the same information that the agents have as well as that agents have perfect information about what they will earn, conditional on their choices. This is both an unreasonable assumption and almost certainly not able to rationalize the data. Notice also that if the econometrician

---

<sup>25</sup>Firms, at least anecdotally, rarely change hurdle rates. For example, *The Economist* reports that “Shell [Oil Company] left its hurdle rates unchanged for two decades until it ‘nudged them down’ in 1997, and now intends to keep them at present levels for years to come.” That said, the failure of one or more major competitors could have a non-trivial impact on the risk, real or perceived, of operating in this market. This would increase the cost of capital and, in turn, the hurdle rate. The intuition for why this affects the analysis is that an increase in risk makes firms value the future less; future profits are less valuable relative to sunk costs paid in the current period, and entry is less likely.

assumed all disturbance terms are unknown to the agents when decisions are made, then it is sufficient to merely minimize the violations of these inequalities. This is still too strong an assumption but will rationalize the data. Relaxing it, however, introduces a classic endogeneity problem: firms enter products when it is less costly to do so. The discussion below addresses this problem.

***Disturbances.***

Three assumptions will identify the sunk costs, which currently can freely vary across product space, firms, and years. The first assumption is that sunk cost disturbances are independently and identically distributed over product space and time.

*Product and Time Disturbances (Assumption III).*

$v_{f,t,x}$  is i.i.d. over  $x$  and  $t$ .

Independence over product space implies that we can re-write as the  $k$ -characteristic specific portion of disturbance  $v_{f,t,x}$  as  $v_{f,t}^k$ , where  $k \in \{con, GWR, COE, CFE, Long\}$ . That is, conditional on  $(f, t)$ , knowing  $v_{f,t}^k$  tells us nothing about  $v_{f,t}^{k'}$  for  $k \neq k'$ .

The second assumption allows for observable heterogeneity in the firm specific portion of the sunk cost disturbances that are based on two important features of the commercial vehicle market. First, congestion and stringent length regulation have made the cabover vehicle ubiquitous in Asia, which may affect brands with their headquarters based in Japan. For example, they may find it cheaper to introduce cabover vehicles, find it more expensive to introduce non-cabover vehicles, or both. For this reason, the constant term and cabover term are allowed to be different for Japan-based brands. Second, the Big Three firms have large assembly operations in the passenger vehicle segment, which are lighter than commercial vehicles. Hence they may have an advantage in introducing light vehicles but a disadvantage in producing heavy vehicles, or both. Because of this, the constant term and GWR term are allowed to be different for Big Three brands. After accounting for these differences, however, I assume the remaining firm specific portion of the sunk cost disturbances is not known by the firms when they make their decisions. For notational ease, denote the portion of disturbances that are and are not in the agents' information sets when

choices are made as  $v_{2,\cdot}$  and  $v_{1,\cdot}$ , respectively.

*Firm Disturbances (Assumption IV).*

For the constant term,  $v_{f,x,t}^{con} = \theta_{Big3}^{con} + \theta_{Japan}^{con} + v_{2,t,x}^{con} + v_{1,f,t,x}^{con}$ . For gross weight rating,  $v_{f,x,t}^{GWR} = \theta_{Big3}^{GWR} + v_{2,t,x}^{GWR} + v_{1,f,t,x}^{GWR}$ . For the cab-over-engine dummy variable,  $v_{f,x,t}^{COE} = \theta_{Japan}^{COE} + v_{2,t,x}^{COE} + v_{1,f,t,x}^{COE}$ . For  $k \in \{CFE, Long\}$ ,  $v_{f,x,t}^k = v_{2,t,x}^k + v_{1,f,t,x}^k$ .

The third assumption states that profit disturbances are not known to the firms when they make their decisions.

*Profit Disturbances (Assumption V).*

$$v_{J_{f,t}, J'_{f,t}, J_{-f,t}}^{\pi} = v_{1, J_{f,t}, J'_{f,t}, J_{-f,t}}^{\pi}$$

The final assumption provides sufficient variation in the instruments to achieve a bound on each side of each parameter.

*Disturbance Support (Assumption VI).*

Let  $z^{g,c} \subset Z$  be the set of demand shifters that comprise the largest share of demand for vehicles with GWR  $g$  and cab type  $c$ .

$(v_{2,t}^{con}, v_{2,t}^{GWR}, v_{2,t}^{COE}, v_{2,t}^{CFE}, v_{2,t}^{Long})$  is bounded such that if  $z_t^{g,c} - z_{t-1}^{g,c} = \operatorname{argmax}\{z_t^{g,c} - z_{t-1}^{g,c}\}$ , then for all  $\tilde{J}$  where  $j(g, c) \notin \tilde{J} \max_f \left\{ \mathcal{E} \left[ \pi(\tilde{J}, \tilde{J} \cup j(g, c), J_{-f,t}) - \theta^{con} - v_{2,t}^{con} - \{c = k\}(\theta^c + v_{2,t}^c) - g(\theta^{GWR} + v_{2,t}^{GWR}) | \mathcal{I}_{f,t} \right] \right\} \leq 0$ .

Taken together, these assumptions provide that the characteristic-specific portion of the sunk costs that vary over time to be observable by firms but not the econometrician. To illustrate, the cabover may be particularly expensive to introduce at time  $t$ , causing firms to add fewer cabovers to their product lineup in this period relative to other cab types. Formally, the product offering decisions,  $J_{f,t}$ , are selected on the  $v_2$  terms, and this will bias estimates if it were ignored. On the other hand, the assumptions provide that knowing the cabover-specific sunk cost disturbance term at  $t$  reveals nothing about the disturbance

term specific to other characteristics at  $t$  or to the cabover at any time other than  $t$ . These assumptions also rationalize the data.<sup>26</sup>

Substituting in to the prior three inequalities with the assumptions yields

$$\begin{aligned}
& \mathcal{E} \left[ \Delta \hat{\pi}(J_{f,t}, J_{f,t} \setminus j, J_{-f,t}) + v_{J_{f,t}, J_{f,t} \setminus j; J_{-f,t}}^{\pi} | \mathcal{I}_{f,t} \right] \\
& + \mathcal{E} \left[ \left( \theta^{con} + \theta_{Big3}^{con} \{Big3\} + \theta_{Japan}^{con} \{Japan\} + v_{2,t,x_j}^{con} + v_{1,f,t,x_j}^{con} \right) \left( \{j \notin J_{f,t-1}\} + \frac{1}{\lambda} \{j \in J_{f,t-1}\} \right) | \mathcal{I}_{f,t} \right] \\
& + \mathcal{E} \left[ x_j^{GWR} \left( \theta^{GWR} + \theta_{Big3}^{GWR} \{Big3\} + v_{2,t,x_j}^{GWR} + v_{1,f,t,x_j}^{GWR} \right) \left( \{j \notin J_{f,t-1}\} + \frac{1}{\lambda} \{j \in J_{f,t-1}\} \right) | \mathcal{I}_{f,t} \right] \\
& + \mathcal{E} \left[ x_j \left( \theta^{COE} + \theta_{Japan}^{COE} \{Japan\} + v_{2,t,x_j}^{COE} + v_{1,f,t,x_j}^{COE} \right) \left( \{j \notin J_{f,t-1}\} + \frac{1}{\lambda} \{j \in J_{f,t-1}\} \right) | \mathcal{I}_{f,t} \right] \\
& + \mathcal{E} \left[ x_j \left( \theta^{CFE} + v_{2,t,x_j}^{CFE} + v_{1,f,t,x_j}^{CFE} \right) \left( \{j \notin J_{f,t-1}\} + \frac{1}{\lambda} \{j \in J_{f,t-1}\} \right) | \mathcal{I}_{f,t} \right] \\
& + \mathcal{E} \left[ x_j \left( \theta^{Long} + v_{2,t,x_j}^{Long} + v_{1,f,t,x_j}^{Long} \right) \left( \{j \notin J_{f,t-1}\} + \frac{1}{\lambda} \{j \in J_{f,t-1}\} \right) | \mathcal{I}_{f,t} \right] \geq 0 \quad (1.12)
\end{aligned}$$

$$\begin{aligned}
& \mathcal{E} \left[ \Delta \hat{\pi}(J_{f,t}, J_{f,t} \cup j', J_{-f,t}) + v_{J_{f,t}, J_{f,t} \cup j'; J_{-f,t}}^{\pi} | \mathcal{I}_{f,t} \right] \\
& - \mathcal{E} \left[ \left( \theta^{con} + \theta_{Big3}^{con} \{Big3\} + \theta_{Japan}^{con} \{Japan\} + v_{2,t,x_{j'}}^{con} + v_{1,f,t,x_{j'}}^{con} \right) \left( \{j' \notin J_{f,t-1}\} + \frac{1}{\lambda} \{j' \in J_{f,t-1}\} \right) | \mathcal{I}_{f,t} \right] \\
& - \mathcal{E} \left[ x_{j'}^{GWR} \left( \theta^{GWR} + \theta_{Big3}^{GWR} \{Big3\} + v_{2,t,x_{j'}}^{GWR} + v_{1,f,t,x_{j'}}^{GWR} \right) \left( \{j' \notin J_{f,t-1}\} + \frac{1}{\lambda} \{j' \in J_{f,t-1}\} \right) | \mathcal{I}_{f,t} \right] \\
& - \mathcal{E} \left[ x_{j'}^{COE} \left( \theta^{COE} + \theta_{Japan}^{COE} \{Japan\} + v_{2,t,x_{j'}}^{COE} + v_{1,f,t,x_{j'}}^{COE} \right) \left( \{j' \notin J_{f,t-1}\} + \frac{1}{\lambda} \{j' \in J_{f,t-1}\} \right) | \mathcal{I}_{f,t} \right] \\
& - \mathcal{E} \left[ x_{j'}^{CFE} \left( \theta^{CFE} + v_{2,t,x_{j'}}^{CFE} + v_{1,f,t,x_{j'}}^{CFE} \right) \left( \{j' \notin J_{f,t-1}\} + \frac{1}{\lambda} \{j' \in J_{f,t-1}\} \right) | \mathcal{I}_{f,t} \right] \\
& - \mathcal{E} \left[ x_{j'}^{Long} \left( \theta^{Long} + v_{2,t,x_{j'}}^{Long} + v_{1,f,t,x_{j'}}^{Long} \right) \left( \{j' \notin J_{f,t-1}\} + \frac{1}{\lambda} \{j' \in J_{f,t-1}\} \right) | \mathcal{I}_{f,t} \right] \geq 0 \quad (1.13)
\end{aligned}$$

---

<sup>26</sup>The  $v_1$  terms rationalize the data, i.e. ensure the model does not over-fit. This implies firms are surprised by the net profits they receive, and occasionally have ex post regret (but overall are still right on average). Technically, this would be satisfied if firms choose to offer  $J_{f,t}^*$  but instead receive  $J_{f,t}$ , which the econometrician observes. For example, upon entering a product  $j$  with  $x_j^{GWR^*}$  GWR, it receives  $x_j^{GWR} = x_j^{GWR^*} + error_{j,f,t}$ . A more detailed discussion that relates disturbance assumptions to the underlying data generating process is found in Pakes (2010).



$$\begin{aligned}
& \mathcal{E} \left[ \Delta \hat{\pi}(J_{f,t}, J_{f,t} \setminus j \cup j', J_{-f,t}) + v_{J_{f,t}, J_{f,t} \setminus j \cup j'; J_{-f,t}}^{\pi} | \mathcal{I}_{f,t} \right] \\
& \quad + \mathcal{E} \left[ \left( \{j \notin J_{f,t-1}\} + \frac{1}{\lambda} \{j \in J_{f,t-1}\} - \{j' \notin J_{f,t-1}\} + \frac{1}{\lambda} \{j' \in J_{f,t-1}\} \right) \right. \\
& \quad \quad \left. \left( \theta^{con} + \theta_{Big3}^{con} \{Big3\} + \theta_{Japan}^{con} \{Japan\} + v_{2,t,x_j}^{con} - v_{1,f,t,x_{j'}}^{con} \right) \right. \\
& \quad + \left( x_j^{GWR} (\{j \notin J_{f,t-1}\} + \frac{1}{\lambda} \{j \in J_{f,t-1}\}) - x_{j'}^{GWR} (\{j' \notin J_{f,t-1}\} + \frac{1}{\lambda} \{j' \in J_{f,t-1}\}) \right) \\
& \quad \quad \left( \theta^{GWR} + \theta_{Big3}^{GWR} \{Big3\} + v_{2,t,x_j}^{GWR} + v_{1,f,t,x_{j'}}^{GWR} \right) \\
& \quad + \left( x_j^{COE} (\{j \notin J_{f,t-1}\} + \frac{1}{\lambda} \{j \in J_{f,t-1}\}) - x_{j'}^{COE} (\{j' \notin J_{f,t-1}\} + \frac{1}{\lambda} \{j' \in J_{f,t-1}\}) \right) \\
& \quad \quad \left( \theta^{COE} + \theta_{Japan}^{COE} \{Japan\} + v_{2,t,x_j}^{COE} + v_{1,f,t,x_{j'}}^{COE} \right) \\
& \quad + \left( x_j^{CFE} (\{j \notin J_{f,t-1}\} + \frac{1}{\lambda} \{j \in J_{f,t-1}\}) - x_{j'}^{CFE} (\{j' \notin J_{f,t-1}\} + \frac{1}{\lambda} \{j' \in J_{f,t-1}\}) \right) \\
& \quad \quad \left( \theta^{CFE} + v_{2,t,x_j}^{CFE} + v_{1,f,t,x_{j'}}^{CFE} \right) \\
& \quad + \left( x_j^{Long} (\{j \notin J_{f,t-1}\} + \frac{1}{\lambda} \{j \in J_{f,t-1}\}) - x_{j'}^{Long} (\{j' \notin J_{f,t-1}\} + \frac{1}{\lambda} \{j' \in J_{f,t-1}\}) \right) \\
& \quad \quad \left( \theta^{Long} + v_{2,t,x_j}^{Long} + v_{1,f,t,x_{j'}}^{Long} \right) \geq 0 \quad (1.14)
\end{aligned}$$

### **Identification.**

The equations above are now additively separable in quantities available to the econometrician, the disturbance terms, and the parameters of interest. Together with sufficient variation in a set of instruments discussed below, these equations provide upper and lower bounds on  $\theta$  and  $\lambda$ .

The intuition for these bounds is as follows. The first step is to bound  $(\theta^{con}, \lambda)$ , conditional on  $\theta^{GWR}$ . Construct four inequalities based on any firm at any time taking the following actions with respect to any GWR conventional cab vehicle: one more product than it did, add one less, drop one more, and drop one less. (Note that the conventional cab is chosen only as an example and because it is the cab type represented by the constant term.) Suppose each action was a possibility. This would create one inequality, or bound, on either side of the parameters for each firm, time, and GWR value. An expectation over those inequalities would average the additively separable disturbance terms across the data to their mean. Since the disturbances are independent of past offerings and independently distributed

over time, they average out to their unconditional mean, which is by construction zero. This requires that each action was possible, a point revised below. The next step is to bound  $\theta^{GWR}$ , conditional on  $(\theta^{con}, \lambda)$ . Construct two inequalities based on the fact that any conventional cab vehicle that was added could have been designed with either one more unit or less unit of GWR. Suppose each firm in each period enters at least one conventional cab vehicle for each GWR. Once again, the additively separable disturbance terms average out to their unconditional mean, which are zero.

In the data, however, all firms do not add, drop, or even offer each GWR cab in each period. At least in the case of the conventional cab, however, one or more firms always offer at least one of each GWR. This means that in all  $t$ , at least one firm could drop one more product of each GWR. Since the part of sunk costs observable to the firm but not the econometrician does not vary by firm, a weighting scheme can be devised to average out the disturbance terms to zero. Nonetheless, there are still periods when *no firm* adds or drops a given GWR conventional cab vehicle. To omit these periods from the expectation would be to select on periods where  $v_{2,t}$  are particularly high, preserving the classic endogeneity/selection problem. To be clear, what precisely is needed is an instrument that selects periods where a given vehicle type is almost surely added by some firm without selecting on the disturbance terms. In fact, the demand shifters satisfy *exactly* this condition. Recall that the microdata reveal that demand for each vehicle type can be tied to a set of exogenous demand shifter in  $Z$ . To condition on sufficiently large increases in this subset of profit shifters is to ensure entry of this product type in this period. Similarly, to condition on a sufficiently large drops is to ensure exit. However, selecting on  $Z$ , or subsets or functions thereof, does not affect the distribution of the disturbances. This provides bounds for  $(\theta^{con}, \theta^{GWR}, \lambda)$ . An analogous routine identifies the remaining parameters.

The objective function is a sum of squared violations of the inequality conditions, weighted by the inverse of the moments. Formally, let

$$Q_n(\theta) = \sum_k \left( \sigma_k(W_i, \theta)^{-1} (m_k(W_i, \theta))_- \right)^2 \quad (1.15)$$

where  $(\cdot)_-$  denotes the negative portion of the quantity within the parenthesis,  $m_k(W_i, \theta)$  denotes moment  $k$  for data  $W_i$ , and  $\sigma_k(W_i, \theta)$  represents the square root of the variance of that moment. A moment  $m_k(W_i, \theta)$  is the expectation of the interaction of instrument  $k$ , given by  $h_k(Z)$  and the left-hand side of the inequalities.  $Q$  is common in recent empirical and econometric work. Standard efficiency concerns suggest over-weighting the most informative moments by scaling with inverse variance. These do not, of course, impact the bias of the estimator, since  $\mathbb{E}[m_k(W_i, \theta)] \geq 0$  is equivalent to  $\mathbb{E}[\zeta_k m_k(W_i, \theta)] \geq 0$  as long as the scaling parameter  $\zeta$  does not vary within moment  $k$ . It can also be seen as analogous to a generalized method of moments estimator where the correlation between the moments is ignored. The parameter search then merely satisfies  $\hat{\Theta} = \arg \min_{\theta \in \Theta} \{Q_n(\theta)\}$ .

#### *Inference.*

I construct sets in which the true sunk cost parameters will lie 95% of the time. Inference based on inequalities is somewhat less straightforward than inference based on equalities—e.g. generalized method of moments—because inequalities provide only one-sided restrictions. The most informative of these is the *least upper bound* and the *greatest lower bound*. As such, these bounds represent a minimum and maximum, respectively, of the moments, rather than an average. This rules out the use of the central limit theory to provide a direct formula for standard errors.

I follow Andrews and Soares [2010] in constructing these sets but with shifted rather than selected moments. The process is summarized below. For details, I refer to their paper (and for readers uninterested in this section, it can be skipped without a loss in continuity). In short, they suggest inverting a test, as in Chernozhukov et al. [2007]. Informally, begin with any suitable parameter guess. Compute the objective function. Take nonparametric bootstrap samples from the underlying data. Compute the objective function for each of these. Form an appropriate critical value at a  $1 - \alpha$  confidence level. Then simply accept any parameter guess for which the objective function is below the critical value. This process results in asymptotically correct confidence intervals.

Andrews and Soares [2010] show, however, this method by itself can have poor power

properties and, in my calculations, it tended to produce very wide confidence sets. The problem stems from cases where some moments are satisfied by a wide margin. Loosely speaking, these moments will cause the econometrician to accept any non-perverse parameter guesses. To illustrate, suppose that all but one moment pertaining to the upper bound of the sunk cost parameter is close to binding while the last moment is very slack. That is, all but one moment provides a moment close to  $y - \bar{\theta}^{Enter} \geq 0$  but the last provides  $1000 \times y - \bar{\theta}^{Enter} \geq 0$ . The last moment is not very informative, and the lower bound should be near  $y$ . Yet, the last moment ensures that  $Q(\theta)$  will be below the bootstrapped objective function values for most  $\theta$  below  $1000 \times y$ .

One solution is to remove such moments from the estimation process, letting the data guide the selection. The inequalities above suggest selecting moments where  $\zeta_{n,k}(\theta) > 1$  where  $\zeta_{n,k}(\theta) = \kappa_n^{-1} n^{1/2} \hat{D}_{n,k}^{1/2}(\theta) \bar{m}_{n,k}(\theta)$ .  $\zeta_{n,k}$  is the slackness measure of moment  $k$  computed over  $n$  observations.  $\bar{m}_{n,k}(\theta)$  is sample moment  $k$ , and  $\hat{D}_{n,k}(\theta)$  comprises the diagonal elements of the sample variance of the moments, i.e.  $\hat{D}_n(\theta) = \text{Diag}(\hat{\Sigma}_n(\theta))$  with  $\hat{\Sigma}_n = n^{-1} \sum_i (m(W_i, \theta) - \bar{m}(W_i, \theta)) (m(W_i, \theta) - \bar{m}(W_i, \theta))'$ . They suggest  $\kappa = (\ln(n))^{1/2}$ .<sup>27</sup> The intuition for this selection is to disregard moments that are sufficiently slack as measured by their standardized distance away from zero.

I test the null hypothesis for all possible values in the parameter space,  $\Theta$ . To speed up the search, a large and sparse grid was started with and then iteratively made smaller and more granular. This was essential. The parameter space covers six dimensions, so even twenty points per parameter equates to nearly 11,400,000 value calculations. Each of these has a bootstrap of size 1,000, although the moments are linear so the calculations are fast. Some quantities can be precomputed and, of course, the computation can be run in parallel.<sup>28</sup>

---

<sup>27</sup>For notational comparability, note that I use  $\zeta$  and  $Q$  in place of  $\xi$  and  $S_{(2)}$ , which appear in Andrews and Soares (2010).

<sup>28</sup>This search was run in MATLAB using the combined efforts of two Dell Precision 7500 terminals, each with two Intel Xeon 3.47GHz processors and 56GB of RAM. This is a powerful machine (as of 2014). The alternative was to run these in parallel across a high powered cluster. Since the inference procedure consists only of affine transforms and sorts of the data, I preferred to run sub-spaces of  $\Theta$  in a loop, with the data organized over

Unlike Andrews and Soares (2010), I shift rather than select the moments (as in Pakes, Porter, Ho, and Ishii (2014)). This requires adding  $\zeta_{n,k}(\theta)$  back to the sample moments under both the null and alternative. For each bootstrap sample, I evaluate  $Q$  and construct a critical value  $\hat{c}_n(\theta_0, 1 - \alpha)$  where  $\alpha = 5\%$ . I accept if the computed test statistic,  $T_n(\theta) = Q(n^{1/2}\bar{m}_n(\theta), \hat{\Sigma}_n(\theta))$  is less than this critical value and reject otherwise. The true parameter vector  $\theta_0$  will lie in the space defined by the intersection of the estimated values of  $[\underline{\theta}, \bar{\theta}]$  and  $[\underline{\lambda}, \bar{\lambda}]$ .

## 1.5 Descriptive Evidence

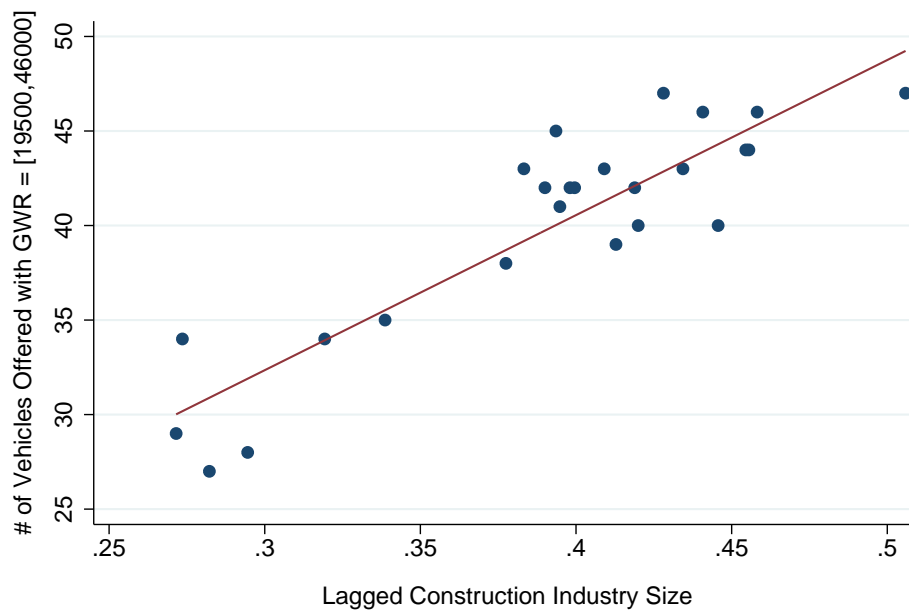
Purchasing patterns from the microdata presented in Section III have already demonstrated strongly heterogeneous preferences among the buyers. What remains to be seen is whether firms are actually adjusting their product characteristics to changes in buyer composition. This would provide support for the notion that firms are solving a problem close in nature to the one presented in Section IV, but would also suggest a set of instruments to identify sunk costs. To illustrate, suppose the construction industry expands rapidly, as it did between 2005 and 2007, and then contracts steeply, as it did in the period subsequent to that. The microdata suggests this group prefers only a small subset of the vehicles produced, so firms adapting to market conditions should expand and contract product offerings tailored to their needs over the same periods. For estimation purposes, it is clear that the introduction of, say, four vehicles of this sort will imply lower sunk costs than three vehicles but higher costs than five vehicles. Algebraically, the expansion in the construction industry can be presented by a change from  $z$  to  $z'$ . If  $j$  is a vehicle that is highly preferred by construction-related industry buyers, then  $\pi(J_{f,t} \cup j, J_{f,t}, J_{-f,t}, z', \cdot) > \pi(J_{f,t} \cup j, J_{f,t}, J_{-f,t}, z, \cdot)$ . If the difference between the left- and right-hand side expressions is large enough,  $f$  or one of its rivals is likely to enter.

Encouraging reduced-form evidence that directly relates the instruments and outcomes

---

observations, moments, and the bootstrap draws.

is presented in Figures 1.2-1.5.

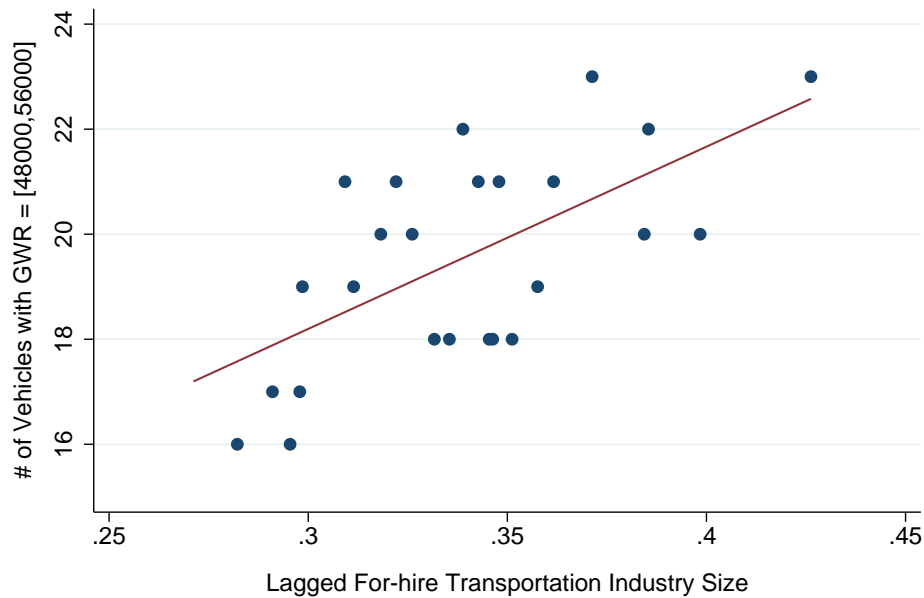


**Figure 1.2:** *Evidence from the construction industry*

I follow the example above and begin with construction-related industries. These buyers account for under 40% of total purchases but over 80% of the sales of vehicles with a GWR between 19,500 and 40,000. Figure 1.2 shows that the number of offerings in this sub-segment is quite closely tied to the industry. In fact, this data was already presented in Table 1.1, although its significance was not obvious at the time. For example, there is a steep increase in medium weight offerings from the early- to mid-2000s and an equally steep decline in the late-2000s.

The freight industries provide the second piece of evidence. These industries again account for only about 40% of total purchases but well over 90% of the sales of vehicles with a GWR above 48,000. Figure 1.3 shows the number of offerings in this segment is again closely linked to the industry.

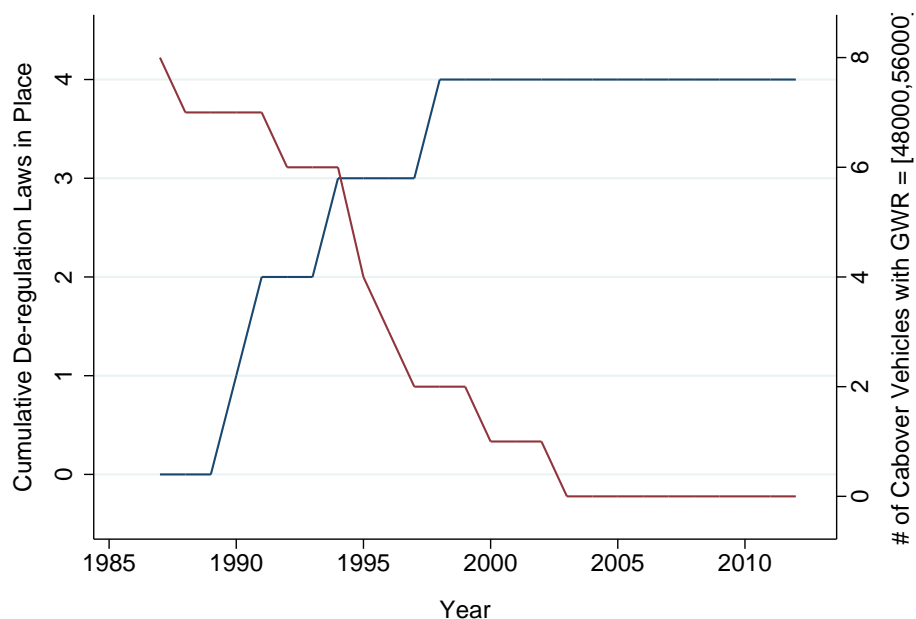
The deregulation of cab length provides the final evidence. For the early and middle part of the 20th century, states independently regulated the use of their highways. One restriction states imposed was a limit on the combined length of vehicle and trailer. This affected freight



**Figure 1.3:** *Evidence from the freight industry*

companies, but had virtually no effect on local service or construction firms, which tend to carry exclusively small loads. Strict length laws advantaged the cab-over-engine relative to conventional vehicles since short cabs translated directly into larger loads and higher revenue. Regulations varied considerably and in many cases created blocks to interstate commerce in some regions of the country. Beginning with the Surface Transportation Assistance Act of 1982, the federal government began to standard the maximum legal load being carried and at the same time deregulated the length of the vehicle pulling it. Heavy cabover sales, once favored by the freight industry in some states, were crushed. Since the process unraveled slowly and modeling its idiosyncrasies are beyond the scope of this paper, I construct an instrument for deregulation from a simple count of the relevant legislative and court decisions.

Figure 1.4 shows the product response to this deregulation. The increasing line, which corresponds to the primary y-axis, represents a count of the relevant administrative, legal, and legislative actions on the length laws. The decreasing line, which corresponds to the secondary y-axis, corresponds to the number of heavy cabover offerings. Although not a



**Figure 1.4:** Evidence from length deregulation

one-to-one mapping, the impact on offerings is clear. There are eight such offerings at the beginning of the panel, but this segment disappears by 2003.

## 1.6 Results

### Demand

Table 1.5 reports the results from estimation of the demand system. Note that coefficients referencing the cab-over-engine, compact-front-end, and long-option are relative to the conventional cab, which is the omitted discrete category. All parameters are estimated very precisely (with the exception of the long-option) and this is especially true for interaction terms, whose precision is greatly aided by the microdata.<sup>29</sup> A few parameters deserve discussion. First, GWR is positive for all buyers. This is an important check of the model, since price is always increasing in GWR and a negative coefficient would imply these

<sup>29</sup>I began with a large set of potential interactions and dropped those that were consistently neither statistically significant nor impactful. Micro-data summary statistics, however, gave a strong indication as to which interactions would ultimately be included.



buyers should choose an alternate (lighter) means of transportation. Moreover, the industry interactions are ordered in precisely the same way as the microdata would suggest. Specialty freight buyers, often called “heavy haul” firms, value GWR the most. Business and personal service firms like bulk couriers and local delivery firms, represented by the constant (since it is the omitted industry type), value it the least.

**Table 1.5: Demand estimates**

	Mean $\beta$		Interactions		
	Estimate	SE	Buyer Type	Estimate	SE
GWR (0,000s lbs.)	0.60***	0.13	SPECIALTY FREIGHT INDUSTRY	21.88***	5.78
			GENERAL FREIGHT INDUSTRY	17.90***	5.35
			HEAVY BUILDING INDUSTRY	10.70***	2.54
			GEN. CONSTRUCTION INDUSTRY	6.68**	2.26
			CONTRACTOR INDUSTRY	3.61***	1.00
Cab-over-engine	-1.85**	0.62	URBAN	13.96***	2.62
			GENERAL FREIGHT INDUSTRY	15.93***	4.43
			GENERAL FREIGHT INDUSTRY X LAW	-4.25**	1.83
Compact-front-end	-24.58**	8.37	LOCAL DELIVERY INDUSTRY	29.67***	7.84
Long-conventional	-0.56	0.39	FREIGHT INDUSTRY X N(0,1)	10.49***	2.05
Cabover X GWR	-3.92***	0.94	---		
Constant	-18.13***	5.28	SPECIALITY FREIGHT INDUSTRY	-89.37***	14.45
			GENERAL FREIGHT INDUSTRY	-67.57***	15.36
			HEAVY BUILDING INDUSTRY	-23.54***	7.71
			GENERAL CONSTRUCTION INDUSTRY	-11.07**	4.87
			CONTRACTOR INDUSTRY	-4.33**	1.86
			LOCAL DELIVERY INDUSTRY	-5.44***	1.21
			N(0,1)	20.74**	7.31
Price (\$000s)	-0.35***	0.06	---		

Note. - Standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Second, the cabover is disliked by the average buyer because of its cramped, bumpy, and noisy ride. These problems are exacerbated by heavy loads, which reflected in the negative interaction with GWR. Yet, the cabover is strongly preferred by urban buyers, who value its agility and visibility, and this is reflected in the large positive interaction with the road density measure. This squares with casual observations: the cabover is an uncommon site in the United States except for areas like downtown Manhattan or Chicago, where these

vehicles are ubiquitous. The impact of the length regulation, which primarily bound freight carriers, is also evident here. The interaction of the freight industry buyer dummy with the cab-over-engine is positive, although the further interaction with length deregulation is very negative.

Buyers dislike price. The coefficient is precisely estimated but difficult to interpret in units of utility. I translate this into more easily understood measures in the following section.

### **Measuring Fit for Demand**

The demand system implies a mean price elasticity of demand equal to 2.23. General Motors suggested an overall price elasticity of demand for passenger vehicles of 1.0 in Berry et al. [2004], so my estimates suggest the commercial vehicle market is more elastic. This is not surprising. Commercial vehicle buyers tend to be small businesses that are price-sensitive and unmoved much by styling, color, or brand prestige.

They also imply an average price-cost margin of 8.95%. Since eight out of nine parent companies are public entities with audited financial statements, we can simply compare against the reported figures. A few caveats are necessary. First, all of these firms operate in either other product markets or other geographic areas or both. Paccar is active in Australia, Isuzu and Hino are based in Japan, Volvo makes motorboat engines, Ford makes passenger vehicles, et cetera. Second, I estimate marginal costs while the firms report average costs. I cannot tell, for example, what proportion of the Selling, General, and Administrative (“SG&A”) line-item costs are sunk rather than marginal and what portion are related to fixed headquarters activities. I should see, however, that my reported price-cost margin falls between the gross margin, which does not include SG&A, and pre-tax operating margin, which does. These bounds are reported in Figure 1.5 . The vertical line represents the average price-cost margin computed from the model, 8.95%, which falls between the upper and lower bounds reported in audited financial statements (averaged over the period 2007 to 2012) for all but one firm. Although SG&A is a significant portion of gross margin, making for wide bounds, these nonetheless provide a sense of fit of the model overall.

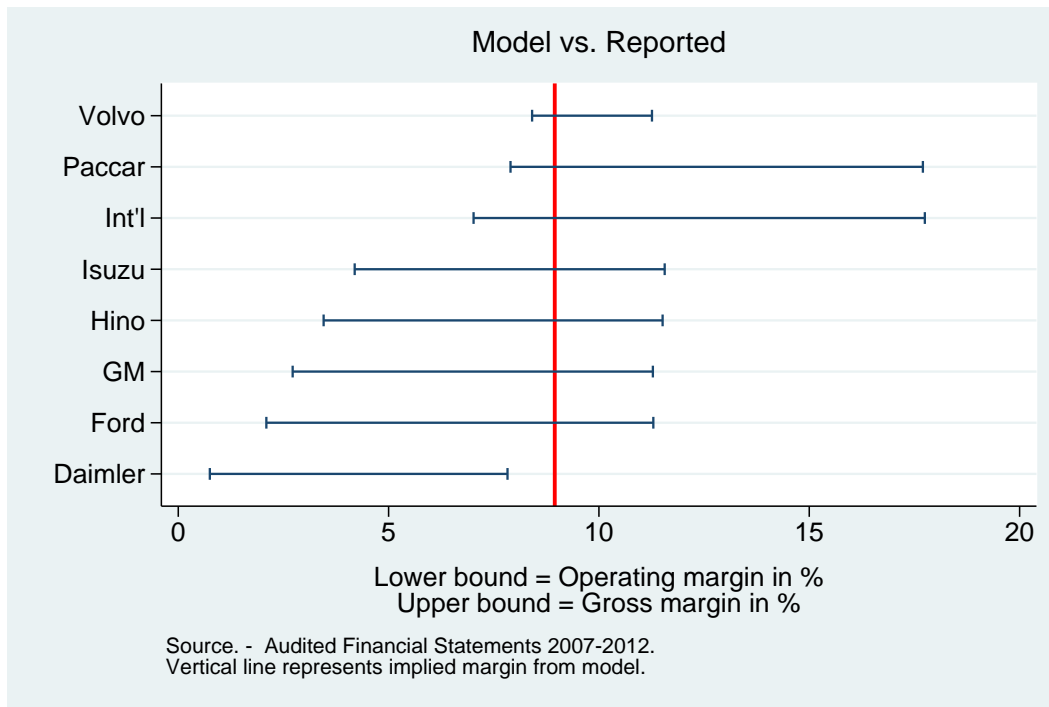
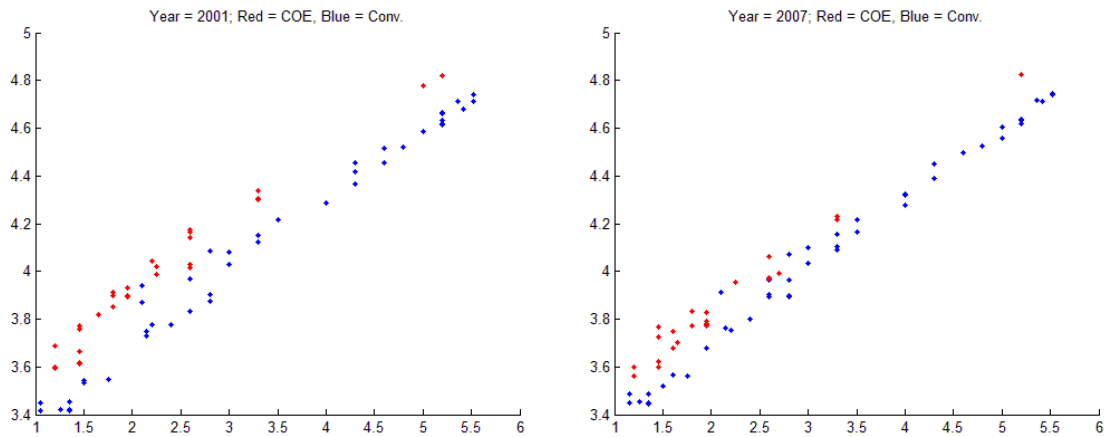


Figure 1.5: Profit margin comparison

## Marginal Costs

Using demand estimates and data, I back out marginal costs and compare them to gross weight rating (the continuous characteristic) conditional on cab type. Figure 1.6 provides two examples from representative “bust” and “boom” years. The left panel is from 2001 while the right panel is from 2007. In both panels, the x-axis measures GWR (in 0,000s of lbs) while the y-axis measures the log of marginal costs (where marginal costs are in \$0,000s). The compact-front-end and long-option do not exhibit much variation in weight, so these were not included. Two things are apparent: the log of marginal costs are linear in GWR and that cab-over-engine vehicle marginal costs are a positive additive shift upwards from the conventional cab. Other years looked similar. This provides some confidence in assuming that the log of marginal costs is linear in the observable regressors.

Table 1.6 now reports the estimated marginal cost coefficients. All are precisely estimated. A 10,000 lb increase in GWR translates to about a 40% increase in marginal cost. Adding the long-option also contributes to higher marginal costs, as do higher wages. The time



**Figure 1.6:** Plot of log of marginal costs on GWR

coefficient is small but positive and significant.

**Table 1.6:** Marginal cost estimates

Marginal Cost Component	Log of marginal costs	
	Estimate	SE
Cab-over-engine dummy	0.176***	0.037
Compact-front-end dummy	-0.187***	0.035
Long-option dummy	0.073***	0.024
GWR (0,000s lbs)	0.392***	0.081
Hourly Wage (\$)	0.018***	0.004
Time	0.009**	0.004
Constant	2.033***	0.391
No. of Observations: 1928		

*Note.* - Robust standard errors given at right.

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

## Sunk Costs

Table 1.7 reports sunk cost estimates of model-level entry and exit. The constant term here refers to a zero-GWR conventional cab. Five things are important to check here. First, confidence intervals that contain the true parameter 95% of the time are both non-empty and reject zero. Second, they imply positive fixed costs for all product configurations I

see in the data. Third, firms with their headquarters in Japan face lower sunk costs to introduce the cab-over-engine but higher sunk costs for all other models. The likely cause is design spillovers from the home market, where the cabover is ubiquitous. Fourth, the Big Three face relatively lower sunk costs to introduce low GWR vehicles but relatively higher sunk costs to introduce high GWR vehicles. The break-even point where the Big Three face essentially identical costs is around 33,000 lbs. Finally, the long option adds to the sunk cost of introduction. It is important to recover a non-trivial positive cost here, since this characteristic is a non-standard extension of the cabin and/or hood that requires the modification of parts and designs.

**Table 1.7:** *Sunk cost estimates*

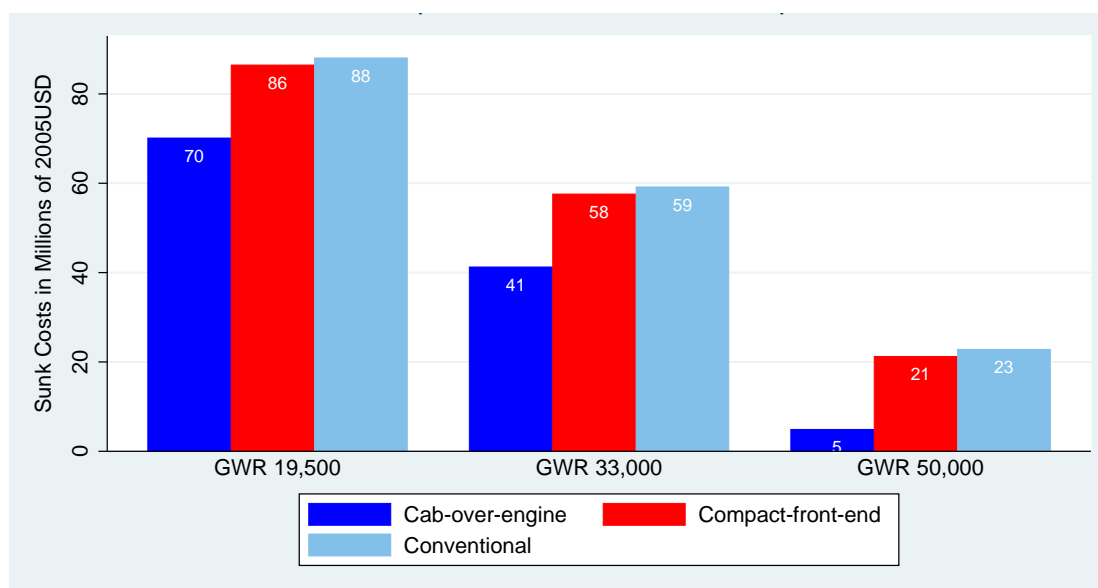
Characteristic	Parameter	Point Estimate	95% Conf. Interval*
constant	$\theta^0$	[ \$129.73 ]	[ \$117.30 , \$144.06 ]
Cab Over Engine	$\theta^{COE}$	[ -\$17.92 ]	[ -\$16.17 , -\$20.58 ]
Compact Front End	$\theta^{CFE}$	[ -\$1.59 ]	[ -\$1.31 , -\$1.75 ]
Long Option	$\theta^{Long}$	[ \$33.89 ]	[ \$30.25 , \$38.70 ]
GWR (0,000s lbs.)	$\theta^{GWR}$	[ -\$21.38 ]	[ -\$19.05 , -\$25.10 ]
constant X BigThree	$\theta^{Big3}$	[ -\$45.22 ]	[ -\$40.77 , -\$51.82 ]
constant X Japan	$\theta^{Japan}$	[ \$6.69 ]	[ \$5.50 , \$8.12 ]
Cab Over Engine X Japan	$\theta^{COE,Japan}$	[ -\$15.00 ]	[ -\$13.46 , -\$17.90 ]
GWR X BigThree	$\theta^{GWR,Big3}$	[ \$14.21 ]	[ \$12.46 , \$16.22 ]
Scaling for Exit	$1/\lambda$	[ -0.386 ]	[ -0.278 , -0.426 ]

*Note.* All figures in millions of constant 2005USD.

\* Probability that the true parameter  $\theta$  lies in this space.

For perspective, Figure 1.7 reports sunk costs across product space for the baseline firms, i.e. those that are neither headquartered in Japan nor a member of the Big Three. As an example, the sunk cost of introducing a conventional cab with a GWR of 33,000 and without the long option is about \$60 million. Sunk costs fall with GWR, which is surprising at first since more rugged vehicles seem to be more complicated to build and design. In practice, however, GWR affects assembly through the quality of the parts—mainly the strength of

the steel chassis, the weight rating and number of axles, and durability of the transmission. These drive marginal costs, which, recall, increase nearly 40% for every 10,000 increase in GWR. Differences in sunk costs across GWR, however, may load on distribution, marketing, selling, and related expenses. Reaching buyers of low-GWR vehicles (e.g. bulk couriers, local delivery businesses, landscapers, and moving companies) may be more costly and challenging than reaching long-haul and heavy-haul freight carriers, who are simply more readily informed and interested in new models, and this could explain the difference.



**Figure 1.7:** Comparisons across product space

For a similar perspective, Figure 1.8 presents sunk costs across the firm types, contrasting Big Three and Japan-headquartered firms against the rest. These illustrate the cost differences across GWR in the former case and cab-over-engine vehicles in the latter.

### Assessing the Capital Budgeting Rule

Sunk costs can be compared against profit estimates, conditional on how long firms kept products in the market, to assess the capital budgeting decisions and implied hurdle rates. The estimation is based on Assumption II, which states that for a product  $j$  offered at  $t$  but not at  $t - 1$ , the agent's expected difference in profits, earned in perpetuity at the hurdle

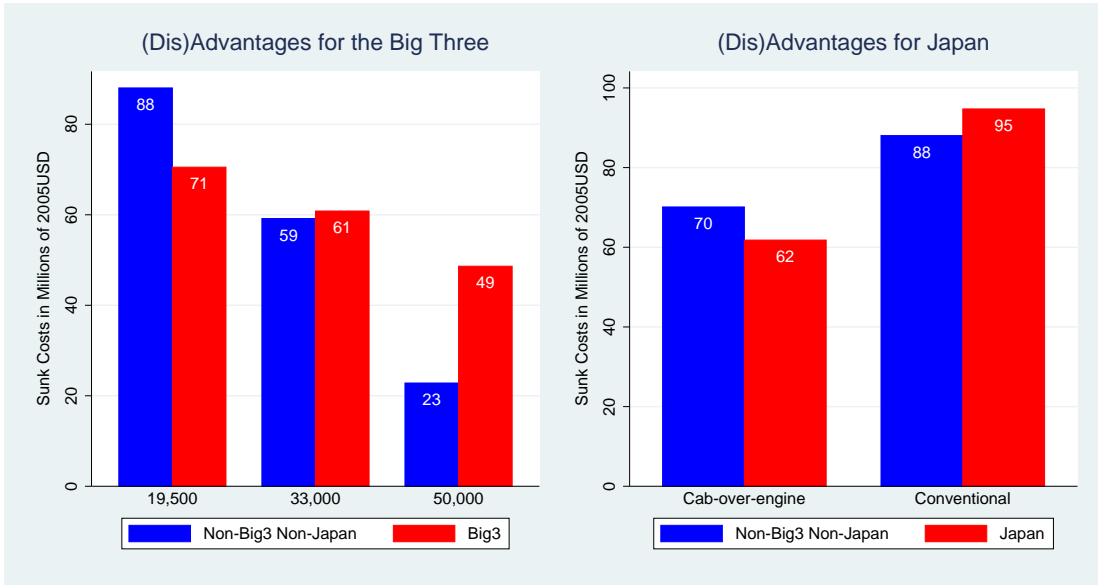


Figure 1.8: Spillovers in setup

rate, weakly exceed the sunk costs. That is,  $\frac{1}{HurdleRate} \times \mathcal{E} [\Delta\pi(J_{f,t}, J_{f,t} \setminus j, J_{-f,t}; z_t, w_t, t) | \mathcal{I}_{f,t}]$  is greater than or equal to  $\mathcal{E} [x_j \tilde{\theta}_{f,t,x} | \mathcal{I}_{f,t}]$ . Realized profits, however, differ from this expectation. Firms do not receive profits in perpetuity, since they may choose to exit  $j$  in the future, although they will receive a scrap value when they do so. They also face different  $J_{f,t}, J_{-f,t}, z_t$ , and  $w_t$  in the future. Thus, the realized discounted cash flows from adding  $j$  that will be offered for  $T$  periods, which firms can then compare to sunk costs, are given by

$$\sum_{\tau=1}^T \frac{1}{HurdleRate^{\tau}} \times \Delta\pi(J_{f,\tau}, J_{f,\tau} \setminus j; J_{-f,\tau}; z_{\tau}, w_{\tau}, t) + \frac{1}{HurdleRate^T} \times \frac{1}{\lambda} \times x_j \tilde{\theta}_{f,t,x_j}$$

That is, in terms of realized cash flows,  $f$  adds  $j$  at  $t$  when

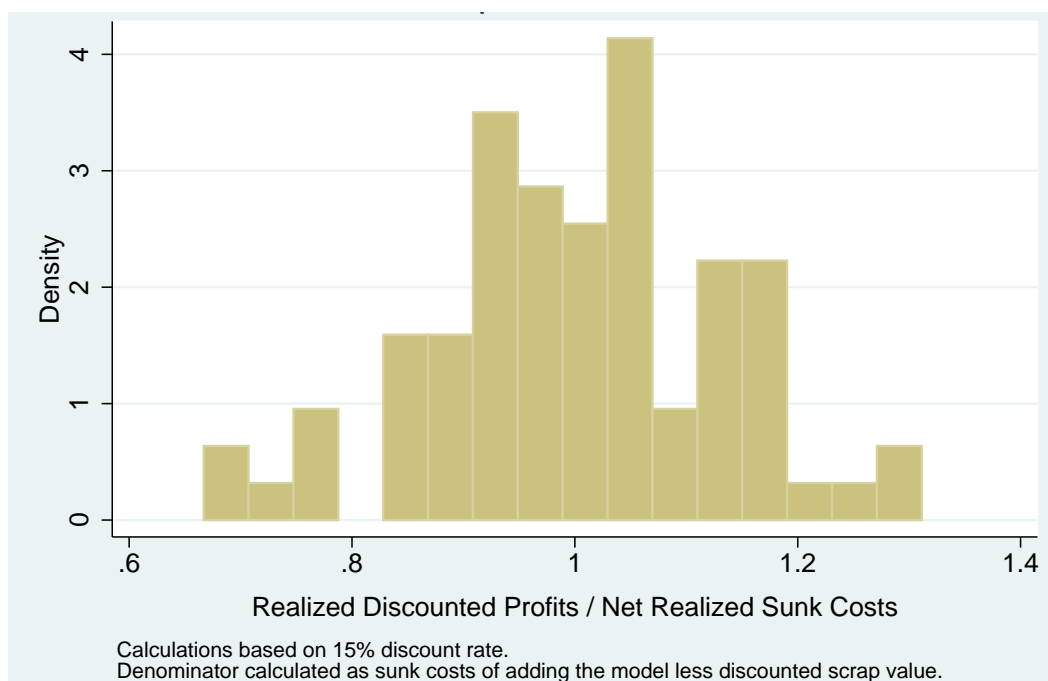
$$\underbrace{\sum_{\tau=1}^T \frac{1}{HurdleRate^{\tau}} \times \Delta\pi(J_{f,\tau}, J_{f,\tau} \setminus j; J_{-f,\tau}; z_{\tau}, w_{\tau}, t)}_{\text{Realized Sum of Discounted Profits}} \geq \underbrace{\tilde{\theta}_{f,t,x}}_{\text{Sunk costs}} - \underbrace{\frac{1}{HurdleRate^T} \times \frac{1}{\lambda} \times x_j \tilde{\theta}_{f,t,x_j}}_{\text{Realized discounted scrap value}}$$

Net present value of sunk costs

In the data, a close relationship between the left-hand side and right-hand side expressions, at reasonable hurdle rates, would imply firms earn what they expect to, on average, using the much simpler capital budgeting rule. Figure 1.9 reports this ratio for three different hurdle rates: 13.26%, 15.26%, and 17.26%. Some products are still offered in the final year of

the data so rather than deal with a more complicated truncation problem, I simply force these products to exit and recover the relevant scrap values.

Figure 1.9 can reject the notion that firms earn radically different profits than they would expect to under the hurdle rate rule. At 15.26%, the distribution of realized ratios is centered close to 1.0. Since the truncation issue slightly understates the left-hand side variable, the true hurdle rate that sets these equal may be slightly higher. This would fall close to the midpoint of the 12% reported in the Poterba and Summers (1995) survey and the 19% reported in the Summers (1989) survey.



**Figure 1.9:** *Earned vs expected discounted profits*

## 1.7 The Impact of the “Bailout”

### Policy Setting

The \$85 billion of federal assistance to GM and Chrysler in 2009 constitutes the largest government bailout of a non-financial industry in modern history. Its causes are debated



but there is mostly consensus on three factors: a global recession beginning in 2008 and prompting a trough in sales, a rise in fuel prices coupled with American manufacturers' focus on trucks and SUVs, and legacy costs from pension and retiree healthcare benefits. By late 2008, there was an immediate fear that GM and Chrysler would default, prompting \$17.4 billion in assistance.<sup>30</sup> Shortly afterward in 2009, the federal government agreed on more funds, bringing the total to \$85 billion.

Whether to provide assistance was hotly contested, split partly along partisan lines, and even became a major US Presidential campaign issue. Republican Presidential Nominee Mitt Romney argued for a market-based solution in a November 2008 Op-Ed in the *The New York Times* titled "Let Detroit Go Bankrupt." Later, in 2012, Barack Obama took credit for its apparent success, saying "I said we're going to bet on American workers and the American auto industry, and it's come surging back."<sup>31</sup> In a 2012 Op-Ed in the *Wall Street Journal*, Robert Barro pointed out that rhetoric supporting the bailout ex post ignored counterfactual policy outcomes. He wrote, "If GM had disappeared, its former workers and other inputs would not have sat around doing nothing. Another company—be it Toyota, Honda or Ford—would likely have taken over its operations."<sup>32</sup>

GM and Chrysler comprise nearly 15% of commercial purchases in 2009, so the bailout is relevant for this segment. Caveats are in order, however. First, debtor-in-possession financing was extremely scarce so the analysis below considers only complete liquidation or a rival firm's takeover, even though more complex arrangements might have been likely. Second, moral hazard is ignored. Firms that expect support in the future may take riskier bets or under-expend effort in unprofitable states of the world. Third, the focus on the commercial segment clearly illustrates why model-level entry and exit matter and is suggestive of what could happen in other markets but is not a comprehensive automotive study. For example,

---

<sup>30</sup>See, for example, Wilcox, Gregory., "Big 3 detail plans for bailouts," *Daily News* 3 Dec 2008., and Isidore, Chris., "Bush announces auto rescue," *CNNMoney* 19 Dec 2008., and Carty, Sharon., "Automakers ask for billions in loans, credit; Two say they're nearly out of money," *USA Today* 3 Dec 2008.

<sup>31</sup>Rosche, Jedd. "Obama quotes Romney's auto bailout op-ed in UAW speech." *Politico* 28 Feb 2012.

<sup>32</sup>Barro, Robert J. "Ryan and the Fundamental Economic Debate." *The Wall Street Journal* 14 Aug 2012.

it ignores general equilibrium effects that are probably small if failure is confined to the commercial segment but problematic if we extend this to the passenger segment, which is more than nine times larger. Fourth, changing an assembly line is presumably quick and painless relative to relocating labor. This may be fine for marginal expansions of product offerings but may prove challenging for a big shift in productive capacity. As a final note, financial distress is rarely random. Often it signals some underlying problems and that the efficient solution is shut down. I will assume—and am aided now by hindsight—that the issues of GM and Chrysler are mainly poor past decisions coupled with a very rare and deep capital drought.

### **Alternate Policies**

The following analysis compares the automotive bailout, i.e. federal support for GM and Chrysler that allowed them to continue operating as independent entities, against three alternate policies. One is liquidation, effectively a removal of the GM and Chrysler brands and products from the market. The other two are acquisitions: one by Ford, which overlapped heavily in product space with the troubled firms, and one by Paccar, which did not.

For each counterfactual policy, I compare the predictions of a model that allows for just prices to re-equilibrate against a model which allows for prices and product offerings to re-equilibrate. There are four broad findings. First, in the case of liquidation, sunk costs are low enough relative to profits to induce entry. Second, in this case, model-level entry and exit have a strong, moderating effect on the impact of liquidation. Finally, with respect to the acquisitions, although the identity of the acquiring firm matters a great deal when model-level entry and exit are ignored, it matters little when they are accounted for.

### **Computing Counterfactual Policy Outcomes**

Assessing *what would have happened* in the event that GM and Chrysler were not rescued by the federal government requires recomputing the product offerings that would result from

the change in the environment. In positioning games, multiple equilibria are the rule rather than the exception. Differentiated product markets with multiple attributes, as in the present setting, feature a large number of potential product offerings and a corresponding large number of potential equilibria. Lee and Pakes (2009) suggest a learning process that results in a distribution of equilibria played as well as potentially reducing the computational burden of this problem. The counterfactual policy outcomes reported below rest on a best response dynamic, which is computed as follows. Begin with product offerings from the prior period and a predetermined order of firm moves. The first firm chooses product offerings that are the best action conditional on what all other firms are currently offering. Their choice updates the product offerings that others observe when making their choices. The second, third, and so on, firms do the same. When the last firm in the ordering has chosen a best response, the order repeats. The process terminates when no firm has any profitable deviations. For each decision, the sunk costs used to compute the best response are based on moves from the 2009 product offerings, not the prior iteration of the learning process.

The rest point of this system is consistent with the necessary conditions that were used in the estimation of sunk costs, so the equilibrium selection process here is internally consistent with the model presented above. The policy analysis below proceeds with an ordering based on market share, with the largest share (“leader”) moving first. Details of the calculations are provided in the Appendix.<sup>33</sup>

## Findings

It is helpful to start with an idea of what products GM and Chrysler offered in the year prior to the decision. Table 1.8 reports the offerings for 2009. GM offers twelve models while Dodge offers four. Both were operating in the lower one-half of the GWR distribution and produced all three types of cabs but did not feature the long/extended option. Several of

---

<sup>33</sup>One alternative method uses random orderings, which will result in a distribution over the outcomes. This is in process.

these models overlap. All of the Chrysler models and over two-thirds of the GM models overlap with offerings by Ford, while none overlap with Paccar (not shown).

**Table 1.8:** 2009 GM and Chrysler Product Offerings

<u>Parent</u>	<u>Brand</u>	<u>Cab Type</u>	<u>Long/Ext. Option</u>	<u>GWR</u>
Chrysler	Dodge	Conventional	No	12,000
Chrysler	Dodge	Conventional	No	13,500
Chrysler	Dodge	Conventional	No	15,300
Chrysler	Dodge	Conventional	No	19,500
GM	GM	Compact Front End	No	11,000
GM	GM	Conventional	No	12,000
GM	GM	Conventional	No	13,500
GM	GM	Compact Front End	No	14,000
GM	GM	Cab Over Engine	No	14,700
GM	GM	Conventional	No	15,300
GM	GM	Cab Over Engine	No	19,500
GM	GM	Conventional	No	19,500
GM	GM	Conventional	No	21,500
GM	GM	Cab Over Engine	No	22,400
GM	GM	Cab Over Engine	No	27,100
GM	GM	Conventional	No	28,000

Liquidation is assessed first. Table 1.9 reports the impact on the most affected and median affected products, ordered by their respective changes in markups. In Panel A, model-level entry and exit are ignored. In this case, all three of the most affected products are owned by Ford, are conventional cab vehicles, and tend to be at the low end of the weight distribution. Markups on these models rise by over 60%. Despite higher prices, their market shares also expand, capturing a subset of buyers who find it difficult to substitute away from low-GWR conventional cabs as well as having a pressing need to purchase a vehicle. In Panel B, model-level entry and exit are accounted for. In stark contrast to the prior results, markups for the most affected products increase only around 10% to 20%. The impact to market share is more muted. These reflect increased sales driven by lower prices that are partially offset by substitution to newly introduced models. The sharp differences in Panel A and B are driven by model-level entry in precisely the places where markups

increase the most.<sup>34</sup> In total seven products enter: two by Daimler, two by International, and one each by Paccar, Volvo, and Hino. Daimler enters direct competitors to the F250 and F450 while International enters a direct competitor to the F350. The other introductions are slightly more dispersed. No products exit.

**Table 1.9:** *Effects of model-level entry on products for liquidation*

---- PANEL A: JUST PRICES RE-EQUILIBRATE ----

Rank	GWR	Cab Type	Brand	Model	Markup Per Model	
					Level Difference	% Chg.
1	12,000	Conventional	Ford	F250	\$2,288	73%
2	13,500	Conventional	Ford	F350	\$2,168	70%
3	15,300	Conventional	Ford	F450	\$1,889	63%
Median	35,000	Conventional	Sterling	Acterra33	\$90	3%

---- PANEL B: PRICES AND PRODUCT OFFERINGS RE-EQUILIBRATE ----

Rank	GWR	Cab Type	Brand	Model	Markup Per Model	
					Level Difference	% Chg.
1	12,000	Conventional	Ford	F250	\$453	18%
2	13,500	Conventional	Ford	F350	\$420	14%
3	15,300	Conventional	Fuso	FM330	\$114	4%
Median	33,000	Conventional	Peterbilt	386	\$0	0%

The policies' distributional affects on buyers are strong. Table 1.10 reports these results. Liquidation has the biggest impact on business and personal service industry firms that reside in areas slightly more dense or urban than average. These firms increase their substitution to the outside good by up to almost 50% when entry is ignored but only 14% when it is accounted for. Unsurprisingly—and reassuringly—this is precisely the subset of firms that the microdata indicated are most likely to purchase low-GWR conventional vehicles. The median impacted buyer, meanwhile, experiences virtually no change. For this reason, the change in total output measured in levels is relatively muted overall. Total output falls 8% when model-level entry is ignored and about 3% when it is accounted for. Nonetheless, this still translates to a nearly 60% drop in the effects of liquidation.

Acquisitions are assessed next. Figure 1.10 reports these results and compares them

<sup>34</sup>See the Appendix for more details on how these are computed.

**Table 1.10: Effects of model-level entry on purchases for liquidation**

Measure	Level changes		Level Difference (1) vs (2)	% Difference (1) vs (2)
	Without Entry/Exit (1)	With Entry/Exit (2)		
<u>No-Purchase Probability</u>				
Most Affected Buyer	-43.2%	-14.0%	-29.2%	-67.5%
Mean Buyer	-1.2%	-0.5%	-0.7%	-57.9%
Median Buyer	0.0%	0.0%	0.0%	-30.8% *
<u>Output</u>				
Total market	-7.6%	-3.2%	-4.4%	-57.9%

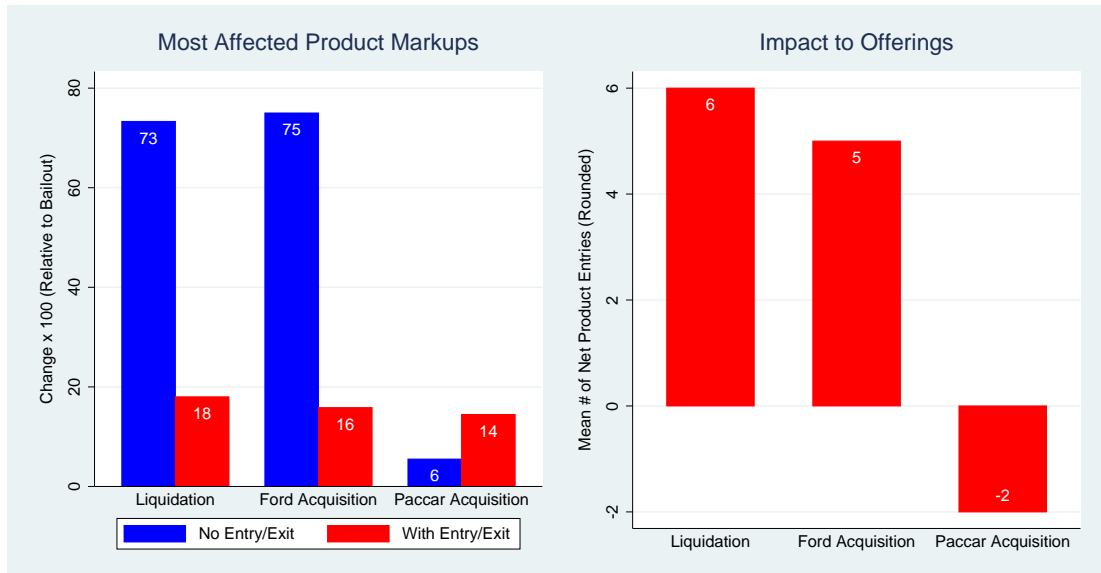
*Most Affected Buyer Type: Industry = Bus. & Personal Service, Urban Measure =  $-.36 \sigma$  below mean*

---

\* The impact to the median buyer in columns (1) and (2) are close to but not exactly zero. For this reason, the "% Difference" column is computable for this case, even though it appears one would need to divide by zero.

with liquidation. The left-hand side graph reports the increase in markups for the most affected products relative to the bailout. The y-axis measures level changes in percents (not to be confused with percentage changes). It shows that when model-level entry and exit are ignored, an acquisition by Ford would resemble liquidation while an acquisition by Paccar would resemble the bailout. It further suggests, however, that when model-level entry and exit are accounted for, it matters little which policy is chosen. Markups for the most affected products rise between 14% and 18%, regardless as to whether GM and Chrysler are liquidated or sold to a rival. The differences between the bailout and the alternate policies are much larger than between the alternate policies themselves. The elimination of two independent owners matters much more than what ultimately happens to the products owned by them, since the sunk costs are low enough that product portfolios will flexibly adjust ex post anyway. These are very different results than one would obtain under high sunk product entry costs. In this case, policymakers would care a great deal about transferring GM and Chrysler products to Paccar or another rival with whom the trouble firms did not previously overlap. The right-hand side graph shows the net number of product entries, i.e. the total number of individual models entered less the total number exited. An acquisition by Ford leads to seven total product introductions and five net

product introductions. Two exits by Ford of duplicate models that it inherited make up the difference. An acquisition by Paccar leads to no entry and two exits of duplicate modes that it inherited. A high number of net entrants in the former two cases drive the large wedge between markups in the left-hand side graph, while the exits by Paccar push prices up slightly.



**Figure 1.10:** *Increased markups induce entry*

Figure 1.11 compares the impacts of the policy changes to buyers by showing the change in the probability of not purchasing a vehicle. The y-axis measures level changes in percents. The left-hand side graph studies the most affected buyers. Predictably, these are tied closely to markups. There is one minor difference when model-level entry and exit are ignored. While markup changes are about even for a Ford acquisition relative to liquidation, purchasing changes are about one-third lower. This reflects idiosyncratic preferences of the buyers for the liquidated products. It is, of course, mechanical in the logit model but also captures the reality of differences in stylizing, dealer relationships and locations, and a host of other factors outside the scope of a tractable model. The right-hand side graph studies the mean affected buyers. As above, the impact of the policies is very slight and the sign of the effect of allowing model-level entry and exit is the same in all cases. Also, as above, the

impact to the median affected buyers (not show) is virtually zero.

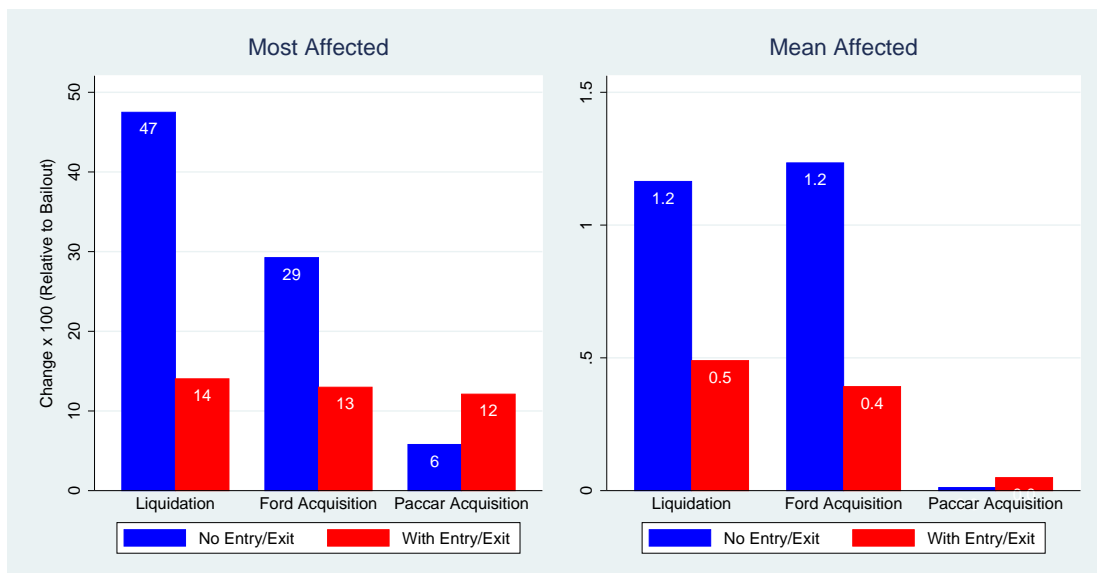


Figure 1.11: Impact to consumers' probability of not purchasing

## 1.8 Conclusions

In markets like the commercial vehicle segment of the US automotive industry, and the automotive industry in general, there has been virtually no firm-level entry and exit for decades and yet frequent product additions and removals. This suggests that even if the cost of entry to startup firms is prohibitively high, the cost to incumbents to enter and exit individual models is sufficiently low for the market to adjust to some policy changes. For this reason, the predictions that come out of models that ignore this fact and allow only prices to re-equilibrate will tend to overstate the impact of policy or market structure changes on markups, profits, and purchases.

Reduced-form evidence indicates that in the commercial vehicle market, producers respond to changes in the composition of demand with changes in the product offerings. For example, as the construction industries expands, the number of product offerings tailored to the preferences of construction buyers expands as well. This indicates that, in



short, sunk costs of introducing these models cannot be so sufficiently high as to discourage this behavior. On the other hand, that some rather than all firms are entering these models indicates that these introductions are not free. These facts provide clear intuition for why the data can bound fixed costs, although to quantify those bounds and predict alternate policy outcomes, I needed a structural model.

Several challenges presented themselves. As with many large, economically important differentiated product markets, commercial vehicles are sold nationally but produced in at most one or two locations nationwide. This ruled out geographic variation on the supply side and, in turn, the use of estimation methods that have previously been exploited in the literature. I instead used the intuition provided by the reduced-form evidence above along with a rich twenty-seven year panel of product offerings to identify the sunk costs. This entailed taking a stand on how managers make potentially very complex decisions, for which I relied on practitioner interviews. To deal with multiplicity, which is the rule rather than the exception in positioning games, I relied only on the necessary—rather than sufficient—Nash equilibrium conditions. To deal with this complication in simulating counterfactual policy outcomes, I used a learning process based on best response dynamics. Taken together, they suggested that sunk costs were low enough to induce entry for policies where markups would have otherwise risen considerably. They demonstrated that in this market, when model-level entry and exit are ignored, it matters a lot which policy is chosen while when they are accounted for, it matters little. Nonetheless, the loss of two independent operating entities increases the overall concentration of ownership in this market. Markups for the most affected products rise slightly, although the impact to the mean and median affected products and buyers, ordered by markup and purchasing changes, respectively, was essentially zero.

## Chapter 2

# Two Period Strategies for Discrete Dynamic Games<sup>1</sup>

### 2.1 Introduction

The forward-looking solution to many discrete dynamic games places a heavy burden on agents. Even when the factors that determine profits are reduced in number and richness, there are still often billions or more of states that agents could keep track of. Markov perfect strategies, i.e. forward-looking strategies that condition directly on the state, require agents keep one parameter in memory for each state. For example, in a simple entry/exit game, incumbents would then need to keep a cutoff value in memory for every possible state, and exit when the scrap value they expect to receive is greater than that cutoff value. A large literature—primarily outside economics—has focused on how to approximate this type of decision rule. The goal is to reduce the burden on agents, but minimize the loss from the approximation.

Faced with such a burden, firms often cut computational corners. Perhaps the most common—and certainly among the most simple—rules is to merely compare the ratio of profits to requirement investment against a "hurdle rate." Frequently managers will also

---

<sup>1</sup>Co-authored with Richard Sweeney

refer to this as a "payback ratio," which makes an equivalent comparison (as we show later in this paper). We call this class of rules "two period strategies." Although there are special cases where two period strategies may be precisely right, they are—like most approximate strategies—going to deviate from an optimal rule in most cases. The first question we ask in this paper is, "By how much?" If the loss that results from employing this rule is quite low, then it will help explain why it is used so frequently.

The second question we ask is, "How does this vary with the underlying characteristics of the market to which it is applied?" The answer to this question has two audiences. First, it may help firms decide whether two period strategies are appropriate given the nature of the industry they operate in. As Jonathan Owen, Director of Operations Research at General Motors writes, "Practitioners...desire an understanding of why a system behaves the way it does, what are the most important parameters, what practical alternative solutions are available in addition to the optimal solution, and how sensitive solutions are to changes in operating conditions" Owen et al. [2011]. Second, if it implies that managers are more likely to employ these strategies, then it may increase economists comfort in assuming this is how firms are making decisions. Since two period strategies are easy to solve for both the agents and econometrician, this would drastically increase the scope of problems that researchers could tackle.

This paper studies two period strategies in the context of the entry/exit game proposed in Pakes et al. [2007] ("POB"). We begin by formally defining a two period equilibrium strategy and compare it to the Markov perfect analog. We then define a loss function, which quantifies the average discounted payoffs that firms lose by employing these strategies. At the base parameters put forth in Pakes et al. [2007], the loss that results from employing two period strategies is very low. The mean loss, where an average is taken across states and weighted by equilibrium play, is typically below 2%. The maximum loss for any state is under 5%.

As the underlying characteristics of the market change, loss tends to move in a predictable way. For example, increasing the discount factor makes agents more patient and increases

the value of the future. All else equal, this should increase the loss, and this is what we see in our simulation exercise. Similarly, we explore the impact of changing the persistence in the underlying profit shifters, the distribution of scrap values, and the distribution of entry fees. Although the relationship of loss with each of these variables is convex, it turns out that the impact is not very large for reasonable values of the variables. For example, loss is convex in the discount factor, and increases five-fold as we move from a discount factor of 0.86 to 0.94. However, very steep increases in loss would only occur at values of the discount factor above what are reasonable for most applied settings, i.e. above 0.925.

## 2.2 Model

We study the simple entry/exit game found in Pakes et al. [2007]. In each period, there are  $n_t$  incumbents and a fixed number,  $\mathcal{E}$ , of potential entrants.  $z_t$  denotes a set of profit shifters that follow a finite-state Markov process.  $\pi(n_t, z_t)$  denotes the per period profits earned by all active firms. At the beginning of each period, firms earn these profits. Next, each incumbent receives a draw from the distribution of selloff, or “scrap,” values, denoted  $\phi$ , and decides whether to exit.  $\phi$  is private information. At the same time, each potential entrant receives a draw from the distribution of entree fees, denoted  $\kappa$ , and decides whether to enter. Let  $x$  denote a realized number of exitors and  $e$  denote the number of entrants.

Incumbents decide whether to exit based on whether  $\delta\phi$ , the discounted scrap value they receive, is greater than  $\delta VC$ , the discounted continuation value. The value of continuing, conditional on continuing (denoted by  $\chi = 1$ ), is equal to

$$VC(n, z; \theta) = \sum_{e, x, z'} \int_{\phi'} V(n + e - x, z', \phi'; \theta) p(d\phi' | \theta) p^c(e, x | n, z, \chi = 1) p(z' | z) \quad (2.1)$$

where  $V$  denotes the value function and  $p^c$  denotes the incumbent’s beliefs about rivals’ play. The value function, conditional on a realized scrap value, is equal to

$$V(n, z, \phi; \theta) = \pi(n, z) + \max \left\{ \delta\phi, \delta VC(n, z; \theta) \right\} \quad (2.2)$$

Potential entrants decide whether to enter based on whether  $\kappa$ , the entry fee, is lower than

$VE$ , the entry value. The value of entering, conditional on entering (denoted by  $\chi^e = 1$ ) is equal to

$$VE(n, z; \theta) = \sum_{e, x, z'} \int_{\phi'} V(n + e - x, z', \phi'; \theta) p(d\phi' | \theta) p^e(e, x | n, z, \chi^e = 1) p(z' | z) \quad (2.3)$$

where  $p^e$  denotes the potential entrant's beliefs about rivals' play.

Let  $F_\phi(\cdot | \theta)$  denote the distribution of scrap values and  $F_\kappa(\cdot | \theta)$  denote the distribution of entry fees. POB assume that  $F_\kappa$  is bounded below by  $\underline{\kappa} > 0$  and that  $\phi$  takes only non-negative values.  $p^e$  and  $p^c$  at time  $t$  depend only on  $(n_t, z_t)$ . Firms make simultaneous moves.  $z$  takes values in  $[1, 2, \dots, \bar{z}]$ . For every value of  $z$ ,  $\pi$  is bounded and, as  $\pi(n, z) \leq 0$  as  $n \rightarrow \infty$ .

Firms make entry and exit decisions based on cutoff rules. This paper will allow these cutoff rules to differ, even in equilibrium, from what firms actually earn. This differs from POB and requires additional notation. Let  $\hat{VC}(n, z; \theta)$  denote the cutoff rule for incumbents and  $\hat{VE}(n, z; \theta)$  denote the cutoff rule for entrants. Incumbents exit at a rate of  $F_\phi(\hat{VC}(n, z; \theta); \theta)$ . Potential entrants enter at a rate of  $F_\kappa(\hat{VE}(n, z; \theta); \theta)$ . Finally, let  $p^{c^*}$  and  $p^{e^*}$  denote the actual probability distribution over entrants and exitors that is realized in equilibrium play (without, thusfar, taking a stand on the nature of that equilibrium). Analogously, let  $VC^*(n, z; \theta)$  and  $VE^*(n, z; \theta)$  denote the continuation value and entry value that are realized in equilibrium play, respectively. To calculate these,  $p^{c^*}$  will be substituted for  $p^c$  in (2.1) and  $p^{e^*}$  will be substituted for  $p^e$  in (2.3).

### 2.2.1 Markov perfect strategy equilibrium

Markov perfect [Maskin and Tirole, 1988] strategies condition directly on  $(n, z)$ . Since firms are homogenous, move simulatenously, and have private information about entry fees and scrap values, they have the same beliefs about rivals' play. Moreover, these beliefs are correct in equilibrium. Thus,  $\hat{VC} = VC^*$  and  $\hat{VE} = VE^*$ .  $p^c$  and  $p^e$  are consistent with incumbents who exit at rate  $F_\phi(\hat{VC}(n, z; \theta); \theta)$  and potential entrants that enter at a rate  $F_\kappa(\hat{VE}(n, z; \theta); \theta)$ .

POB note that existence is guaranteed, while uniqueness is not—and in fact, multiplicity

may be the rule rather than the exception in many entry/exit games of this form. They further note that equilibrium play will generate a finite-state Markov chain in  $(n, z)$ . No matter the initial state, play will wander into a subset of states (the “recurrent class”) and stay there forever.

There are a large number in even simple entry/exit models, so keeping in memory one cutoff value for each state can be quite burdensome on agents. In their simple single-location model Monte Carlo, POB consider only 3 potential entrants, 45 values for their scalar profit shifter, and 3 values for a growth rate (that determine transitions of the profit shifter), yet arrive at 4,185 states. Merely adding a second profit shifter that takes on an equal number of values expands this multiplicatively to 188,325 states—for a model that is far too simple to resemble real markets.<sup>2</sup> In differentiated product markets, e.g. those considered in Berry et al. [1995] or Nevo [2001], this requires an unreasonable amount of memory.

The point here is not blanket criticism of Markov perfect equilibria. For example, firms’ behavior may be consistent, or nearly consistent, with Markov perfect strategies, even if the econometrician is unsure about what generates this behavior. As the saying goes, agents “don’t need auction theory to submit what often look like equilibrium bids.” Firms could, for example, learn through repeated play, as in Fudenberg and Levine [1998] and Fershtman and Pakes [2012].<sup>3</sup> This does, however, beg the question of what managers actually do. In survey data (Summers [1987], Poterba and Summers [1995], Graham and Harvey [2001], Moore and Reichert [1983]) and industry anecdotes (Alden and Smith [1992], Owen et al. [2011], and the first chapter of this dissertation), managers do, in fact, approximate the solution to complicated dynamic problems.

---

<sup>2</sup>In truth, firms need only keep track of cutoff values for states in or in the immediate vicinity of the recurrent class, although this still requires keeping track of over 600 cutoff values in the base case model of POB.

<sup>3</sup>This would still not resolve the memory issue.

### 2.2.2 Two period strategy equilibrium

This paper studies an alternative set of strategies that requires agents to keep track of only one parameter, the hurdle rate, which is denoted  $H$ . Incumbents set cutoffs such that  $\hat{V}C(n, z; \theta) = \frac{1}{H^c} \pi(nz)$ . Potential entrants set cutoffs such that  $\hat{V}E(n, z; \theta) = \frac{1}{H^e} \pi(nz)$ . We call these two period strategies. They are “two period” in the sense that they are analogs of the early empirical literature on entry (Mazzeo [2002], Seim [2006]). In the first period of these models, firms decide whether to enter, paying a fixed cost in the event that they do. In the second period, entrants earn profits while others receive zero payoff. These papers exploited cross-sectional data by assuming all firms are inactive prior to the start of play, that they are stuck with their decisions forever, and that observed choices reflect a long-run “steady state” of the market. However, these strategies “repeat” over an infinite horizon game.

Two period strategies restrict agents to condition only on  $\pi(n, z)$ , whereas the Markovian strategies allow firms to be fully flexible across  $(n, z)$  couples. This lifts the heavy burden discussed above, but comes at a cost: the cutoff rules will generally not equal the mean discounted payoffs that firms earn in equilibrium. That is,  $VC^*(n, z; \theta) \neq \hat{V}C(n, z; \theta)$  and  $VE^*(n, z; \theta) \neq \hat{V}E(n, z; \theta)$ .

Two period equilibrium strategies are defined as follows. Incumbents choose parameters to maximize

$$\Lambda^c(H^c) = \sum_{n,z} \left[ \left[ 1 - F_\phi(\hat{V}C(n, z; \theta)) \right] \mathbb{E}(\phi | \phi > \hat{V}C(n, z; \theta)) + F_\phi(\hat{V}C(n, z; \theta)) VC^*(n, z; \theta) \right] p(n, z; \theta)$$

where  $p^c$  equals  $p^{*c}$  and reflects rivals’ cutoffs based on two period strategies. Potential entrants choose parameters to maximize

$$\Lambda^e(H^e) = \sum_{n,z} \left[ F_\kappa(\hat{V}E(n, z; \theta)) \left[ \hat{V}E(n, z; \theta) - \mathbb{E}(\kappa | \kappa < \hat{V}E(n, z; \theta)) \right] \right] p(n, z; \theta)$$

These expectations weight by the frequency of visits in equilibrium, given by  $p(n, z; \theta)$ .

The use of two period strategies lifts the memory burden imposed by Markovian strategies but sacrifices expected payoffs in equilibrium. If these losses are small, the incentive to keep track of a complicated strategy is small, and two period strategies will

appeal to agents. The simulation exercise below addresses precisely this issue. If the losses remain small as underlying market parameters change, this could help explain the apparent widespread use of these strategies.

Before proceeding, we formalize the notion of loss. Let  $\Delta^e(VC^*, \hat{V}C)$  denote the loss to employing  $\hat{V}C$  instead of  $VC^*$ , weighted by frequency of play in equilibrium, which equals

$$\sum_{n,z} \left[ \left[ 1 - F_\phi(VC^*(n, z; \theta)) \right] \mathbb{E}(\phi | \phi > VC^*(n, z; \theta)) + F_\phi(VC^*(n, z; \theta)) VC^*(n, z; \theta) \right] p(n, z; \theta) - \Lambda(H^c)$$

Define  $\Delta^e(VE^*, \hat{V}E)$  analogously.

Finally, note that one could formally microfound this model if firms face a cost of employing more complicated strategies. Suppose, for example, that firms face a cost  $c$  per parameter (or cutoff value) stored in memory. If agents are forced to choose between two period and Markov perfect strategies, the former would arise endogenously if  $(S - 1) \times c > \max\{\Delta^c, \Delta^e\}$ , where  $S$  is the number of states.

## 2.3 Base Case Performance

### 2.3.1 Parameters

We begin by characterizing the equilibrium strategies and loss function at a base case set of market parameters that are essentially identical to those in the POB one-location model. As in their setup,  $\phi$  is distributed exponential with parameter  $\sigma_\phi$ . The only exception is that our  $\kappa$  is distributed log-normal whereas theirs is distributed such that  $f(\kappa = r) = a^2(r - 1/a)e^{-a(r-1/a)}$ .<sup>4</sup> Profits are given by  $\pi(n, z) = 2Z^2/(1 + n)^2$  and  $z = \log(Z)$  where  $z_t$  follows a second-order Markov process such that  $z_{t+1} = z_t + g_{t+1}$ . The base case parameters of the market are given by:

- $\delta = 0.9$

---

<sup>4</sup>This allows us to easily construct mean values of a truncated  $\kappa$  distribution, which is required in Equations (5) and (7). Unfortunately,  $\kappa$  is now unbounded from below, the number of potentially active agents is unbounded from above, and we cannot insure a finite state space. This is easily amended; for now, note only that we set our maximum  $n$  sufficiently high that the probability of entry there is very low, so the problem is insubstantial.



- $\mathcal{E} = 3$
- $\bar{z} = 45$
- $\sigma_\phi = 0.75$
- $\mu_\kappa = 2.125$
- $\sigma_\kappa = 0.4$
- $g \in \{-0.05, 0, 0.05\}$

### 2.3.2 Equilibrium Parameters

In the base case, the equilibrium cutoffs were determined by  $H^c = 8.88\%$ . In the authors' limited experience, the parameters tended to converge to this unique set regardless of where the algorithm started. That said, starting values ranged only from 3% and 20%. As in the Markov perfect case, multiplicity may very well be the rule rather than the exception, but the results thus far do not support that.<sup>5</sup>

### 2.3.3 Loss

Loss is remarkably low at this set of market parameters and equilibrium strategies. All results presented below weight by equilibrium play. However, we consider the problem only from the standpoint of incumbents; we leave to future work the entry decision. The mean loss is about 1.85%. The maximum loss, given any state, is about 4.73%.

The state where we find the maximal loss is at  $(n,g,z)=(17,3,45)$ . Since profits are quadratic in  $Z$ , it is not surprising that the maximum difference is where  $Z$  takes its largest value. Moreover, maintaining the high level of  $Z$  next period is determined by  $g$ , so it is not surprising that the maximum loss occurs where  $g = 3$ , i.e. where  $Z$  is guaranteed to stay high.

---

<sup>5</sup>Simultaneous moves with complete information would clearly generate the opposite conclusion.

## 2.4 Varying Market Parameters

At the base parameters, the two period strategies perform quite well. The next question is how they vary as the underlying market parameters vary.

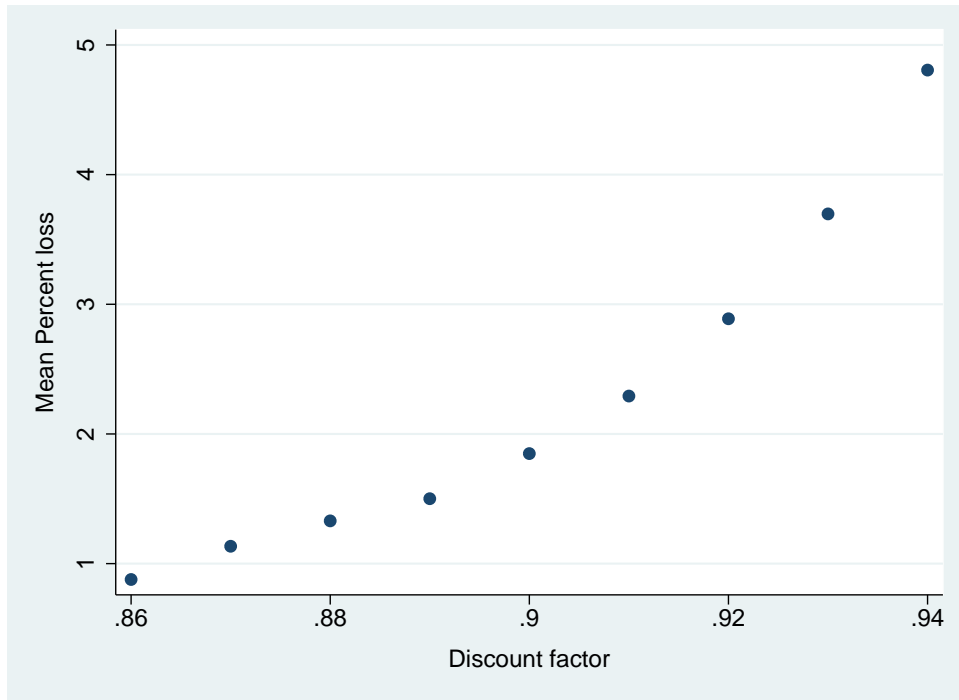
### Discount Rate

We begin by varying the discount rate,  $\delta$ . Lowering the discount rate increases the importance of the future and widens the gap between two period and Markov perfect strategies. Figures 2.1 and 2.2 report how mean and maximal loss, respectively, vary with the discount rate. For this exercise, we varied  $\delta$  between 0.86 and 0.94. Over these parameters, the mean loss varies between 0.88% and 4.81%. The maximum loss varies between 2.64% and 8.73%. In both cases, and in line with our intuition, loss increases as delta increases. The relationship is convex, although discount factors much above 0.925 are unlikely to be reasonable for most industrial settings. Mean and maximum loss, however, are still under 4% and 8%, respectively, at a discount factor of 0.93. Thus, the discount factor does not, by itself, hinder the performance of the two period strategies.

### Stochastic Process of Profit Shifter

Next, we vary the stochastic process that governs  $Z$ . In POB, there is persistence in  $g$  for values away from the boundary. In the POB base case, when  $g = 0.05$  today, tomorrow  $g$  takes a value of 0.05 with 75% probability and 0 with 25% probability. When  $g = -0.05$  today, tomorrow  $g$  takes a value of -0.05 with 75% probability and 0 with 25% probability. As persistence increases, ignoring the value of  $g$  becomes more costly and hurts the performance of the two period strategies.

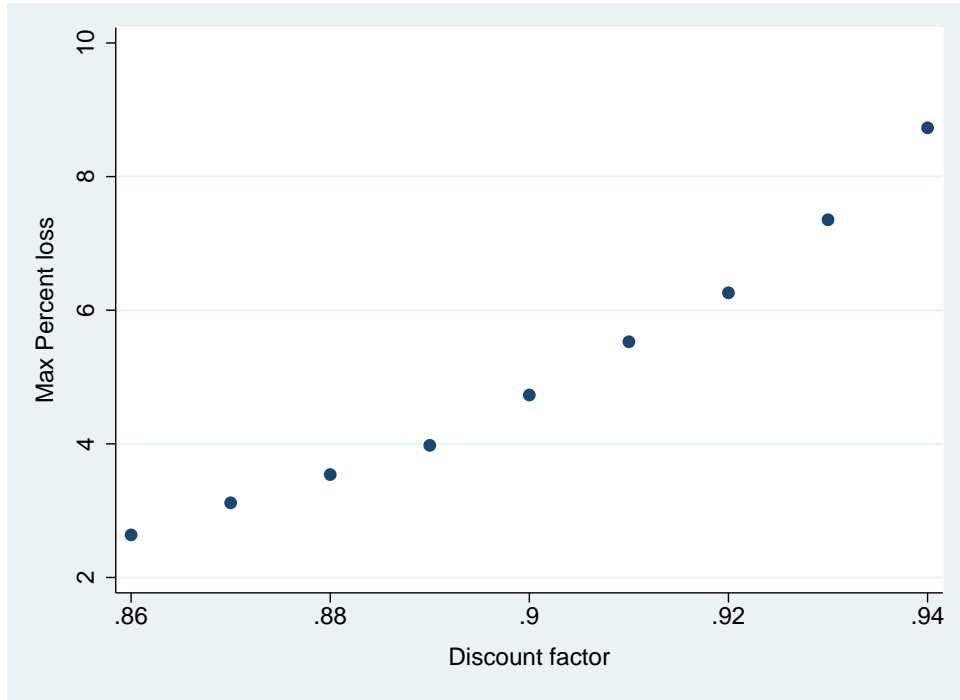
We varied the 75% persistence in the base case to between 65% and 95%. The impact on mean loss is relatively muted. Figure 2.3 reports this result. Mean loss increases and is convex in the degree of persistence, varying between 1.64% and 2.77%. The impact on maximum loss is greater. Figure 2.4 reports this result. Maximum loss varies between 3.78% and 8.10%. This is not surprising; the impact of increasing the persistence does not effect



**Figure 2.1:** *Impact of discount rate on mean loss*

one-third of the states, i.e. all of those for which  $g = 0$ . Moreover, it does not impact states for which  $Z$  takes either 0 or its maximum value. Finally, it has little substantive impact on the portions of the state space where profits are insensitive to  $Z$ . That is, profits are quadratic in  $Z$ , so the profit function—and by extension, the value of continuing—are much more sensitive to high  $Z$  values.

Persistence in the growth rate of the underlying profit shifters near 100% seems unrealistic in most applied settings. On the one hand, this suggests that loss is not very sensitive to the underlying persistence in  $Z$ . On the other hand, we allow here only for one-unit movement in the profit shifters. These results may not hold for larger (but still predictable) swings in  $Z$ . For example, if reaching some state guaranteed a  $Z$  value near zero in the subsequent period, loss would be very large.

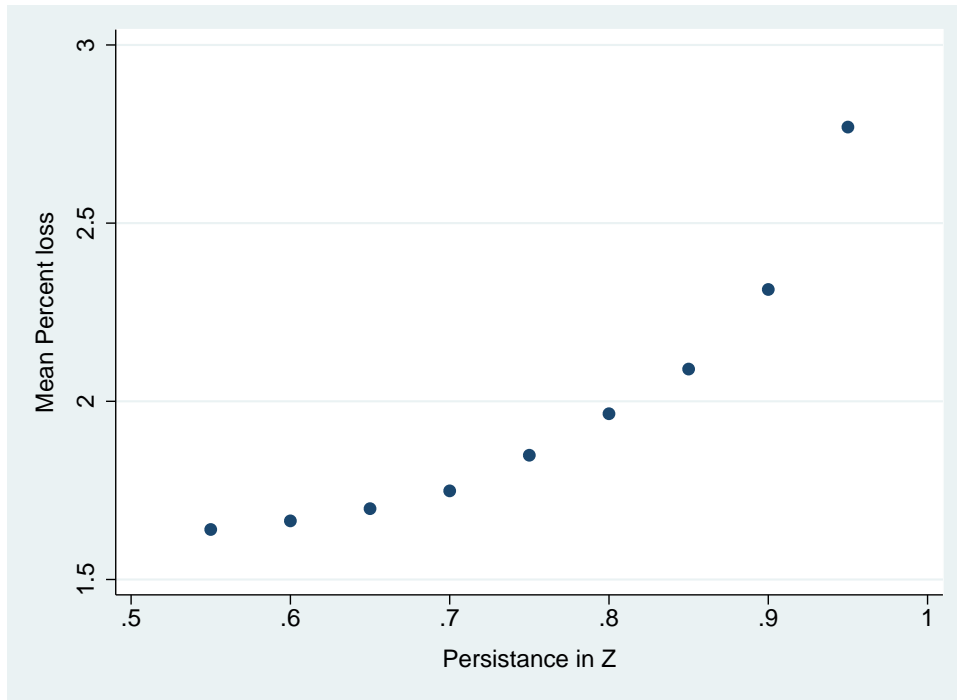


**Figure 2.2:** *Impact of discount rate on maximum loss*

### Scrap Value Distribution

Recall scrap values are distributed exponentially with a parameter of  $\sigma_\phi$ . Changes in the parameters non-linearly affect the equilibrium in complicated ways, but varying  $\sigma_\phi$  has at least one obvious effect: it provides more variation in exit rates across states. This provides more opportunity for the restricted strategies to “go wrong.” Note that the base case value of  $\sigma_\phi = 0.75$  does not encourage much exit: the distribution across states is bimodal at values of 0% and 13% and has support only up to 30%. Higher  $\sigma_\phi$  increase the expected draws of  $\phi$ , make firms more likely to exit, and reduce the mass at, for example, zero entry.

We vary  $\sigma_\phi$  between 0.6 and 1.0. Recall that  $\sigma_\phi$  represents the inverse of the distance parameter governing the exponential distribution, so increasing  $\sigma_\phi$  lowers the average scrap value received and, in turn, the probability of exit. Figures 2.5 and 2.6 report the impact on mean and maximum loss, respectively. In line with our intuition, loss increases as average scrap values increase (and  $\sigma_\phi$  decreases). Mean loss varies between close to zero and 4.11%. Maximum loss varies between 1.06% and 10.04%. The relationship between loss and  $\sigma_\phi$  is



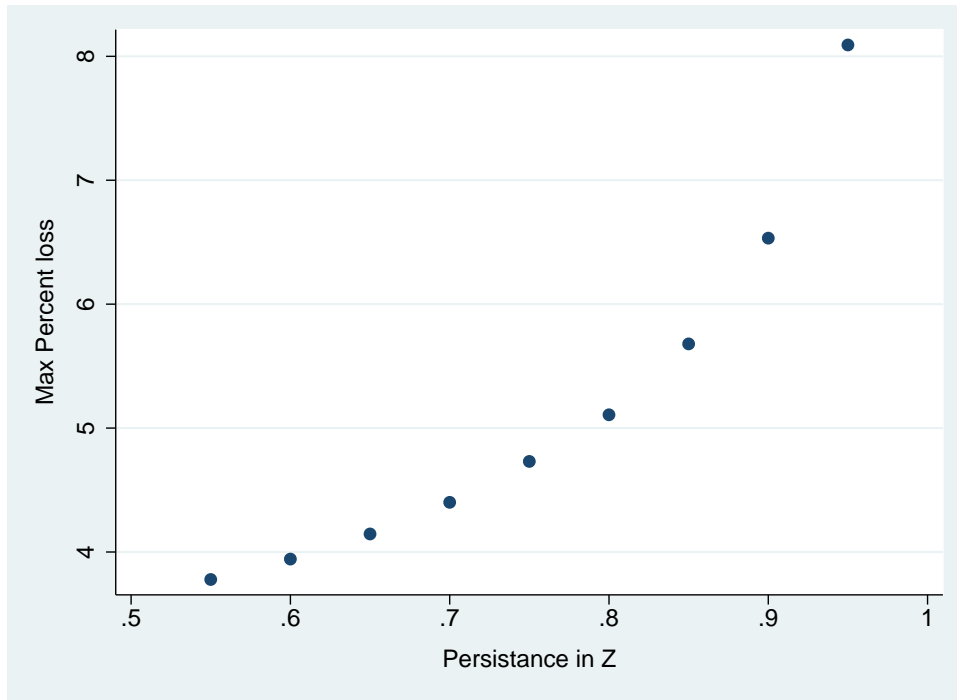
**Figure 2.3:** *Impact of persistence on mean loss*

convex down (and between loss and average scrap value is convex up).

### Entry Fee Distribution

Entry fees are distributed log-normally, where the underlying normal distribution has mean  $\mu_\kappa$  and standard deviation  $\sigma_\kappa$ . The sign of increasing  $\sigma_\kappa$  is unclear. On the one hand, it makes entry less likely, and slows down the degree to which the market can "correct" after winding up at a unusually high or low value of  $n$ , conditional on  $Z$ . On the other hand, lower levels of entry making incumbency more valuable, and if this enters non-linearly into the problem, then it is impossible to sign the effect. It turns out that higher  $\sigma_\kappa$  has a positive, linear, but small effect on loss. Figures 2.7 and 2.8 report the impact of changing  $\sigma_\kappa$  on mean and maximum loss, respectively. Mean loss varies between about 1.5% and 2.2% while maximum loss varies between about 4.4% and 5.2%.

Since  $\sigma_\kappa$  near 0.5 imply entry rates that approach 0% in many states, it is quite surprising that the relationship is not concave. That said, the impact on loss is sufficiently small here



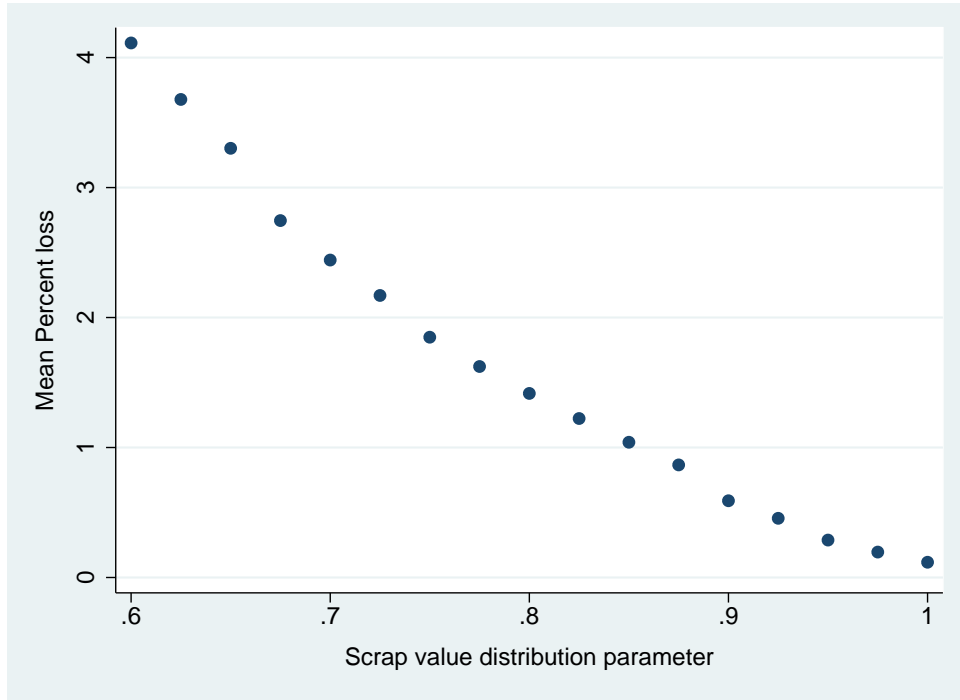
**Figure 2.4:** *Impact of persistence on maximum loss*

that one can safely assume this will, by itself, impact the performance on the two period strategies.

## 2.5 Extensions

The results above provide preliminary but compelling insight into how two period strategies perform relative to Markov perfect strategies. There are obvious extensions. The first is to consider entry values. Entry rates are lower than incumbency rates and vary more across states. Thus, entry may be more sensitive to restrictions on the degree of forward-looking behavior. *VE* differs from *VC* only in the perceptions of rivals' behavior, so qualitatively these results should not vary much, although this is an open and important question.

Throughout these results, mean and median loss was surprisingly low. If this result holds for a more robust analysis, it begs the question of where dynamics are important in this context. Surely there exists a transition process for the profit shifters where the strategy



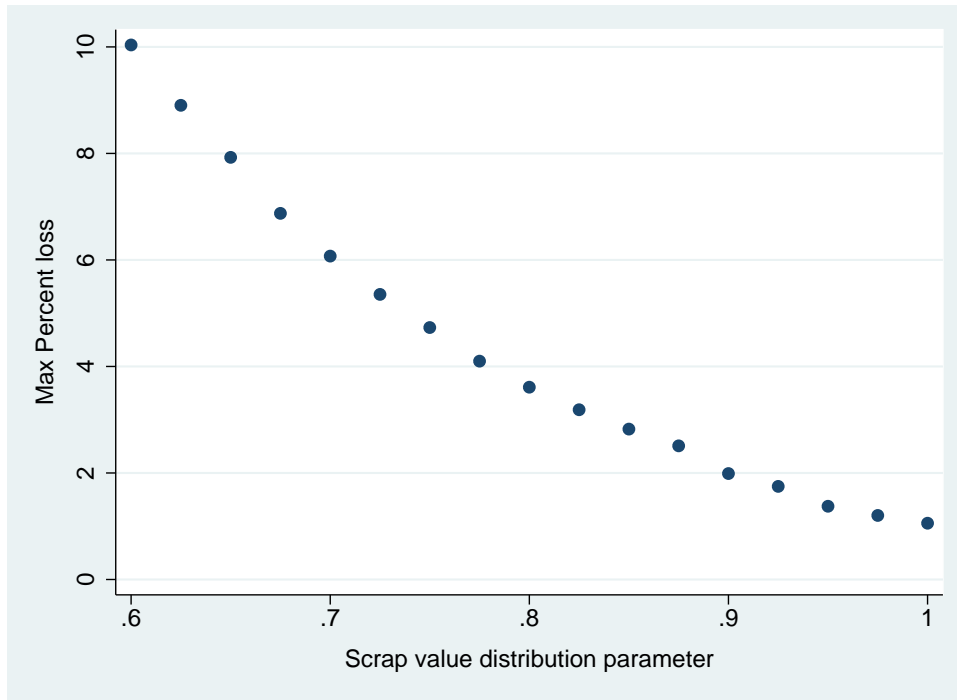
**Figure 2.5:** *Impact of scrap value distribution parameter on mean loss*

restriction has bite.  $Z$  shifts by at most one unit each period in the specifications above. If, for instance, the process instead produced multi-unit shifts—that agents could anticipate—this would create substantial problems for two period strategies. On the other hand, large shifts are infrequent in most applied settings—or, at a minimum, applied settings for which the research is correctly trying to fit a stationary process. Most prior empirical work on dynamic games has featured only small shifts in the underlying profit shifters over time,<sup>6</sup> aside from policy changes and other such discrete and observable events. In any case, bounded transitions of  $Z$  and reasonable values of  $\delta$  suggest that the gap between cutoffs derived from restricted strategies and the mean discounted payoffs that firms actually receive cannot be too far apart. This suggests an upper bound for the performance difference between Markov perfect and two period strategies. An analytic upper bound for loss across states merits further consideration, even if a closed-form solution for mean loss is unlikely to exist.

The results only considered parameter-by-parameter robustness tests. A more com-

---

<sup>6</sup>See, for example, Ryan [2012], Collard-Wexler [2013], Blonigen et al. [2013], Roberts et al. [2012].



**Figure 2.6:** *Impact of scrap value distribution parameter on maximum loss*

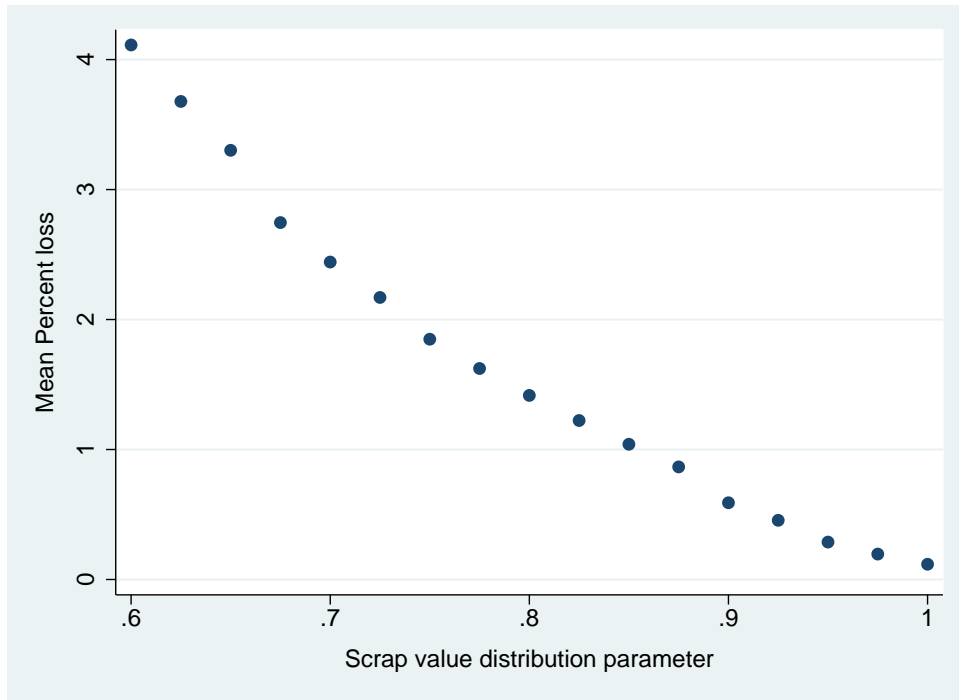
prehensive tests, which we leave to future work, could consider changing two or more parameters at a time impact the results. For example, increasing  $\sigma$  and  $\delta$  individually increase loss, but not substantially so. If they are both allowed to increase, and these effects compound, it may be more problematic.

Finally, this analysis should extend also to games of positioning or with additional controls. The former of these fixes the number and identity of firms, but allows their product choices to differ (Nosko [2010], Eizenberg [2014]). The latter of these can include investment in quality or capacity (Ryan [2012], Collard-Wexler [2013], Sweeney [2015]).

## 2.6 Conclusion

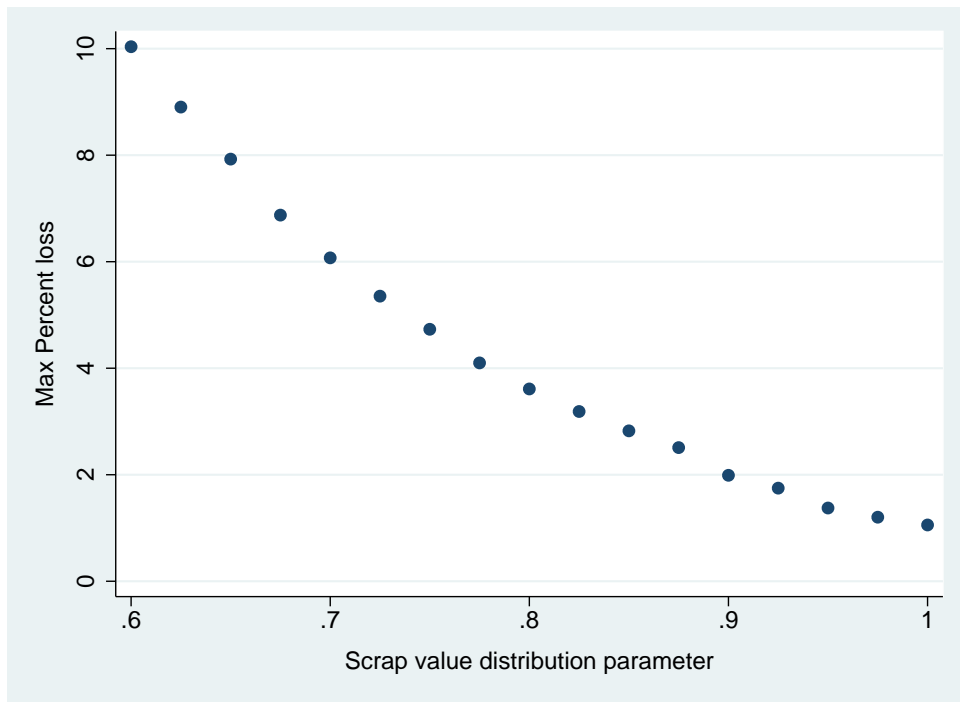
This chapter explored the use of two period strategies for discrete dynamic games. Capital budgeting surveys, management interviews, and industry anecdotes suggest this rule of thumb, which maps closely to "hurdle rate" and "payback period" ratios, indicate these





**Figure 2.7:** *Impact of scrap value distribution parameter on mean loss*

strategies are quite common in practice. It found that the loss of mean discounted payoffs that agents suffer when employing them in a simple entry/exit game are quite low. It also found they tend to be insensitive to changes in the underlying market characteristics.



**Figure 2.8:** *Impact of scrap value distribution parameter on maximum loss*

## Chapter 3

# The Impact of Money on Science: Evidence from Unexpected NCAA Football Outcomes

### 3.1 Introduction

Investments in science can generate large social returns. Scientific discoveries have eradicated diseases, reduced famine, increased labor productivity, and supported national defense. However, scientific laboratories and experiments are expensive to run and research funds are often the key limiting factor in scientific advancement. Together these facts make the level of R&D investment a central concern of university administrators and public policymakers. There is urgency about this issue in the United States where the federally-financed share of university research has fallen over the last forty years and recent recession-induced budget cuts have slashed states' investment in academic research. These developments prompted the America COMPETES Acts of 2007 and 2010 [National Research Council, 2012], which called for a doubling of funds to basic science, and a potential reauthorization of the Act in 2014. Despite its salience, the question of whether policymakers fund a socially-optimal level of R&D remains open, largely because there are few estimates of the causal impact of

research expenditures on scientific discovery.

This paper estimates the dollar elasticity of research output across American universities. It addresses two empirical challenges. First, research grants tend to be awarded to more productive institutions. This endogeneity causes parameter estimates to be upward biased. Second, expenditure data that includes long-term projects of various lengths and lags will make tying money to outcomes difficult. For example, construction of the multi-billion dollar Large Hadron Collider began ten years prior to the first experiments. This errors-in-variables problem causes estimates to be downward biased. The first of these problems suggests—at a minimum—controlling for institution-specific effects, although this tends to amplify the bias from the second. To solve both problems, we exploit an exogenous shifter of marginal research funds between universities: the unexpected success of college football teams.

Identification relies on the fact that football team performance impacts cash flow to the university and, in turn, the funds available for research. Even if unobserved school-specific factors that drive research output also influence football team success, they are unlikely to influence *unanticipated within-season* changes to team success. We measure football team success using the Associated Press Top 25 Poll and use the difference between post-season and pre-season vote counts as the instrumental variable. Since the individual voting results of the Top 25 Poll are made public, and the professional sportswriters who vote have a significant reputational stake in properly forecasting teams' quality and teams' true prospects, the difference between post-season outcomes and pre-season expectations can be treated as random.<sup>1</sup>

Three aspects of this relationship aid greatly in obtaining results. The first is the degree to which swings in football fortunes impact overall school finances. Since the late 1980s college football has generated tens of billions in cash flow to American colleges and universities. One of the more prominent examples is the University of Texas at Austin: in 2013 its football

---

<sup>1</sup>Readers unfamiliar with the context can consider an injury to a key player as the sort of random shock underlying this variation.

team generated more revenue than the majority of professional National Hockey League teams.<sup>2</sup>

At Louisiana State University, football revenue is nearly a third of total tuition receipts. A large portion of this revenue is ploughed back into the athletic department, but a sizable part is returned to the school's general account in the form of unrestricted funds. In addition, a successful football season on the field usually translates to a successful fundraising campaign off the field. For example, Texas A&M University raised more money the night after its star quarterback Johnny Manziel's Heisman Trophy win than it typically raises in a month, in turn setting records for quarterly and annual alumni giving (Holmes, pers. comm., July 11, 2014). The second is that this source of funds is highly volatile, which means that administrators are likely to treat these changes as temporary windfalls rather than opportunities to start long-term projects. The third is that much of a team's success is, in fact, quite unpredictable. This is empirically true in our data and a fact to which college football fans can attest.

We use a two-stage least squares (2SLS) specification to estimate the impact of money on scientific output, which we measure in four ways. When the output measures are scholarly articles and the citations that accrue to them, we estimate dollar elasticities of 0.31 and 0.59, respectively. When the output measures are new patent applications and the citations that accrue to them, we estimate dollar elasticities of 1.91 and 3.30, respectively. These estimates contrast sharply with non-IV estimates under the same specification, which tended toward zero. All calculations are made controlling for time and school effects and school-specific time trends. The non-IV results closely resemble prior work by Adams and Griliches [1996, 1998] and would lead to underinvestment in scientific research, with two important caveats as these results apply to policy-setting. First, since the predicted variation in research expenditures is linear in the instrument, unexpected losses hurt research budgets as much as unexpected wins help them. The transfers of research budgets between schools are assumed

---

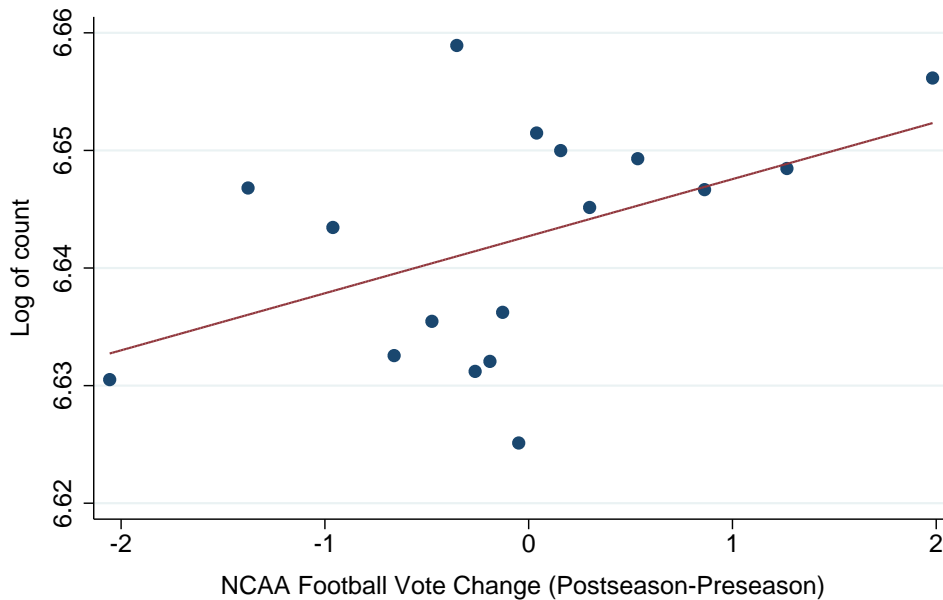
<sup>2</sup>Smith, Chris. "College Football's Most Valuable Teams 2013: Texas Longhorns Can't Be Stopped." *Forbes*, December 18, 2013. <http://www.forbes.com/sites/chris-smith/2013/12/18/college-footballs-most-valuable-teams-2013-texas-longhorns-cant-be-stopped> (accessed December 2, 2014).

zero sum, so if the schools are merely trading scarce output-producing assets—like highly productive scientists—then an aggregate increase in research expenditures may have no impact at all, even though our results predict a strong positive impact of expenditures on each *individual* institution. This seems unlikely, since small and perhaps temporary budget shocks are unlikely to result in long-term and expensive commitments like hiring. Moreover, these high output faculty would bring federal grants along with them, a point we address below and do not see in the data. How precisely universities and the scientists within allocate funds is beyond the scope of this paper but nonetheless an important question we leave to future research. Second, the main specifications take an explicit stand on the timing of football, funding, and publishing. Misspecification can bias the coefficients, so we provide support for the assumption and discuss the factors underlying the temporal relationships.

Figures 3.1-3.2 illustrate the reduced-form relationship. The x-axis in each shows unanticipated football success, measured by within-season changes in Associated Press voting. In Figure 3.1, the y-axis in the top panel represents the log of the count of scholarly articles published and in the bottom panel represents the log of the count of the citations that accrue to those articles. In Figure 3.2, the y-axis in the top panel represents the log of the count of new patent applications and in the bottom panel represents the log of the count of the citations that accrue to those applications. We remove school and year fixed effects as well as school-specific time trends from the variables on both axes. The x-axis has been standardized across polls by standard deviation and lagged appropriately. The positive impact of unexpected football outcomes on all four measures of scientific discovery are positive (and significant at 95%).

We can strengthen the causal interpretation of this relationship with an exogeneity check. Since we observed research funding from federal and non-federal sources separately, we can assess the impact of unexpected football outcomes on both independently. Of course while research funding coming from non-federal sources should be affected by football, those coming from federal sources should not. We find strong evidence for this fact in the data.

This paper contributes to several literatures. In measuring the elasticity of university

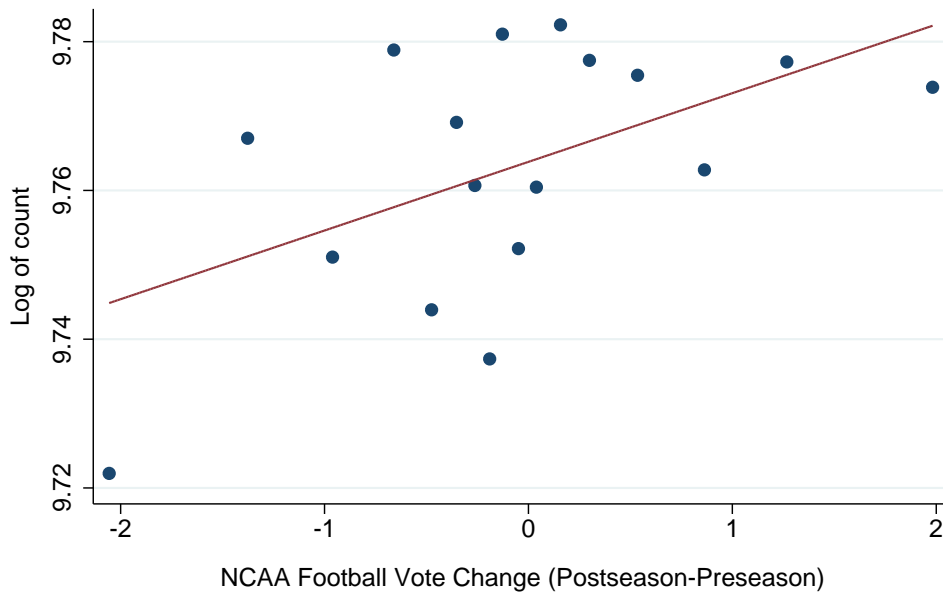


Source: Thomson Reuters for publication data and Associated Press for football data.

**Figure 3.1:** Football and scholarly articles

research expenditures, it follows closely in the footsteps of Adams and Griliches [1996, 1998]. Their cross-sectional OLS specification combines observations over their panel and finds a dollar elasticity of 0.5 when the outcome measure is the number of scholarly articles and 0.6 when the outcome measure is the number of citations that accrue to them. However, when they include university fixed effects to control for institution-specific unobservables, elasticities fall by 80% and are no longer separate from zero. They conclude, *“To date we have little hold over changes in financial and other circumstances that bring about a change in the stream of a university’s research output.”* This is precisely the issue we wish to address.

The scope of our study extends beyond scientific publishing to patenting behavior. After the passage of the Bayh-Dole Act in 1980 allowed academic institutions to retain ownership of inventions developed through federally funded research, it incited a strong growth in academic patenting and patent licensing [Henderson et al., 1998, Sampat et al., 2003, Hausman, 2013]. Pakes and Griliches [1980, 1984] were first to consider patents as an outcome of interest. They found a positive relationship with lagged investment and



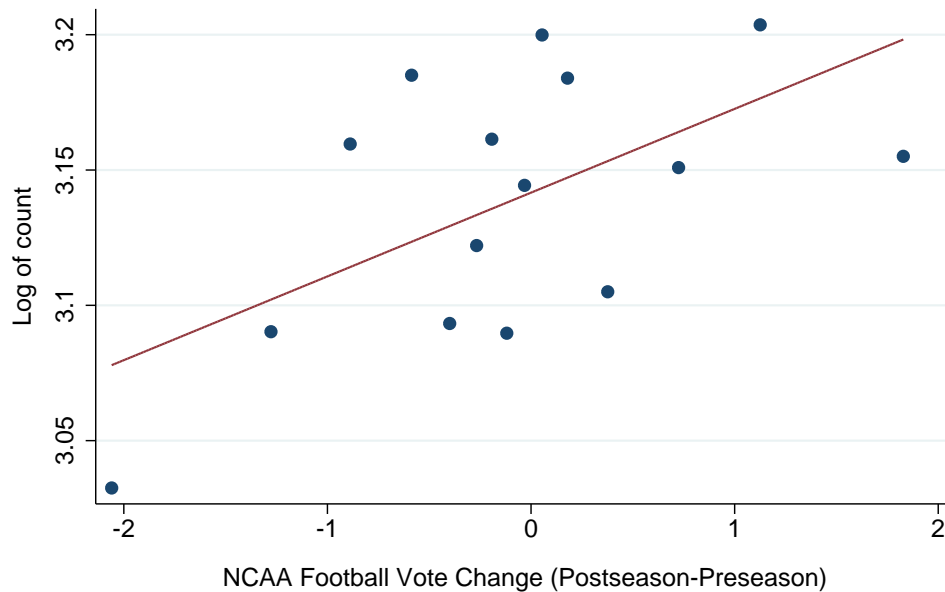
Source: Thomson Reuters for publication citation data and Associated Press for football data.

**Figure 3.2:** Football and scholarly article citations

knowledge stocks in firms.<sup>3</sup> Relatedly, Azoulay et al. [2014] study the impact of government research grants on private sector pharmaceutical and biotech firms. They exploit institutional features of the granting institution to address endogeneity issues and find that a \$10 million increase in government funding generates 3.3 additional patents. Jaffe [1989] spawned a related stream of papers that measured whether R&D efforts spillover to local private firms. The focus on spillovers, however, led this paper and those that followed to focus on exogenous shifts to university research activity rather than university research spending *per se*. As an example, Hausman [2013] uses the Bayh-Dole Act to credibly demonstrate these spillovers on a host of private-sector outcomes like profits and employment. In addition, we shed more light on the mechanisms utilized by academic institutions to fund scientific R&D. The roles of government [Nelson, 1959, Jaffe, 1989, Henderson et al., 1998] and private industry [Mowery and Rosenberg, 1989, Cohen et al., 1998, Wright et al., 2014] have been extensively studied, while the role of science philanthropy only recently started to attract

<sup>3</sup>See Griliches et al. [1988] for a survey of the early literature.



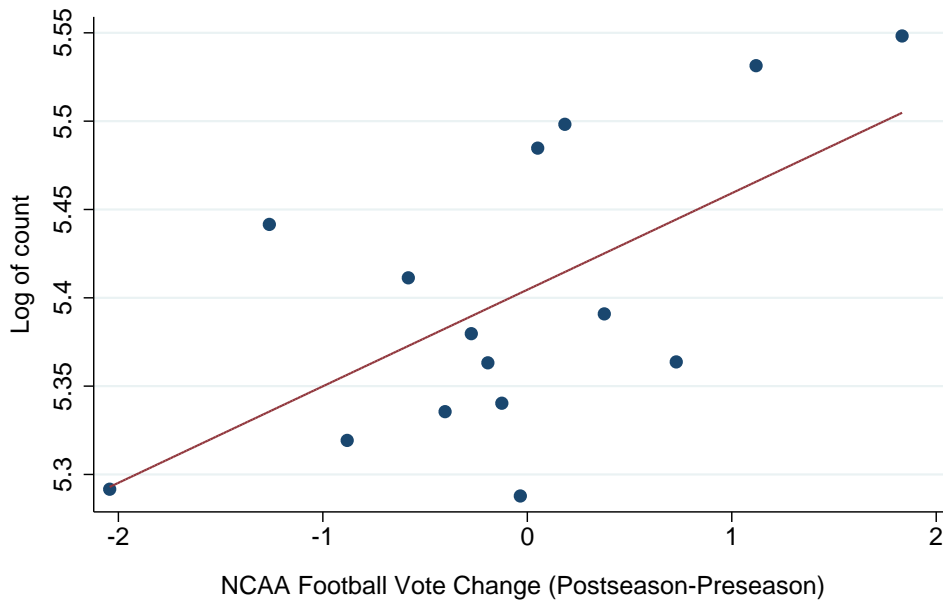


Source: Thomson Reuters for patent data and Associated Press for football data.

**Figure 3.3:** *Football and new patent applications*

more attention [Murray, 2013].

This paper also contributes to recent literature using athletic outcomes for identification. Card and Dahl [2011], for example, study how external cues precipitate violence by showing that domestic abuse rises in cities where the local NFL team suffers an unexpected loss. Anderson [2012] asks whether schools are justified in their large investments into college sports and uses the difference between realized outcomes and betting spreads to show that winning attracts students and donations. Meer [2013] tests habit formation in charitable giving by using prior years' athletic success as an instrument for past giving. The tie between athletics and donations was established previously in Meer and Rosen [2009] using university microdata.



Source: Thomson Reuters for patent data and Associated Press for football data.

**Figure 3.4:** Football and patent citations

## 3.2 US College and University Research

### Spending Levels

Colleges and universities conduct more than 15% of total research and development in the United States, which totaled \$450 billion in 2013. They also account for more than 50% of basic science expenditures [Battelle Memorial Institute, 2013]. These institutions historically relied heavily on the federal government for funding, although the federally-funded share of research has fallen from 78% to 67% over the past four decades. Private funding from corporations has stayed essentially flat, despite wide year-to-year variation. Institutionally-sourced funds have partially compensated, rising from 11% to almost 20% over the same period [National Science Foundation, 2013]. Survey data also suggests this increase is insufficient: 84% of US academic researchers expressed concern over the reduction in US federal R&D funding. For comparison, consider the following: despite the widely publicized shortage of qualified R&D staff in the United States, only 48% of researchers listed this issue as a concern [National Research Council, 2012].

Congress and the White House have taken notice and begun to act on these concerns. In 2005, Congress asked the National Research Council to prepare a plan that would ensure American competitiveness in science and technology. Congress then provided bipartisan support for the America COMPETES Act, which President Bush signed into law in 2007. The Act emphasizes investment in the science, technology, engineering, and mathematics fields and authorizes a doubling of National Science Foundation (NSF) grants for many fields by 2011. In 2009, Congress requested a follow-up report. Two years later, President Obama signed a reauthorization of the bill, the America COMPETES Act of 2010. The budget sequestration process of 2013 reignited this debate. Ultimately, the authorized funding increases were not realized. This has prompted the Director of the National Institute of Health to worry that “we will lose a generation of young scientists” and that “a lot of good science just won’t be done.”<sup>4</sup>

As of the time of writing, Congress has proposed but not yet passed another reauthorization of the Act. A central driver of this debate is uncertainty about where the funding level stands in relation to the optimal social level of R&D, which itself turns on the underlying return on research investment. Measuring this return requires a more detailed examination of the funding process.

### **3.2.1 How Research is Funded**

Private colleges and universities fund their operations primarily through tuition, federal grants, philanthropic donations, and auxiliary enterprises (like healthcare and athletics). For public universities, state appropriations also account for a significant share of incoming cash flow. Schools then use these funds mainly for student instruction, research, administration, and running the auxiliary programs. They budget operational expenses on an annual basis<sup>5</sup>

---

<sup>4</sup>Vergano, Dan. “Science Faces Sequestration Cuts.” USA Today, February 25, 2013. <http://www.usatoday.com/story/tech/sciencefair/2013/02/25/budget-nih-collins/1947277/> (accessed December 2, 2014).

<sup>5</sup>We use the word “operational” to separate these from capital expenses, like construction projects, which are likely to be budgeted long ahead of time.

and typically follow a June rather than December fiscal year end to synchronize with the course-year calendar.<sup>6</sup> The unused portion of funds are rarely allowed to carry over to the next year (Porter, pers. comm., July 22, 2014).

The budgeting process is complicated by widespread earmarks. Strict guidelines on how funds can be spent are attached to a large portion of incoming cash flow, creating a distinction between “restricted” and “unrestricted funds.”<sup>7</sup> Some earmarks are obvious: an NSF grant will go directly to the project for which it was awarded and state appropriations will directly subsidize instruction of in-state residents. Other earmarks are not so straightforward. A multi-million dollar donation by a wealthy single donor or a foundation could carry with it the requirement that it be used to extend hours at an art museum or gymnasium, increase a particular genre of books in the library, or expand the student center. For example, in 2010 Harvard University received a restricted gift of \$50 million from the Tata family to fund two new buildings on the business school campus.<sup>8</sup> For both bookkeeping and flexibility reasons, these unrestricted funds are a precious commodity.

For research, unrestricted funds are often the “source of last resort.” They are needed when costs run over, other sources fall short, or faculty are too new to have attracted sufficient grant money. That is, although federal and state funds still account for the majority of university-led R&D, they are frequently too slow or inflexible to handle the immediate and diverse needs of academic scholars. In the absence of unrestricted funds to close the gap, research is often put on hold. Murray [2013] identifies philanthropic donations as one possible channel for research institutions to fill funding gaps and provides a great example: in 2008, when the fiscal crisis forced the State of California to reduce funding to the UC Berkeley’s Radio Astronomy Lab and federal government cut funding for Allen Telescope Array, Microsoft’s Paul Allen stepped in and donated funds to ensure continuous

---

<sup>6</sup>There are a few exceptions to the June fiscal year end but these too end in the summer months and are immaterial for our discussion.

<sup>7</sup>We thank Kyle Welch for bringing this to our attention.

<sup>8</sup>Walsh, Coleen. “Business School Announces Tata Gift; Two Initiatives.” *Harvard Gazette*, October 14, 2010. [http://news.harvard.edu/gazette/story/2010/10/hbs\\_gift/](http://news.harvard.edu/gazette/story/2010/10/hbs_gift/) (accessed December 2, 2014).

operation of the facility.<sup>9</sup> Combined with other, both large and small philanthropic gifts, unrestricted funds can also allow for scientific research to continue when federal financial support for science does not deliver. Other sources of unrestricted funds include auxiliary operations, like athletics and healthcare, housing, and tuition (primarily for out-of-state residents in the case of public schools). Since it enables us to identify and precisely estimate the elasticity of research output, football's contribution is covered in detail below.

## How Football Contributes Financially

"We took direct dollars from the athletic budget and put it into academic programs."

E. Gordon Gee, 11th and 14th President, Ohio State University<sup>10</sup>

Football contributes to unrestricted university finances in two ways. The first channel is auxiliary revenues. Since the late 1980s, Division I NCAA football has generated over \$10 billion in sales. For perspective, Table 3.1 provides the top 20 college football teams in terms of revenue. The sheer size of these programs is staggering, especially in relation to professional teams. For example, The University of Texas at Austin earns nearly \$110 million in revenue. For comparison, this figure is 40% higher than the median professional hockey team and on par with the median professional basketball team. On a per game basis, it is 5 times larger than both of these and about 7 times larger than the median baseball team. Their size relative to total tuition is also quite large. More than half of the schools on the list have football programs that are more than 20% of the total tuition receipts. At Louisiana State University and Agricultural and Mechanical College (LSU), the University of Nebraska-Lincoln, and the University of Oklahoma Norman Campus, this figure is nearly a

---

<sup>9</sup>Murray [2013] emphasizes three key points about science philanthropy: that it is mostly channeled into restricted funds, that it heavily favors translational science, and that it generally does not strive to fill funding gaps. However, it is important to note that her study focuses on large philanthropic gifts (>\$1M) at top 50 research institutions and provides only one part of the funding equation. Some of these large philanthropic gifts happen to be unrestricted, and universities also collect many small philanthropic gifts which are usually unrestricted in nature. For example, the Harvard Alumni Association webpage provides opportunities for alumni to donate directly to the various university funds, most of which are unrestricted.

<sup>10</sup>"Dropping The Ball: The Shady Side Of Big-Time College Sports," The Bob Edwards Show (Washington D.C.: Public Radio International, January 4, 2015).

third. Football also dwarfs other athletics in this sample. With only three exceptions, football contributes more to athletics revenue than all other sports combined. This is generally true outside of the current sample, too. Despite the popularity of college basketball, for example, its financial importance pales in comparison to football across virtually all US schools.

**Table 3.1: Football and university finances**

	Football			# Seasons	
	Revenue	As a percent of all Athletics Revenue	Tuition Receipts	Ranked 1 or 2	Unranked
( NCAA Football: 12 game season )					
The Univ of Texas at Austin	\$109,400,688	66%	23%	2	9
The Univ of Alabama	88,660,439	62%	25%	2	9
Univ of Michigan-Ann Arbor	81,475,191	66%	9%	1	5
Univ of Notre Dame	78,349,132	72%	29%	3	10
Univ of Georgia	77,594,300	79%	23%	1	9
Auburn Univ	75,092,576	73%	26%	2	8
Univ of Florida	74,820,287	58%	23%	4	2
Louisiana State Univ A&M College	74,275,838	63%	32%	2	10
Univ of Oklahoma Norman Campus	69,647,986	56%	31%	1	9
Univ of Arkansas	61,492,925	62%	8%	0	12
Ohio State Univ-Main Campus	61,131,726	49%	4%	4	4
Pennsylvania State Univ-Main Campus	58,722,182	56%	9%	1	8
Univ of Washington-Seattle Campus	56,379,534	66%	32%	1	13
Univ of Nebraska-Lincoln	55,866,615	64%	16%	3	5
Univ of Iowa	55,648,679	52%	20%	0	12
The Univ of Tennessee	55,359,423	50%	17%	1	8
Univ of Oregon	53,982,076	66%	13%	1	12
Texas A & M Univ-College Station	53,800,924	69%	27%	0	10
Univ of Wisconsin-Madison	50,641,993	35%	8%	0	12
Michigan State Univ	47,869,615	60%	5%	0	16
( For comparison)					
Median Pro Football Team (16 game season)	\$269,000,000				
Median Pro Baseball Team (162 game season)	214,000,000				
Median Pro Basketball Team (82 game season)	139,000,000				
Median Pro Hockey Team (82 game season)	80,500,000				

Source: NCAA.org (college athletic revenue), US Dept of Education National Center for Education Statistics (tuition), Associated Press (team rankings), and Forbes.com (professional sports revenue).

A share of these revenues are returned to the general university fund and ultimately support academic endeavors. For example, in 2012, the Louisiana State University team pledged over \$36 million over 5 years to support the school's academic mission. In 2005, the

Notre Dame football used \$14.5 million of its post-season bowl winnings to fund academic priorities. From 2011 to 2012, the University of Florida team gave \$6 million to cover shortfalls in university funding [Dosh, 2013]. From 2012 to 2013, the University of Texas - Austin gave \$9.2 million of its \$18.9 million back to the university fund while the University of Nebraska - Lincoln did the same with \$2.7 million of its \$5.2 million surplus [Lavigne, 2014].

The second channel is alumni contributions. Football success is a major catalyst for philanthropic fundraising shocks [Meer and Rosen, 2009, Anderson, 2012]. For example, Texas A&M University raised more money the night after its freshman quarterback, Johnny Manziel, won the Heisman Trophy than it typically raises in a full month. That year, the school announced it received a record-setting \$740 million in donations (Porter, pers. comm., July 22, 2014). The university chancellor John Sharp highlighted the significant role college football played in their fund-raising efforts, stating, "Football is one heck of a megaphone for us to tell our story".<sup>11</sup> Schools also can directly tie athletic privileges to academic donations. Stinson and Howard [2010, 2014] document how one large Midwestern school makes donors of academic gifts over \$3,000 eligible to buy season tickets.

Football success, and most likely its financial contribution, are quite volatile. This is shown in the rightmost two columns of Table I. Sixteen of the twenty schools have competed for the national championship over the panel 1987 to 2012. On the other hand, every team was unranked at least twice over the panel, and many were unranked more than ten times. These reversals of fortune are important because variation in teams' rank provides the underlying variation for our identification. For example, a surprising 11-0 record of Boise State University football team in 2004-2005 resulted in an marked increase in university donations, a 66% increase in sales of university merchandise at the bookstore, and a 60% increase in sales of the subsequent year's seasons tickets [Grant et al., 2008].

---

<sup>11</sup>Troop, Don. "Texas A&M Pulls in \$740-Million for Academics and Football." *The Chronicle of Higher Education*, September 16, 2013. <http://chronicle.com/blogs/bottomline/texas-am-pulls-in-740-million-for-academics-and-football/> (accessed December 2, 2014).

## 3.3 Data

### 3.3.1 Sources

We draw data from four sources. The first is vote data from the Associated Press (AP) Top 25 Poll, which we use to construct our instrumental variable. The poll surveys sixty-five sportswriters and sports broadcasters. Each provides a ranking for the top twenty-five teams from NCAA Division I. Each team receives 25 points for each 1st place vote, 24 points for each 2nd place vote, and so forth, and the votes are aggregated over survey responses.<sup>12</sup> The AP publishes the vote totals of all teams. Ballots are collected weekly through the season, with results made public and published at the end of the week. We measure the within-season change in team quality by subtracting pre-season votes from end-of-season votes. Polls varied slightly in the number of voters and, in 1987 and 1988, the number of points allocated, so we normalize the measure by standard deviation. This data is widely disseminated each week of the season and has a special place in college football; unlike professional sports or other college athletics, which rely on playoffs and divisional rank and record, polls were the sole source of determining an NCAA football champion until 2013.<sup>13</sup> At least three other polls are widely published, although the AP Poll is the best known. Moreover, although they are closely correlated, the other major polls had obvious limitations for our setting.<sup>14</sup> The relevant time variable for this data is the fiscal year in which a season is wholly contained. Fiscal years coincide with the academic calendar for schools in our data.

The second component is academic publishing data. Thomson Reuters Web of Science collects this for their Incites database product. We extract a count of the scholarly articles published and a count of the citations that accrue to those articles (up to the date of data

---

<sup>12</sup>The exception is for 1987 and 1988, where voters ranked only the top 20 teams. For these polls, teams received 20 points for each 1st place vote, 19 points for each 2nd place vote, and so forth.

<sup>13</sup>In 2014, a playoff system was instituted.

<sup>14</sup>The BCS Poll, for example, did not cover our full sample. The Coaches Poll could, hypothetically, be contaminated by strategic voting. Other polls were much less widely known and relied upon.



retrieval). Observations are specific to a calendar year, institution, and academic discipline. Since the instrument only has variation at the institution-year level, we aggregate up to this level by taking a sum over all science disciplines, excluding social sciences and medicine.<sup>15</sup> Although including the latter two categories improves power in our first stage, it can bias our estimates away from the elasticities of interest.<sup>16</sup>

The third component is US patent application data. Thomson Reuters collects this for their Thomson Innovation database. It allows us to identify university patentees better than the raw USPTO patent records. We use the browse feature in Thomson Innovation Assignee/Applicant search field to identify all possible university name variations together with unique 4-letter Assignee Codes identifying one of approximately 22,300 patenting organizations worldwide. This enables us to count and aggregate patent applications wherever a college or university appears as an assignee or applicant on the patent record. Again, we extract a count of new patent applications filed and a count of the citations that accrue to those patents (up to the date of data retrieval). Although patents are assigned into technological classes, there is no clear map to academic disciplines. Thus, we aggregate up to the institution-year level by taking a sum over all classes. We assemble this data on a fiscal year basis. More details on patent dataset construction are provided in the Appendix.

The final component is university research expenditure data. The National Science Foundation (NSF) collects this data annually in their Higher Education Research and Development Survey (prior to 2010, called the Survey of R&D Expenditures at Universities and Colleges). Responses are carefully reviewed and verified as needed.<sup>17</sup> The survey

---

<sup>15</sup>For the Incites database, this includes physics, chemistry, mathematics, computer science, biology and biochemistry, microbiology, plant and animal science, agricultural science, geoscience, environmental science, and ecology.

<sup>16</sup>Social sciences are not central to the current policy debate. They also tend to have longer and more dispersed publication lags relative to non-social sciences, which will bias our coefficient estimates downward (unless we take a much stronger stand on the timing). In the same vein, medical research will include a large number of development applications relative to the other natural sciences. Unrelated to these, we are also forced to exclude space science, which includes astronomy but is dominated by aerospace and aeronautical engineering.

<sup>17</sup>In two cases where we needed clarification, the NSF had also asked for them. This gave us confidence that the data was thoroughly reviewed and validated by the NSF. Ronda Britt at the National Center for Science and

is an annual census of all institutions spending at least \$150,000 in separately budgeted R&D. The data is broken down by federal and non-federal sources as well as by disciplines. Our first expenditure measure is tied to scholarly articles, so as with the Thomson Incites data, we take a sum over all science disciplines, excluding social sciences and medicine.<sup>18</sup> Our second expenditure measure is tied to new patent application filings, which are not discipline specific, so we take a sum over all non-social science, engineering, and medical disciplines. This data is on a fiscal year basis.<sup>19</sup>

### **Panel Length and Scope**

The instrument is based on the difference between post-season and pre-season votes. Since the median NCAA team receives zero votes, using the universe of teams would result in a very large number of zero values. So that the schools are selected agnostically and the instrument has power, we simply order the teams by the sum of the absolute value of their vote changes and select the top forty schools for our panel. This is exactly  $\frac{1}{3}$  of the 120 Division I teams. The only caveat is that if there are heterogeneous treatment effects, our estimates pertain only to schools with large football programs. The resulting list is very diverse. It includes private (e.g. Stanford, Notre Dame) and public (e.g. Alabama, Nebraska) institutions as well as relatively small (e.g. Boise State) and large (e.g. Texas) ones. The magnitude of our estimates are not very sensitive to the size of the panel.<sup>20</sup>

---

Engineering Statistics was particularly helpful. Our main issue was missing values for Boise State University prior to 1992 and in 2005 and 2006. In the earlier years, the institution was below the survey threshold. For the later two years, the NSF followed up with the school and confirmed it made an error in reporting due to a personnel change. We omit these years from our analysis, although the results are robust to dropping this institution entirely.

<sup>18</sup>For the NSF data, this includes physics, chemistry, and mathematics and statistics, computer science, biological sciences, and other life sciences, agricultural sciences, geosciences, oceanography, atmospheric sciences, and earth sciences.

<sup>19</sup>This includes all departments from the first measure, as well as medical sciences (including clinical medicine, immunology, pharmacology and toxicology, and molecular biology), engineering (including aerospace, chemical, civil, electrical, materials, mechanical, and other), interdisciplinary and other sciences, and astronomy (which, along with aerospace engineering, would be classified as "space science" in the Incites database).

<sup>20</sup>Clearly, however, shrinking the list far below forty simply limits power throughout specifications while expanding the list far beyond forty can introduce enough zeros to the instrument to weaken it.

The beginning of the panel coincides with the start of the “modern era” of college football, which traces back to the 1984 Supreme Court ruling on *NCAA v. Board of Regents of the University of Oklahoma*.<sup>21</sup> Prior to the ruling, the NCAA restricted the number of games that could be broadcast, threatening non-complying schools with an association-wide boycott. In 1981, two schools challenged the NCAA’s authority and in 1984, the Burger court ruled that the NCAA violated antitrust laws by controlling television broadcasting rights. Effectively, schools and their conferences were now free to negotiate directly with broadcasters. Broadcast networks treated the first year or two as a trial for the new arrangement, but by 1987 the number of televised games and the exposure of the league surged, leading to an unprecedented financial gain. That year featured the highly contentious Fiesta Bowl, which became one of the most watched college games in history, and marks the start of our panel of football outcomes.<sup>22</sup> Data on scholarly articles begin on the same date, while the patent data begin in 1996. Although we observe data for earlier periods, the international harmonization of the United States patent system in the early 1990’s created a large spike in the number of filings and seemingly increased the overall level of patenting. If the response of patenting behavior to research funding was different prior to 1996, and the goal is to recover parameter estimates that are informative for current policymaking, then including data on filings prior to 1996 will lead to the wrong parameter estimates. The dataset ends in 2011. While 2012 data was available for our outcome measures, scholarly articles and patents have had so little time to attract citations and the resulting drop off is so steep that these additional points create essentially only noise. Moreover, there is a chance that patent applications filed in 2012 have not yet been recorded as of the writing of this paper. This leaves 23 and 16 years of observations for scholarly articles and patents,

---

<sup>21</sup>See *NCAA v. Board of Regents of the University of Oklahoma*, 468 US 85 (1984).

<sup>22</sup>The game pitted Penn State against a heavily-favored University of Miami. The pre-game antics of Miami, including dressing in military fatigues for the flight to the game, and controversial remarks by both sides at a joint team dinner the night before the game contributed to wide-spread media attention. For the first time in history, a sitting US President (Ronald Reagan) was interviewed at the halftime show. Penn State won 14-10. The national press coverage of the players, coaches, their backgrounds, and the developments leading up to the game are all common in the “modern era” but were unheard of prior to 1987.

respectively.

### 3.3.2 Summary Statistics

First, we summarize the data by institution. There are forty in total. Texas A&M - Main Campus spends the highest amount on non-social non-medical science research, at \$132 million, followed by the University of Georgia. The mean level is \$49 million. Texas A&M - Main Campus also spends the most in total non-social science and engineering, at \$259 million, followed closely by the University of Wisconsin - Madison. The mean level is \$98 million.

The University of Illinois at Urbana-Champaign publishes the highest number of scholarly articles, at an average of over 3,000, followed by the University of Wisconsin - Madison. The average number is 820. The University of Illinois at Urbana-Champaign also has the highest average number of related citations, at an average of over 90,000, followed by Stanford University. The average number is 22,336. Stanford University tops the list of new patent application filings and the citations that accrue to them, at 156 and 3,429, respectively. The University of Texas - Austin is second, with 117 and 2,205, respectively. The mean levels are 34 and 570, respectively.

Next, we summarize the data by year. The average level of non-social non-medical science expenditures grows from \$24 to \$75 million from 1987 to 2011, while the average level of total non-social science and engineering expenditures grows from \$41 to \$165 million over the same period. This is an average compound growth rate of 5% for the former funding measure and 6% for the latter. The funding measures are monotone increasing over the panel, with a few exceptions. In 1994, 2004, and 2010, both funding measures drop relative to the year before (in nominal terms). These years directly follow the peak unemployment periods of the last three US recessions.<sup>23</sup> Over the same period, scholarly articles grow at 3.2% while patent applications grow more than twice as fast, at 7.2%. There is considerable

---

<sup>23</sup>Peak unemployment hit 7.8% in 1992, 6.3% in 2003, and 10.0% in 2009 for the 1990s recession, 2000s recession, and "Great Recession," respectively. Source: "Business Cycle Expansions and Contractions." NBER Website. March 5, 2015.

variation across schools within each year, but both variables tend to increase monotonically over the panel. The time series of citations is more complicated, since the amount of time other work has to cite these articles and applications is falling over the panel. Both citations are monotonically increasing up to and including 1998 and then monotonically decreasing after and including 2005. In any case, all main empirical specifications below include year fixed effects, so these issues should not present a problem.

## **3.4 Empirical Model**

### **3.4.1 Overview**

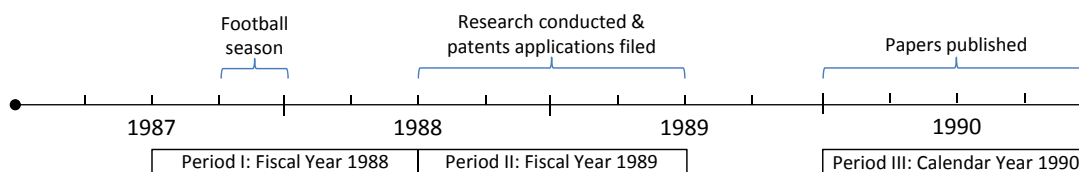
We aim to better inform policymakers and administrators about the impact to scientific output from an additional dollar of investment in university research. Estimating this requires addressing two empirical issues. The first comes from the fact that high quality institutions attract big grants as well as big ideas. This causes parameter estimates to be upward biased and suggests that, at a minimum, removing the institution-specific means and time trends from the data is required. However, this still leaves open the question of endogeneity and, as Adams and Griliches [1998] note, probably exacerbates the second issue, an errors-in-variables problem. When the data include long-term projects with multi-year payoffs, tying research outcomes to the expenditures that generated them becomes difficult. Even if a tight causal relationship exists, estimating it can be impossible without information that the econometrician rarely has access to. In the case of the multi-billion dollar Large Hadron Collider at CERN, construction began ten years prior to the first experiments. In the case of the Stanford Linear Accelerator Center, researchers still benefit from portions of the initial \$114 million investment in 1961.

The solution is to find a quantity in the data generating process that shifts only marginal research funds and yet is not correlated with the time-varying quality of the institution. To achieve this, we use unexpected NCAA football outcomes. Unexpected wins, for example, shift out research funds and, in turn, drive scientific discovery. Football presumably has

a negligible effect on the ability of a school to conduct cutting edge research, and so is excluded from the outcome variables except through funding—especially one or two years into the future.

## Timing

Our assumptions regarding the temporal relationship between the variables are as follows: football outcomes impact the level of research in the subsequent period and scholarly articles in the period subsequent to that. Since patent applications usually need to be filed with the USPTO prior to discussing findings in a public forum, i.e. seminar or conference, the patent filings are typically concurrent with the research. Figure 3.5 illustrates these relationships using our first year of data.



**Figure 3.5:** *Timing of football, research, and publishing*

The first period is fiscal year 1988. This period covers regular season football, which is played in the fall of 1987, as well as post-season football, which is played in January of 1988. Football outcomes impact incoming unrestricted funds during this period, including playoff “bowl” proceeds, alumni donations, and the pre-sale of the next season’s seats and broadcasting rights. Changes in these incoming funds are budgeted out and spent in the following period, fiscal year 1989. Research is conducted. Alongside or immediately following the research, scientists file patent applications, which must legally precede any dissemination of the findings. The final period is calendar year 1990. Successful research carried out in the second period will be published in journals during this period. The temporal relationship between football, expenditures, and publishing is an assumption we discuss in detail in a later section.

### 3.4.2 Specification

The first stage assesses the relationship between the instrument and the endogenous regressor. Specifically, we estimate the following:

$$\text{LogNonFedExpenditures}_{i,t} = \alpha_0 + \alpha_1 \text{Football}_{i,t-1} + \mu_i + \delta_t + \gamma_i t + v_{i,t} \quad (3.1)$$

where  $i$  denotes institution,  $t$  denotes the fiscal year,  $\text{LogNonFedExpenditures}$  denotes the log of non-federal research expenditures,  $\text{Football}$  denotes the difference between postseason and preseason Associated Press votes (standardized across polls),  $\mu$  and  $\delta$  denote school and time dummies, and  $\gamma$  captures the school-specific time trend (omitting the superscripts). We use first stage estimates,  $(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\mu}_i, \hat{\delta}_t, \hat{\gamma}_i)$ , to generate predicted values for  $\text{LogNonFedExpenditures}_{i,t}$ , denoted  $\widehat{\text{LogNonFedExpenditures}}_{i,t}$ .

To estimate the dollar elasticity of scientific output, we regress the log of each of our four output measures on the predicted values from the first stage. The four estimating equations are given by the following:

$$\text{LogArticles}_{i,t} = \beta_0^1 + \beta_1^1 \widehat{\text{LogNonFedExpenditures}}_{i,t-1} + \kappa_i^1 + \phi_t^1 + \lambda_i^1 t + \epsilon_{i,t}^1 \quad (3.2)$$

$$\text{LogArticleCites}_{i,t} = \beta_0^2 + \beta_1^2 \widehat{\text{LogNonFedExpenditures}}_{i,t-1} + \kappa_i^2 + \phi_t^2 + \lambda_i^2 t + \epsilon_{i,t}^2 \quad (3.3)$$

$$\text{LogPatents}_{i,t} = \beta_0^3 + \beta_1^3 \widehat{\text{LogNonFedExpenditures}}_{i,t} + \kappa_i^3 + \phi_t^3 + \lambda_i^3 t + \epsilon_{i,t}^3 \quad (3.4)$$

$$\text{LogPatentCites}_{i,t} = \beta_0^4 + \beta_1^4 \widehat{\text{LogNonFedExpenditures}}_{i,t} + \kappa_i^4 + \phi_t^4 + \lambda_i^4 t + \epsilon_{i,t}^4 \quad (3.5)$$

where  $\kappa$  and  $\phi$  denote school and time controls and  $\lambda$  captures the school-specific time trend (again, omitting superscripts). (2) and (3) use lagged expenditures while (4) and (5) use contemporaneous expenditures. For identification, we require that  $\text{Football}_{i,t-1}$  is uncorrelated with  $\epsilon_{i,t+1}^1$ ,  $\epsilon_{i,t+1}^2$ ,  $\epsilon_{i,t}^3$ , and  $\epsilon_{i,t}^4$ , and that  $\text{Football}_{i,t-1}$  is a sufficiently strong

predictor of  $\text{LogNonFedExpenditures}_{i,t}$ . We cluster our standard errors at the university level, which allows for arbitrary correlation of the unobservables within a university over time, i.e. the squared sum of regressor and error are required to have the same distribution across clusters. In theory, we could also allow for arbitrary correlation of the unobservables within-year in the same specification, as utilized in Petersen [2009], but this is too steep a requirement of the data. As one of our robustness checks, we tested an alternative specification that clustered at the year level and resulted in smaller standard errors, so we did not include these results (although they are available on request).  $\beta_1$  is the parameter of interest.

We also estimate the dollar cost of a patentable idea (or, to be precise, an idea that the researcher and institution deem worthy of a patent application). To translate the elasticity estimate into level changes, we multiply the reciprocal of this elasticity—roughly the percent change in research expenditures required per one percent change in patent applications—by the average ratio of expenditures to applications. Thus, the cost estimate equals

$$\frac{1}{N} \frac{1}{T} \sum_i \sum_t \hat{\beta}_1^3 \frac{\text{NonFedExpenditures}_{i,t}}{\text{Patents}_{i,t}}$$

where  $N$  is the number of schools.

There are two potential concerns about the exclusion restriction. Both seem small. One occurs if unexpected college football outcomes drive research outcomes, whether in the laboratory or publication process. This includes the case where, for example, football success may attract researchers that are inherently more productive on average. Previous work like Anderson [2012] has shown that the undergraduate student body does improve after teams win. This same argument is unlikely to hold for graduate students and faculty. Another occurs if a third factor simultaneously improves both football and research outcomes, but goes unnoticed by the Associated Press voters. Since the reputation and career prospects of these sports writers and broadcasters depend on the accuracy of their predictions and their perceived access to information, this also seems unimportant. Nonetheless, evidence in the “Exogeneity Check” section provides further support for the instrument.



## 3.5 Results

### 3.5.1 From Football to Money

Our first stage results assess the relationship between unexpected football outcomes and research expenditures. Table 3.2 reports these results. Each specification includes university and time fixed effects as well as university-specific time trends. Columns 1-2 show that a one thousand unit change in the vote difference would increase non-medical non-social science expenditures by 3.3% and total non-social science and engineering expenditures by 2.6%. These estimates are significant at 99.6% and 97.5% levels, respectively. Columns 3-4 show that this same change would increase non-medical non-social science expenditures by \$1.775 million and total non-social science and engineering research expenditures by \$1.966 million. The first of these estimates is significant at the 99.3% level, but the second is not precisely estimated. The fit of these specifications range between 95% and 98%, which is not surprising given the large number controls.<sup>24</sup>

**Table 3.2:** *The impact of football on non-federal research expenditures*

	(1)	(2)	(3)	(4)
1000 Vote Change	0.0327*** (0.0108)	1,775*** (618.3)	0.0256** (0.0110)	1,966 (1,226)
Constant	115.9*** (0.185)	1.053e+07*** (10,376)	124.2*** (0.190)	2.983e+07*** (19,981)
School FE	x	x	x	x
Year FE	x	x	x	x
School Time Trend	x	x	x	x
R-squared	0.965	0.956	0.976	0.963
Observations	949	949	949	949
Number of Clusters	40	40	40	40

*Note.* - Standard errors, clustered at university level, in parenthesis. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

These estimates square with stylized facts about college football and finances. A one

<sup>24</sup>Raw vote differences are used here so the coefficients can be easily interpreted. In the 2SLS results below, vote differences are normalized across years to improve comparability and improve power.

thousand vote change is approximately equal to, for example, a move from 17th place to 1st place or from an unranked position to 10th place. The comparison is imperfect, but this translates to a \$60 million revenue change in Table 3.1. Our discussions with administrators suggest that roughly five to ten percent of cash flow changes find their way back to university research, and translate into somewhere between \$3 million and \$6 million of additional funding. Since funds are shared between social and non-science departments, and since higher revenues translate to higher costs—for example, hiring more security guards at games to monitor larger crowds at games—then our estimates are in line with what one would expect.

### **3.5.2 From Money to Scholarly Articles**

To assess the impact of money on science, we begin with the relationship between research expenditures and academic publishing behavior. Table 3.3 reports these results. This table, as well as the three that follow, present the OLS estimates in the first five columns and the 2SLS estimates in the latter five. Our main findings are in the final column. We find that the dollar elasticity of scholarly articles is 0.310, after controlling for school and time fixed effects as well as school-specific time trends. The instrument, lagged unexpected college football success, provides exogenous variation to science and engineering research expenditures sourced from the university. The estimate is significant at 99.2%. The F-statistic of the accompanying first stage is 10.98.

The sharp contrast with the OLS results is striking. Our main elasticity estimate is nearly ten times what results from an OLS specification with the same level of controls, which would lead policymakers to underestimate the returns to funding scientific research and presumably under-invest in it. One potential issue is that the instrument is identifying a local average treatment effect that is substantively higher than the average elasticity of the sample schools (or sample school-years). This would happen if the sensitivity of the schools' budgets to football outcomes are correlated with the schools' elasticity. If anything, we would expect this to go in the opposite direction—with schools transforming dollars to

**Table 3.3: The impact of money on scholarly articles**

	OLS Specifications					IV Specifications				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Lagged Log Expenditures	0.456*** (0.0962)	0.438*** (0.105)	0.438*** (0.0419)	0.0658** (0.0280)	0.0349** (0.0171)	0.596 (0.486)	0.603 (0.502)	0.442 (0.481)	0.471 (0.326)	0.310*** (0.117)
Constant	1.930* (1.038)	2.220* (1.168)	0.719** (0.289)	3.559*** (0.208)	1.313*** (0.189)	0.485 (5.046)	0.449 (5.437)	1.158 (4.690)	0.981 (3.326)	1.233 (1.210)
School FE			x	x	x			x	x	x
Year FE		x		x	x		x		x	x
School Time Trend					x					x
Observations	869	869	869	869	869	869	869	869	869	869
R-squared	0.422	0.433	0.956	0.983	0.994	0.382	0.382	0.956	0.966	0.989
Number of Clusters	40	40	40	40	40	40	40	40	40	40
F-Stat: 1st Stage						10.98	10.98	10.98	10.98	10.98

Note. - Standard errors, clustered at university level, in parenthesis. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

discoveries at the highest rates also being the schools whose budgets are least affected by football.

More likely the issue is a rather serious errors-in-variables problem for the OLS. The difficulty in temporally tying budgets to discoveries is at the heart of the problem. For example, portions of the \$114 million investment in the Stanford Linear Accelerator, built in 1961, generated research for years afterwards. In projects like this, operating expenses may precede experiments for many years, weakening the link between research budgets and articles in the subsequent year and attenuating the elasticity estimates.

To explore this point further, we remove institution specific controls. In fact, removing the institution-specific time trend alone results in a nearly tenfold increase in the estimated elasticity (without much relative change in precision). It is, unfortunately, impossible to say whether the sharp rise is attributable to the mitigation of the errors-in-variables problem or to the re-introduction of institution specific unobservables that drive both expenditures and scientific output. Removing the institution or year fixed effects does not further change the estimates much. It seems that whatever the relative contribution of the errors-in-variables problem or the omitted variable bias may be, their combined effect varies in a complicated way—over time and within the institution.

Our pooled OLS estimates are at the bottom end of those found by Adams and Griliches [1998]. In the presence of only time fixed effects and three high-level institutional controls (for top ten public university, top ten private university, and other private university), they find a dollar elasticity of scholarly articles of between 0.4 and 0.7. The school fixed effects also have the same impact on their OLS results that they have on ours: they find an elasticity of roughly zero.

An important caveat follows for policymakers that wish to use this figure to predict returns to an aggregate national increase in research funding. Since the predicted variation in expenditures is linear in the instrument, unexpected losses hurt research budgets as much as unexpected wins help them. Thus, our instrument transfers money between schools rather than shifting aggregate annual spending up or down. If these transfers are merely luring

scarce assets between institutions, and if these scarce assets—like high-output scientists—are inelastically supplied in the short-run, then our estimates are uninformative about how scientific output responds to aggregate funding increases.<sup>25</sup> This is unlikely. Small and temporary shifts in funding do not drive expensive and long-term commitments like faculty hiring. Faculty also, by casual observations, are not perfectly mobile. Finally, successful scientists tend to attract federal grants so their movement would shift federal research budgets, but our exogeneity check below reveals this is not the case. Instead, conversations with administrators and researchers suggested an increase in materials purchases and technical staff hires. They also suggested the latter tend not to have or be in pursuit of an advanced degree, since adding doctoral students and post-doctoral fellows are typically—like faculty—long-term and costly commitments. Exploring precisely how universities or the scientists within allocate these funds is beyond the scope of this paper but an important question we leave to future research.

### 3.5.3 From Money to Scholarly Article Citations

Next, we consider the citations that accrue to the aforementioned articles. Table 3.4 reports these results. We find that the dollar elasticity of article citations is 0.590. The result is significant at 97.2%.

The larger coefficient on citation-weighted articles squares with intuition. Scholars make extensive-margin decisions about whether to take on more projects. They also make intensive-margin decisions about how much to invest in those they already plan to take on. The article count can be thought of as this extensive margin while the citation-weighted count captures both. To see this, considering the limiting case where researchers facing windfall funding invest only in improving projects they already plan to take on: the number of articles would show no change while the citation-weighted count would fully-reflect the investment. The fact that the citation weighted estimate is close to twice the no-weight

---

<sup>25</sup>An aggregate increase aimed at universities may draw scientists away from the private sector, although crowding out hardly seems like a policy goal. It may also draw scientists from abroad, but again this hardly seems like a first order policy goal.

**Table 3.4: The impact of money on scholarly articles' citations**

	OLS Specifications					IV Specifications				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Lagged Log Expenditures	0.374*** (0.135)	0.435*** (0.137)	-0.243*** (0.0697)	0.0743 (0.0535)	0.0574 (0.0370)	0.651 (0.659)	0.798 (0.716)	1.059 (1.331)	0.426 (0.502)	0.590*** (0.268)
Constant	5.897*** (1.450)	4.080** (1.518)	7.900*** (0.480)	4.700*** (0.395)	18.05*** (0.654)	3.037 (6.813)	0.166 (7.739)	-2.084 (12.97)	2.914 (5.120)	16.07*** (2.736)
School FE			x	x	x			x	x	x
Year FE		x		x	x		x		x	x
School Time Trend					x					x
Observations	869	869	869	869	869	869	869	869	869	869
R-squared	0.179	0.349	0.850	0.962	0.985	0.081	0.191	0.594	0.954	0.974
Number of Clusters	40	40	40	40	40	40	40	40	40	40
F-Stat: 1st Stage						10.98	10.98	10.98	10.98	10.98

Note. - Standard errors, clustered at university level, in parenthesis. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

estimate suggests scholars are splitting the investment across these margins.

This is, of course, only one possible interpretation, and Adams and Griliches [1998] propose two interesting alternative views. The first is that as Ph.D. students become junior faculty at smaller schools, papers derived from their doctoral work will be incorrectly attributed to the school that hired them. However, this problem will be partially corrected if these papers happen to cite scholars at their degree-granting institution. The second is that larger programs tend towards basic research, which is more likely to have “hit” papers. We prefer our interpretation of the relative magnitudes since the ratio of articles to citations is robust to the inclusion of institution fixed effects and institution-specific time trends.

The errors-in-variables problem discussed in the preceding section again seems an issue for OLS specifications. Although the pooled OLS specification yields a relatively precise estimate of 0.374, the addition of the full set of controls yields an imprecise estimate of 0.057. These results are, again, near the bottom end of the Adams and Griliches [1998] range. They find a dollar elasticity of article citations of between 0.6 and 0.9 without school fixed effects but close to zero effect with school fixed effects.

### **3.5.4 From Money to Patents**

To assess the impact of money on translational and applied science output, we assess the relationship between research expenditures and patenting behavior. Table 3.5 reports these results. We expand our funding data to encompass total non-social science and engineering disciplines, rather than non-medical non-social science only.<sup>26</sup> We find that the dollar elasticity of patent applications is 1.91. This result is significant at 96.2% and the corresponding first stage F-statistic is 11.18. This elasticity is surprisingly high and implies increasing returns to research spending, i.e. for each proportional increase in research expenditures, new patent applications will rise by more than 1%. The contrast with the OLS coefficients are even more striking than in the case of scholarly publications. Here, the

---

<sup>26</sup>The University of Oregon was the only school in our sample without an identifiable Assignee/Applicant name or DWPI Assignee code in Thomson Innovation. Since the University of Oregon is also the only school in our sample without the School of Engineering, we drop it from the patent analysis portion of the paper.

elasticity from the main specification is between two and three times the precisely-estimated OLS coefficient. Moreover, it is nearly one hundred times the imprecisely estimated OLS estimate with a full set of controls and more than twice the upper bound of the 95% percent confidence interval around that estimate.

We also estimate the dollar cost of generating a patentable idea. This entails dividing the ratio of non-federal research spending to patents by the elasticity estimated above, and averaging across schools and, where applicable, time. Using only the most recent years' spending-to-patent ratios yields a cost of \$2.612 million. Using all years' ratios yields a cost of \$2.975 million. University patenting has increased steadily since the Bayh-Dole Act of 1980, which allowed universities to retain ownership over their publically-funded intellectual property. Thus, the first figure should better predict the response to a current policy change. Despite broader coverage in terms of disciplines and a longer panel, these figures are quite close to—although slightly lower than—the roughly \$3.3 million cost estimated in Azoulay et al. [2014].

### **3.5.5 From Money to Patent Citations**

Finally, we assess the relationship between research expenditures and patent citations. Table 3.6 reports these results. Our main specification yields an elasticity of 3.30, and this result is significant at 98.8%. This estimate is nearly twice the elasticity on the count of patent applications, again suggesting researchers are roughly splitting their time between launching new projects and improving the quality of existing projects. Although the outcome data comes from an entirely separate source than the scholarly article data, it is reassuring to see the ratio of documents to their accrued citations be the same for both scholarly articles and patent applications.

Nonetheless, this suggests strongly increasing returns to research investment at the margin and may be surprising. However, when one considers the large amount of fixed investment, both in terms of faculty and facilities, then if the bottleneck for research—as recent press has indicated—is at the funding level, these elasticities both seem reasonable.



**Table 3.5: The impact of money on patent applications**

	OLS Specifications					IV Specifications				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Lagged Log Expenditures	0.690*** (0.114)	0.657*** (0.112)	1.017*** (0.103)	0.536*** (0.193)	0.0207 (0.313)	-0.0790 (0.960)	0.0460 (0.762)	2.014** (0.901)	1.916** (0.791)	1.914** (0.922)
Constant	-4.713*** (1.321)	-4.780*** (1.266)	-6.886*** (0.791)	-3.759** (1.396)	-2.775 (1.920)	4.037 (10.89)	3.003 (8.900)	-20.08** (9.757)	-19.29** (8.839)	-19.23* (10.36)
School FE			x	x	x			x	x	x
Year FE		x		x	x		x		x	x
School Time Trend					x					x
Observations	607	607	607	607	607	607	607	607	607	607
R-squared	0.401	0.425	0.836	0.850	0.909	0.125	0.125	0.752	0.798	0.851
Number of Clusters	39	39	39	39	39	39	39	39	39	39
F-Stat: 1st Stage						11.18	11.18	11.18	11.18	11.18

Note. - Standard errors, clustered at university level, in parenthesis. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

**Table 3.6: The impact of money on patents' citations**

	OLS Specifications					IV Specifications				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Lagged Log Expenditures	0.609*** (0.0977)	0.746*** (0.128)	-0.872*** (0.258)	0.816*** (0.288)	0.129 (0.229)	-0.0488 (1.374)	-0.110 (1.209)	2.251 (1.954)	2.539*** (1.120)	3.302*** (1.314)
Constant	-1.527 (1.112)	-2.077 (1.410)	9.529*** (1.982)	-2.006 (1.984)	1.322 (1.460)	5.954 (15.60)	5.366 (14.20)	-20.71 (21.15)	-25.97*** (12.50)	-35.06*** (14.72)
School FE			x	x	x			x	x	x
Year FE		x		x	x		x		x	x
School Time Trend					x					x
Observations	604	604	604	604	604	604	604	604	604	604
R-squared	0.174	0.395	0.695	0.805	0.857		0.067	0.235	0.760	0.770
Number of Clusters	39	39	39	39	39	39	39	39	39	39
F-Stat: 1st Stage						11.18	11.18	11.18	11.18	11.18

Note. - Standard errors, clustered at university level, in parenthesis. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

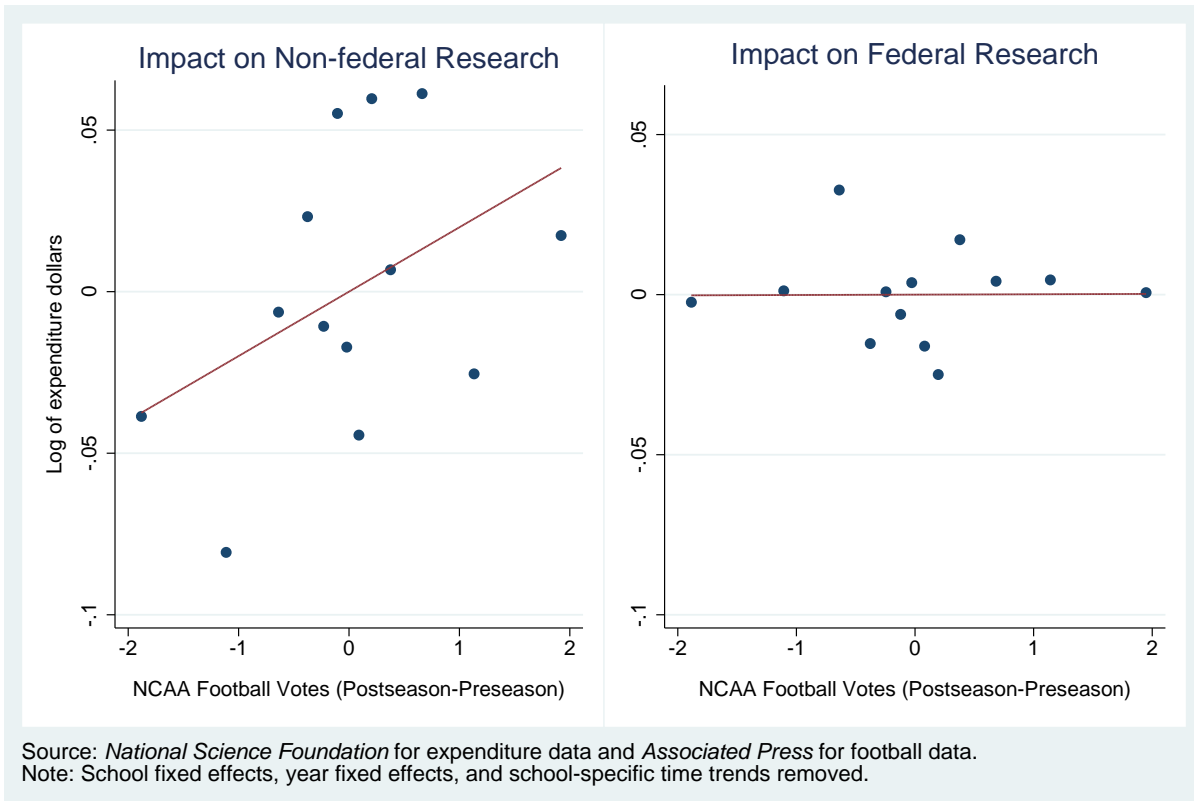
Of course, while one would need to estimate the real returns from academic patents to figure out whether the government and universities are funding the right level of research, our results appear to support arguments for an increase in research spending.

### 3.5.6 Exogeneity Check

College football should impact only research funds provided by non-federal sources and have no effect on funds provided by the federal government. We observe the dollars contributed from these sources independently in the data and use this to strengthen the exogeneity argument for our instrument. That is, if some unobservable factor that varied by institution and time was driving unexpected football success, scientific discovery, and non-federal research funding, then it is likely to show up in federal research funding as well. Figure 3.6 addresses this potential confound. The y-axis shows research expenditures, by source, while the x-axis shows the instrument. School and year effects, as well as a schools-specific time trend, have been removed from both.

The left panel reports the strong, positive relationship between the instrument and research expenditures sourced from non-federal entities. This is merely the graphical representation of the first-stage results. The right panel, in contrast, reports the lack of any relationship between the instrument and federally-sourced research expenditures.

In fact, we can statistically separate the effect on these two sources. To do so, we pool together federal and non-federal data, so that an observation is university-year-source specific. We interact the full set of controls with a dummy variable for non-federal expenditures so that our university and year fixed effects as well as our university-specific time trends are source specific. The standard errors are clustered at the university-source level. Table 3.7 reports the results of this exercise. In the first column, the left-hand side variable is the log of non-medical non-social science expenditures. The instrument has a precisely estimated zero effect on the federally-sourced portion of expenditures. That is, the coefficient is not significant and the 95% confidence interval spans a relatively narrow range of -0.75% to 0.78%. In contrast, the coefficient on the interaction term—representing the impact of the



**Figure 3.6:** *Impact of football on federal vs. non-federal funding*

instrument on the non-federally-sourced portion of expenditures—is positive and significant at over 99%. The second column shows analogous results when the log of total non-social science and engineering expenditures is used as the left-hand side variable.<sup>27</sup>

We interpret this as support for the instrument. As discussed in the “Empirical Model” section, there are two potential issues with identification. The first occurs when unexpected football success causes success in the research or publication process. The second occurs when a third unobserved factor simultaneously drives football and research outcomes. For example, a charismatic new college president could enhance both football and faculty recruiting.<sup>28</sup> Both issues seems *prima facie* unproblematic, but Figure 3.6 lends additional

<sup>27</sup>We thank James Lee for suggesting this specification.

<sup>28</sup>Our restriction to unanticipated football outcomes would further require that the star football recruits either go unnoticed by poll respondents in the pre-season poll, or join after the pre-season poll is completed.

**Table 3.7:** *Exogeneity check for instrument*

	(1)	(2)
Non-Federal Dummy	35.24*** (0.0738)	24.69*** (0.0679)
Instrument	0.000113 (0.00383)	-0.000298 (0.00374)
(Non-Federal Dummy) x Instrument	0.0190*** (0.00615)	0.0163** (0.00672)
Constant	-23.12*** (0.0517)	-11.64*** (0.0528)
School FE	x	x
Year FE	x	x
School Time Trend	x	x
R-squared	0.982	0.984
Observations	1,905	1,905
Number of Clusters	80	80

*Note.* - Standard errors, clustered at university level, in parenthesis. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

proof. If the instrument causes, or is the result of, an aggregate productivity shock at the university-year level, then scientists could attract more grant money, which the data rejects. Moreover, if the instrument enables the university to recruit more productive faculty, or is the result of a factor that enables the same, then these star faculty should bring with them large federal grants. This, too, is rejected by the data.

### 3.5.7 Discussion of Timing

The empirical specification considers a particular temporal relationship between football, funding, and publishing. Misspecification of this relationship can result in biased estimates relative to the policy-relevant parameters. Below we show that although the impact of football extends to periods other than what the model strictly specifies, the impact of funding does not. This leaves the estimates unbiased. Last, we discuss factors that drive the timing of the expenditures-publications relationship, which may strike social scientists as compressed.

We first assess misspecification in the first stage. Our analysis assumes that football in period  $t$  mainly impacts funding in period  $t + 1$ . However, football can impact funding at  $t$  if, for example, football inspires some immediate and directed donations to academic endeavors. It may also impact funding at  $t + 2$  and later if team success at the end of the season influences the starting point of success for future seasons. It should not, of course, impact funding at  $t - 1$ , a period prior to the football season. Table 3.8 reports the relationship between funding and the seven prior years' instrument values as well as the current year and following year instrument values. A full set of controls are included. The first column reports the impact on the log of non-medical non-social science expenditures while the second reports the impact of the log of total non-social science and engineering expenditures. These results confirm our intuition about the first stage timing. In both cases, funding is most strongly impacted by one-year lagged instrument values and is not meaningfully impacted by the future value of the instrument. There is a non-trivial impact from the instrument in the contemporaneous period and the instrument lagged more than one year. However, the coefficients tend to drop monotonically in terms of both magnitude and significance.

Expenditures, on the other hand, largely impact output in a single year, so the estimates are unbiased despite the impact of football being spread out. Table 3.9 reports this result for academic publishing. The left-hand side variable is the log of scholarly articles. The right-hand side variables are predicted values of leading, contemporaneous, and one- to three-year lagged research expenditures. In line with the model, the largest and only significant coefficient is on once-lagged expenditures. Table 3.10 reports the result for the case where the publishing measure is new patent applications. In line with the model once again, the largest and only significant coefficient is on contemporaneous year expenditures.

That funding impacts subsequent-period scholarly articles may strike readers as too fast, but several factors explain this. Recall that research budgets are reported on a fiscal year basis while output is recorded on a calendar year basis, so there is an added six month gap between funding and publishing periods. Many projects are bottlenecked due to

funding. In these cases, budgets will be spent soon after they are replenished, so the actual lag we measure may be close to two years rather than one. Furthermore, non-social non-medical science is, at least anecdotally, faster to conduct than social science research. NSF grant data provides evidence of this. For example, among all grants given out in 2000, 2005, and 2010, 26% resulted in the original grant year for physics-related proposals while only 5.4% of economics-related proposals did. Among grants over the same time period that resulted in at least one publication, 46% resulted in journal publications in the original grant year for physics-related proposals while only 13% of economics-related proposals did. The publication process is correspondingly fast. The receipt-to-acceptance time for manuscripts published in non-social non-medical science journals is around a third of those published in economics journals. For example, the average time between first submission and final publication was between 13 and 22 weeks at the five main journals of the American Physics Society (Pattard [2010]) but 62 weeks at the *American Economic Review* (Moffitt [2009]).<sup>29</sup> Also, natural science disciplines in particular tend to rely more on short papers, proceedings, and letters, which have much shorter review times. For example, “rapid communication” section of the five main journals of the American Physics Society has an average receipt-to-acceptance time of between 9 and 15 weeks and a minimum of only two days.

The timing of patent filings is driven by different factors. To ensure intellectual property protection of an idea, scholars must submit their new patent application to the USPTO ahead of giving seminars or conference talks. On top of that, patents are faster to write and assigned a “publication date” before the review process. Taken together, it is not surprising that the data indicates scientists file in the same period that research budgets are replenished. With respect to the timing of patent filings, our results are in line with earlier work in different settings. For example, Hall et al. [1986] studied manufacturing industry investments to R&D from 1972 to 1979 and stated, “R and D and patents appear to be dominated by a contemporaneous relationship.”

---

<sup>29</sup>This comparison is drawn from 2008.

### 3.6 Conclusion

Unanticipated within-season football success impacts school-sponsored research, providing rich exogenous variation that identifies the impact of money on science. An instrumental variable approach is important to study this relationship for two reasons. First, large grants are typically awarded to institutions that would otherwise attract big ideas, so an approach that ignores this endogeneity will recover upwardly biased parameters. Second, funding data include long-term projects with payoffs to researchers over many years, making it difficult to tie shifts in spending to shifts in scientific outcomes and creating an errors-in-variables problem that attenuates estimates toward zero. Our approach yields a dollar elasticity of scholarly articles at 0.31 and of article citations at 0.59. It also yields a dollar elasticity of new patent applications at 1.91 and of patent citations at 3.30. If citations are rough measure of quality, then these results suggest researchers are splitting their time between launching new projects and improving the quality of existing projects. We find it costs universities, at the margin, approximately \$2.6 million to generate an idea worthy of filing a patent application.

The inclusion of school specific controls, i.e. fixed effects and a school specific time trend, improved the first stage power but ultimately did not change the elasticity much for the 2SLS specifications. Their inclusion sharply reduced OLS estimates, which tended toward zero for all outcome measures. This highlights the importance of using instruments to “pick out” marginal expenditure shifts that can be tied to scientific outcomes. Without them, this exercise would understate the returns to university R&D and lead policymakers to under-invest, which highlights the need to an instrumental variable that provides rich short-term variations in funds.



**Table 3.8: First stage timing**

Dep. Var.:	Non-Federal Non-Medical Non-Social Science Expenditures (1)	Non-Federal Non-Social Science and Engineering Expenditures (2)
Instrument: 1 Year Ahead	0.00334 (0.00729)	-0.00177 (0.00772)
Instrument: Contemporaneous	0.0206* (0.0105)	0.0115 (0.00756)
Instrument: 1 Year Prior	0.0316*** (0.0113)	0.0242*** (0.00853)
Instrument: 2 Year Prior	0.0316** (0.0148)	0.0246** (0.0108)
Instrument: 3 Year Prior	0.0328** (0.0128)	0.0227** (0.0109)
Instrument: 4 Year Prior	0.0221 (0.0139)	0.0132 (0.0129)
Instrument: 5 Year Prior	0.0213 (0.0129)	0.00820 (0.0104)
Instrument: 6 Year Prior	0.0162 (0.0128)	0.00554 (0.00892)
Instrument: 7 Year Prior	0.0116 (0.0144)	0.00822 (0.0126)
Constant	7.831*** (0.0633)	8.146*** (0.0626)
School FE	x	x
Year FE	x	x
School Time Trend	x	x
R-squared	0.977	0.984
Observations	673	673
Number of Clusters	40	40

*Note.* - Standard errors, clustered at university level, in parenthesis. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

**Table 3.9: Second stage timing: scholarly articles**

	(1)	(2)	(3)	(4)	(5)
1 Year Leading Expenditures	0.0204 (0.107)				
Contemporaneous Expenditures		-0.00767 (0.0976)			
1 Year Lagged Expenditures			0.310*** (0.117)		
2 Year Lagged Expenditures				-0.225 (0.249)	
3 Year Lagged Expenditures					-0.0297 (0.176)
School FE	x	x	x	x	x
Year FE	x	x	x	x	x
School Time Trend	x	x	x	x	x
R-squared	0.993	0.993	0.989	0.990	0.994
Observations	949	909	869	829	789
Number of Clusters	40	40	40	40	41

Note. - Standard errors, clustered at university level, in parenthesis. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

**Table 3.10: Second stage timing: patents**

	(1)	(2)	(3)	(4)	(5)
1 Year Leading Expenditures	1.209 (1.999)				
Contemporaneous Expenditures		1.914** (0.922)			
1 Year Lagged Expenditures			0.0623 (1.118)		
2 Year Lagged Expenditures				-1.448 (1.419)	
3 Year Lagged Expenditures					-0.558 (1.649)
School FE	949	909	869	829	789
Year FE	0.993	0.993	0.989	0.990	0.994
School Time Trend	x	x	x	x	x
R-squared	0.892	0.851	0.911	0.867	0.904
Observations	609	607	607	607	607
Number of Clusters	39	39	39	39	39

Note. - Standard errors, clustered at university level, in parenthesis. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

# References

- James Adams and Zvi Griliches. Measuring Science: An Exploration. *Proceedings of the National Academy of Sciences*, 93(23):12664–12670, 1996.
- James Adams and Zvi Griliches. Research Productivity in a System of Universities. *Annales d'Economie et de Statistique*, (49-50):127–162, 1998.
- Jeffrey M Alden and Robert L Smith. Rolling horizon procedures in nonhomogeneous markov decision processes. *Operations Research*, 40(3-supplement-2):S183–S194, 1992.
- Michael L. Anderson. The Benefits of College Athletic Success: An Application of the Propensity Score Design with Instrumental Variables. June 2012.
- Donald W. K. Andrews and Gustavo Soares. Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica*, 78(1):119–157, 2010. ISSN 1468-0262. doi: 10.3982/ECTA7502.
- P. Azoulay, John Graff-Zivin, J. S., D. Li, and B. Sampat. Public R&D Investments and Private Sector Patenting: Evidence from NIH Funding Rules. Working paper. August 19 2014.
- Patrick Bajari, C. Lanier Benkard, and Jonathan Levin. Estimating dynamic models of imperfect competition. *Econometrica*, 75(5):1331–1370, 2007. ISSN 1468-0262. doi: 10.1111/j.1468-0262.2007.00796.x.
- Battelle Memorial Institute. 2014 Global R&D Funding Forecast, 2013.
- Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890, 1995.
- Steven Berry, James Levinsohn, and Ariel Pakes. Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of Political Economy*, 112(1):68, 2004.
- Bruce A Blonigen, Christopher R Knittel, and Anson Soderbery. Keeping it fresh: Strategic product redesigns and welfare. Technical report, National Bureau of Economic Research, 2013.
- Severin Borenstein. Hubs and high fares: dominance and market power in the U.S. airline industry. *The RAND Journal of Economics*, pages 344–365, 1989.
- Timothy F Bresnahan and Peter C Reiss. Entry and competition in concentrated markets. *Journal of Political Economy*, pages 977–1009, 1991.

- David Card and Gordon B. Dahl. Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior. *The Quarterly Journal of Economics*, 126(1):103–143, 2011.
- Victor Chernozhukov, Han Hong, and Elie Tamer. Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, 75(5):1243–1284, 2007.
- W.M. Cohen, R. Florida, L. Randazzese, and J. Walsh. *Industry and the Academy: Uneasy Partners in the Cause of Technological Advance*. in Noll, R., ed., *Challenges to Research Universities*, 1998. Brookings Institution.
- Allan Collard-Wexler. Demand fluctuations in the ready-mix concrete industry. *Econometrica*, 81(3):1003–1037, 2013.
- Ulrich Doraszelski, Gregory Lewis, and Ariel Pakes. Just starting out: Learning and price competition in a new market. Technical report, Working paper, Harvard University, 2014.
- K. Dosh. *Saturday Millionaires*. Wiley. New York., 2013.
- Michaela Draganska, Michael Mazzeo, and Katja Seim. Beyond plain vanilla: Modeling joint product assortment and pricing decisions. *QME*, 7(2):105–146, 2009.
- Alon Eizenberg. Upstream innovation and product variety in the US home PC market. *The Review of Economic Studies*, 81(3):1003–1045, 2014.
- Ying Fan. Ownership consolidation and product characteristics: A study of the US daily newspaper market. *The American Economic Review*, 103(5):1598–1628, 2013.
- Chaim Fershtman and Ariel Pakes. Dynamic games with asymmetric information: A framework for empirical work\*. *The Quarterly Journal of Economics*, 127(4):1611–1661, 2012.
- Henry Ford and Samuel Crowther. *My life and work: In collaboration with Samuel Crowther*. Cornstalk Publishing Company, 1922.
- Drew Fudenberg and David Levine. *The theory of learning in games*, volume 2. MIT press, 1998.
- John R Graham and Campbell R Harvey. The theory and practice of corporate finance: Evidence from the field. *Journal of financial economics*, 60(2):187–243, 2001.
- R. Grant, J. Leadley, and Z Zygmunt. *The Economics of Intercollegiate Athletics*. World Scientific Publishing Company, Singapore, 2008.
- Zvi Griliches, Ariel Pakes, and Bronwyn H. Hall. The Value of Patents as Indicators of Inventive Activity. NBER Working Paper No. 2083. September 1988.
- B. H. Hall, Zvi Griliches, and Jerry A. Hausman. Patents and R and D: Is there a lag? *International Economic Review*, 27(2):265–283, June 1986.
- Naomi Hausman. University Innovation, Local Economic Growth, and Entrepreneurship. Working paper. June 2013.

- R. Henderson, A. B. Jaffe, and M. Trajtenberg. Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965 - 1988. *Review of Economics and Statistics*, 80(1):119–127, February 1998. doi: 10.1162/003465398557221.
- V Joseph Hotz, Robert A Miller, Seth Sanders, and Jeffrey Smith. A simulation estimator for dynamic models of discrete choice. *The Review of Economic Studies*, 61(2):265–289, 1994.
- Adam B. Jaffe. Real Effects of Academic Research. *The American Economic Review*, 79(5):pp. 957–970, 1989. ISSN 00028282.
- Paula Lavigne. College Sports Thrive Amid Downturn. *ESPN Outside the Lines*, 2014.
- Eric Maskin and Jean Tirole. A theory of dynamic oligopoly, i: Overview and quantity competition with large fixed costs. *Econometrica: Journal of the Econometric Society*, pages 549–569, 1988.
- Rosa L Matzkin. Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models. *Econometrica: Journal of the Econometric Society*, pages 239–270, 1992.
- Michael J Mazzeo. Product choice and oligopoly market structure. *RAND Journal of Economics*, pages 221–242, 2002.
- Robert L McDonald. Real options and rules of thumb in capital budgeting. *Project Flexib*, 2000.
- Jonathan Meer. The Habit of Giving. *Economic Inquiry*, 51(4):2002–2017, 2013. ISSN 1465-7295. doi: 10.1111/ecin.12010.
- Jonathan Meer and Harvey S. Rosen. The Impact of Athletic Performance on Alumni Giving: An Analysis of Microdata. *Economics of Education Review*, 28(3):287–294, June 2009.
- Robert A. Moffitt. Report of the Editor: American Economic Review. *American Economic Review*, 99(2):660–70, 2009. doi: 10.1257/aer.99.2.660.
- James S Moore and Alan K Reichert. An analysis of the financial management techniques currently employed by large US corporations. *Journal of Business Finance & Accounting*, 10 (4):623–645, 1983.
- Eduardo Morales, Gloria Sheu, and Andrés Zahler. Gravity and extended gravity: Estimating a structural model of export entry. 2011.
- David C. Mowery and Nathan Rosenberg. *Technology and the Pursuit of Economic Growth*. Cambridge University Press, 1989.
- Fiona Murray. *Evaluating the Role of Science Philanthropy in American Research Universities*, volume 13 of *Innovation Policy and the Economy*, pages 23 – 59. National Bureau of Economic Research, 2013.
- National Research Council. *Research Universities and the Future of America: Ten Breakthrough Actions Vital to Our Nation's Prosperity and Security*. The National Academies Press, Washington, D.C., 2012.

- National Science Foundation. Higher Education Research and Development Survey (HERD) - 2012 survey cycle. Technical report, 2013.
- Richard R. Nelson. The Simple Economics of Basic Scientific Research. *Journal of Political Economy*, 67(3):pp. 297–306, 1959. ISSN 00223808.
- Aviv Nevo. Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2): 307–342, 2001.
- Chris Nosko. Competition and quality choice in the CPU market. *Manuscript, Harvard University*, 2010.
- Jonathan H Owen, Robert R Inman, and Dennis E Blumenfeld. Practical modeling in automotive production. In *Planning Production and Inventories in the Extended Enterprise*, pages 249–269. Springer, 2011.
- Ariel Pakes and Zvi Griliches. Patents and R&D at the Firm Level: A First Report. *Economics Letters*, 5(4):377 – 381, 1980. ISSN 0165-1765. doi: [http://dx.doi.org/10.1016/0165-1765\(80\)90136-6](http://dx.doi.org/10.1016/0165-1765(80)90136-6).
- Ariel Pakes and Zvi Griliches. *Patents and R&D at the Firm Level: A First Look*, volume Title: R & D, Patents, and Productivity, pages 55 – 72. University of Chicago Press, 1984.
- Ariel Pakes, Michael Ostrovsky, and Steven Berry. Simple estimators for the parameters of discrete dynamic games (with entry/exit examples). *The RAND Journal of Economics*, 38(2): 373–399, 2007.
- Ariel Pakes, Jack Porter, Kate Ho, and Joy Ishii. Moment inequalities and their application. *Econometrica*, 83:315–334, 2015.
- Thomas Pattard. How to publish your work in the physical review. February 2010.
- Martin Pesendorfer and Philipp Schmidt-Dengler. Asymptotic least squares estimators for dynamic games. *The Review of Economic Studies*, 75(3):901–928, 2008.
- Mitchell A. Petersen. Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. *The Review of Financial Studies*, 22(1):pp. 435–480, 2009. ISSN 08939454.
- Amil Petrin. Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy*, 110(4):705–729, 2002.
- James M Poterba and Lawrence H Summers. A ceo survey of us companies' time horizons and hurdle rates. *Sloan Management Review*, 37(1):43–53, 1995.
- Mark J Roberts, Daniel Yi Xu, Xiaoyan Fan, and Shengxing Zhang. A structural model of demand, cost, and export market selection for chinese footwear producers. Technical report, National Bureau of Economic Research, 2012.
- Stephen P Ryan. The costs of environmental regulation in a concentrated industry. *Econometrica*, 80(3):1019–1061, 2012.

- Bhaven N. Sampat, David C. Mowery, and Arvids A. Ziedonis. Changes in University Patent Quality After the Bayh-Dole Act: a Re-examination. *International Journal of Industrial Organization*, 21(9):1371 – 1390, 2003. ISSN 0167-7187. doi: [http://dx.doi.org/10.1016/S0167-7187\(03\)00087-0](http://dx.doi.org/10.1016/S0167-7187(03)00087-0). The economics of intellectual property at universities.
- Joseph Alois Schumpeter. *Socialism, capitalism and democracy*. Harper and Brothers, 1942.
- Katja Seim. An empirical model of firm entry with endogenous product-type choices. *The RAND Journal of Economics*, 37(3):619–640, 2006.
- Jeffrey L. Stinson and Dennis R. Howard. Athletic Giving and Academic Giving: Exploring the Value of SPLIT Donors. *Journal of Sport Management*, 24(6):744–768, November 2010.
- Jeffrey L. Stinson and Dennis R. Howard. The Value of Split Donors. The NACDA Report 1, National Association of Collegiate Directors of Athletics, February 2014.
- Lawrence H Summers. Investment incentives and the discounting of depreciation allowances. In *The effects of taxation on capital accumulation*, pages 295–304. University of Chicago Press, 1987.
- Richard Sweeney. Environmental regulation, imperfect competition, and market spillovers: The impact of the 1990 clean air act amendments on the us oil refining industry. 2015.
- Andrew Sweeting. Dynamic product positioning in differentiated product markets: The effect of fees for musical performance rights on the commercial radio industry. *Econometrica*, 81(5):1763–1803, 2013.
- Elie Tamer. Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies*, 70(1):147–165, 2003.
- Ravi Varadhan and Christophe Roland. Simple and globally convergent methods for accelerating the convergence of any em algorithm. *Scandinavian Journal of Statistics*, 35(2): 335–353, 2008.
- Wards. Automotive yearbook, 1986-2013.
- Brian D. Wright, K. Drivas, Lei Zhen, and Stephen A. Merrill. Technology Transfer: Industry-funded Academic Inventions Boost Innovation. *Nature*, 507(7492):297–299, March 2014.

# Appendix A

## Appendix to Chapter 1

### A.1 Constructing empirical distribution of buyers

To construct the empirical distribution of buyers, I match state-level observations on the road density measure (from the US Department of Transportation *Highway Statistics*) to state-level observations on industry (from the US Census *County Business Patterns*). Weights in the distribution are based on the number of employees rather than establishments or companies since it's the employees that operate vehicles, not the fictitious legal entities that employ them. This data provides that, for example, the New York State-based construction industry accounts for 0.26% of all employment in the buyer industries for 2011 and that this group of potential buyers face surrounding roads that are classified 75% urban (based on total road mileage).

Industries vary in the portion of employees that will operate vehicles. This is presumably quite high in the freight transportation industry but low in, for example, construction. I scale industry weights in the empirical distribution to match the average vehicle ownership in the microdata. For example, if freight firms account for 30% of commercial vehicle owners in the microdata but 15% of buyers in the US Census data, the weight of commercial vehicle buyers in the empirical distribution is doubled (technically, it would scaled by  $2.43 \approx \frac{.3}{1-.3} \times \frac{1-.15}{.15}$  were the other industries to scale proportionately and in the opposite direction).



Buyers also vary in unobservable ways and these are drawn from independent standard normal distributions.

## A.2 Calculating market size.

The market size for each year,  $M_t$ , is constructed as a product of a mean market size over the panel and a scaling factor for each year. First, to compute the mean market size over the panel, write the total units sold in  $t$  as  $q_t$  such that  $q_t = \sum_j q_{j,t}$ . Set mean market size  $\bar{M}$  to a level such that the average “inside share” across the years equals  $\frac{\frac{1}{T} \sum_t q_t}{\bar{M}}$ . Second, to compute the scaling factor, write the scaled industry-time specific employment levels (described in the Appendix section immediately above) as  $y_{l,t}$  at  $t$  and let  $y_t = \sum_l y_{l,t}$ . Set the scaling factor for  $t$  so that the change in market size is proportional to the change in  $y_t$ . That is, set the scaling factor  $\tilde{M}_t$  such that  $\tilde{M}_t = \min\{q_t\} + (q_t - \min\{q_t\}) \frac{\max\{q_t\} - \min\{q_t\}}{\max\{y_t\} - \min\{y_t\}}$ , which yields  $M_t = \bar{M} \times \tilde{M}_t$ .

## A.3 Computing and selecting equilibrium

This section provides computational details on selecting out equilibria using a learning process based on best response dynamics. The order of decisions is based on 2009 market share, aggregated to the parent level, which is where the decisions are made. With GM and Chrysler eliminated from the set of individual owners, there are seven remaining parent companies. The ordering of their market shares is as follows: Ford (26%), Daimler (18%), International (17%), Paccar (14%), Volvo (7%), Isuzu (3%), Hino (1%). Best response dynamics simply begin with the set of product offerings inherited from the previous period and then cycle through the firms, updating the set of product offerings following each choice. The cycle terminates when no profitable deviations can be found.

Candidates for a best response action by a firm  $f$  is any possible  $J_f$  and is of the size  $2^{\mathbb{J}}$ , where  $\mathbb{J}$  is the total number of distinct product types. Large portions of the potential action space will never be a (conditional) best response and can be ruled out *ex ante*,

speeding up computations considerably. First, four of the product types are heavy cab-over-engine vehicles. The impact of length deregulation has reduced demand for the vehicles to effectively zero in 2010, conditional on any other product being offered over 33,000 lbs GWR. Recalling Figure V, not a single vehicle of this type has been offered in the entire six year span preceding 2010. This alone reduces the number of potential actions by any firm to approximately 500,000. Second, recall firms headquartered in Japan introduce the conventional cab at a higher sunk cost than their rivals. For this reason, it turns out that these firms do not find it profitable to add any conventional cab vehicle over 33,000 lbs GWR conditional on at least three competing firms offering conventional cab vehicles over 40,000 lbs GWR. Third, recall the Big Three firms find it cheaper to introduce lighter vehicles and more expensive to introduce heavier vehicles relative to their rivals. For this reason, it also turns out that Ford (the only remaining Big Three firm) does not find it profitable to offer any cab over 48,000 lbs GWR conditional on at least three competing firms offering conventional cab vehicles over 48,000 lbs GWR. In my experience to date, Volvo, Paccar, International, and Daimler never completely vacate the heavy conventional cab sub-segment, so these restrictions are always in effect.

The fourth restriction employs a different logic, which is based on the number of offerings. If a firm finds it more profitable to offer  $n$  products than  $n + 1$  products, then it will always find it more profitable to offer  $n$  than  $n + k$ ,  $k > 1$ .

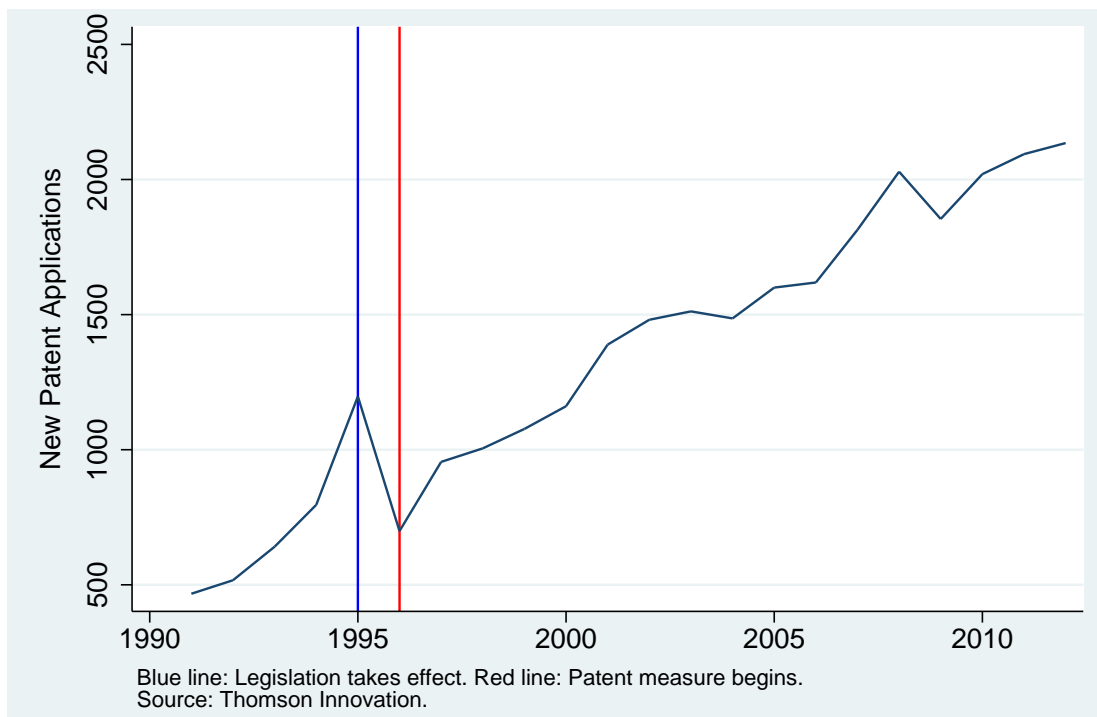
## Appendix B

### Appendix to Chapter 3

#### B.1 Allocating patent filings to universities

Identifying cohesive patent portfolios and patent applicants and assignees can be a difficult task. Numerous variations in names of patent-seeking institutions appear in USPTO records caused by either the variation in patent-prosecuting law firms or human errors and incorrectly spelled names. For example, there are 157 variations of assignee/applicant names grouped under the “University of California” umbrella in our Thomson Innovation patent sample. These names range from “*The Regents of the University of California*”, “*University of California Berkeley*”, “*University of California Los Angeles*”, “*The Regents of the University of Caliofornia*”, to the “*The Regents of the University of California*”. In addition, to better identify university owned patents and patent applications, we utilize DWPI assignee classification available in Thomson Innovation: a unique 4-letter identifying code assigned to approximately 22,300 international patentees. For example, The University of California is assigned a unique 4-letter code “REGC”, and in order to retrieve all patent records assigned to the University of California, we query Thomson database for all variations of assignee/applicant string grouped under “University of California” and associated with “REGC” assignee code for earliest patent priority years 1996-2011. It is important to note that, while we collect patent data starting with 1987, our panel officially starts in 1996. We

start the panel in 1996 because of the effects that the international harmonization of the United States patent system in early 1990's had on university patenting behavior. As shown in Appendix Figure I, one of the patent law amendments with a significant impact was the introduction of provisional patent applications in June 1995.<sup>1</sup>



**Figure B.1:** *Patent applications filed over time*

Since the introduction of provisional patenting provided a convenient solution for academic researchers faced with publish-or-patent-first dilemma, it resulted in a sharp increase in university patent applications. A published article, a conference presentation, or even as much as a conversation describing an invention before a patent is filed represents a public disclosure, and can deem that invention unpatentable. To the extent that the scientific work in academia is first and foremost driven by article considerations in peer-reviewed

---

<sup>1</sup>As described in 35 U.S.C. §111, a provisional patent application allows an applicant to file an application with specification only, and without any formal patent claims, oath, declaration, or any prior art disclosures. A provisional patent application establishes an early patent filing date, but does not evolve into an granted patent unless the applicant converts it into a full patent application within twelve months. It effectively allows an applicant to lock-in a patent priority date, without being subjected to the cost of a regular patent application filing.

journals, provisional patent applications are an exceptionally good fit for this environment as they enable the university to lock an early priority date, while providing additional 12 months for inventors to publish, disseminate and improve the invention. Universities use provisional patent applications to reduce uncertainty surrounding market value of inventions and make a more informed decision of whether to prosecute full patents. Indeed, many university Technology Transfer Offices laud provisional patent applications as the first order of business after being informed of a new invention.<sup>2</sup>

Prevalence of provisional patenting in academia was the main reason behind our decision to stop our panel with patent applications filed in 2011. Since provisional patent applications take 12 months before they are published, patent application data from 2012 would be missing all provisional patents applied for in that year, and would result in a truncated patent count.

We use priority dates rather than application dates to count patent records because priority years most closely correspond with the date when the invention was *first applied for*. While the patent priority date is most often no different than the regular patent application date, in a case of a converted provisional application, a priority date will be earlier than the regular application date. This is especially the case when a divisional or a continuation application was filed. In addition, we use the DWPI Patent Family list available in Thomson Innovation database to assign all retrieved patent records to unique groups sharing the same priority application. This enables us to more closely identify patent groups surrounding the same invention and ensures that we do not overcount patent records in the sample. Each DWPI Patent Family is counted only once, and all forward patent citation counts are aggregated on a DWPI Patent Family level.

---

<sup>2</sup>For example, see Office of Vice President for Research at Penn State (<http://www.research.psu.edu/patents/protect-your-invention/what-happens-after-submission>), Innovation and New Ventures Office at Northwestern University (<http://www.invo.northwestern.edu/process/assessment-patents>), Northeastern University Center for Research Innovation (<http://www.northeastern.edu/research/cri/inventors/commercialization-process/>), or Boston University Technology Development (<http://www.bu.edu/otd/for-researchers/technology-transfer-process/patentapp/>).

### **B.1.1 Disaggregating state educational systems**

To further exacerbate the problem of allocating patent applications to universities, some university systems do not specify campus locations where the invention was made when filing their patents. For example, almost 75% of all patent applications from University of California System in our sample are assigned to “The Regents of The University of California” without any additional information about invention-originating campus. Consequently, we do not know if the invention was made at The University of California at Berkeley, The University of California at Los Angeles, or any other campus in the system. Since our instrument works on the individual campus level only, and does not propagate through the whole system, we need to allocate patent applications to individual campuses within the university. In other words, unexpected success of The University of California at Berkeley football team will not impact R&D expenditures at The University of California at Los Angeles, and vice versa. To rectify this problem, we use the inventor’s home address information provided on US patent records and use Google Maps API to calculate the “by car” estimated travel time from inventor’s home address to every campus in the university system. We then systematically examine the patent portfolio and count a patent as originating at a specific university campus if at least one inventor lives less than 26 minute drive from that campus.<sup>3</sup>

---

<sup>3</sup>This distance is based on the 2013 US Census American Community Survey estimate of mean travel time to work in the United States of 25.8 minutes. Our results are robust to different travel times: travel times of 15, 20 and 30 minutes did not cause any significant changes in our outcomes.