Essays on the Structural Estimation and Analysis of Corporate and Industrial Markets

by

Moshe Aharon Cohen

B.A. Economics, Hebrew University of Jerusalem (2004)
LL.B., Hebrew University of Jerusalem (2004)
M.A. Economics, Hebrew University of Jerusalem (2005)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2009

Signature of Author .....
Department of Economics
September 2009

Certified by .....
Glenn Ellison
Gregory K. Palm (1970) Professor of Economics
Thesis Supervisor

Certified by ....
Whitney Newey
Jane Berkowitz Carlton and Dennis William Carton Professor of Microeconomics
Thesis Supervisor

Certified by .....
Victor Chernozhukov
Professor of Economics
Thesis Supervisor

Accepted by .....
Esther Duflo
Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics
Chairman, Departmental Committee on Graduate Studies

Essays on the Structural Estimation and Analysis of Corporate and Industrial Markets

by

Moshe Aharon Cohen

Submitted to the Department of Economics
on September 2009, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

The three chapters which comprise my thesis are a collection of essays on the analysis of the corporate governance and airline markets and of the federal appellate structure.

In Chapter 1, I use a discrete choice framework to analyze state design and firm choice of the implications of incorporation: corporate governance laws, corporate taxes and court structure. Firms – differentiated on ownership, management, industry concentration, financial profile and unobservable dimensions – freely choose their preferred state of incorporation or reincorporation. The revealed preference embedded in this observable choice is used as window into the heterogeneous preferences within and across firms, yielding several findings: For example, I find, surprisingly, that firms are very responsive to incorporation and franchise taxes. In addition, on average, firms like antitakeover statutes, but, consistent with an agency story, firms with an institutional shareholder block and venture capital backed firms dislike them. On average, firms dislike mandatory governance statutes restricting managerial power and facilitating the representation of minority shareholders, but these laws are less restrictive for the choice of firms in concentrated industries. All firms dislike well functioning courts, consistent with a litigation deterrence motive. The recovered firm preferences are then taken to the simulation of recently proposed federal reforms aimed at centralizing the domicile implications and restricting firm choice. They are also related to the documented differential returns earned by firms with better internal governance in the 1990s, as well as to other (new) trading strategies that would have yielded abnormal returns in the 2000s.

Chapter 2 begins with the observation that airlines choose the domestic markets – city pairs – they serve and the prices they charge given the structure of their network and the networks of rival airlines. I cast this choice into a dynamic oligopoly entry game to recover airline fixed and variable operating costs, entry costs, and profits, using a panel of 20 quarters of DB1B and T-100 Domestic Segment Data. These estimates are then used to analyze the strategic and cost saving effects of hubs, and LCC. I find that hubs are

valuable to consumers and increase the variable profits of the hubbing airline, but when including fixed costs their desirability is much less clear. LCC, and especially Southwest and JetBlue are especially attractive to consumers, have lower marginal costs and have a strong negative impact on the profits of the incumbents in the markets they serve.

In Chapter 3, using data on all federal civil trial and appellate cases from 1992-2003, I show that appeals are generally rejected and, for some case categories, can have negative expected net present value. Appellate outcomes can be further related to the trial decision being by judge or jury, the identity of the prevailing party (plaintiff or defendant, US or private), and the form of representation. Some of these factors influence the propensity to appeal, however, others, including whether trial was by judge or jury, go in the opposite way. I discuss the implications of these findings for the modeling of the incentives to appeal and settlement breakdown, and for appellate reform.

Thesis Supervisor: Glenn Ellison
Title: Gregory K. Palm (1970) Professor of Economics

Thesis Supervisor: Whitney Newey
Title: Jane Berkowitz Carlton and Dennis William Carton Professor of Microeconomics

Thesis Supervisor: Victor Chernozhukov
Title: Professor of Economics

Contents

To my parents, Kenneth and Sharon, and my grandparents
Olga, Edward, Hanna and Benjamin.

Acknowledgements

I have benefitted from and am grateful for the unparalleled quality of the MIT faculty. My advisors, Glenn Ellison, Whitney Newey and Victor Chernozhukov have all provided continual, dedicated, insightful and incisive advice and support throughout my thesis writing years. They have always made themselves available, placing their advising role at the top of their busy schedules. Their advising has greatly improved the quality of the work in my thesis and will undoubtedly contribute to the work I do in the future. Their different teaching, research and advising styles have synthesized to provide me with a rich example of academic excellence. I thank Jerry Hausman who has practically been a fourth advisor. His prompt dedication and involvement in my work, despite his rich and demanding schedule, have been astonishing. I am grateful to Stewart Myers whose advice and support have been invaluable in my transition from economics to finance as well as on the academic job market. His advice has significantly improved the first chapter of my thesis and will undoubtedly be very helpful in my future work. For helpful comments, suggestions and support, I also thank Michael Greenstone, Nancy Rose, Stephen Ryan, and the rest of the faculty at the MIT Department of Economics and at the MIT Sloan School of Management.

I thank Justice Dalia Dorner of the Supreme Court of Israel for whom I had the pleasure of clerking immediately prior to my doctoral education. Justice Dorner took me under her wing and shared the skills, intellect, and tremendous experience she has accumulated, which are second to none. She has been a mentor to me and for that I am forever indebted.

I am grateful to my friends and colleagues at MIT for the social and academic support they have given me. They have enriched my graduate experience and have contributed to my academic training. I especially thank Nicolás Arregui, Arthur Campbell and Alejandro Drexler, my roommates and classmates, and my officemates, Florian Ederer, Jesse Edgerton and Johannes Spinnewijn. I also especially thank Daniel Gottlieb for his close friendship and for the useful comments and suggestions he has given me for all of the chapters in my thesis and for my job market presentation. I thank Paul Scrimpf for useful suggestions,

especially regarding computational programming, and Konrad Menzel for his econometrics suggestions. Finally, I thank Emily Gallagher for her administrative help and for her continuous support.

I am indebted to Ori Aronson for our deep friendship and for his continuous support, since our first year in law school together, some 10 years ago. I also thank Yuval Abrams for his friendship, support and useful advice. My academic plans and decisions have been shaped, in no small part, by the discussions we have been having since our earlier years of high school.

Finally, I thank my family. My unique relationships with all of my siblings Kitra, Naftali, Sarit, Shlomtzion, Shimmy and Yael have each contributed to me in their own way. I thank my maternal grandparents, Olga and Edward Simon, for their absolute dedication and involvement in my life, for all they have given me, and for all that we have shared. I am also grateful to my paternal grandparents Benjamin and Hanna Cohen for the example they have set for me, and for their personalized warmth and caring. Lastly, I am eternally grateful to my parents, Kenneth and Sharon, for their love, caring, and dedication. Both of my parents are extraordinary, idealistic people, whose life example is a moral compass that has guided me thus far and which I aim to follow.

My achievements are to the credit of my parents and grandparents and, consequently, I dedicate this work to them.

Chapter 1

Revisiting Corporate Governance Regulation: Firm Heterogeneity and the Market for Corporate Domicile

1.1 Introduction and Motivation

This paper focuses on state design and firm choice of corporate governance law. Corporate governance law is the set of rules in legislation or judicial precedent that "govern" the internal and external agency problems that exist within a corporation. They govern areas such as takeovers, shareholder voting rules and managerial liability, and provisions such as supermajority requirements and antitakeover defenses. These laws represent the policies instituted to combat the divergence of interests within a firm and to align managerial and shareholder interests.

There are three crucial facts that form the background for this paper: First, the majority of corporate governance issues are **not** regulated by the federal government; the 51 US jurisdictions design their own distinct corporate environments. Second, state corporate governance regimes are applied to a firm by way of incorporation in a certain state. Third, firms are free to choose in which state to incorporate (or reincorporate[1]). The location of incorporation does not need to be related to the firms' business locations. However, the consequence of incorporation is submittal to the complete corporate governance regime in a state. Firms generally cannot choose which provisions in a state's governance regime they will be subject to - they must take all the provisions as a package. This package generally includes a choice of the local judicial forum, since personal jurisdiction over the internal conflicts in the firm generally goes to the state of incorporation. It also includes the taxes that states impose as a consequence of incorporation within them.

State corporate governance laws, together with federal securities law, form the background against which firms exercise their residual freedom to design their corporate charters[2]. It is the combination of the laws and the firm charter that forms the firms' complete governance structure. The responsibility for the regulation of corporate governance has

[1] American corporate law generally requires manager initiation and shareholder approval of a reincorporation decision, upon which firms can then change their state of incorporation freely.

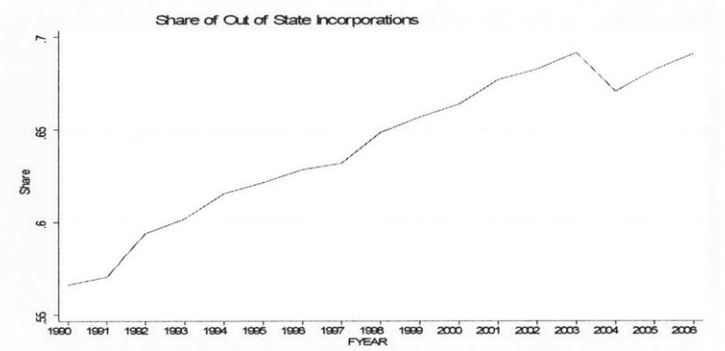
[2] In this paper, for convenience, I often use the term charter to refer to the bylaws and all other internal governance tools used by the firm.

been repeatedly granted to the states only to be taken back by the federal government in times of crisis (see Mallette and Spagnola, 1994), and thus it is likely, given the current economic crisis that the decentralized nature of corporate governance law will be revisited once again. However, the current legal environment[3] engendered unprecedented diversity in state corporate governance laws and consequently in the internal governance of firms. There is also wide diversity in firm ownership and management and in industry structure. These combined introduce diversity and complexity in the response to policy.

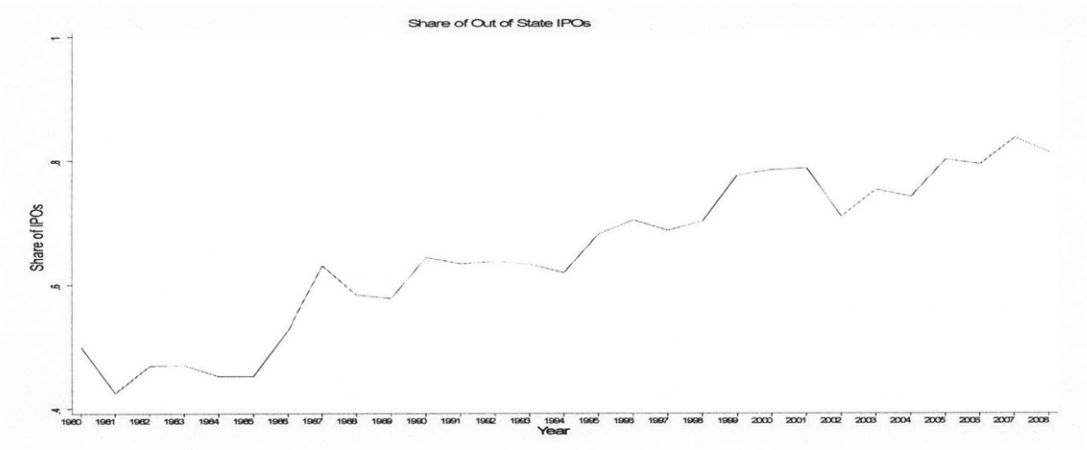
This paper begins by characterizing the diversity in state incorporation implications. I find that there is substantial heterogeneity in the laws, taxes and performance of the relevant state courts. Over the course of the entire time period covered (1990 – 2007), there is legislative activity across many states, although at much lesser levels than that which was present in the late 1980s and early 1990s.

I then find that there is much heterogeneity in the choices of firms as well. Figure 1 (sourced at the data discussed below) shows that there is a significant tendency to incorporate in one of the 51 headquarter jurisdictions, as well as a growing trend - both amongst IPO firms and the stock of public firms as a whole - for firms to shop for their preferred incorporation venue. There is significant variance in the choice of the state of incorporation both in the time series and in the cross section. Many states make little effort and have little success in retaining firms headquartered therein and especially in recruiting firms headquartered elsewhere. However, there are several incorporation "hot spots", most notable amongst which are Delaware, Nevada and Maryland, and a number of states that make sizeable efforts to increase the stock of firms they attract. Contrary to what is commonly believed, it is not all about Delaware - the leader in incorporation

[3] Mallette and Spagnola, 1994, discuss how states surrendered their common law authority to the federal government in the Great Depression era, took much of it back in the merger wave of the 1960s, were halted by the Supreme Court *Edgar* vs. *Mite* (1982) decision which struck down protectionist statutes as constricting interstate commerce, and then given new freedom in *CTS Corporation* v. *Dynamics Corporation of America* (1987) where the Supreme Court accepted review of the invalidation of Indiana's state takeover statute. It is this decision that is seen as largely resposible for the huge wave of litigation, particularly concentrated in the late 80's and early 90's.



(a) Full Stock



(b) IPO

Figure 1-1: Trends in Out of State Incorporations.

shares - which, in fact, recently has seen a significant decline in its shares. The dispersion of incorporation shares in figure 7 in the appendix, and in the revenues from incorporation, franchise and even corporate income taxes, in the cross section and the time series, nicely displays the variance and concentration in incorporation choices[4].

The combination of the variance in the laws, court characteristics and taxes that make up the incorporation implications, as well as in the choices of the various firms, allows for the recovery of the revealed preferences embedded in firm choice and the analysis of the following two questions:

The first is how firms choose their incorporation state, or, in other words, what matters to firms in their choice of corporate governance. This question can be divided into two parts. The first is what drives the average firm's choice, or in other words which of the incorporation implications: which laws, court features and taxes motivate firms in their choice of incorporation state and what are the relative magnitudes of these different incorporation implications. The second is what causes firms to choose differently, or in other words why is it not (as is commonly perceived) "all about Delaware" - why is there variance in firm choice. The differential firm choices are seen as a window into the firm and allow for the analysis of the different features of firms that are instrumental in their different choices in the selection of their governance regime.

In order to exploit this information revealed in the preferences displayed by the different firms, a novel dataset with firm and incorporation characteristics is assembled and then a random coefficient discrete choice model is specified. In the model, incorporation is treated as a "product" that the states design, differentiated along all of the dimensions of the implications of incorporation, including the direct "price" - the tax implications incorporation imposes on each firm. In every one year period, each heterogeneous firm chooses its preferred "product" by choosing to incorporate, to remain incorporated, or to reincorporate in one of the 51 US jurisdictions. Firms are decomposed into their

[4] Figures on the dispersion of these measures will be available in an online appendix.

ownership patterns, director characteristics, industry concentration, financial profiles, the geographical location of their headquarter states, and the residual unobservable dimensions of heterogeneity within them. The choice of incorporation state is seen to be made based on the preferences - resulting from these dimensions of firm heterogeneity - for the laws, court characteristics and taxes that makeup the incorporation implications. I find that all incorporation implications, the laws, the court characteristics and the incorporation taxes matter. My findings are thus separated into these three groups of incorporation implications:

Regarding the corporate governance laws, I find mean preferences consistent with the expected (narrow) managerial preferences within a firm. Firms generally like antitakeover legislation, dislike mandatory laws restricting managerial privileges and dislike laws restricting the flexibility in making shareholder payouts. This "agency" view of mean firm preferences is reinforced by looking at individual firm heterogeneity. I find, using a regression discontinuity approach and the structure of my model, that, controlling for selection endogeneity, firms with sophisticated shareholders - institutional shareholders or venture capital backed firms - and sizeable holdings in the firm, dislike antitakeover laws. In other words, the mean preference for antitakeover laws is reversed in firms in which shareholders are more powerful, thus reinforcing the view that when shareholders have more authority they choose a more convenient regime for takeovers. Firms in more concentrated industries do not display statistically significant different preferences for antitakeover laws, but, contrary to the mean firm, are not influenced by the existence of mandatory laws restricting managerial freedoms. Unobserved heterogeneity plays an important role for these laws, for which preferences are dispersed. In addition, there is a life cycle pattern in these preferences, whereby older firms have an even stronger preference for antitakeover laws, but are less constrained by the mandatory laws mentioned. Interestingly, the characteristics of managers and directors are not instrumental in incorporation decisions.

The mean firm preference is for "bad" or backloaded courts, at the trial or appellate

level. Here, looking inside the firm, I find that firms that are significantly held by institutional shareholders exhibit an even stronger preference for "bad" courts at the appellate level, but prefer well functioning courts at the trial level.

Firms dislike taxes to such a degree that they respond to tax changes by reincorporation even though the increases in taxes are very small. I show, both in my aggregate model - which accounts for the changes over time in the tax rates and in the individual liability of a given firm when its tax base changes - and specifically regarding two tax changes in 2003, that firms are highly sensitive to the incorporation taxes, despite their relatively small magnitudes. State corporate income taxes, which are related to incorporation only in very limited settings, but suspected to be highly manipulated, are found to not be instrumental in the incorporation choice.

The recovery of these firm preferences allows for the treatment of the **second question**: the impact of federal interventions in state governance laws. Consistent with some of the plausible federal reforms frequently discussed in the media[5], I simulate the choices that would be made under counterfactual policies which limit the variance in legal structures across jurisdictions, eliminating the key antitakeover laws and imposing mandatory limitations on managerial freedoms. I find that the laws change the aggregate shares and impact the desirability of some of the popular incorporation alternatives. However, consistent with the findings above, many of the incorporation implications matter. This can help or hurt the various states. For example, the state of Delaware, the most popular incorporation hot spot, increases its share of firms; however, Maryland and Nevada decrease theirs. Furthermore, since the patterns of incorporation are related to the characteristics of the firm, any change would impact the distribution of shareholders holding the firms incorporated in the various states. Thus, for example, a change in the distribution of antitakeover laws causing a jurisdiction that previously had antitakeover laws to no longer have

[5] See for example claims made by (then) Senator Obama in the second presidential debate (Belmont University, 10/2/2008); and Carl Ichan, "Capitalism Should Return to Its Roots" (WSJ, 2/9/2009).

them would be a drawback for the mean firm, but would attract firms with sophisticated shareholders. When several changes are made simultaneously, it is the combination of the specific firm preferences that shapes the overall firm specific effect, which often differs substantially from that of the mean firm.

Finally, I connect these findings to the study of performance via returns earned by stock market trading strategies. I relate this work to previous findings on the connection between governance and performance and show the potential for other (new) strategies - of sorting firms on some of the additional firm heterogeneity dimensions discussed - to be tested against the background of this work. I find that trading strategies based on the antitakeover and mandatory laws alone would have yielded abnormal returns in the 2000s. Holding a zero cost portfolio, buying firms incorporated in states with many antitakeover laws, and selling (short) firms in states with few antitakeover laws, would have yielded a monthly abnormal return of 36 basis points. Alternatively holding a portfolio buying firms in states with many of the mandatory laws and selling those in states with few mandatory laws would have yielded abnormal returns of 46 basis points per month.

These findings relate to the existing literature regarding the impact and importance of corporate governance, which largely rely on abnormal returns earned by the "better" governed firms. This paper finds an interaction between the many dimensions of firm heterogeneity and corporate governance in legislation. However, there is specificity in the form of the interaction. Thus, for example, while the concentration of industry affects the preferences for some laws, it is not instrumental regarding antitakeover laws. As will be shown, these findings can be seen as complementary pieces in the puzzle of the impact of governance rules on firms.

Furthermore, while this work addresses the divergence in preferences in the context of the incorporation choice, my findings - as well as other potential findings concerning other dimensions of firm heterogeneity - can be related to the broader issue of how these dimensions affect our concept of the firm - as an equilibrium between all these competing

constituencies. Thus, for example, my findings regarding institutional shareholder preferences relate to the wide body of literature analyzing the impact of the general increases in institutional shareholder holdings (see for example Aghion et al 2007). Similarly, my findings on venture backed firms relates to the literature on the firms' life cycles in general and the venture capital cycle in particular (see for example Gompers and Lerner 2004).

On the methodological side, this paper extends the canonical framework of a random coefficients discrete choice problem in that individual firm choice and many degrees of firm heterogeneity are observable. The product is the sum of observable characteristics and thus the decomposition is very straightforward.[6] In addition, there is variation in all characteristics over time, which facilitates the controls for unobservables, and there is wide variation in the base, schedule and rates of taxes across jurisdictions. The treatment of endogeneity is done using instruments based on a regression discontinuity design, exploiting the randomness inherent in the inclusion in the S&P 500 index, where the highly non-linear choice structure is accounted for using a control function approach and, alternatively, in the appendix, using two other methods of moments. Counterfactual policy analysis relates not only to the changes in the aggregate shares, but also to the makeup of firm heterogeneity in the various jurisdictions, under the various policies.

The paper is structured as follows. I begin, in section 2, by reviewing the relevant legal and corporate finance literatures upon which I build in this work. Then, in section 3, I present and describe the data and characteristics that are inputs into my model and methodology. Section 4 discusses the model and estimation strategies. In section 5 I turn to my results. I first present and discuss the results from my full model. I then grant incorporation taxes special treatment and provide a reduced form analysis of two tax changes in 2003. Section 6 discusses the simulation of counterfactual policies. Section 7 discusses firm performance and trading strategies. Section 8 concludes and discusses some future extensions.

[6] This, compared to the canonical examples of cars and cereals which are challenging to decompose.

1.2 Related Literature

1.2.1 The Choice of Law

The freedom offered to states in the design of law and the freedom offered to firms in the choice of incorporation state has spawned a sizeable legal literature. This active literature has not reached any consensus, despite the some 30 years over which it has developed. Indeed, some see there to be "genius", active competition between states (or at least some of them) to attract firms (for example Romano, 1985, 2006); some see the state of Delaware - which has a sizeable share of the publicly traded firms - to have won and there now not to be any "competition" (for example Kahan and Kamar (2002), and Bebchuk and Hamdani (2002)) and some see the "competition" to be more with the interventions (actual or potential) of the federal government. Furthermore, within the competition camp as well, there is a debate as to whether this competition is positive - in the sense that it is inducing states to generate "good" laws that promote firm (and shareholder) value ("A race to the top") (Winter (1977), Easterbrook(1983), Fischel (1982)) - or negative - in that it may induce states to cater to management, which potentially has more power in making the incorporation choice, and may warrant federal intervention ("A race to the bottom", Nader (1976), Cary (1974)).

Furthermore, Delaware's success has received specific attention. Daines (2001) finds that controlling for firm financials, the share of director and officer ownership, and industry and year dummies, Delaware firms have a higher Tobin's Q (in the cross section years $1981-1996$). Daines' explanation is that these firms are more vulnerable to takeovers and that this increases firm value. Subramanian (2004) finds this effect be decreasing and after 1996 eliminated (due to what he sees as antitakeover movements in Delaware combined with a friendlier takeover attitude in the late 90s). In analyzing the incorporation decision of firms at IPO, Daines (2002) looks at the influence of some of the legal variation in states, and finds only the existence of national law firms to be a significant driving force (albeit

with questionable exogeneity). Marcel Kahan, in more recent work (2006) looks explicitly at the effects of some of the state laws on the aggregate retention rate of states, i.e. the proportion of firms located in the state that at IPO, in the years 1990 – 2002, also choose to incorporate (or remain incorporated) there and finds this yearly retention rate to be related to the governance laws, and to a court quality ranking.

1.2.2 Does Governance Matter?

There is now a growing corporate finance literature focusing on the final governance structure, the firm bylaws (that evolve from the legislative structure), that answers this question in the affirmative. However, this is still a largely researched and debated question. The standard approaches in this literature have been to use the stock market valuation of publicly traded firms with heterogeneous governance mechanisms to evaluate internal governance. These approaches generally pre-specify the "better" rules, identify the firms operating with such rules in their charter, and then rank and price them relative to (comparable) firms with (pre-specified) "worse" governance rules. Accordingly, these approaches ask which arrangements and structures are correlated with higher shareholder value. This is done in two ways: The first is to look at a cross section of prices and attribute some of the added value of Tobin's Q to the existence of the "better" governance measures (for example Brown and Caylor (2006)). The second, more influential approach, followed (or led) by Gompers et al (2003, GIM), is the construction of profitable trading strategies based on buying the "best" corporate governed firms (in terms of shareholder rights) and selling the "worst" corporate governed firms. GIM also find these "better" firms to have higher value, profits, sales and growth, lower capital expenditures and fewer corporate acquisitions. Once again, the "best" and "worst" are determined (in what is now termed the GIM index), using a cumulative score for the presence of the good and the absence of the bad governance mechanisms out of the pool of 24 such mechanisms in their data. This approach has been further refined. Cremers and Nair (2005) find the premiums to

result from buying and shorting firms with high ownership concentration (of large public pension fund blockholders). Giroud and Mueller (2008) sort based on the competitiveness of the industry and find that the GIM index matters primarily for firms in noncompetitive industries. Masulis et al (2007) connect these governance provisions to acquirer returns and find acquirers with "better" GIM firms, and those operating in more competitive industries, to experience higher abnormal returns. Note that these approaches often add a Delaware dummy as an additional characteristic to test the importance of this popular choice of venue, and this dummy is generally not significant.

While the results in these studies are very interesting, they present a challenge in that it is the difficult to model and understand the origin of these provisions. Better firms may have what are seen to be better internal bylaws, and it is very difficult to model and thus isolate the effects of all the provisions in a firm and attributes of a firm that make it better. Core et al (2006) push these points to examine the source of the mispricing of firms with better GIM governance, and find that it cannot be attributed to shareholder rights (in terms of analyst forecasts or surprise earning announcements), and conclude that, most likely, it is a correlation with one of the "pricing" puzzles of the 90s. Bebchuk et al (2005), criticize the Gompers measure claiming, based on some empirical evidence, that only six of the measures drive the results. Brown and Caylor (2006) say it is other measures, including new internal governance provisions as well as the ownership composition. Furthermore, both the nature of the choice (i.e. what bylaws can be chosen), and the implications of the choice (i.e. the manner in which the choice will be enforced in courts) are affected by the legal environment. Thus, firms may be choosing a jurisdiction because of its taxes, legal environment, or laws[7] and this may imply the choice of some of the other internal provisions. As explained above, the choice of internal governance is by no means independent from the choice of incorporation.

[7]Note that including some or all of the laws is insufficient in that laws represent the entire equilibrium result of legislative policy, which is implemented by the courts. However, bylaws resulting from laws may indeed require a different weighting.

The approach followed here is thus to combine the insights from both these literatures. My model takes the attributes of firms found to be correlated with the differential performance of better governed firms (as well as other measures not yet examined in the literature, and the laws, which often are mechanically related to the bylaws) and relates them to the fundamental choice of law and incorporation venue. I specify a model aimed at capturing the relatively well defined key factors in the incorporation choice, and thus facilitate the isolation of the preferences for the different laws and product characteristics. Thus, for example, my finding of a shareholder-manager divide in the preference for governance law, enriches the finding that governance matters differentially for firms with more or less powerful shareholders. My finding on firm preferences changing based on the concentration of industry complements the finding of governance mattering differentially in the presence of more competition. And my finding on the differential preferences of firms with different IPO characteristics suggests further trading strategies that could be explored. Furthermore, as shown below, the choice of law in itself suggests trading strategies that do, especially in recent years, produce abnormal returns.

More broadly, introducing heterogeneity into the choice of incorporation produces insights on the formal and real authority of the firm (i.e. Aghion and Tirole, 1997, Aghion et al 2007), especially regarding some of the recently important trends in institutional ownership, market power and the role of venture capital, which are interesting in themselves. My findings relate firm preferences to this important decision of incorporation choice, while exploiting the structure of incorporation choice to analyze diverging interests within a firm. Finally, as mentioned, this methodology is required in order to evaluate the impact of counterfactual policies.

1.3 Data: The State Characteristics and the Attributes of Firm Heterogeneity

In this section I describe the data used to capture the features of the incorporation package as well as the attributes of firms relevant to their incorporation choices: What motivates firms in their incorporation or reincorporation choices? Heron and Lewellen (1998) detail some of the reincorporation motives expressed by management in proxy statements, identifying them to be the establishment of takeover defenses, the reduction in director liability, obtaining legal flexibility and predictability, achieving tax and franchise fee savings, reconciliation of operating and legal domicile and (although to a lesser degree) the facilitation of acquisitions. Similarly, I conducted an online search for "advice" on where to incorporate, to capture the practical and informal discussions of the important determinants of firm choice. This search yielded many legal firms and organizations recommending that a corporation consider its home state, and also other options such as Delaware, Nevada, and Wyoming, in their comparison of the (pro and anti) business laws, the level of advancement of the legal systems, the other identity and characteristics of the other successful firms incorporated therein, and the "prestige" the various incorporation choices may carry, as well as the tax (franchise tax and state corporate income tax) "mentality" and costs of being incorporated away from home[8]. Indeed, these considerations are - to a large degree - an expression of preferences for governance mechanisms - both substantial and procedural, and consequently it is on them that I collect my data and focus my analysis in the "demand estimation".

What motivates states in the selection of laws, process and taxes? States undoubtedly prefer to have higher tax revenues. However, they are bureaucracies serving many masters or constituents with varying (often opposing) objectives. Indeed, there have been very few legislative changes in governance laws in the past 18 years (the period of my analy-

[8] For examples see www.mynewcompany.com/whichstate.htm; http://www.bizfilings.com/products/articles/which_state_to_inc.asp; and www.incnow.com/faq.shtml (all visited on 6/13/08).

sis). States do, however, undoubtedly care about their budgets. Tax revenues are not insignificant. The leading example is of course the state of Delaware, whose revenues from franchise taxes alone account for roughly one third of its total revenue, thus constituting a major piece of its budget. These revenues are almost entirely profits. Roberta Romano (1998) estimates the costs of serving the incorporated firms in the state of Delaware to be under 3% of these revenues. Indeed, the recent franchise tax increase it made in 2003 was explicitly motivated by the desire to shrink its expected budget deficit. As explained below, state corporate income taxes - while not directly related to incorporation - may be influenced by incorporation choices (in the most obvious sense where firms choose incorporation and location simultaneously, but even when they make separate decisions for these two choices). State corporate income taxes (SCIT) are sizeable. Furthermore, states do actively promote themselves (and their relative advantages) as a convenient place to incorporate (and locate)[9].

1.3.1 Public Firms

In this work I focus on publicly traded firms, which are easier to collect data for and are traded. Compustat data treats incorporation and location as scalar variables which are updated to the most recent value and consequently cannot be used for the time series. Thus, the main data source on firm incorporations and reincorporations is the actual SEC filings, as contained in the monthly SEC disclosure CD's. This data comes from the 10Ks and 10Qs filed by all public firms. For consistency, I pulled the data out of one CD a year beginning in 1990. I supplemented this data with compustat back-tapes. I use compustat financial information when available and the financial information in these filings otherwise[10]. Thus, the thousands of reincorporations identified from the SEC filings

[9] See for example http://www.wyomingcompany.com/ (visited 7/23/08).

[10] The only further potential step to track all reincorporations would be to look at the SDC merger data and isolate mergers in which the accounting survior is the new firm created solely for the purpose of reincorporation. This is not likely to be common, but I plan to explore this robustness check in later versions of the paper.

are a subset of the full universe of reincorporations. In my data, I track firms using their cusips, tickers and gvkeys. This implies that only reincorporations leaving one of these intact are analyzed[11]. This excludes some movements, but keeps those that are more likely to be directly related to a preference for the new state's product and not the result of other restructuring done with different motivations and for other objectives[12]. Finally, I supplement this data with the complete records of IPO data from Thompson SDC. IPO data adds more variables (such as whether the firm was venture capital backed and the share of insiders), and narrows in on firms that are at the critical private-public juncture. I find roughly 2500 movements in my sample period. This is consistent with previous literature using other approaches (such as sampling and checking a subset of the firms)[13].

1.3.2 Private Firms

Ideally, we would like to have a complete universe of public and private firms. Private firms are smaller and thus are likely to respond differently to tax changes. However, the agency problems faced by a (smaller) private firm and consequently the governance of private firms are decidedly different. It would be especially interesting to track private firms as they go through the initial IPO process. Unfortunately, the lack of mandatory reporting for private firms greatly hampers the data collection process. Using Dun and Bradstreet data, which tracks the larger private firms, I collected a full cross section of the most recent private firm information, and find (see figure 2) patterns different from those in public firms (compare Dammann and Schundeln (2007) that use similar data) . However the collection and full analysis of a panel of private firms and the determinants of private firm incorporation choices and their governance choices is left for future research.

[11] This will not be the case if there is a merger in which the accounting survivor is a new firm created for the reincorporation process.

[12] Following Daines 2001, the legal literature often drops all financial and utility firms, since many of them face additional regulations and laws. I have experimented both with and without these firms and have found (similar to the finding in the legal literature) that the results are similar in virtually all the analysis in this paper. In the specifications that follow I generally choose to use the full sample.

[13] Compare Rauh 2006 suggesting that about 5% of firms reincorporated over a 13 year period.

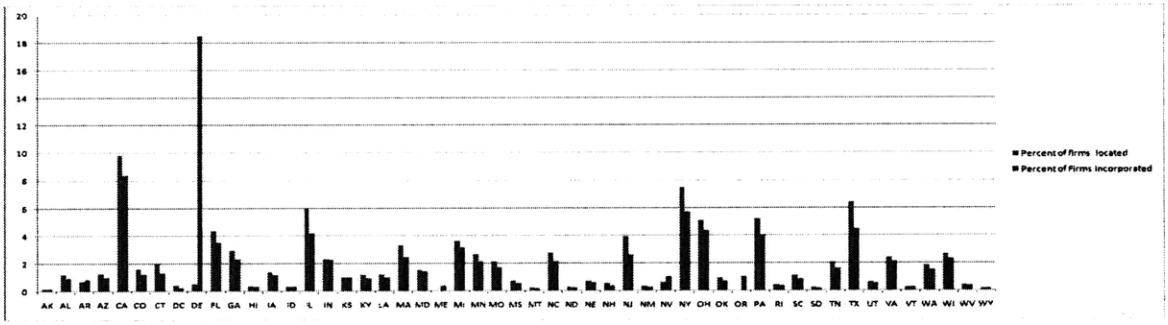


Figure 1-2: Shares of Private Firms Located and Incorporated in the 50 States.

1.3.3 Price - Incorporation and Franchise Taxes:

Overview

Taxes are the most direct price imposed on and paid by firms. States choose the taxes imposed on all firms operating or domiciled therein. As a result, there is considerable variance in the tax implications of incorporation in the different states. There are two main taxes that are generally directly related to the choice of incorporation: the incorporation or organization tax and the franchise tax. Both of these taxes generally have one of several unusual bases: the amount of the firm's actual or assumed par value capital, the number of authorized shares, or the total paid in capital. There is also variation across states and over time in the maximum amount of tax levied and in the manner in which these taxes are calculated. The difference between the taxes is that the first is paid upon incorporation and any increase in this base. Thus, if a firm decides to increase its number of authorized shares (the maximum number of shares management can issue without obtaining further shareholder approval) it will generally pay (in states imposing an incorporation tax with this base) a tax which is a function of the number of *additional* authorized shares. There is, therefore, an implied penalty for moving to a different state that charges such a tax, since in doing so the firm will have to pay the incorporation tax rate on the full amount.

Franchise taxes are computed from a similar base, and paid annually.

Indeed, the unusual base of authorized shares, which can vary considerably from the number of shares issued, is largely arbitrary. The number of authorized shares is in the financial statement, but is not collected by compustat or any other electronic database I am aware of. It does, perhaps, somewhat reflect the power and flexibility management has in making large expenditures without issuing debt, and in combating takeovers. However, it is difficult to see the connection between this firm *choice* of base and the potential implications it has on the balance of shareholder-management power. It also seems to be rather easy to manipulate. There is, however, a methodological advantage of using tax as a price in that there is significant price variation, stemming not only from the cross section and time series variation, but also from different firms being subjected to different prices (across the different jurisdictions), depending on their tax base and their location in the previous period. Compare Kahan and Kamar (2001) who discuss some of this price variation.

The variance in the incidence of these taxes on incorporated firms stems not only from the variance in rates, but also - more importantly - from the fact that in many states these taxes do not differentiate in the rates charged to domestic firms - incorporated in the state - and foreign firms - incorporated in a different state - but conducting business in the state. In these states, the base is (for domestic and foreign firms) the capital which can be attributed to the state. Thus, the tax base is often related proportionally to the place of income generation, a location assumed exogenous to the incorporation choice in this work. Thus, for many states (and consequently for many choice alternatives) the choice of incorporation conditional on fixed business locations, will not entail any tax implications at all. Finally, these taxes are very small. Even for the large public firms considered, the taxes generally do not exceed a few hundred thousand dollars a year, and are often much less. In fact, most jurisdictions do not impose a tax at all.

Do these taxes matter? The legal literature generally treats these taxes as *di minimus*

(and thus they have not been analyzed) as they generally do not exceed several hundred thousands of dollars a year for large firms (and are often much less). However, if these taxes do not matter, they are a distortion free way for states to collect hundreds of millions of dollars (at least) in taxes. Alternatively, if they do matter, or, in other words, if they are (at least in some jurisdictions) high enough to have real effects on (at least some) firm behavior, if firms have non-zero elasticities to these taxes, these elasticities can be used to price the preferences for the governance arrangements that firms care about. Indeed, franchise taxes do not seem to be irrelevant even for public firms. In Delaware, the maximum annual franchise tax is now (after a 10% increase in 2003) $\$165,000$. Firms that have a minimum of \$660M in assets and \$26.4M in authorized shares will pay the maximum tax. However, since the tax base depends on authorized (not issued) shares, firms with a high ratio of authorized to issued shares (a common phenomenon) could have significantly fewer assets and still pay the maximum rate. In the case of Delaware, the Bar commonly bemoans the adverse effects of tax increases on the number of incorporations. Firms also cite differential franchise tax rates as a reason for migrating out of Delaware. And finally, perhaps consequently, Delaware invests significant resources in justifying the taxes, claiming investments in improving the quality of its system, particularly in the time periods close to tax increases (compare Barzuza 2004). Indeed, we would expect that a source of revenue responsible for such a large share of the state budget would be carefully calibrated.

Incorporation and Franchise Tax Rates: Data

I manually constructed the time series of all incorporation and franchise taxes for the 51 US jurisdictions, by locating the state laws in which the taxes are imposed and then looking back at all their amendments since 1990. The CCH research network and their (older) paper volumes were useful in this regard, as well as the Lexis-Nexis and Westlaw databases. Appendix 3[14] details the incorporation and franchise tax rates for the jurisdictions that

[14] Will be available on my website.

impose them differentially for firms incorporated therein. In cases where incorporation has no effect on the tax, I omit the rates. When the cap on the tax is lower than 10,000 dollars I just list the cap. I have gone back as far as 1990 and thus list any previous rates that may have been in effect since then. Note that while the tax is often a small percent of the base, the base for the public firms is often in the hundreds of millions or billions. Data for the firm tax base is taken from compustat (or disclosure CDs). Note, however, that the authorized shares are approximated by the total number of shares issues. Thus the tax amounts are biased downward. The necessary assumption made here is therefore that there is no clear systematic bias (correlated with the price elasticities) in the gap between authorized and issued shares.

1.3.4 SCIT: A Product Characteristic

SCIT: Background

The third tax considered, the State Corporate Income Tax (SCIT), is different in that it is not closely linked to incorporation. The SCIT is apportioned between the states with which the firm has "nexus", which, while generally established by incorporation, can also be established by having property or a place of business in the state (a commonplace reality when conducting business therein) - using a formula weighting sales, employment and property. The apportionment base largely overlaps with the nexus base. Non business income is taxed at the home state. This home state is also not necessarily the state of incorporation. Furthermore, some states have a "throwback rule", [15] which stipulates that if there is no tax in the states in which the income is generated (usually where the product or service is sold), then the income is "thrown back" for tax purposes, to the state from which the product or service was shipped or provided. Here too, the determination

[15] Arizona, Connecticut, Delaware, Florida, Georgia, Iowa, Kentucky, Louisiana, Maryland, Michigan, Minnesota, Nevada, New Jersey, New York, North Carolina, Ohio, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Virginia, Washington, West Virginia, and Wyoming do NOT have this rule.

of this state will likely be related to the place of income generation and not incorporation. There is a physical presence criterion. Finally, there is the issue of the difference in reporting criteria across states. For example, Delaware does not tax intangible property, which incentivizes firms to establish a subsidiary in Delaware with trademark rights and to funnel significant earnings into this (untaxed) subsidiary. These types of phenomenons have encouraged states [16] to pass combined reporting laws by which all income from all subsidiaries is reported together and taxed together. This may facilitate more deductions of losses across firm parts, but also combats questionable transfers of income to subsidiaries located in favorable (tax) jurisdictions. However, once again, this is all related to the corporate income base, and therefore should not directly affect incorporation.

There is, however, likely a connection between the location and domicile of a corporation. These decisions are often made together. In other words, firms may weigh the incorporation features jointly with the location features. The SCIT also do affect many of the "tax ratings" of the states, which do not explicitly and transparently separate taxes by their varying incidences (i.e. by whether they require physical activity in the state). They contribute to the overall "feel" of the tax burden therein - a factor which may have an effect as well[17]. Firms may also "punish" states that are taxing them at high rates by leaving. Furthermore, there are situations in which a firm may not have nexus to the state (other than through incorporation), but may still owe the state a significant amount of taxes should nexus be established. Such may be the case if sales made in the state are the only connection to it[18]. In these cases incorporation may indeed have an effect. Finally, it is commonly perceived that there is gaming of the system in a myriad of ways, some of which may indeed relate to this choice of incorporation.

[16] Alaska, Arizona, California, Colorado, Hawaii, Idaho, Illinois, Kansas, Maine, Minnesota, Montana, Nebraska, New Hampshire, New York (recently adopted), North Dakota, Oregon, Texas, Utah, Vermont and West Virginia (recently adopted), have combined reporting laws.

[17] See for example www.learnaboutlaw.com (visited 7/31/08)

[18] The Commerce Clause in the Constitution is generally seen as prohibiting states to tax firms from other states soley due to their selling to residents of the state.

SCIT: Data

The SCIT rate were manually collected and coded from the hard copies of each year's Book of States. A look at the data in Appendix 4[19], reveals wide variance in the cross section of rates, and many changes over time.

1.3.5 State Laws

Background

The governance laws take the form of directly granting shareholders more voting power and say in the corporation and limiting the benefits and discretion of management and also of making takeovers - which can be seen as alternatives to the current management - more or less difficult. See definition of all laws in appendix A1. State laws can be categorized into those that are more or less pro shareholder rights. However, there is some theoretical ambiguity in the overall potential effects of the laws (compare Kahan (2006)). Thus, for example, provisions like control share cash-out or fair price provisions, protect shareholders, but in doing so also make takeovers more costly. There is a trade-off and such provisions are harmful only if their bite is primarily regarding shareholder wealth enhancing takeovers. It would indeed be surprising to find so many uniformly good or uniformly bad laws successfully passed in many states. It is of course possible that at least some of these laws would increase shareholder premiums conditional on a takeover occurring and in expectation be value enhancing. However, the prevalent opinion in the legal literature (which likely shapes perceptions in firms), largely resulting from event studies of the stock market reaction to the passing of takeover laws, is that the laws are shareholder wealth decreasing (see Romano, 1983, Karpoff and Malatesta 1989).

Furthermore, there is variation in the status of laws - mandatory or enabling - and even within the enabling laws there are differences in terms of the quorum required within a firm

[19] Will be available on my website.

to opt out of them. I stress, however, that even flexible enabling laws are likely to have effects. Firstly, opting out of these laws is very rare (see for example Subramanian 2001). Secondly, we know from much other economic research that defaults behaviorally do matter (compare for example Madrian and Shea, 2001). Corporate law - by and large - is enabling, and few claim that corporate law is irrelevant. Laws represent the political consensus reached in the state and the foundation and direction upon which other legislation and judicial decisions and interpretations are made. These laws serve as the background for negotiations between managers and shareholders (when there are - as there often are - conflicts of interest).

Data

The current versions of state laws can be found in Lexis and Westlaw. Westlaw is particularly good in that it tracks many of the changes over time in the laws and so it facilitates the construction of a panel. Given the incompleteness of these sources, as well as the varying structure and language of the laws in the 51 US jurisdictions, I used a variety of other sources as well: For some of the laws, the Model Business Corporations Act: Annotated, has (incomplete) comments on the states adopting the provisions of, or similar provisions to, those included in this codex. Many of the laws related to takeovers are also tracked by the State Takeover volumes published by the IRRC (see for example, Pinnell 2000). There are also up to date takeover watch databases (including for example SharkRepellent.net), which track some of the laws relating to takeovers. Following the previous legal literature, and my own preliminary disaggregate research, I found it useful to reduce the dimensionality and exposition of my results by combining two groups of laws into indices:

The first, the ATS index, is composed of the five antitakeover statutes found to be important by Bebchuk and Cohen (2003). These include control share acquisitions, expanded constituencies, fair price, business combination and poison pill endorsement laws. As explained in the appendix, these laws, while often relating to shareholder rights more

generally, offer protections to firms from takeovers. Control share acquisition laws require a disinterested shareholder vote to grant a new large shareholder voting rights. This shareholder vote is an impediment to takeovers, but does provide some protection to shareholders against coercive bids, and thus some see this law as positive despite its antitakeover nature. Expanded constituency laws grants management discretion to consider other firm constituencies, such as employees and suppliers when considering a takeover offer. This provides an easily manipulable legal base directors can use to resist value enhancing takeovers[20]. Fair price provisions limit the range of prices bidders can pay in two tiered offers and thus reduce the bargaining power of bidders. Shareholders are more likely to resist takeovers since they do not risk facing a significantly lower price in the second round. This constrains potentially beneficial acquirers in situations in which the stock price is in decline. Business combination laws impose a moratorium on certain transactions between large shareholders and the firm, unless the transaction is approved by the board of directors. This grants management the power to limit the benefits and synergies of mergers and thus reduces the overall desirability of the takeover. Finally, poison pill endorsements are a seal of approval given by the state for the use of poison pills, which are a host of mechanisms that grant the holders of target stock the ability to make takeovers more difficult. Poison pills are seen as a crucial component in modern takeover resistance strategies.

The second, the MAND index, follows, Kahan 2006, and includes four laws relating to the shareholder-manager balance of power, where states differ in the flexibility given to firms to opt out of them. Following Kahan 2006, states are coded as having the law if they impose the provision as a mandatory rule. These include cumulative voting, limits on loans to officers to directors, the restriction of limits on the personal liability of

[20]This interpretation follows that common in the legal literature. However, recent research has suggested that there are situations in which broadening the firms' objective can be beneficial (see for example Allen and Gale 2002, Allen, Carletti and Marquez 2007). The assumption thus made here is that the laws as they are written and applied in the US (or at least as they are construed by firms and shareholders), are prone to managerial manipulation. Future research is needed to better determine the precise effects of such laws and the scenarios in which they are desirable.

directors, and merger vote majority requirements. Cumulative voting allows shareholders to concentrate their vote and thereby facilitates the ability of minority shareholders to elect directors. These provisions are thus seen to increase shareholder rights. Limits on loans to officers and directors often impose personal liability on the recipients of loans or procedural requirements for the approval of these loans. The restriction of limits on the personal liability of directors are laws which do not allow firms to eliminate the personal liability of directors for a breach of duty. Finally, merger vote majority requirements are a limitation on the procedure by which mergers are approved[21].

Following Wald and Long, 2007, I also track the laws relating to the requirements of asset to liability ratios to make payouts to shareholders (found there to influence firm incorporation decisions). Finally, I look at the presence of laws recognizing actions of managers that are made outside the scope of their authority, ultra vires, as "firm actions" for which the firm is responsible. In the absence of these laws, claims could be made regarding the voidability of corporate actions beyond the scope of the charter. Thus the existence of these laws imposes more responsibility on corporations for the actions of their agents[22].

Indeed, while most of the changes in these laws were made in the late 1980s and early 1990s, there are changes over the entire time period studied. One of the most interesting changes concerns the special rules regarding loans made to officers and directors. The federal intervention in the passing of the Sarbanes-Oxley act introduced a general prohibition on such loans, thereby imposing this restriction across jurisdictions. This is an example of a (limited) federal intervention which has the effect of reducing the variance

[21] This law is somewhat different in that it could also be conceivably coded as an antitakeover statute. However, the difference here is that a state is coded as having the law if it imposes it as a mandatory provision.

[22] I also experimented with other laws including antigreenmail restrictions, compensation restrictions, control share cash out provisions, the adoption of the Model Business Corporation Act, severance pay, and labor contract provisions (see definitions in the appendix). However, I omit these from the analysis since they generally were not significant, or lacked sufficient cross section variability. I plan to revisit some of these laws in future research.

in incorporation implications, an intervention we are likely to see more of if the current regulatory proposals materialize. I return to this in section 6 below in discussing policy proposals.

1.3.6 Court Quality

Background

Much of the hype for the corporate law hot spots, and for Delaware in particular, concerns the relative quality of their court systems. Delaware boast a unique five member chancery court which has exclusive jurisdiction over, and hence specialization in, corporate law disputes. Furthermore, some claim that the Delaware court contributes to Delaware's supremacy by administering law that is predictable but not easily replicable.

Data

Ideally, to capture the benefits of the better systems, we would like measures for the overall quality of the decisions (i.e. for whether it was the "correct" decision), for the time it took to administer them, and relatedly, for the expenses that were required to get these "correct" decisions. Obviously, these measures are not available. There have been a number of studies assessing and comparing the quality of the state courts (see Choi et al 2008 for a review). These studies employ different methodologies and do not reach similar conclusions. I approximate for the quality or nature of the legal systems using two databases with proxies that seem most relevant for the questions at hand (given the data limitations): The first includes the Chamber of Commerce ranking and score, which are based on surveys of senior lawyers (in house counsel) at large corporations (with annual revenues of over 100M). These measures are commonly used (see for example Dammann and Schundeln 2007, and Kahan 2006) to rate the states, however, they suffer from several limitations. The first, which is more technical, is that they only go back to 2001 (and hence my panel is just for 2001 – 2007), and there are some differences in the survey

methodology and scoring even over this time period. The second is that they do not relate specifically to corporate law, but rather to more general categories of laws (such as torts and contracts, criminal law, and so on). The third - and what invites much criticism in the literature - is that given that they originate exclusively from the in house legal counsel at large firms, they are likely biased towards the preferences of management in these firms (to which the legal counsel often report). The second database comes from the State Court Statistics Project, which is conducted by the National Council for State Courts (NCSC) and disseminated by the ICPSR (I collected the most recent data from the NCSC website and thus have the years $1993-2005$). I look at both the appellate and trial level statistics: I include the following measures (for both the appellate and trial levels, compare Dakolias, 1999): The first is the ratio of civil cases disposed (whether by throwing out the case or deciding it for or against the appellant) to civil cases filed (a "clearance ration" and an "appeals clearance ratio", respectively). For states with more than one appellate court, I average the measures. The second is the clearance ratio for all cases (not just civil). I experimented with this measure since many courts have jurisdiction over many areas of law, and hence their workload and efficiency may be influenced by the caseload in all of these areas of law. The third is the ratio of the total number of judges in the courts to the total resident population. Finally, I include the ratio of appeals that were successful (where the decision was reversed or modified) to those that were not (where the appeal was dismissed or the trial level decision was affirmed). All measures at time t are used to analyze the (consequential) behavior at time $t+1$.

These measures proxy for how efficient the systems are at getting rid of cases (how back-loaded they are), where, for many of these disputes, the time the case is in trial is a very significant cost determinant (and, consequently, likely to significantly influence litigation behavior). These measures are by no means constant (even though in equilibrium the workload must be balanced). Courts tend to have "better" and "worse" years in handling their workloads. This can obviously also be influenced by the number of judges that are

in office (which motivates the inclusion of the number of judges per resident population measure separately). Finally, the appeal success rate is an (imperfect) measure for a variety of litigation climate indicators, including how much the trial level courts are respected (by the higher court levels) as well as how likely appeals are to be filed (although this is a more complicated equilibrium result influenced by the success of settlements out of courts, beliefs, etc.).

1.3.7 Firm Characteristics - Decomposition of the Heterogeneity

Having described the main characteristics of the incorporation product, I now move to discuss the data sources used and the motivation for the construction of the observed firm heterogeneity. In addition to the data on firms and IPOs discussed above, these relate to the structure of firm ownership, the industry concentration, and the characteristics of its management and directors:

Ownership

The dramatic changes in the percentage of institutional holdings alters the balance of power between shareholders and management. Institutional owners are generally seen as more sophisticated owners, especially when they hold significant shares, and thus are more likely to have a stronger say in the firm. Institutional investors are at least partially responsible for the "greater involvement of boards of directors and shareholders" (Holmstrom and Kaplan 2003). As mentioned, their presence has also been found to be related to the abnormal returns earned by firms with better internal governance[23]. Accordingly, I model the heterogeneity across several ownership dimensions:

[23] Interestingly, the growing strength of sophisticated investors dampens the concern for agency problems between the shareholders and management and increases concerns of agency problems between larger and smaller shareholders.

Institutional Holdings: Data

Thompson's Reuters CDA/Spectrum Institutional (13f) Holdings, has data for the stock holdings of all institutions managing 100M dollars or more[24]. Using this data, Cremers and Nair (2006, CN) look at two measures of internal governance - the percent held by the firm's largest institutional block-holder (which are shareholders with more than 5%), and the percent held by the 18 largest public pension funds. CN see public pension funds to be more "free from conflicts of interest and corporate pressure" and as "aggressive shareholder activists" (compare Guercio and Hawkings (1999)). They also see institutions holding larger shares to have incentives "to monitor the management and pay for part of the gains that occur through takeovers" ...potentially being "crucial to facilitate" and thus working "in tandem with the market for corporate control". Following their work, I construct four measures of ownership, including their two measures[25], as well as the fractional ownership by all institutional investors, and the total fractional ownership of blockholders with more than 1%. The reporting periods differs by institution (it ranges from quarterly to yearly), and thus when there is more than one reporting quarter they are averaged.

Indeed, the existence of such institutional investors implies a selection by them, which is an endogeneity concern in the sense that while it is interesting to see which stocks are picked by institutions, we would also like to randomly assign them to different firms and trace these firms' differential choices and performance. This endogeneity has recently been discussed and dealt with by using the inclusion in the Standard & Poor's (S&P) 500 as an instrument (see Aghion, Van Reenen and Zingales, 2008 and Sapra, Subramanian and Subramanian 2008), an approach I follow below.

[24] There may be some omissions for small holdings under 200,000 dollars.

[25] I was able to identify 15 of their funds in my data.

Director Holdings and Characteristics

The IRRC Director's database contains director level data yearly from 1996 – 2006 from which I aggregated (to the year level), the following variables: the average director age; the percent of reported Asians, African Americans, Whites, Hispanics, and Native Americans; the proportion of women, the average number of other major boards the directors are on; the average number of years served (I control for the firm age); the average number of years left (if there is a fixed term); total shares held; total voting power held; the proportion of the directors that are linked to the firm; the proportion of independent directors; the proportion up for election; the proportion that attended less than 75% of the meetings; the proportion that own less than 1%; and the proportion that are grandfathered upon retirement/tenure.

This data is supplemented with data from the "Corporate Library" (which goes from 2001 – 2007) on the CEO compensation and characteristics, total number of directors, and the overall compliance levels with SOX and with the loan requirements in SOX.

1.3.8 Internal Governance

Complete data on governance provisions in firm charters is taken from the IRRC database for the years 1990 – 2006. This is the data used by Gompers (2003) expanded to 2006 (they used the data up to 1998) and is generally published on a biannual basis. This data does not cover all of the publicly traded firms (it generally covers several thousand a year). It is "derived from a variety of public sources including corporate bylaws and charters, proxy statements, annual reports, as well as 10-K and 10-Q documents filed with the SEC. The IRRC universe is drawn from the S&P 500 as well as the annual lists of the largest corporations in the publications of Fortune, Forbes, and BusinessWeek. The IRRC sample was expanded by several hundred firms in 1998 [and has been expanding consistently since then through additions of some smaller firms and firms with high institutional-ownership levels...even in 1990 the IRRC tracked more than 93 percent of the total capitalization

of the combined New York Stock Exchange (NYSE), American Stock Exchange (AMEX), and NASDAQ markets." (Gompers 2003). As mentioned there is a GIM index which is publicly available. See appendix A2 for a description of the variables. The most recent sample is taken from the SharkRepellent website which tracks more firms (although does not go back historically).

1.3.9 Firm Industry Concentration

Previous studies have not found industry controls to produce any clear or meaningful conclusions. However, the recent work by Giroud and Mueller (2008) suggests that governance may matter (more) in noncompetitive industries. Thus a final characteristic of the firm is the concentration of the industry in which it operates, the Herfindahl index of the SIC code, as provided by compustat or the U.S. Bureau of Census (which accounts for private firms as well). I experimented with the 2, 3 and 4 digit codes, but generally use the 3 digit code, following the existing literature.[26]

1.4 Modeling The Demand

1.4.1 Formulation and Specification

Firms have the following utility function from each potential state of incorporation:

$$u_{ijt}(x_{jt}, y_{jt}, p_{jt}, s_{it}, \xi_{jt}) = x_{jt}\beta_i + y_{jt}\gamma_i - \alpha_i p_{ijt} + \phi g_{ijt} + \eta m_{jt} + \xi_j + \Delta\xi_{jt} + \varepsilon_{ijt}. \quad (1.1)$$

x_j is a vector of state laws, which includes the two indices ATS and MAND as well as the payout and ultra vires laws and the state corporate income taxes; y_j is the vector of court qualities; p_{ij} are the franchise taxes, and the incorporation taxes (note that these taxes are

[26] We do however see a potential caveat in merely analyzing the concentration of the industry. If indeed the threat to management is the driving force, then a measure for how competitive the market for (comparable) managers should be sought. Some (especially high ranking) positions across different industries, compete in the same market for managers.

firm specific). g_{ij} is the geographical distance - the physical distance from the incorporated state to the home state; m_j is a dummy variable indicating whether this product is the "home state" for the firm. Naturally, all characteristics have a time subscript as well, reflecting their changes over time. ε_{ijt} is the commonly used logit error. Firms receive independent draws from a type-two extreme value distribution in each period.

As is common in these specifications, the individual specific coefficients capture the heterogeneity in firms and the (plausible) variation in tastes, for the laws and process. Note that the price and geography characteristics, by construction, differ for different firms. ξ_j are the unobserved benefits from being incorporated in the system (commonly assumed to be enjoyed by all firms)[27]. In essence, this is a state of incorporation (product) fixed effect. It captures the average (firm specific or aggregate) preference for the system. The $\Delta\xi_{jt}$ is then the time-specific deviation from the average ξ_j. In this respect this application nicely matches the characteristic based demand approach. Products (in my model and as observed by firms) are a bundle of characteristics, and these characteristics vary over time.

I began with a specification with 51 product choices (all US jurisdictions) and an outside option of incorporating abroad. However, this model was very difficult to estimate, given that in my database of public firms there are very few states that attract a significant number of out of state firms. Thus, in order to achieve convergence of my estimators I limit the choice set. I assume that firms choose to incorporate at home, in one of 10 out of state options, or in the outside option, which is anything else. Including the home state as one of the options allows me to keep most of the variation in the product characteristics of all jurisdictions (and the absence of these characteristics abroad). This difference in the choice sets adds to the variation that comes from firms switching their incorporation choices, since structurally similar firms face different choices given their exogenous physical

[27] Note that the lack of a firm specific subscript i, on ξ_{ji}, or a model for the distribution of preferences for the unobservable is restrictive only as we depart from a completely flexible preference structure for the observed characteristics. Compare Nevo, 2000.

location. The 10 out of state options include any state that has more than 1% of the firms (at any point in my sample period) - which includes Delaware, Nevada, Maryland, Florida, Colorado, and Massachusetts - as well as California - home to many firms, and known for having very little takeover legislation, and New York, also home to many firms. I also include Pennsylvania, which historically had more firms, and Wyoming, which has made efforts recently to promote itself. As can be seen in Figures 6 and 7 in the appendix most of these states have seen significant shifts in the shares of firms they attract over my sample years, from the general public firm pool, and from IPO firms.

Firms are assumed to observed all product characteristics and weigh them in their location choice. We assume the utility from the outside option is:

$$u_{i0t} = \xi_{0t} + \pi_0 D_{it} + \sigma_0 \nu_{i0t} + \omega_0 m_{0t} + \varepsilon_{i0t}. \tag{1.2}$$

Following the standard assumptions we normalize ξ_0 to zero, thus the benefits from incorporation in one of the 11 choice options are relative to the normalized outside option of going elsewhere. $u_{i0} = 0$ so the utilities represent the difference between the particular chosen good and the outside good. This assumption implies that when choosing one of the states not accustomed to hosting out of state firms or to incorporate abroad, firms are doing something different than what my model captures, which I normalize to a utility of zero[28].

Note, that much of the firm heterogeneity is observed. Thus, we can model the random coefficients as:

$$\begin{pmatrix} \alpha_i \\ \beta_i \\ \gamma_i \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} + \pi D_i + \Sigma \nu_i \text{ with } \nu_i \sim N(0, I_{k+1}).$$

[28] I experimented with several other measures of the outside good, including dropping firms making choices outside these 11 options, and using incorporation abroad, or incorporation in one of these states as an outside option, and found my results to be similar.

The D_i captures the firm structure heterogeneity in financial profile, ownership, and industry, as discussed above. The v_i capture unobservable firm heterogeneity (where Σ captures the scale), i.e. other components not captured in the D_i vectors. This of course is much more flexible and general than a nested logit model, where the home vs. one of the other products "nest" is captured with the inclusion of the "home" dummy variable for the home product.

To simplify notation, we define $x_j = (x_j, y_j)$, the variables for which there are random coefficients, $y_j = (g_j, m_j)$, and $p_{ij} = (p_{ij}, g_{ij})$ the variables for which there are no random coefficients, and:

$$\begin{aligned} \theta^1 &= (\beta, \gamma) \\ \theta^2 &= (\eta) \\ \theta^p &= (\alpha, \phi) \\ \theta^o &= (\pi) \\ \theta^u &= (\Sigma) \end{aligned}$$

So:

$$u_{ij}(x_j, y_j, \xi_j) = \sum_k x_{jk}\theta^1_k + \sum_h y_{jh}\theta^2_h + \xi_j + \triangle\xi_{jt} + \theta^P p_{ijt} + \sum_{kr} x_{jk} D_{ir} \theta^o_{rk} + \sum_{kl} x_{jk}\nu_{il}\theta^u_{kl} + \varepsilon_{ijt}$$

Thus, each characteristic with a random coefficient has $(1+R+L)$ coefficients: the average coefficient, R coefficients on the observable firm structure demographics and L coefficients on the unobservables. For simplicity $L = 1$ so we have one unobservable per product characteristic.

Given the variation in the product characteristics over time, I include 11 dummies for each of the choice options. These dummies subsume the ξ_j, thus restricting the endogeneity concerns to the time specific - product specific unobservable, $\triangle\xi_{jt}$, not captured by and

related to the variables in my model. Thus, the specification becomes:

$$u_{ijt}(x_j, y_j, D_i, \xi_j, \triangle\xi_{jt}) = S_j + \sum_k x_{jk}\theta_k^1 + \sum_h y_{jh}\theta_h^2 + \triangle\xi_{jt} + \theta^P p_{ijt} + \sum_{kr} x_{jk} D_{ir}\theta_{rk}^o + \sum_{kl} x_{jk}\nu_{il}\theta_{kl}^u + \varepsilon_{ijt}$$

where S_j are the state dummies.

Firms that choose state j are those for which state j provides them with the highest utility, i.e. those belonging to the set:

$$A_{jt}(x_{.t}, p_{.t}, \delta_{.t}; \theta) = \{(D_i, p_{ij}, \nu_i, \varepsilon_{it}) | u_{ijt} \geq u_{ilt} \quad \forall l \in J\}. \tag{1.3}$$

Thus, the individual choice probabilities are:

$$\Pr(j|D_i, \theta, \delta) = \int_v \frac{\exp[S_j + \sum_k x_{jk}\theta_k^1 + \sum_h y_{jh}\theta_h^2 + \triangle\xi_{jt} + \theta^P p_{ijt} + \sum_{kr} x_{jk} D_{ir}\theta_{rk}^o + \sum_{kl} x_{jk}\nu_{il}\theta_{kl}^u]}{1 + \sum_q \exp[S_q + \sum_k x_{qk}\theta_k^1 + \sum_h y_{qh}\theta_h^2 + \theta^P p_{iqt} + \triangle\xi_{qt} + \sum_{kr} x_{qk} D_{ir}\theta_{rk}^o + \sum_{kl} x_{qk}\nu_{il}\theta_{kl}^u]} f(\nu) d(\nu).$$

1.4.2 Endogeneity: Discussion

Identification comes from many firms selecting from a wide menu of characteristic bundles, where there are changes over time in the product characteristics, and also much variation in the product choice sets faced by the different firms. Indeed, firms come from all jurisdictions (and abroad). Identification of the standard deviation of the random coefficients comes both from individuals switching in response to changes in the characteristics, as well as from structurally similar individuals facing different choices sets. This is the result of the changes over time in the options afforded to structurally similar firms, as well as a result of my model design whereby the comparisons made are with a varying home option and a fixed set of 10 out of home alternatives. The proportion that switch (or behave differently) characterize the shape of the distribution of the unobserved heterogeneity in the preferences.

The ability to include state of incorporation dummies controls for the ξ_j which is the chief source of endogeneity in these models, reduces most of the endogeneity concerns:

The prices, the variation in tax rates[29], likely reflect the advantages different states offer to the firms located therein, however, as mentioned, most of the price variation is in the differential choice sets and base of the different firms. The infrequency of tax changes reduces the need (and in fact ability) to instrument for taxes.

There are standard "default" IO instruments that come from the structure of the setup (see BLP, 1995, Hausman et al 1994, and Hausman 1996). These include (in context of our application) the observed characteristics of the states that are assumed to be exogenous, and the sum of the values of the same characteristics of the products offered by other states[30]. These instruments present particular challenges here in that they rely heavily on the structure of the game played by the states, a structure very difficult to specify, given the stagnation in state action. Accordingly, I was unable to use them to explain the broad variation in the taxes paid (given the state base and rate as well as the individual firm's tax liability). I experimented with using the variation in overall state tax revenue, since these influence the tendency to change the price, but are (likely) not directly related to changes in the unobserved product characteristics captured in the $\triangle\xi_{jt}$, as they are chiefly motivated by the many other budgetary factors. However, here too, unsurprisingly, I was unable to fit a significant first stage. Taxes are thus treated as exogenous, as are the state laws and court structure (compare the discussion in Nagar et al 2005).

The demographics are, in the context of this model, a decomposition of the heterogeneity of the firm. However, econometrically, they are no different from product characteristics in that their interaction with the characteristics need not be correlated with the $\Delta\xi_j$. As mentioned above, ownership by institutional shareholders is likely to be endogenous, in the sense that while sophisticated shareholders help police management and shape firm pref-

[29] As mentioned only the incorporation tax and the franchise tax are treated as prices. SCIT is a characteristic.

[30] Note that the third set of instruments in BLP, the sum of the characteristics of the products offered by other states, as well as the instruments offered in Hausman et al (1994) and in Hausman (1996), the price of the same product in other markets, are not relevant here, since each state offers one unique product. The difference in price here is a form of (third degree) price discrimination in that firms cannot choose different products at different prices within a given state.

erence, they may also tend to choose firms that are expected to perform better given their being in a jurisdiction which receives a particularly favorable time-specific shock. Here, as mentioned above, I use the instrument proposed by Aghion et al (2007), inclusion in the S&P 500. Inclusion in this index has a large random component (there can only be 500), unrelated to the fundamental performance of the firm, but the assumption is that nonetheless it is the inclusion in the S&P 500 itself that generates a kick in institutional ownership. Firms included in this index attract institutional funds for a variety of reasons[31]. Thus I use a regression discontinuity approach whereby I include a flexible function of market value [32] and a dummy for inclusion in this index. The assumption is that (flexibly) controlling for market value, being in or out of the index is largely random and thus this variation can be used to look at the effects of randomly increasing the share of institutional shareholders. Indeed, I will assume that this is the key endogeneity correction needed. However, the methodology outlined here can easily accommodate the treatment of endogeneity in any of the other product characteristics or firm attributes[33].

1.4.3 Estimation

I outline and follow the control function approach proposed by Imbens and Newey (2008), and implemented by Blundell and Powell (2004), and Petrin (2006)[34]. According to this approach, we write the endogeneity treatment as follows:

[31] Openly indexed funds are more likely track it, managers (in open and closed funds) are benchmarked against this index, and fiduciary duty laws influence such portfolio selection. See Aghion et al 2007.

[32] I chose levels to control for the linear relationship with market value that discretely breaks with the discontinuity at the index. Controls are added for a power series of market value.

[33] Endogeneity of the director and manager characteristics was not dealt with in detail given my finding below of their not being instrumental in the incorporation decision.

I note that another potential instrument for institutional ownership is whether firms payout dividends (given the rigidity in such decisions). I plan to explore this instrument further in future versions of this work.

[34] See Appendix E for two alternative treatment structures.

Assume that for the endogenous institutional ownership D_{eit}

$$D_{eit} = E[D_{ei}|z_{ijt}] + \zeta_i \qquad \text{(FS)}$$

where z_{ijt} are the all exogenous variables (and instruments). Note that this includes all characteristics of all choices.

Form

$$\hat{\zeta}_i = D_{ei} - \hat{D}_{ei}$$

by taking functionals of the residuals from the estimation of the first stage. Table 2 shows that all ownership measures are significant and that the first stage works well[35]. The R squared are relatively high[36].

We include and estimate an $S_j f(\hat{\zeta}_i)$ for each product. The significance of these product specific residuals is evidence of endogeneity, assuming the exclusion restrictions on the instruments are valid. In addition, assuming one scalar error per product, a well defined inverse for D_{ei}, the general single equilibria across markets assumptions, and that conditional on the $S_j f(\hat{\zeta}_i)$ we are left with a similar specification and logit error, the inclusion of this control variable essentially "controls" for the parts of the endogenous regressors that are correlated with the $\triangle\xi_{jt}$, allowing for consistent estimation of the coefficients, and the direct use of maximum simulated likelihood.

First, we construct the following likelihood:

$$L(D;\delta,\theta) = \sum_{i=1}^{N}\sum_{t=1}^{T}$$
$$\log\int_v \sum_{j=1}^{J}\left(\frac{\exp[S_j+\sum_k x_{jk}\theta_k^1+\sum_h y_{jh}\theta_h^2+\triangle\xi_{jt}+\theta^P p_{ijt}+\sum_{kr} x_{jk}D_{ir}\theta_{rk}^o+\sum_{kl} x_{jk}\nu_{il}\theta_{kl}^u+S_j f(\hat{\zeta}_i)]}{1+\sum_q \exp[S_q+\sum_k x_{qk}\theta_k^1+\sum_h y_{qh}\theta_h^2+\theta^P p_{iqt}+\triangle\xi_{qt}+\sum_{kr} x_{qk}D_{ir}\theta_{rk}^o+\sum_{kl} x_{qk}\nu_{il}\theta_{kl}^u+S_q f(\hat{\zeta}_i)]}\right)^{1(jit)} f(\nu)d(\nu).$$

Then we directly maximize the sample analog:

[35] There is of course a different first stage for each specification of the model. Table 2 includes controls for the litigation variables.

[36] Over 15% for the 1% block and 24% for the total institutional share meaure, 15% for the pension block and 17% for the 5% block measure.

$$SL = \sum_{i=1}^{N} \sum_{t=1}^{T} \log \frac{1}{R} \sum_{r=1}^{R} \sum_{j=1}^{J} \left(\frac{\exp[S_j + \sum_k x_{jk}\theta_k^1 + \sum_h y_{jh}\theta_h^2 + \theta^P p_{ijt} + S_j f(\hat{\zeta}_i) + \sum_{lr} x_{jtl} D_{ir} \theta_{rl}^o + \sum_{ko} x_{jtk} \nu_{ito} \theta_{ko}^u]}{1 + \sum_q \exp[S_q + \sum_k x_{qk}\theta_k^1 + \sum_h y_{qh}\theta_h^2 + \theta^P p_{iqt} + S_q f(\hat{\zeta}_i) + \sum_{lr} x_{qtl} D_{ir} \theta_{rl}^o + \sum_{ko} x_{qtk} \nu_{ito} \theta_{ko}^u]} \right)^{1(jit)} f(\nu) d(\nu).$$

This is the probability of observing (all of) the choices in the data, given the structure above. The right-hand-side does not have an analytical solution (given the assumed normal distribution for v) and has to be simulated. Generally we average over R draws from the assumed (normal) distribution, using different methods (Halton draws, Halton draws with some extra noise, and just plain random draws with noise) to ensure proper coverage of the domain of integration. We then obtain estimates of S_j, and the θs, controlling for clustering and a host of starting points, and directly test and control for endogeneity.

Table 2

(OLS regression clustered at the state level, controlling for all characteristics of all products)

	1% Block	5% Block	Total Inst. Share	Pension Block
	(1)	(2)	(3)	(4)
SP500	.098***	.013**	.240***	.018***
	(.017)	(.005)	(.037)	(.001)
Market Value(billions)	-.003***	-.002***	-.0004	-.00008***
	(.0005)	(.0001)	(.002)	(.00003)
Controls	Yes	Yes	Yes	Yes
Obs.	128002	128002	128002	128002

1.5 Results and Discussion

In this section I begin by looking specifically at the full model and then look specifically at two tax changes in 2003 to confirm the tax results and to "zone in" on the populations for which taxes are likely to most matter.

1.5.1 Demand Estimation

As mentioned above, the huge dimensionality of the data precludes the possibility of discussing all combinations of specification. Thus, in what follows, I will try to give a representative sample of the results. Note that the estimates are identified up to scale, given the normalization above. Thus, the focus should be on their relative magnitudes (ratios) rather than on their absolute size. Table 5 below details the summary statistics for the relevant variables in the specifications discussed below.

Table 1.1: Summary statistics

Variable	Mean	Std. Dev.	N
ATS	3.094	1.597	1921327
MAND	0.623	0.781	1921327
Payout Restrictions	0.925	0.316	1921327
Ultra Vires	0.560	0.496	1921327
Institutional Ownership	0.151	0.207	1938536
Industry Concentration	0.2	0.209	1938536
Age (Founding)	24.501	24.788	428787
Age (IPO)	6.233	5.571	798019
SCIT	0.065	0.035	1823545
Incorporation Taxes (Thousands)	34.065	229.691	1938536
Venture Backed	0.155	0.362	849924
Insiders After IPO	36.606	21.626	318099
Clearance Ratio (Appeals)	1.079	0.261	1249885
Clearance Ratio (Trial)	0.843	0.185	1361846
Distance (Thousands)	1.344	1.063	1752226

Tables 1A and 3A present the several specifications. Note that with the inclusion of additional firm structure measures, the sample size decreases and thus, while the qualitative results are similar, the coefficients are not identical.

Preferences for States of Incorporation and Incorporation Characteristics

First, while most of the product fixed effects are insignificant, the fixed effects for the Home state and Delaware products are large, positive and significant. This reflects the

preferences for being incorporated at home and in Delaware being the most common, but also suggest that despite the many variables in the specifications I present, there are still many residual unexplained characteristics that influence firm choice. These can include for example networks effects, the "prestige" of the Delaware jurisdiction and its unique court, or a general reluctance to explore outside options in the case of firms preferring to stay at home.

Secondly, regarding the laws, on average, firms strongly prefer to have takeover statutes (the ATS index laws), and have a negative preference for the MAND index. This implies that, on average, firms prefer to not be constrained in the election of directors (through cumulative voting), or in the ability of managers to receive loans[37]. Similarly they prefer to have the option to excuse directors of personal liability and to not be constrained in the merger vote majority requirements. These combined suggest that, on average, managers have a strong say in the "firm" choice and the ability to insulate themselves.

As to the preferences for the court process, I find that the only variables which matters significantly to be the clearance ratios at the appellate and trial levels[38]. As mentioned, these are proxies for the overall expected delay in the courts system when cases are back-loaded from previous years. Firms seem to prefer busier, more backloaded courts. This could be explained by the insulation and deterrence an inefficient court system may offer, since it implies that cases will take longer (something large deep pocketed firms can afford), and thus justice, if and when served, will be costly. Interestingly, I do not find the Chamber of Commerce ranking or score, as well as the other variables discussed above to significantly influence the choice of jurisdiction[39]. These results (my findings regarding the clearance rates and the lack of significance of the other measures) contrast with claims

[37] Although, as mentioned above, this ability is now uniformly banned by the SOX legislation.

[38] The specifications presented below are for the trial court congestion measure. The appellate measure is of a similar sign, but would require an argument based on the marginal deterrence of appeals. When both variables (and their interactions) are included I lose some power and the significance levels are lower.

[39] When adding the score and rank variables and thus reducing the sample size significantly we lose the significance of this variable. However, it is unclear that this last result is not simply a sample size problem, not reflecting the preference for this characteristic.

made in the literature regarding the importance of the legal system to firm choice. However, I note that these results, which are the results of the variation across all jurisdictions, do not imply that the unique corporate tribunal in Delaware does not play a special role[40].

I was unable to identify the two incorporation taxes separately, likely due to their multicollinearity, given their shared base, and thus I combined them. The incorporation tax variable represents the total tax implications for choosing a given jurisdiction, given the firm's location in the previous period. I computed the taxes using the rules in play in year $t-1$ with the tax base in year t. My reasoning is that given that I draw my sample once a year I cannot expect firms to have moved (in the data) in the year in which taxes change. Thus, in the context of the model above, my timing assumption is that at the beginning of each period firms observe the tax rates in each jurisdiction and form an expectation of what their tax base will be[41]. I also experimented with a tax measure based on last year's base and rate and found the results to be similar

As expected the demand is downward sloping. Note that given the large variance in tax schedules, the tax measure is not a rate, but rather a total (firm specific price) in thousands of dollars. Any significant tax increase (or initial imposition of tax) could conceivably raise the tax liabilities by much more than several thousands and thus see a significantly large firm response. Indeed, it seems that this **choice** of jurisdiction greatly hampers the ability to collect high taxes, thus potentially explaining some of the state lethargy in this area.

These results are consistent with my findings below regarding the 2003 tax increases. However, the magnitude of the coefficients is much smaller. I suspect that, indeed, the main response to taxes comes when there is a more salient change (as was the case regarding

[40] Furthermore, we may worry about this variable being endogenous in that a massive migration to an attractive jurisdiction may bring with it an added burden to the courts that they have trouble accommodating. However, given that, for the majority of states, corporate law disputes represent only a fraction of the total burden on the courts, and that we generally do not observe any massive firm migration in a short period of time, it is unlikely that the burden on the courts will be endogenously determined.

[41] Note that this tax base is generally based on measures controlled by the firm, such as the number of authorized shares.

the two most popular jurisdictions for "jurisdiction shoppers"), and less when the firm's base increases due, for example, to the increase in its authorized shares. In addition, as mentioned, there is often a penalty for moving in that firms will be required to pay the incorporation tax on their entire tax base (as apposed to on an increase in base if they stay in the same jurisdiction). When taxes are used to price the laws we get a rough NPV (using a 5% yearly discount rate) for the ATS laws of $3.8M$ (for the average β) and negative $8.3M$ for the MAND laws. This suggests that while these laws matter, they do not matter very much. These are potentially intriguing conclusions. However, given that the taxes reflect the aggregate variation it is of course possible that taxes matter a lot for a subset of firms. In any case, it is clear that the laws do not matter all that much for a large segment of the firm population.

The SCIT are not significant when controls are added. As mentioned above, this is not surprising, and suggests that the large amount of gaming of the differential rates does not manifest in firm incorporation choices.

Finally, as expected the distance (measures in thousands of miles) negatively influences the incorporation decision. Indeed, Nevada is aptly termed "Delaware of the west" reflecting the general reluctance to have the incorporation state far away. A distant incorporation state would undoubtedly raise the firm's costs in that administrative requirements as well as any court proceedings would require a long and costly trip to the domicile state.

Firm Ownership

As mentioned, I experimented with 4 measures. The results on all the general institutional holdings variables were similar and indeed stronger as the size of the institutional holdings increased. Thus the results for the 5% block are generally of a larger magnitude than those of the 1% block, which in turn are generally larger than those of the overall share. The total institutional shares measure was generally not significant. The public pension fund measure was generally not significant either (and at times not even of the same

sign), thus questioning the link between these incorporation decision results and the above mentioned results on abnormal returns. I thus chose to present the results from the 1% block specification, which seemed the most representative. As mentioned, this is the total share of the firms' stock held by institutions that have shares of at least 1%. I discuss the results here, assuming my endogeneity controls are valid, and that I accurately capture the preferences of these "sophisticated" shareholders, and then return to discuss the different endogeneity corrections in section 5.2 below.

Overall, institutional shareholders display a stronger reverse preference for the anti-takeover laws (the sign is negative and the absolute magnitude of the interaction is larger than the mean and so the combined effect, which is the sum of the two, is negative). This suggests that these shareholders prefer managers to be policed by the potential takeover threat and to not be insulated by law. Indeed, much research has shown that takeovers often benefit target shareholders (and punish bad managers). Thus, this result can be seen to reflect evidence of a key area in which there is a divergence of interests between shareholders and managers, where it takes sophisticated shareholders, with a large enough stake in the firm, to enforce the shareholder interests against those of managers. However, this raises the question of why previous findings have found governance to especially matter in firms with a larger fraction of institutional shareholders (see for example Cremers and Nair 2005). If indeed these firms can have an impact on the governance choices of the firms they hold, or on firm decisions more broadly, why would the existence of internal governance mechanism be so important?[42]

Interestingly, institutional investors express an additional, particularly large, negative preference for the mandatory index. This can be explained in that when shareholders have more power they do not need, nor desire, a rigid structure constraining their choice of mechanisms. This would imply that institutional shareholders still require a takeover friendly

[42] We may also wonder why it is chiefly in firms held by public pension funds, a measure not found to be particularly important here.

legislative environment, but do not desire restrictive laws in other areas[43]. However, this result does suggest a potential caveat in including the merger voting supermajority in the MAND grouping (one based on Kahan 2006). This requirement may indeed be more relevant to the antitakeover group and may explain some of the size of this coefficient. Future versions of this paper will include more results looking into these indices.

Taken as a whole, it seems that while institutional shareholders have a significant influence on the firm, they do not simply echo a collective "shareholder wealth increasing" interest, and are not a perfect substitute for governance mechanisms. There may be a non-monotonicity in the effect of institutional shareholders on firm value, where when institutional shareholders gain too much of the firm's share their objectives may change and they may have a detrimental effect on firm decisions. We must thus also seriously consider the potential for agency problems between the more concentrated and powerful shareholders and the more dispersed less informed shareholders (see for example Nagar et al 2005). Thus, for example the more powerful contingencies may not like the mandated cumulative voting mechanisms which can grant dispersed shareholders more of a say in the firm[44]. More broadly, if indeed the current trend continues and institutional shareholders become increasing powerful and dominant in the public exchanges, we must consider the complex effects their presence has on firms and firm objectives, both at the firm level - in the formulation of firm strategy - and in the design of regulatory policy.

Finally, these investors express seem to prefer well functioning courts, when measured at the trial level. This suggests, that institutional shareholders prefer to be in better policed environments, when measured at the trial level. In contrast, I find that, at the appellate level, these investors exhibit an even a stronger (than mean) negative preference for the key court characteristic. This result suggests a congruence in preferences there.

[43] It is also possible that sophisticated shareholders may feel they have an advantage in jurisdictions with MAND laws in that they can better police management (compared to competing firms with a more dispersed ownership structure).

[44] Here too a disaggregate analysis of the laws in the MAND index may prove useful.

In future versions of the paper I plan to examine if indeed this can be reconciled with an aim to deter value reducing external litigation and to facilitate the deterrence ensuing from the potential to take management to court. This finding implies however that the choice of incorporation jurisdiction is indeed sensitive to internalizing the differential effects of judicial jurisdictions facing different burdens.

Industry Concentration

Here too, I experimented with several measures. I found the results using 3 and 4 digit SIC codes to be similar[45]. Following Giroud and Mueller (2008) I use the 3 digit measure for the results presented below[46]. I find that industry concentration particularly matters regarding the MAND index. I do not find evidence of firms "surrendering" their takeover protections when operating in concentrated industries. Firms with market power seem to behave similarly to the "average" firm regarding antitakeover provisions. Firms in more concentrated industries have a strong positive preference for these rules. However, firms in concentrated industries are not influenced by the MAND laws when making their incorporation choices. Their preference for the MAND laws is positive and strong when compared to the "average" firm, and the overall effect washes out. This may reflect more of an awareness of shareholders (and an incentive to act on it) of the need to curb managerial behavior when the market disciplining forces are absent. However, when taken together with the mean preferences, we may interpret this as *managers* being less constrained in industries with less product market competition and thus having the freedom to choose their incorporation jurisdiction irrespective of these laws (compare Cuñat and Guadalupe, 2005). I believe this finding, as well as its interaction with the findings in the asset pricing literature mentioned, does indeed merit further inquiry.

[45] However, there were differences when using the broad 2 digit measure.

[46] I also drop observations for which the industry concentration is above 97.5%, as they likely reflect industries which are too narrowly defined.

Financial Profile

The main variables explored here are the age of the firm (since its foundation), and the time since its original IPO. Table 3A shows that both the firms' age since its founding, as well as the time since its original IPO both have a similar small but statistically significant influence on firm preferences over laws. Older firms have an additional positive preference for the ATS index, although the results are much stronger when measuring age since IPO. These firms are also more comfortable with the MAND laws. This could reflect a congealment around these internal practices. In other words, the older and more established firms, may not require the same degree of flexibility in their governance. This, of course, could reflect both a vintage effect - in that the newer firms are different and desire or require more flexibility, and also a life cycle explanation, whereby as the firm matures, its needs will change and it will move to another jurisdiction. My data on both the date of the firm's founding as well as on the firms' original IPO comes from both compustat as well as from the SDC data. When looking at the movements of firms in my data I find that indeed a significant share are by firms that underwent an IPO since 1980. This is thus suggestive evidence that there is some life cycle behavior in incorporation choices.

Indeed, when looking at the trends amongst IPO firms, it is clear that there is a strong tendency towards Delaware incorporation, even as its shares are in decline. My anecdotal evidence, from discussions with managers in the venture capital industry and in some of the data collection agencies mentioned, suggests that indeed Delaware incorporation (a jurisdiction with a MAND value of 0) has become a default for many firms when initially incorporating[47].

When looking specifically at the younger firms which underwent an original IPO after 1980, I find that whether or not the firm was venture capital financed significantly influences its preferences. Other variables explored, such as the IPO financing, and the share of

[47] This, however, is consistant with both a vintage and life cycle explanation, since all it requires is for firms to initially prefer low MAND jurisdictions.

insiders before and after IPO did not seem to matter. The results regarding venture backed firms echo those of institutional shareholders (although are of a smaller magnitude). This is unsurprising in that venture capital investors often carry a significant stake in the firm, and similarly possess the experience and sophistication surmised to influence the preferences of institutional investors above. It can thus be seen as further evidence supporting the potential of a formal and real authority divide in the firm decision making process. However, it is clear that there is much diversity in the objectives of the various venture capital investment policies, and thus the effect captured is likely an average effect from amongst significantly varying venture capital effects[48].

Other Firm Heterogeneity

Neither of the two main director measures considered - the total number of directors and the percent of independent directors - are significant. The one exception is the interaction of the total number of directors with the MAND legal index, which is marginally significant. However, the paucity of data, once including these measures, changes the estimates significantly, and significance is lost even for the mean β on the MAND index itself. Thus, these results should be not be taken as conclusive evidence of director composition not mattering. I used these two measures following the previous literature (see for example Masulis et al (2007)). However, similarly, I did not find any of the other measures mentioned above to be significant in the incorporation decision[49].

In addition, I experimented with two measures of compliance with the Sarbanes Oxley legislation (compare Chhaochharia and Grinstein 2007). The first, is a compliance score, tracking compliance with the CEO and CFO certification of published financials, loans compliance, designation of a financial expert on the audit committee, compliance with the standards for overall director independence, and compliance with the standards of commit-

[48] The significance level of the venture capital measure is lower as well, likely a cause of the dispersion in funds.

[49] Given these findings I did not proceed to treat concerns of endogeneity in these measures.

tee independence. The second refers specifically to the compliance with the requirement to eliminate loans to directors and executives. While this data is available for the year 2003 I tracked the firms for which it was available in the years before and the years after, experimenting with different time windows. Once again, these were not significant in the incorporation choice, but significantly restrict the sample size.

Unobserved heterogeneity

As can be seen, the standard deviation is significant only for the MAND index (and not for ATS or the legal index)[50]. As mentioned above, identification comes from firms switching and from the difference in the choice sets of structurally identical firms. The magnitude of the MAND σ is particularly large thus suggesting that $100 * (1 - \Phi(\frac{-\beta_k}{\sigma_k})) = \tilde{} \ 30\%$ of the firms have a positive valuation for the index. The ATS unobserved heterogeneity suggests that a tiny fraction of firms have a negative preference for the ATS laws. This suggests that controlling for the unobserved heterogeneity is important, but primarily for the MAND measure, given the large amount of observed firm structure heterogeneity.

Furthermore, note the importance of the unobserved heterogeneity particularly concerning the MAND index. In Table 2A we see that when the σs are dropped the mean β changes dramatically in magnitude and in fact is no longer significant. The interactions with the observed heterogeneity change significantly as well. These results strengthen the need to estimate a distribution around the random coefficients, even in the presence of much observed heterogeneity, not only to get realistic substitution patterns (as pointed out by the original BLP, 1995, models), but to get the point estimates right as well.

[50] Compare previous literature, such as Nevo 2001, where the results on the standard deviations are generally insignificant.

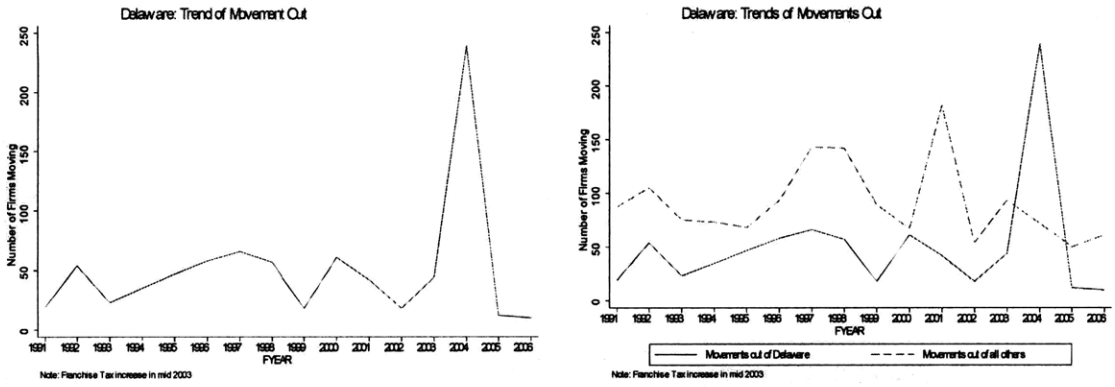
1.5.2 Endogeneity Corrections

Control Function

Table 2A gives the base specification with and without endogeneity corrections. Column 1 repeats the base specification from table 1A, column 2 has unobserved heterogeneity, while column 3 does not. Both columns 2 and 3 do not have endogeneity corrections. This table shows that despite the ability to capture much of the product specific unobservable with the product fixed effects, the controls are needed. Indeed, when added, the product specific residual functionals enter significantly for most products and their joint significance is easily established. This, in the context of the control function approach can be seen as a test confirming the importance of an endogeneity correction. The magnitude of the legal coefficients is influenced as well. The mean ATS β is about 60% its value without the corrections. The interaction with ownership (for which we instrumented) is also about half the absolute magnitude for the ATS index and about 66% the magnitude for the MAND index, reflecting the standard bias towards zero. Interestingly, the interaction with the industry characteristics is influenced by these controls as well. Finally, there are some changes in the size of the product specific fixed effects, especially regarding those of DE and HOME which are both decreased. Note that these patterns do not apply to the other coefficients and thus this is not merely an issue of scale.

1.5.3 Robustness: Incorporation and Franchise Taxes: The Effects of the 2003 Tax Increases on Firm Incorporation Choice

This wide price variation is incorporated in my model above, where I find that indeed firm demand is downward sloping. However, it is interesting to "zone in" on recent tax increases in Delaware (where the tax was raised by 10%) and Nevada (where a franchise tax was introduced) and apply a standard difference in difference approach. This is useful for confirming the results regarding the high sensitivity to taxes, as well as for understanding



(a) Delaware Movements

(b) Delaware Movements and Reference

Figure 1-3: Flows Out of States.

the populations most affected by such changes.

Figure 3 displays the flows of public firms out of Delaware and Nevada since 1991 (my data start in 1990), plotted on their own and against the flows out of all other jurisdictions. These are firms that remain in the data (so they remain (active) public firms). As can be seen, there is a spike in 2004, which, given the tax increases in mid 2003, is the expected time frame for the response. It is interesting to note that the "movers" in both jurisdictions are generally smaller firms with a lower market value and net income, and in Delaware they were still paying close to the maximum tax rate[51] (in Nevada there was no tax prior to 2003).

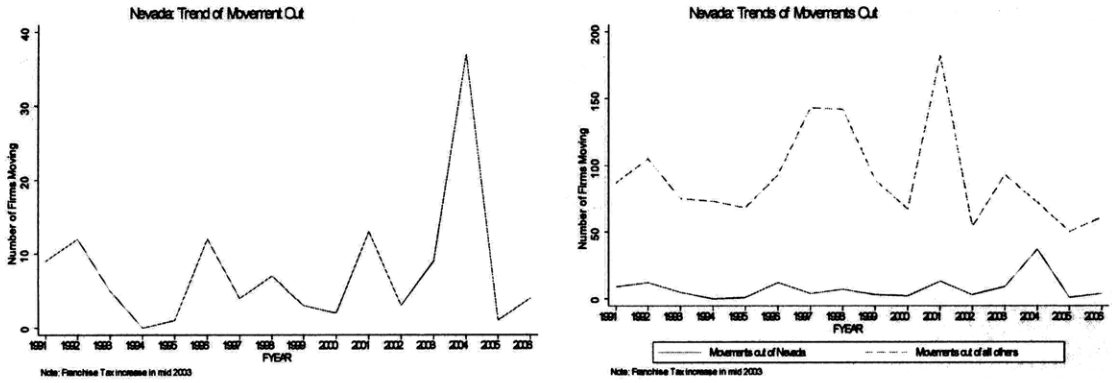
[51] Of course they may have been paying the maximum rate, since the tax based used (shares issued) is downward biased.

Table 1.2: Summary statistics (Non-Movers - Delaware)

Variable	Mean	Std. Dev.	N
Marketvalue	2382.289	12099.217	2671
Employees	7.104	37.753	3191
Net Income	72.832	652.752	3247
Franchise Tax Last Period	130066.355	49936.876	3342

Table 1.3: Summary statistics (Movers - Delaware)

Variable	Mean	Std. Dev.	N
Marketvalue	528.866	1471.514	111
Employees	1.622	3.992	167
Net Income	-13.418	266.814	171
Franchise Tax Last Period	110629.984	56584.013	172



(a) Nevada Movements

(b) Nevada Movements and Reference

Figure 1-4: Flows Out of States.

Table 1.4: Summary statistics (Non Movers - Nevada)

Variable	Mean	Std. Dev.	N
Marketvalue	638.002	1833.133	132
Employees	1.343	8.031	385
Net Income	2.766	89.335	478

Table 1.5: Summary statistics (Movers - Nevada)

Variable	Mean	Std. Dev.	N
Marketvalue	93.469	175.769	10
Employees	0.085	0.251	30
Net Income	-0.22	6.991	35

Table 1

	P(move)	P(move)	P(move)	P(move)
	(1)	(2)	(3)	(4)
Nevada-2004	-	.040***	-	.024***
		(.010)		(.007)
Delaware-2004	-	.056***	-	.039***
		(.012)		(.009)
NV-After	.010**	-	.007	-
	(.004)		(.004)	
DE-After	.019***	-	.018***	-
	(.006)		(.006)	
Firms	All	All	"Foreign"	"Foreign"
Year FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Observations	201,790	201,790	120,240	120,240

This is naturally confirmed in the difference in difference estimation in Table 1, where I look at the average marginal effects on the logit probability of moving out of the Nevada and Delaware jurisdictions respectively. Controls include all state fixed effects and year fixed effects, and I cluster by state. The difference in difference coefficients are the interactions. Columns 1 and 2 include all firms, while columns 3 and 4 look specifically at firms incorporated outside of their home state. Columns 1 and 3 difference the entire before and

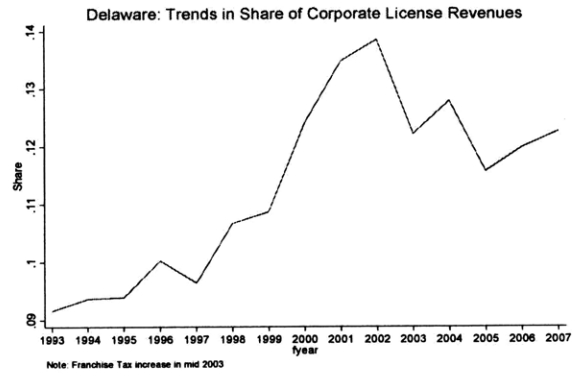


Figure 1-5: Trends in Share of Delaware's Corporate License Revenues.

after periods, whereas columns 2 and 4 focus on the effect in 2004, the year immediately following the tax increase. All difference in difference coefficients are significant. However, it seems that most of the treatment effect (the effect of the taxes) is concentrated in 2004. In most cases these firms go back home. Of the 239 firms that moved out of Delaware in 2004, 236 moved their incorporation back to their home state (2 moved to Maryland, and one to Nevada). Similarly, of the 37 firms that moved out of Nevada in 2004, 31 moved back to their home state as well, thus reinforcing the notion of home incorporation being a default for firms. Note that the other changes in Delaware's corporate law in 2003[52] were not, to the best of my knowledge, fraught with controversy. Finally, looking, in figure 5, at the aggregate share of Delaware's corporate license revenues in fiscal years ending in June of 1993 – 2007 shows the declining revenues that increased only temporarily in 2003 (or the fiscal year ending in June of 2004), only to decline with the massive exodus of firms.

[52]These were the limited expansion of the court of chancery's matter of jurisdiction and jurisdiction over executive officers, and the resolution of the ambiguity regarding shareholder and director rights to inspect the corporate books and records. See http://corp.delaware.gov/2003amends.shtml, visited 7/25/2008.

1.6 Counterfactual Policies

As mentioned, the advantage of specifying and estimating firm preferences in the manner proposed is that these preferences can be used to make out of sample prediction regarding different counterfactual policies. The advantage of having individual level data is that we can easily examine not only the impact on the aggregate shares, but also on the many degrees of firm structure. This can be seen as a partial equilibrium simulation in that we assume the distribution of firm structure remains constant and simulate the predicted choices firms (with this structure) will make under different policies. We can examine the differential behavior of firms with different firm structure heterogeneity. However, a "general equilibrium" variant would also require a model for how the specific policy change impacts firm structure. Thus, for example, if indeed sophisticated shareholders are selecting firms partially due to the firms' choice of laws, a simulation of the effect of centralizing some of the governance legislation will then change the distribution of ownership across firms. This will change the structure of the counterfactual market even if we assume that the preferences remain as we estimated them. Similarly, a radical change in the market may change the preferences captured (but not decomposed) by the fixed effects.

In what follows I simulate the effect of a simple counterfactual policy: the effect of eliminating the ATS statutes (setting ATS to zero for all firms) and imposing all of the MAND laws (setting MAND to 4) for all firms. I look at the year 2006 and compare the predicted probabilities produced, given the 2006 data, with those under the uniform legal policy described. I measure the "average policy effect", the difference in the expected probability of choosing a particular alternative before and after the policy change, thus explaining the change in the average choice, accounting for endogeneity.

As expected, there are significant changes in firm shares. Most notably, the share of firms predicted to be incorporating at home decreases by 8%, as does the share of firms in Nevada (down 58%), and Maryland (down 57%), while the share of firms incorporating in California increases by 120%. This suggests that California is losing many firms due to

its lack of takeover legislation and that part of the reason firms stay at home is the more convenient laws. Maryland and Nevada are capturing market share due to their "manager friendly" laws. Delaware, on the other hand, increases its share by 8%. To a degree, this dispels the notion, quoted above, that Delaware is attracting firms due to its "loose laws". In fact Delaware has few takeover laws and when the divergence in laws is made to disappear it does better. Once again, this is but one simple counterfactual exercise. I plan to examine more policies in later versions of this paper[53].

1.7 Incorporation Choice and Firm Performance

As mentioned, we can price firm preference over laws, in terms of the differential taxes. However, the full price would include the implied reduction in firm value from being in a suboptimal legislative environment. This would require a model and estimation strategy to isolate and compare the firm's choice with the counterfactual of the optimal choice for that given firm. Thus, for example, an analysis of the returns earned by movers would require a decomposition of the reasons for a move (or carefully chosen trading portfolios), and the complicated apportionment of any, firm specific, differential returns to any difference in laws. Indeed, this is clearly a challenging exercise, one which is beyond the scope of this paper. In the section below, I relate my findings to the analyses of performance common in the asset pricing corporate governance literature (regarding firm bylaws), which are done, largely, by tracking (abnormal) stock returns. Furthermore, I suggest how new trading strategies can easily be formed using the additional variables introduced in the analysis above.

Clearly, as discussed throughout, and as I have found in my data, the variables in my model are related to the internal governance of a firm. The laws are often mechanically related to the bylaws, and many of the firm structure variables in my analysis were explicitly

[53] Updated versions will be posted online.

chosen based on prior research. However, my findings enrich the analysis and interpretation of the findings regarding the link between these variables and firm performance. Thus, for example, it seems that institutional shareholders have a say in the firm and thus can shape the internal governance. If indeed this is the case, it is interesting that it is chiefly amongst these firms that governance is found to matter (in the asset pricing literature). Similarly, it is interesting that industry concentration was particularly important regarding the laws in the MAND index, and not those in the ATS index. Furthermore, the finding regarding venture backed firms is but one example of a new way to sort firms before comparing the differential performance of firms with different governance regimes.

In what follows I conduct one of the most simple abnormal returns exercises. As mentioned, the GIM (and now commonplace) methodology is to construct an index of governance, and update it when new data is available. Market value weighted portfolios are formed for each group and normalized, so the trading strategy replicated is buying democratic firms (with few pro managerial provisions) and selling (short) dictatorial firms (every month). I follow the same methodology with both of my legal indices, while updating them in the year after any laws are passed for the firms in that respective jurisdiction. Data on prices is taken from CRSP. The differential monthly performance (i.e. differential monthly stock market returns) of firms in jurisdictions with high values of the indices and firms in jurisdictions with low values for the indices are then regressed against 4 commonly used factors, as proposed by Fama and French 1993, and augmented by Jagadeesh and Titman (1993) and Carhart (1997)). These include the returns to the market portfolio, and to three portfolios that capture the size, book-to-market ratio and momentum effects. Data on the value of these factors is taken from Kenneth French's website. The abnormal return is then the constant in these regressions, representing the return above and beyond what are seen as the key factors moving stock market returns[54].

[54] My estimates are not identical to theirs, likely due to the information on some firms being hand-collected and matched to CRSP by them. I use their index and the advantages of my panel in finding firms over time (even when they changes one of their identifiers, such as their ticker).

Following the previous literature it is useful to separate my time period into the period of the 1990s, and the later period. Indeed, the GIM index no longer "works" in the later period and as suggested by various data suite teams[55], and in various public proxy voting recommendation pieces (see for example the ISS US Proxy Voting Guidelines Summary, 2000 – 2007), much of these findings have been internalized. Thus, for example there is a growing trend for firms to relinquish their poison pill protections and classified boards. If indeed this is the case, the core legal defaults may matter once again (and may matter more), especially for newer firms or firms with provisions open to periodic votes, which may "suffer" the interventions of newly informed shareholders. The trends in institutional ownership as well would induce similar results, in that they are likely influential not only in the choice of jurisdiction, but also (perhaps even more importantly) in the bylaw structure as well. Indeed, the post 1990s period in general and post Enron and Sarbanes-Oxley era in particular are seen by many to be different, due to the more suspicious attitude towards management. Takeovers have rebounded to their 1980's levels, but they are now less hostile and less leveraged, and thus may be faced with different (or less) internal protections in firms.

As displayed in table 6, I present the trading strategy for both of the indices, since 2000 and since 2001. In unreported regressions, I find that the constant in these specifications is significant only in the period since 2000 (or 2001) but not for the full period and not in the 1990s[56]. Interestingly, I find that firms in jurisdictions with more takeover protection perform better (the coefficient implies that 36 basis points could be earned per month by using this strategy), suggesting the legislative protection was shareholder wealth increasing. However, those in jurisdictions with a higher value of MAND perform better as well. And so the results are mixed.

These results are no more than interesting facts and should be interpreted with care.

[55] Compare Sharkrepellent "Research Insight" reports consistently documenting a large trend amongst existing public firms and IPOs to reduce their takeover protections.

[56] When excluding all financial services firms the abnormal returns are slightly higher.

My speculations regarding the reason for the post 2000 years is no more than a speculation and must be carefully analyzed. However, this - as well as the analysis of other (new) trading strategies suggested by my findings - is an obvious extension for future research.

Table 6

Legal Index Trading Strategies

	ATS (high-low)	ATS (high-low)	MAND (high-low)	MAND (high-low)
Const.	.363**	.361*	.461**	.450*
	(.168)	(.196)	(.216)	(.240)
MKT	-.293***	-.295***	-.075	.079
	(.050)	(.051)	(.065)	(.062)
SMB	-.238***	-.341***	.055	-.059
	(.052)	(.052)	(.066)	(.064)
HML	.122*	.203***	-.197**	-.092
	(.067)	(.062)	(.086)	(.076)
UMD	.057	.040	-.192***	-.112**
	(.041)	(.045)	(.052)	(.055)
Obs.	72	84	72	84

1.8 Summary, Discussion, and Future Work

This paper exploited the choice that the American legal structure offers states and firms to recover the differing firm preferences for the characteristics of the incorporation bundle. We found that there is significant lethargy in state activity. However, given the "home bias" or the preference firms have for remaining incorporated in their home state, there is still wide variation in the incorporation choice. We then addressed the issue of what matters to firms in their choice of governance regime, in the choice of incorporation. As discussed above, the choice offered by the US regulatory structure allows for the recovery

of much information embedded in the preferences revealed by firms:

Beginning with the taxes, I showed that, unsurprisingly, there is no evidence that state corporate income taxes, which generally are not directly related to incorporation, play a role in incorporation choices. This alleviates concerns that the large suspected abuse and manipulation of the differing state tax rules extend to the incorporation freedoms. Moving to incorporation taxes I found, using both reduced form evidence from specific tax changes and in the aggregate model, that at least a subset of firms are very responsive to taxes. This may serve as an explanation for the fact that incorporation taxes are so low. However, when viewing taxes as a substitute for governance laws, a simple calculation has firms valuing governance laws at several million dollars, a miniscule amount given the size of the public firms analyzed. This suggests that for some firms the laws do not matter all that much. Note that this may be important in the overall state calculus given the large number of (smaller) private firms, which may be responsive in a manner similar to the smaller public firms. However, clearly, for larger firms, this elasticity of willingness to pay taxes-to governance, is much higher. Furthermore, the large response to tax changes may reflect some gaming behavior between firms and states as a deterrent for future tax initiatives. Finally, it is clear that while taxes are a direct (measurable) price in that the incorporation (or reincorporation) choice has direct tax implications, the multitude of effects the components of the incorporation package have on the various firm constituencies are likely to be significantly larger. In other words, the coefficient which is an average across firms comprising of various types of variation in the data cannot explain the full importance and interaction of the taxes with the specific firm characteristics. Laws may (differentially) impact the market value of different firms (particularly those that do *not* move) and this may serve as a more powerful incentive in the choice of incorporation. However, a different model is needed in order to isolate and relate the impact of the various laws and incorporation bundle characteristics on the heterogenous firm values, and I leave this for future research.

I then moved on to the core of the paper - governance legislation. I showed that the preference structure for governance laws is complex, and, unsurprisingly, very much related to the internal and external characterization of the firm. The average preferences of the firm can often be seen to reflect the (narrow) interests of management: On average, firms dislike antitakeover laws and mandatory laws restricting the flexibility to grant managers more power and to limit their liability. Similarly, there is a marginally significant distaste for restrictions on payouts to shareholders. Thus, there are indeed clear patterns in the collective preferences of firms. However, these preferences are not uniform. When institutional shareholders as well as venture capital - both of which can be seen as sophisticated shareholders - have significant stakes in the firm, they express a clear distaste for antitakeover laws. This may reflect the findings that takeovers often benefit shareholders and punish inefficient management. It may also reflect the prevailing view that these laws are shareholder wealth reducing[57]. Firms with more market power are similar to the average firm in their preferences for antitakeover litigation, however, they have a different intensity of preferences for mandatory laws restricting managerial power. This suggests some specificity in the trade-off between product market competition and corporate governance regulation in law. I did not find any of the director characteristics with which I experimented to have power in explaining the choice of the incorporation package. The age of the firm, both since its founding and since its IPO, affects the intensity of preferences as well. This can be seen as evidence of a "life cycle" development of firm preferences, or of a recent shift in the preferences of younger firms.

In addition to the analysis of the observed heterogeneity, this paper shows that accounting for unobserved heterogeneity is important as well. I find, particularly concerning the mandatory laws analyzed, that there is still much residual variance in firm preferences,

[57] This view is the result of many event studies, many of which were conducted in the 80's (see for example Bhagat and Romano, 2002). From a theoretical point of view, even if antitakeover laws make takeovers more difficult, they may increase the premiums to shareholders when takeovers are successful and have a positive effect overall.

the omission of which would obscure the findings. This further confirms the importance of the methodological approach. Furthermore, this paper shows and compares several endogeneity corrections for the selection of firms by institutional shareholders and finds that the omission of these controls significantly biases the findings towards zero. Indeed, the analysis of observed and unobserved firm heterogeneity makes a more general point regarding the need to view firms as a sum of their inner components, both as a technical manner - to disentangle the conflicting effects - as well as a necessary step to obtain a more complete view of "firm" choice.

I then moved on to the other pieces of the incorporation package. I find that firms generally, and institutional shareholders increasingly, dislike efficient courts, in that they prefer to be under the jurisdiction of court systems which have accumulated a large number of cases from previous years, which, consequently, can be expected to perform more slowly. Institutional shareholders seem to prefer well functioning courts, at least at the trial level, once again highlighting the divide in preferences[58].

To complete the analysis of the incorporation bundle, I find, as expected, a significant distaste for the geographical distance. This suggests the potential for regional incorporation "hot spots" and may explain some of the reason for Nevada emerging as a "Delaware of the West".

Taking a step back, the more general point made is that, as shown in the counterfactual simulations and as suggested by my findings, policy towards firms does not and will not have a uniform impact. Despite the lethargy in state legislation, there is considerable variance in the incorporation implications across states. However, sorting does and will occur when firms are given the freedom to choose. This sorting is likely to be largely related to the effective authority within the firm, and not solely to the differential manner in which heterogeneous firms can increase their value. In other words, the optimistic view, which

[58] As mentioned the preferences of institutional shareholders for delay at the appellate level are reversed, albeit with marginal significance levels. I plan to explore the source of this difference in looking at trial and appellate courts. The average preference is for "slower" courts at both levels.

would hope for firms to spatially match their needs to different niches provided by different states, presupposes a firm collective that maximizes a shared objective, a view severely challenged by the findings in this paper. Reform of firm choice can be made in one of two ways. The first is to increase the freedom afforded to firms by not restricting them to any bundled incorporation package, but rather allowing them to piece their corporate governance regime together on their own. This can be done by separating the choice of judicial forum from the incorporation decision, thus allowing firms, for example, to adjudicate in Delaware, using California law, or even by slicing the choice of law more finely to allow firms to choose to be subject to different jurisdictions for different areas of corporate governance law. The second is of course to reduce the freedom given to firms either by linking incorporation to the physical location, or, more easily, by imposing uniform federal regulation. The findings in this paper suggest that more choice would not necessarily seep down to firms with weaker shareholders and thus would miss many of the desired beneficiaries. Furthermore, this would require an accounting for the resulting incentives states have to develop their regimes, including a potentially complicated allocation and apportionment of the incorporation taxes. However, centralizing corporate governance law would eliminate choice from all firms, a result that may indeed be too dramatic and require too much foresight from the policy maker, which would be required to effectively consider the differing needs of the different firms. Furthermore, it is clear that the current makeup of firm heterogeneity is merely a snapshot of an ongoing process. Firms, both public and private, are in a state of flux as the composition of ownership and management and the structure of industries are rapidly changing. The current economic crisis is contributing to this restructuring as well, as crises often do. Thus, policy must also be sensitive to the expected future changes, both those resulting directly from the impeding regulations as well as those naturally arising from the changing economic conditions. Firms with a high concentration of sophisticated shareholders should be closely monitored as they shift the location of the agency issues to the potential majority-minority shareholder divide.

Furthermore, as mentioned, the counterfactual analysis may be more a summary of the current dynamics than a reliable predictor of what will happen with policy reform. The measures of laws may be, to some degree, capturing the sentiment in the various jurisdictions and not the laws per se. Sentiment and forward looking expectations undoubtedly play a large role and are difficult to account for when extrapolating the results to inform future proposed policies.

It seems obvious that the taxation of incorporation should be revisited. First, facilitating the recovery of more tax revenue may serve as an incentive for states to more actively consider the design of their systems. Second, the use of taxes as a price may even be a potential means to screen firms based on their differential needs. A restructuring of the current base (likely an antiquated historical artifact) to make it more salient, to relate it to the agency issues within the firm, and to prevent an easy escape from taxes with a reincorporation move, would then be required.

Finally, there is indeed a conceptual link between what is seen to matter for the "upper level" of governance - the choice of litigation regime - and the overall preferences of firms as reflected in their charters. Thus, the effects of being in jurisdictions with different corporate governance laws can be taken directly to stock market trading strategies, either generally, as shown in the paper, or more narrowly, by zeroing in on particular firm cohorts. However, the findings in the incorporation analysis should be seen as complementary to the findings on the general count of governance provisions in firm charters and to impact the interpretation of these latter results. Thus, for example, we may question the finding that governance matters more to firms with a high level of institutional shareholders, if indeed these are firms where shareholders have more of a say and can directly monitor management, as suggested here. Similarly, we may question the intensity of any trade-off between market competition and governance, given the lack of a significant shift in preferences for the ATS laws, and a modest shift for the MAND laws. More broadly, any claim regarding the effects of the passing of governance laws on firms (see for example

Giroud and Mueller 2008) must include an accounting for the collective firm choice to remain under the jurisdiction of such laws and not to reincorporate. A natural development in the corporate governance literature will be to more carefully analyze the components of the broad indices used (the GIM index being an obvious example), their relationship to the legal environment, and their differential impact on different firms. This extension of the study of the link between governance and performance should also be made, of course, regarding other real measures of performance and firm choice, such as the choice of investments and capital structure and firm profitability.

This paper can be seen as a preliminary step towards the consideration of the manner in which federal regulation should be made. In this vein, the extensions to this paper are straightforward. First, more work is needed in modeling the political economy of the supply side. Much progress can be made in understanding the current distribution of incorporation bundle characteristics, as well as in designing policy that can motivate state innovation, if indeed the states' objective function can be modeled. Second, as noted throughout the paper, the large dimensionality of the data I collect in terms of the many attributes of firm structure allows for a very large number of specifications testing a broad range of issues regarding the behavior of firms differentiated across these dimensions. An exhaustive examination of all the degrees of observed heterogeneity in my data is too broad for the scope of this single paper. Indeed, as mentioned, my results suggest that there are more dimensions of firm heterogeneity that have significant explanatory power. Thus, more aspects of the data can be examined and the data can be augmented to specifically analyze other dimensions of heterogeneity. Similarly, while I focus on a broad range of legislation, there are still rules that I have not considered. Thus, for example, while I find that financial firms behave similarly to the overall pool of firms, I plan to explore the specific rules relating to them and the manner in which they (differentially) impact firms[59]. Indeed, as the new

[59] For example, Maryland has become increasingly popular amongst REITs and so it would be interesting to see whether there may be judicial and legislative developments to impact this.

policy proposals of the incoming administration crystallize, the methodology presented here and the counterfactual simulations can be tailored in furtherance of their analysis. Furthermore, as noted throughout, while the results of this paper are largely based on an aggregate analysis, more "micro" studies (similar to the one done above regarding the legislative changes in 2003 can be done).

As a next step, private firms should be analyzed. These firms vary to a considerable degree both during their "private" life and in their decision to go public. The collection of a more representative panel of private firms allows for a closer look at the timing of the IPO decision and its response to regulation of trading exchanges (a complementary form of regulation to that studied in this paper). The structure of such an analysis parallels the structure used here. However, private firms face different governance and control issues and have different agency problems. Thus, we should expect to see different preferences exhibited between public and private firms and in private firms amongst themselves and this may serve as a test for the model presented here for incorporation motives. Finally, while my focus has been primarily on the US, this market structure of competition over firms and firm choice is becoming increasingly relevant in Europe as well. The research outlined above can be nicely paralleled in the study of the evolving European markets and ultimately in the global market as well.

Bibliography

[1] Aghion, Philippe and Jean Tirole, (1997), "Formal and Real Authority in Organizations," 105(1), *The Journal of Political Economy*, 1 – 29.

[2] Aghion, Philippe, John Van Reenen and Luigi Zingales, (2007), "Innovation and Institutional Ownership", mimeo.

[3] Allen, Franklin and Douglas Gale, (2002), "A Comparative Theory of Corporate Governance", mimeo.

[4] Allen, Franklin, Elena Carletti and Robert Marquez, (2007), "Stakeholder Capitalism, Corporate Governance and Firm Value", *Working Paper 07-39, Wharton Financial Institutions Center*, University of Pennsylvania.

[5] Ali, Ashiq, Sandy Klasa, and Eric Yeung, (2008), "The Limitations of Industry Concentration Measures Constructed with Compustat Data: Implications for Finance Research", mimeo.

[6] Barzuza, Michal, (2004), "Price Considerations in the Market for Corporate Law," *Cardozo Law Review*, 26, 129.

[7] Barzuza, Michal, (2008), "Delaware's Compensation," *Virginia Law Review*, 94(3), 521 – 573.

[8] Bebchuk, Lucian Arye, and Assaf Hamdani, (2002), "Vigorous Race or Leisurely Walk: Reconsidering the Competition over Corporate Charters," *The Yale Law Journal*, 112(3), 553 – 615.

[9] Bebchuk, Lucian Arye, and Alma Cohen, (2003), "Firms' Decisions where to Incorporate," *The Journal of Law and Economics*, XLVI, 383 – 425.

[10] Bebchuk, Lucian, Alma Cohen and Alan Ferrell, (2005), "What Matters in Corporate Governance?", *Harvard Law School Working paper.*

[11] Berry, Steven, James Levinsohn, and Ariel Pakes (BLP), (1995), "Automobile Prices in Market Equilibrium", *Econometrica*, 63(4), 841 – 890.

[12] Berry, Steven, James Levinsohn, and Ariel Pakes, (2004), "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, 112(1), 68 – 103.

[13] Bhagat, Sanjai, and Roberta Romano (2002) "Event Studies and the Law: Part II: Empirical Studies of Corporate Law", 4 *American Law and Economics Review*, 380.

[14] Blundell, Richard, and James Powell, (2004),"Endogeneity in Semiparametric Binary Response Models, *Review of Economic Studies* 71(3),655 – 679.

[15] Brown, Lawrence D. and Marcus L. Caylor, (2006), "Corporate Governance and Firm Valuation," *Journal of Accounting and Public Policy*, 25, 409 – 434.

[16] Burda, Martin, Matthew Harding, and Jerry Hausman, (2008), "A Bayesian Mixed Logit-Probit Model for Multinomial Choice," mimeo.

[17] Bureau of National Affairs, "Multistate Tax Report: 2003 Survey of State Tax Departments," 10(4), April 25, (2003).

[18] Carhart, Mark M. (1997), "On Persistence in Mutual Fund Performance", *The Journal of Finance*, 52(1), 57 – 82

[19] Cary, William L., (1974), "Federalism and Corporate Law: Reflections upon Delaware," *Yale Law Journal*, 83, 663.

[20] Chernozhukov, Victor, Guido Imbens, and Whitney Newey, (2007), "Instrumental Variable Estimation of Nonseparable Models", *Journal of Econometrics*, 139, 4 – 14.

[21] Chhaochharia, Vidhi, and Yaniv Grinstein, (2007), "Corporate Governance and Firm Value: The Impact of the 2002 Governance Rules", The Journal of Finance, LXII (4), 1789 – 1825.

[22] Choi, Stephen J., G. Mitu Gulati, and Eric A. Posner, (2008), "Which States Have the Best (and Worst) High Courts?", *University of Chicago, Law & Economics Working Paper* No. 405.

[23] Core, John, (1997), "On the Corporate Demand for Directors' and Officers' Insurance," *Journal of Risk and Insurance*, LXIV, 63 – 87.

[24] Core, John, (2000), "The Directors' and Officers' Insurance Premium: An Outside Assessment of the Quality of Corporate Governance," *Journal of Law, Economics, and Organization*, XVI, 449 – 477.

[25] Core, John E., Wayne R. Guay and Tjomme O. Rusticus, (2006) "Does Weak Governance Cause Weak Stock Returns? An Examination of Firm Operating Performance and Investors' Expectations," *The Journal of Finance*, LXI(2), 655 – 687.

[26] Cremers, K.J. Martijn and Vinay B. Nair, (2005), "Governance Mechanisms and Equity Prices", *The Journal of Finance*, LX(6), 2859 – 2894.

[27] Cuñat, Vicente, and Maria Guadalupe, (2005), "How Does Product Market Competition Shape Incentive Contracts?," *Journal of the European Economic Association* (2005)

[28] Daines, Robert, "Does Delaware Law Improve Firm Value?", (2001), *Journal of Financial Economics*, 62, 525 – 558.

[29] Daines, Robert, and Michael Klausner, (2001), "Do IPO Charters Maximize Firm Value? Antitakeover Protection in IPOs," *Journal of Law, Economics, and Organization*, XVII, 83 – 120.

[30] Daines, Robert, (2002), "The Incorporation Choices of IPO Firms," *New York University Law Review*, 77, 1559 – 1611.

[31] Dakolias, Maria, (1999), "Court Performance Around the World: A Comparative Perspective", 2 *Yale Human Rights and Developments Law Journal*, 88 – 142.

[32] Dammann, Jens, and Matthias Schundeln, (2007), "The Incorporation Choices of Privately Held Corporations," *Law and Economics Research Paper No. 119*, The University of Texas School of Law.

[33] Dear, Jake, and Edward Jessen, (2007), ""Followed" Cases and Leading State Cases", 1940 – 2005, 41 *U.C. Davis L. Rev.* 683.

[34] Dubé, Jean-Pierre, Jeremy T. Fox, and Che-Lin Su, (2008), "Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation", mimeo.

[35] Easterbrook, Frank, (1983), "Antitrust and the Economics of Federalism,"*Journal of Law and Economics*, 26, 23.

[36] Eckbo, B. Espen, (1990), "Valuation Effects of Greenmail Prohibitions," XXV, *Journal of Financial and Quantitative Analysis*, 491 – 505.

[37] Edwards, Franklin (2004), "U.S. Corporate Governance: What Went Wrong and Can It Be Fixed?", in *Market Discipline Across Countries and Industries.*

[38] Fama, Eugene F., and Kenneth R. French (1993), "Common Risk Factors in the Returns on Bonds and Stocks", *Journal of Financial Economics*, 33, 3 – 53.

[39] Fischel, Daniel, (1982), "The 'Race to the Bottom' Revisited: Reflections on Recent Developments in Delaware's Corporation Law," *Northwestern University Law Review*, 76, 913.

[40] Fox, William F. and LeAnn Luna, (2002),"State Corporate Tax Revenue Trends: Causes and Possible Solutions," *National Tax Journal*, LV (3), 491 – 508.

[41] Gartman, Grant A., (2000), *State Antitakeover Law,* Investor Responsibility Research Center, Washington D.C.

[42] Garvey, Gerald T., and Gordon Hanka, (1999), "Capital Structure and Corporate Control: The Effect of Antitakeover Statutes on Firm Leverage," *Journal of Finance*, 54, 519 – 546.

[43] Giroud, Xavier and Holger M. Mueller, (2008), "Does Corporate Governance Matter in Competitive Industries?" mimeo.

[44] Gompers, Paul and Josh Lerner (2004), *The Venture Capital Cycle*, MIT Press (second edition).

[45] Gompers, Paul A. and Andrew Metrick, (2001), "Institutional Investors and Equity Prices," *Quarterly Journal of Economics*, 116(1), 229 – 259.

[46] Gompers, Paul, Joy Ishii and Andrew Metrick, (2003), "Corporate Governance and Equity Prices", *Quarterly Journal of Economics*, 107 – 155.

[47] Guercio, Diane D., and Jennifer Hawkins,(1999) "The Motivation and Impact of Pension Fund Activism",*Journal of Financial Economics,* 52, (1999), 293 – 340.

[48] Hansmann, Henry, (2006), "Corporation and Contract", *American Law and Economics Review*, 8, 1 – 19.

[49] Hahn, Jinyong, and Jerry Hausman, (2003), "Weak Instruments: Diagnosis and Cures in Empirical Econometrics,"*American Economic Review*, 93(2), 118.

[50] Hausman, Jerry, and William E. Taylor, (1981), "Panel Data and Unobservable Individual Effects", *Econometrica*, 49(6), 1377 – 1398.

[51] Hausman, Jerry, G. Leonard, and J.D. Zona, (1994), "Competitive Analysis with Differentiated Products", *Annales D'Economie et de Statistique*, 34, 159 – 180.

[52] Hausman, Jerry, (1996), "Valuation of New Goods Under Perfect and Imperfect Competition," *The Economics of New Goods, Studies in Income and Wealth* (ed. by T. Bresnahan and R. Gordon, Chicago: National Bureau of Economic Research), 58.

[53] Heron Randall A. and Wilbur G. Lewellen, (1998), "An Empirical Analysis of the Reincorporation Decision," *Journal of Financial and Quantitative Analysis*, 33(4), 549 – 568.

[54] Holmstrom, Bengt, and Steven N. Kaplan, (2001), "Corporate Governance and Merger Activity in the United States: Making Sense of the 1980s and 1990s," *Journal of Economic Perspectives*, 15(2), 121 – 144.

[55] Holmstrom, Bengt, and Steven N. Kaplan, (2003) "The State of U.S. Corporate Governance: What's Right and What's Wrong?" *ECGI Finance Working Paper*, 23/2003.

[56] Imbens, Guido W., and Whitney K. Newey (2008), "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity", mimeo.

[57] Jagadeesh, Narasimhan, and Sheridan Titman, (1993), "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency", *Journal of Finance*, 48, 65 – 91.

[58] Kahan, Marcel, and Ehud Kamar, (2001), "Price Discrimination in the Market for Corporate Law," *Cornell Law Review*, 86, 1205 – 1256.

[59] Kahan, Marcel and Ehud Kamar, (2002), "The Myth of State Competition in Corporate Law," *Stanford Law Review*, 55(3), 679 – 749.

[60] Kahan, Marcel, (2006), "The Demand for Corporate Law: Statutory Flexibility, Judicial Quality, or Takeover Protection," *Journal of Law, Economics and Organization*, 22(2), 340 – 365.

[61] Karpoff, Jonathan M., and Paul H. Malatesta, (1989), "The Wealth Effects of Second-Generation State Takeover Legislation," *Journal of Financial Economics*, 25, 291–322.

[62] Klausner, Michael, (1995), "Corporations, Corporate Law and Networks of Contracts," *Virginia Law Review*, 81, 757 – 852.

[63] Knittel, Christopher R. and Konstantinos Metaxoglou (2008), "Estimation of Random Coefficient Demand Models: Challenges, Difficulties and Warnings", *NBER Working Paper* 14080.

[64] Lambert, Richard A., and David F. Larcker, "Golden Parachutes, Executive Decision-Making and Shareholder Wealth," *Journal of Accounting and Economics*, VII (1985), 179 – 203.

[65] MacKerron, John A., (1993), "A Taxonomy of the Revised Model Business Corporation Act," *UMKC Law Review*, 61, 663 – 691.

[66] Madrian, Brigitte and Dennis F. Shea, (2001) "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior," *Quarterly Journal of Economics*, 116, 1149 – 1187.

[67] Mallette, Paul and Robert Spagnola, (1994), "State Takeover Legislation: The Political Defense", *SAM Advanced Management Journal*, 59(3), 15.

[68] Masulis, Ronald W., Cong Wang and Fei Xie, (2007), "Corporate Governance and Acquirer Returns", *The Journal of Finance*, LXII, 4, 1851 – 1889.

[69] Nader, Ralph, Mark Green, and Joel Seligman, (1976), *Taming the Giant Corporation*, New York, W.W.Norton.

[70] Nagar, Venky, Kathy Petroni and Daniel Wolfenzon, (2005), "Governance Problems in Close Corporations", *Working Paper.*

[71] Nevo, Aviv, (2000), "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand", *Journal of Economics and Management Strategy*, 9(4), 513 – 548.

[72] Nevo, Aviv, (2001), "Measuring Market Power in the Ready-to-Eat Cereal Industry", *Econometrica*, 69(2), 307 – 342.

[73] Newey, Whitney, (1990), "Efficient Instrumental Variable Estimation of Nonlinear Models," *Econometrica*, 58, 809 – 837.

[74] Petrin, Amil, (2005), "The Use of Control Functions to Identify Demand when Errors are Non-Additive", *University of Chicago Working Paper*

[75] Petrin, Amil (2006), "Control Function Corrections for Unobserved Factors in Differentiated Product Models", mimeo.

[76] Petrin, Amil, and Kenneth Train, (2006), "Control Function Corrections for Omitted Attributes in Differentiated Product Markets," *Quantitative Marketing and Economics*,

[77] Pinnell, Maria Carmen S.,(2000), *State Takeover Laws* (Washington, DC: Investor Responsibility Research Center Inc.

[78] Rauh, Joshua, D. (2006), "Own Company Stock in Defined Contribution Pension Plans: A Takeover Defense?", *Journal of Financial Economics*, 81(2), 379-410.

[79] Roe, Mark, (2003), "Delaware's Competition", *Harvard Law Review*, 117, 588 – 644.

[80] Romano, Roberta, (1985), "Law as a Product: Some Pieces of the Incorporation Puzzle," *Journal of Law, Economics and Organization*, 1(2), 225 – 283.

[81] Romano, Roberta, (1993), *The Genius of American Corporate Law* (AEI Press).

[82] Romano, Roberta,(1993), "Public Pension Fund Activism in Corporate Governance Reconsidered," *Columbia Law Review*, 93, 795 – 853.

[83] Romano, Roberta, (1998), "Empowering Investors: A Market Approach to Securities Regulation," *The Yale Law Journal*, 107(8), 2359 – 2430.

[84] Romano, Roberta, (2006), "The State as a Laboratory: Legal Innovation and State Competition for Corporate Charters", *Yale Journal on Regulation*, 209–

[85] Sapra, Haresh, Ajay Subramanian and Krishnamurthy Subramanian, "Corporate Governance and Innovation: Theory and Evidence", (2008), mimeo.

[86] Shleifer, Andrei, and Robert Vishny, (1986), "Greenmail, White Knights, and Shareholders' Interest," *Rand Journal of Economics,* XVII, 293 – 309.

[87] Shleifer, Andrei, and Robert Vishny, (1989), "Management Entrenchment: The Case of Manager-Specific Investments," *Journal of Financial Economics,* XXV, 123 – 140.

[88] Shleifer, Andrei, and Robert Vishny, (1997), "A Survey of Corporate Governance," *Journal of Finance*, LII, 737 – 783.

[89] Law Suit Climate, *State Liabilities Ranking Studies* series (2003 – 2008), The US Chamber of Commerce

[90] Subramanian, Guhan, (2001), "The influence of Antitakeover Statutes on Incorporation Choice: Evidence on the "Race" Debate and Antitakeover Overreaching", *Harvard NOM Working Paper.*

[91] Subramanian, Guhan, (2004), "The Disappearing Delaware Effect", *Journal of Law, Economics and Organization*, 20, 32 – 59.

[92] Villas-Boas, J. Miguel and Russel S. Winer, (1999) "Endogeneity in Brand Choice Models," *Management Science*, 45(10), 1324 – 1338.

[93] Wald, John K. and Michael S. Long, (2007), "The Effect of State Laws on Capital Structure," *Journal of Financial Economics*, 83(2), 297 – 319.

[94] Winter, Ralph K., Jr., (1977), "State Law, Shareholder Protection, and the Theory of the Corporation," *The Journal of Legal Studies*, 6, 251.

Table 1A

Variable	Means (β's)	Standard Deviations (σ's)	Interactions with Firm Structure: Ownership	Interactions with Firm Structure: Industry Concentration
Laws				
ATS	.314***	.001	-.478***	.049
	(.096)	(.002)	(.164)	(.042)
MAND	-.687***	1.31***	-1.19***	.493***
	(.208)	(.115)	(.406)	(.102)
Payout Restrictions	.037	-	-	-
	(.277)			
Ultra Vires	-.353	-	-	-
	(.292)			
Courts				
Clearance Ratio	-.75**	.066	1.43**	.711
	(.364)	(.045)	(.712)	(.422)
Taxes				
Incorp. Taxes	-.002***	-	-	-
	(.0004)			
SCIT	-3.28	-	-	-
	(4.35)			
Distance	-.272***	-	-	-
	(.083)			
Product Fixed Effects				
HOME	4.37***		FL	-1.21
	(.232)			(.201)
DE	5.49***		CO	-.16
	(.340)			(.313)
NV	.11		MA	-.625***
	(.458)			(.209)
MD	-.017		NY	-.035
	(.287)			(.435)
CA	-.413*		PA	.071
	(.358)			(.442)
WY	-1.75***		-	-
	(.595)			

ATS is an index of Antitakeover laws, MAND is an index of mandatory laws, Payout Restriction indicates the minimum asset to liability ratio required to make a shareholder payout. Ultra Vires indicates whether ultra vires actions are recognized in the state. Clearance Ratio is the ratio of cases accumulated

to cases disposed of in trial courts in the previous year. Incorp. Taxes are the total tax liability for the firm (incorporation and franchise taxes) as resulting from the choice of jurisdiction. SCIT is the state corporate income tax rate in the previous year. Distance is (in thousands of miles) from the home state to the incorporation state. Ownership is measure as the percent of shares held by institutions with at least a 1% block in the firm. Industry concentration is calculated based on the herfindahl index using the 3 digit sic code. This specification uses 780,824 observations. Endogeneity is controlled for using the control function approach described in the text. Standard errors are clustered at the home state level. ***,**,* indicate significance at the 1%,5% and 10% levels respectively.

Table 2A

Variable	(Endogeneity Correction)	(2)	(3)
HOME	4.37***	3.68***	3.49***
	(.233)	(.248)	(.25)
DE	5.49***	4.54***	4.32***
	(.34)	(.305)	(.293)
NV	.11	.48	.496
	(.458)	(.438)	(.467)
MD	-.016	.920**	.699*
	(.287)	(.379)	(.409)
ATS			
Mean	.314***	.194**	.181**
	(.97)	(.078)	(.075)
Ownership	-.478***	-.242***	-.21***
	(.712)	(.078)	(.074)
Industry Concentration	.05	.210***	.202***
	(.042)	(.068)	(.061)
σ	.002	0.0007	-
	(.002)	(.001)	
MAND			
Mean	-.688***	-.639***	-.19
	(.208)	(.214)	(.127)
Ownership	-1.19***	-.8***	-.67***
	(.406)	(.266)	(.174)
Industry. Concentration	.493***	.305***	.210***
	(.103)	(.111)	(.071)
σ	1.31***	1.14***	-
	(.115)	(0.174)	
Taxes			
Incorp. Tax	-.002***	-.002***	-.002***
	(.0004)	(.0003)	(.0003)
SCIT	-3.28	-2.07	-1.80
	(4.350)	(4.36)	(4.38)
Distance	-.272***	-.250***	-.224***
	(.084)	(.070)	(.074)
Payout Restrictions	.037	.015	-.058
	(.277)	(.223)	(.200)
Ultra Vires	-.353	-.268	-.173
	(.292)	(.257)	(.275)

Variable descriptions are the same as under table 1A, and the same variables are used. Not all variables

are reported for ease of exposition. Here, the first column has the control function endogeneity correction, while the second and third do not. In addition, column three does not account for unobserved heterogeneity in the random coefficient design.

Table 3A

Variable	(1)	(2)
ATS		
Mean	.446***	.368**
	(.157)	(.159)
Ownership	-.752***	-.679***
	(.218)	(.225)
Industry Concent.	.098	.039*
	(.041)	(.063)
Venture Backed	-.035	-.052**
	(.026)	(.026)
Age (since IPO)	.011***	-
	(.004)	
Age (since founded)	-	.004***
		(.001)
σ	.009	.003
	(.007)	(.003)
MAND		
Mean	-.861**	-.718*
	(.419)	(.410)
Ownership	-1.58***	-1.58***
	(.450)	(.492)
Industry Concent.	.710***	.637***
	(.140)	(.242)
Venture Backed	-.157***	-.123***
	(.050)	(.047)
Age (since IPO)	.041***	-
	(.001)	
Age (since founded)	-	.011***
		(.003)
σ	1.49***	1.37***
	(.216)	(.173)
Clearance Ratio		
Mean	-1.35**	-1.29**
	(.53)	(.525)
σ	0.012	0.048
	(.07)	(.059)
Taxes		
Incorporation	-.002*	-.002**
	(.0008)	(.0009)
SCIT	-6.32	-3.87
	(6.73)	(6.15)
Observations	298,473	220,930

Variable descriptions are the same as under table 1A. In addition, Venture Backed is a dummy variable indicating whether the firm had an initial IPO backed by venture capital, and two alternative age variables are added. These additional variables are cumulative to all variables in the model. Endogeneity corrections are using the control function approach as explained in the text.

1.9 Appendix E

As mentioned, once the instrument is identified, it can be incorporated in my highly non-linear model in two additional (related) ways:

1.9.1 GMM

The first is analogous to the BLP (1995, 2004) literature. Define δ_{jt} to be the time specific state specific fixed effect which captures this $\triangle\xi_{jt}$ as well. In other words we now have a state time specific effect. Thus yielding the specification of:

$$u_{ijt}(x_j, y_j, D_i, \xi_j, \triangle\xi_{jt}) = \delta_{jt} + \theta^p p_{ijt} + \sum_{lr} x_{jtl} D_{ir} \theta^o_{rl} + \sum_{ko} x_{jtk} \nu_{ito} \theta^u_{ko} + \varepsilon_{ijt}.$$

And, separating out the endogeneity of the ownership demographic, we have:

$$u_{ijt}(x_j, y_j, D_i, \xi_j, \triangle\xi_{jt}) = \delta_{jt} + \theta^P p_{ijt} + \sum_{lr} x_{jk} D_{ir} \theta^o_{rk} + \sum_{l} x_{jl} D_{ie} \theta^o_{el} + \sum_{ko} x_{jtk} \nu_{ito} \theta^u_{ko} + \varepsilon_{ijt}.$$

or:

$$u_{ijt}(x_{jt}, y_{jt}, \xi_j, \Delta\xi_{jt}) = \delta_{jt}(\cdot) + \mu_{ijt}(x_{jt}, p_{ij}, \nu_i, D_i; \alpha, \theta^o, \theta^u) + \varepsilon_{ijt}$$

where,

$$\begin{aligned} \delta_{jt} &= x_{jt}\theta^1 + y_{jt}\theta^2 + \triangle\xi_{jt}, \\ \mu_{ijt} &= [x_{jt}] * (\theta^o D_i + \theta^u \nu_i) + \theta^p p_{ijt}. \end{aligned}$$

Now, note that with aggregate data, by construction, the only observed choice variable is the shares, and thus it is these shares that have to be inverted, using a contraction mapping, to match the predicted shares with the observed shares (or minimize their distance). In this application this is avoided due to the richness in my data, containing individual

choice. We proceed as follows:

First, we construct the following likelihood:

$$L(D;\delta,\theta) = \sum_{i=1}^{N}\sum_{t=1}^{T}\log\int_v \sum_{j=1}^{J}\left(\frac{\exp[\delta_{jt}+\sum_{lr}x_{jk}D_{ir}\theta^o_{rk}+\theta^P p_{ijt}+\sum_l x_{jl}D_{ie}\theta^o_{el}+\sum_{ko}x_{jk}\nu_{io}\theta^u_{ko}]}{1+\sum_q \exp[\delta_q+\sum_{lr}x_{jk}D_{ir}\theta^o_{rk}+\theta^P p_{iqt}+\sum_l x_{jl}D_{ie}\theta^o_{el}+\sum_{ko}x_{jk}\nu_{io}\theta^u_{ko}]}\right)^{1(jit)} f(\nu)d(\nu).$$

This is the probability of observing (all of) the choices in the data, given the structure above, similar to the construction mentioned in the paper. The difference is that here all variables that are not firm specific are soaked up in the state-time fixed effect. Similarly, we maximize the analogous simulated likelihood:

$$SL = \sum_{i=1}^{N}\sum_{t=1}^{T}\log\frac{1}{R}\sum_{r=1}^{R}\sum_{j=1}^{J}\left(\frac{\exp[\delta_{jt}+\sum_{lr}x_{jk}D_{ir}\theta^o_{rk}+\theta^p p_{ijt}+\sum_l x_{jl}D_{ie}\theta^o_{el}+\sum_{ko}x_{jk}\nu_{io}\theta^u_{ko}]}{1+\sum_q \exp[\delta_q+\sum_{lr}x_{jk}D_{ir}\theta^o_{rk}+\theta^p p_{iqt}+\sum_l x_{jl}D_{ie}\theta^o_{el}+\sum_{ko}x_{jk}\nu_{io}\theta^u_{ko}]}\right)^{1(jit)} f(\nu)d(\nu).$$

This yields estimates of the $\delta_{jt}, \theta^0, \theta^u, \theta^P$. Now the $\hat{\theta} = \{\theta^1, \theta^2\}$ are recovered by forming GMM moments:

Define

$$w_{jt} = S_j + \sum_k x_{jk}\theta^1_k + \sum_h y_{jh}\theta^2_h$$

and construct the residuals:

$$\omega_{jt}(\theta) = \bar{\delta}_{jt} - w_{jt}\bar{\theta}.$$

Next, we define $H(Z_{jt})$ to be a function of instruments (following Newey, 1990). Here, since the errors are linear, we use a simple series with interactions between the $z_1, z_2, ..., z_k$.

Finally, we construct the moments as

$$g_{jt}(\theta) = H(Z_{jt})\omega_{jt}(\theta),$$

where the number of moments grows with the number of terms we construct in the series in $H(\cdot)$.

Now, by assumption, at the true θ_0:

$$E[g_{it}(\theta_0)] = 0.$$

Thus, define:

$$\hat{g}(\theta) = \frac{1}{TJ} \sum_{jt} g_{tj}(\theta)$$

the GMM estimate is:

$$\hat{\theta} = \arg_{\hat{\theta}} \min \hat{g}(\theta) \hat{\Omega}^{-1} \hat{g}(\theta)',$$

where $\hat{\Omega}$ is consistent estimate of $E[g_{jst}(\theta_0) g_{jst}(\theta_0)']$. This estimate is obtained by first using the initial estimate of $\tilde{\Omega} = I$ and then, computing and using $\tilde{\theta}$ to form the sample analog of $E[g_{jst}(\theta_0) g_{jst}(\theta_0)']$ as

$$\hat{\Omega} = \frac{1}{T} \sum_{j=1}^{J} \sum_{s=1}^{S} \sum_{t=1}^{T} g_{jt}(\hat{\theta}) g_{js}(\hat{\theta})',$$

where J is the number of products, $S = T =$ the number of periods. Essentially, to control for serial correlation of the errors, we interact each period's $g(\cdot)$ with all time periods (including its own period).

For simplicity, we bootstrap the standard errors of the estimates (while drawing from the state clusters), constructing an outer-loop on the entire estimation procedure.

The MPEC Approach

Implementation of the maximization is done using in Matlab and AMPL. The advantage of using AMPL - which is mathematical software designed for optimization with a large number of constraints - is the use of automatic differentiation, the richness of optimizers and the ability to reduce the simulation error arising from the multiple estimation steps. Indeed, this procedure skirts some of the criticisms in Dubé, 2008, and Knittel, 2008.

An alternative formulation of this problem is to combine all the steps into the following estimation problem:

Define:

$$\frac{1}{R}\sum_{r=1}^{R}\sum_{j=1}^{J}[\frac{\exp[\delta_{jt}+\theta^{P}p_{ijt}+\sum_{kr}x_{jk}D_{irt}\theta^{o}_{rk}+\sum_{kl}x_{jk}\nu_{ilt}\theta^{u}_{kl}]}{1+\sum_{q}\exp[\delta_{qt}+\theta^{P}p_{iqt}+\sum_{kr}x_{jk}D_{irt}\theta^{o}_{rk}+\sum_{kl}x_{jk}\nu_{ilt}\theta^{u}_{kl}]}]^{1(jit)}=s_{ijtr}$$

and then minimize the GMM moments, subject to the first order conditions from the MSL problem holding at the solution. Formally:

$$\min_{\theta}\hat{g}(\theta)\hat{\Omega}^{-1}\hat{g}(\theta)'$$

s.t.

$$\frac{d}{d\delta_{jt}}=\sum_{i=1}^{N}\sum_{t=1}^{T}(1-s_{ijtr})^{1(jit)}+(\sum_{r=1}^{R}\frac{1}{R}\frac{1}{1+\sum_{q}\exp[\delta_{qt}+\theta^{P}p_{iqt}+\sum_{kr}x_{jk}D_{irt}\theta^{o}_{rk}+\sum_{kl}x_{jk}\nu_{ilt}\theta^{u}_{kl}]})^{1-1(jit)}$$

$\forall j$

$$\frac{d}{d\theta^{P}}=\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{J}p_{ijt}(1-s_{ijtr})^{1(jit)}+(\sum_{r=1}^{R}\frac{1}{R}\frac{p_{ijt}}{1+\sum_{q}\exp[\delta_{qt}+\theta^{P}p_{iqt}+\sum_{kr}x_{jk}D_{irt}\theta^{o}_{rk}+\sum_{kl}x_{jk}\nu_{ilt}\theta^{u}_{kl}]})^{1-1(jit)}$$

$$\frac{d}{d\theta^{o}_{rk}}=\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{J}(x_{jk}D_{irt}-\sum_{m=1}^{J}x_{mk}D_{irt}s_{imtr})^{1(jit)}\qquad\forall rk$$

$$\frac{d}{d\theta^{u}_{kl}}=\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{J}[(x_{jk}\nu_{ilt}-\sum_{m=1}^{J}x_{mk}\nu_{ilt}s_{imtr})]^{1(jit)}\qquad\forall kl.$$

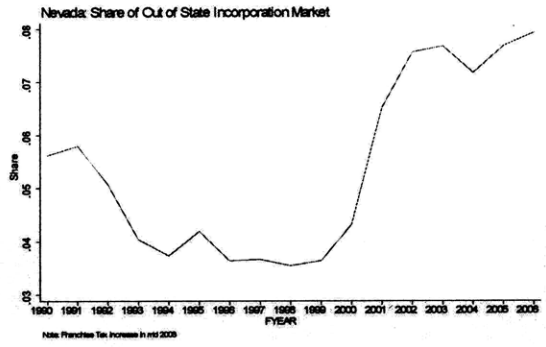
A comparison of the estimates from all three approaches will be detailed in an updated version of this appendix.

1.10 Additional Figures

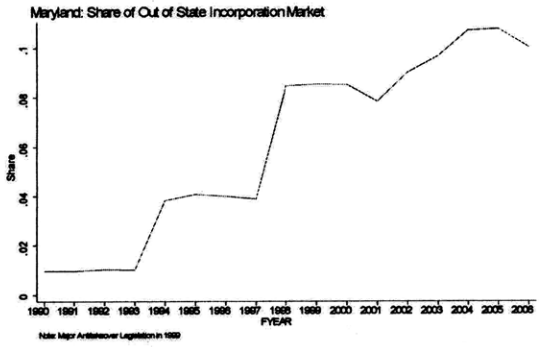
1.10.1 Figure 6: Trends in shares of Incorporations

1.10.2 Figure 7: Trends in share of IPOs

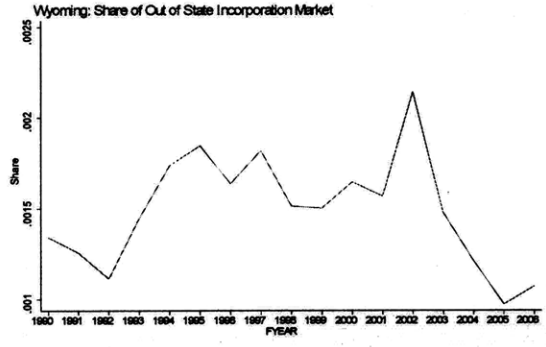
(a) Delaware Share: Out of State Incorporations



(b) Nevada Share: Out of State Incorporations

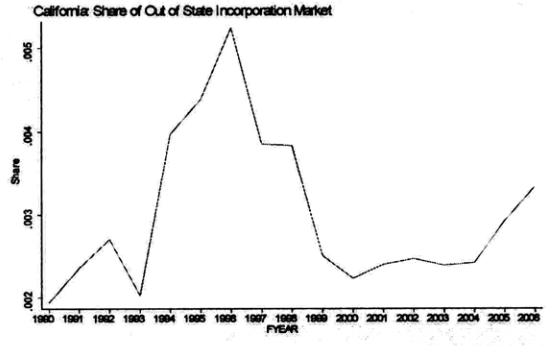


(c) Maryland Share: Out of State Incorporations



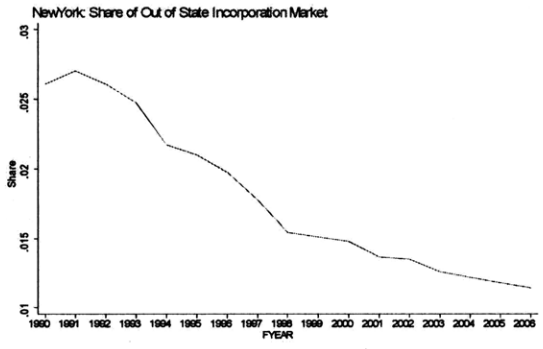
(d) Wyoming Share: Out of State Incorporations

(e) Pennsylvania Share: Out of State Incorporations



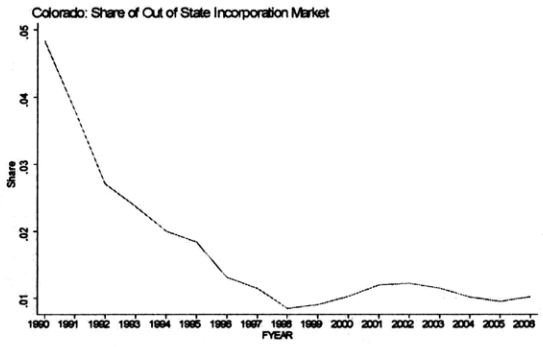
(f) California Share: Out of State Incorporations

Figure 1-6: Trends in Shares of Out of State Incorporations



(g) New York Share: Out of State Incorporations

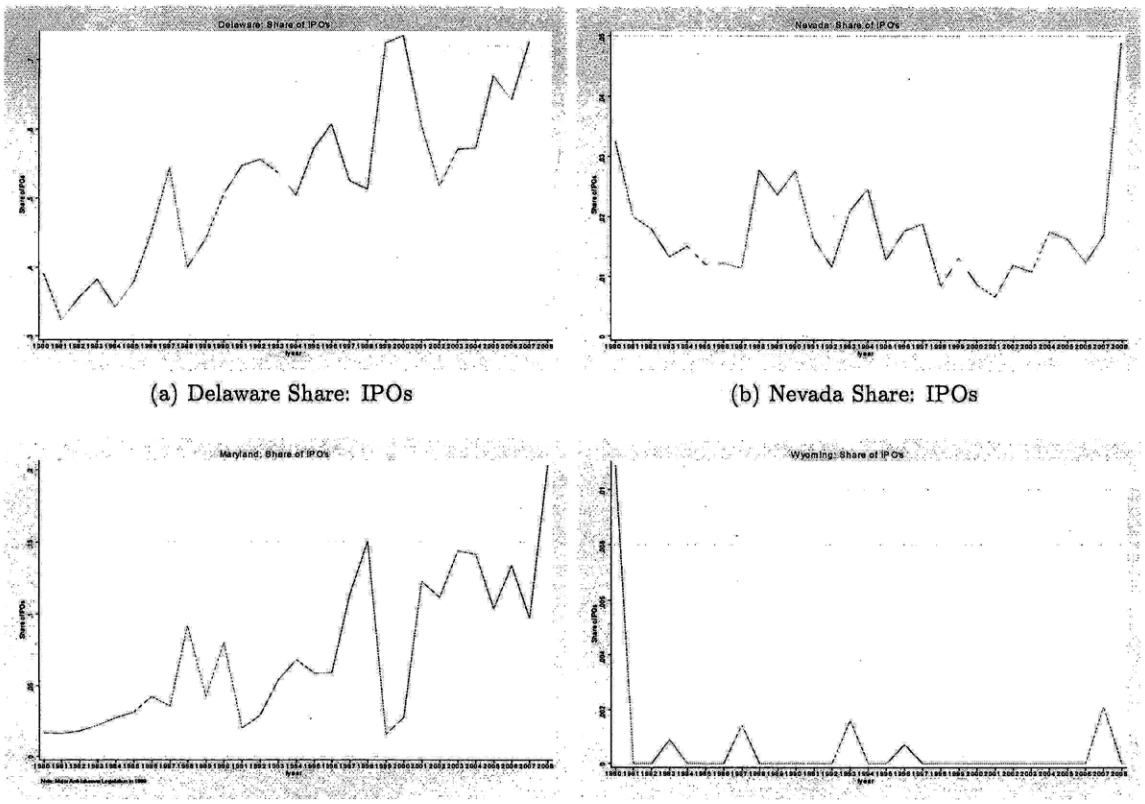
(h) Massachusetts Share: Out of State Incorporations



(i) Colorado Share: Out of State Incorporations

(j) Florida Share: Out of State Incorporations

Figure 1-6: Trends in Shares of Out of State Incorporations



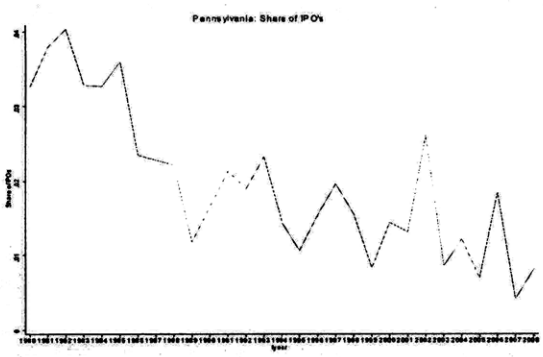
(a) Delaware Share: IPOs

(b) Nevada Share: IPOs

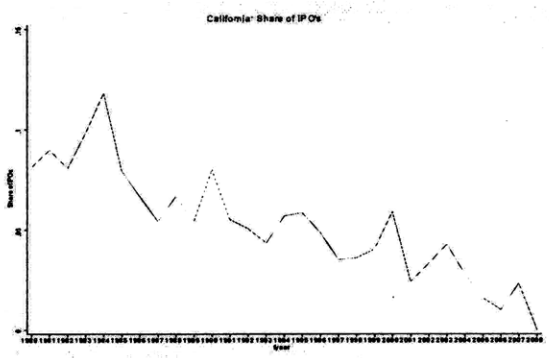
(c) Maryland Share: IPOs

(d) Wyoming Share: IPOs

Figure 1-7: Trends in Shares of IPOs



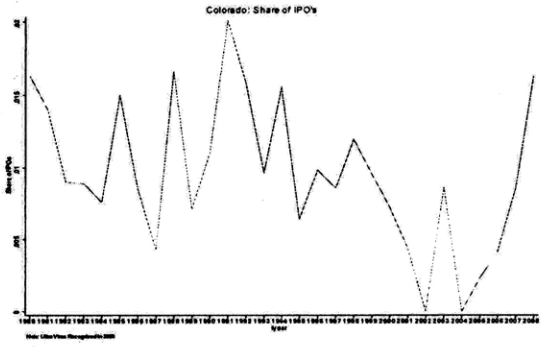
(e) Pennsylvania Share: IPOs



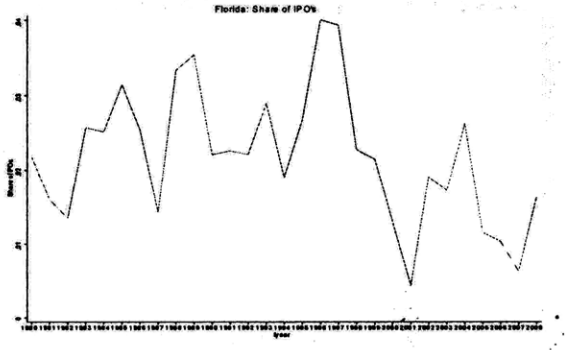
(f) California Share: IPOs

(g) New York Share: IPOs

(h) Massachusetts Share: IPOs



(i) Colorado Share: IPOs



(j) Florida Share: IPOs

Figure 1-7: Trends in Shares of IPOs

1.11 Appendix A1 - Description of State Laws

When possible I quote the definition in GIM. In these cases, the definition will be followed by "(GIM)". Otherwise, these definitions will be culled from a variety of sources including SharkRepellent (which is current and thus from which I also can track any changes in state laws since GIM), the ISS Proxy Voting Manual, and the State Takeover Law Handbook. I separate the laws into those used in the specifications above and other explored, but found not to be significant or to not have sufficient cross section variation.

1.11.1 ATS LAWS

Control Share Acquisition Provisions

These are provisions that "require a majority of disinterested shareholders to vote on whether a newly qualifying large shareholder has voting rights. They were in place in 25 states by September 1990", where 4 states added their provision in 1990 and one more added its provision 1991." (GIM). In Arizona, Florida, Idaho, North Carolina, South Carolina, Tennessee and Washington they apply to out of sate corporations as well. It essentially requires that a bidder obtain shareholder votes or risk not being able to use the acquired stock to obtain control. This is one of the five provisions Bebchuk and Cohen (2003) see to be the central ones to incorporation choice. It is seen to be beneficial even by some apposed to ATS since it protects shareholders against coercive offers, without granting managers the ability to resist bids.

Director Duties

These provisions also termed **Expanded Constituency Provisions**, "allow directors to consider constituencies other than shareholders when considering a merger. These constituencies may include, for example, employees, host communities, or suppliers. This provision provides boards of directors with a legal basis for rejecting a takeover that would

have been beneficial to shareholders. Thirty-one states have Directors' Duties laws allowing similar expansions of constituencies, but in only two of these states (Indiana and Pennsylvania) are the laws explicit that the claims of shareholders should not be held above those of other stakeholders [Pinnell 2000]."

Fair Price Provisions

"Fair-Price provisions limit the range of prices a bidder can pay in two-tier offers. They typically require a bidder to pay all shareholders the highest price paid to any shareholder during a specified period of time before the commencement of a tender offer, and do not apply if the deal is approved by the board of directors or a supermajority of the target's shareholders. The goal of this provision is to prevent pressure on the target's shareholders to tender their shares in the front end of a two-tiered tender offer, and they have the result of making such an acquisition more expensive. Also, 25 states had Fair-Price laws in place in 1990 (with 3 of them passing their laws in 1990, MAC), and two more states passed such laws in 1991." (GIM). They limit the bargaining power of bidders in that the risk for shareholders not tendering in the first round, of obtaining a low price in the second round, is mitigated.

Freeze Out Provisions

Also termed **Business Combination Statutes**. These provisions impose "a moratorium on certain kinds of transactions (e.g., asset sales, mergers) between a large shareholder and the firm, unless the transaction is approved by the Board of Directors. Depending on the state, this moratorium ranges between two and five years after the shareholder's stake passes a pre-specified (minority) threshold." (GIM) In effect the laws limit the benefits of takeovers in that the synergies in the case of mergers or other restructuring cannot take place immediately.

Poison Pill Endorsements

This is a seal of approval given by the state for the use of poison pills (see their definition below). This in effect provides a layer of protection should the pills be challenged. The endorsement does vary by state in its degree. Thus, for example New York and North Carolina are clear that such plans are still subject to judicial review to ensure shareholder interests are considered and that the freedom to use these pills is not unlimited.

1.11.2 MAND LAWS

Cumulative Voting

Six states have mandatory provisions requiring election by cumulative voting (see entry in appendix A2 below). Other states allow the firms to choose.

Limits on Loans to Directors and Officers

Most states permit loans to directors and officers, subject to self dealing constraints. Four states have special rules holding directors personally liable for the loans or have procedural requirements (such as shareholder approval). As mentioned, this is a particularly interesting provision, given that now, post Sarbanes-Oxley, there is a general prohibition on such loans.

Restrictions on Limits to the Personal Liability of Directors

Some systems permit firms to eliminate personal liability of directors for a breach of duty (see entry below). Six states do not permit this, or do so on grounds narrower than Delaware.

Merger Vote Majority Requirements

Seven states require that mergers be approved by two-thirds of the shareholders, and do not permit the company to adopt a lower threshold. Other states require a regular or two-third majority but allow firms to vary the percentage in their certificate of incorporation. And indeed (Kahan 2006) finds most firms with the choice to, opt out.

1.11.3 Payout Restrictions

We use Wald and Long's (2007) coding of the minimum asset to liability ratio required to make payout to shareholders (which clearly affects leverage decisions). These laws are very stable over our time period, but they document them having significant effects on manufacturing firms' location (and reincorporation) choices.

1.11.4 Ultra Vires Recognition

These laws recognize actions taken by the firm (through its agents) as firm action, even when the actor overstepped the boundaries of their position in the firm charter. This imposes more responsibility on the firm over its actors. Most states have generally chosen to recognize such acts. Jurisdictions that do not, offer firms a limitation on the breadth of their liability.

1.11.5 Other Laws Tested but Not Used

Anti-Greenmail Restrictions/Profit Recapture

These restrictions refer to "a transaction between a large shareholder and a company in which the shareholder agrees to sell his stock back to the company, usually at a premium, in exchange for the promise not to seek control of the company for a specified period of time. Antigreenmail provisions prevent such arrangements unless the same repurchase offer is made to all shareholders or approved by a shareholder vote. Such provisions are thought

to discourage accumulation of large blocks of stock because one source of exit for the stake is closed, but the net effect on shareholder wealth is unclear [Shleifer and Vishny 1986; Eckbo 1990]. Five states have specific Antigreenmail laws, and two other states have "recapture of profits" laws, which enable firms to recapture raiders' profits earned in the secondary market. We consider recapture of profits laws to be a version of Antigreenmail laws (albeit a stronger one)... states with Antigreenmail laws tend to pass them in conjunction with laws more clearly designed to prevent takeovers [Pinnell 2000]. Since it seems likely that most firms and states perceive Antigreenmail as a takeover "defense," we treat Antigreenmail like the other defenses and code it as a decrease in shareholder rights." (GIM)

Compensation Restrictions

These laws prohibit the establishment of irregular, compensation increases, during takeover contests, such as new golden parachutes (see below) provisions. However they are of limited efficacy since they generally do not apply to the period before the takeover contest (see Mallette and Spagnola, 1994).

Control Share Cash Out Provisions

These provisions enable "shareholders to sell their stakes to a "controlling" shareholder at a price based on the highest price of recently acquired shares. This works something like fair-price provisions (see above) extended to non-takeover situations. These laws were in place in three states by 1990 with no additions during the decade" (GIM), (or thereafter). Naturally, they discourage takeovers in that they are potentially made more expensive with this option extended to shareholders.

Model Business Corporation Act (MBCA)

This is a complete codex written by the American Bar Association. A number of states have adopted the MBCA and so there clearly are network benefits from sharing its laws.

However, four of the five largest states and Delaware have not adopted it. There is also considerable variation in the time the MBCA was adopted (although largely before the 90s). It is therefore more a proxy for a form of legal network effects and less an indicator for having particular laws. Naturally, there are many alternatives to adopting the MBCA. Non-adopting states deviate from its provisions in a variety of ways. It is thus a difficult "law" to analyze.

Severance Pay and Labor Contract Provisions

These provisions are used to by states to protect employees in the event of a takeover. Severance pay ensures severance to the employees, while the labor contract provisions requires that, post-takeover, no collective bargaining or employment contract be terminated without the explicit consent of all parties to the contract. Massachusetts and Rhode Island had both provisions, although the former has been invalidated by federal courts in that it is preempted by the Employee Retirement Income Security Act (ERISA). Pennsylvania still has both, while Delaware and Illinois only have the latter. (See *Simas v. Quaker Fabric Corp. of Fall River*, 6 F.3d 849 (1st Cir. 1993) and *United Paperworks International Union Local* 1468, *et al, v. Imperial Home Decor Group*, 1999 WL 1115761 (D.R.I. 1999). I thus counted all state with at least one of the two as having this type of protection.

1.12 Appendix A2 - Description of Bylaw Characteristics

1.12.1 Blank Check Preferred Stock

"Stock over which the board of directors has broad authority to determine voting, dividend, conversion, and other rights. While it can be used to enable a company to meet changing financial needs, its most important use is to implement poison pills or to prevent takeover by placing this stock with friendly investors. Because of this role, blank check preferred stock is a crucial part of a "delay" strategy. Companies that have this type of preferred

stock but require shareholder approval before it can be used as a takeover defense are not coded as having this provision in our [the IRRC] data. This stock, when issued, gives directors the broad authority to establish voting, dividend, conversion and other rights. This flexibility is good for facilitating the adjustment to changing financial conditions. However it also grants the authority to issue stock necessary for the implementation of defenses, including anti-takeover defenses."(GIM)

1.12.2 Bylaw and Charter Amendment Limitations

These limit "shareholders' ability to amend the governing documents of the corporation. This might take the form of a supermajority vote requirement for charter or bylaw amendments, total elimination of the ability of shareholders to amend the bylaws, or the ability of directors (beyond the provisions of state law) to amend the bylaws without shareholder approval." (GIM)

1.12.3 Classified Board

"A Classified Board (or "staggered" board) is one in which the directors are placed into different classes and serve overlapping terms. Since only part of the board can be replaced each year, an outsider who gains control of a corporation may have to wait a few years before being able to gain control of the board. This slow replacement makes a classified board a crucial component of the Delay group of provisions, and one of the few provisions that clearly retains some deterrent value in modern takeover battles [Daines and Klausner 2001]."(GIM) This provision is mandatory in Massachusetts (unless opted out by the directors), as of 1990, and enabling in other states.

1.12.4 Common Stock Redemption Rights

They are similar to poison pills in that they allow for the sale of stock back to the firm at a premium price, if another shareholder acquires a significant share of the stock through a

tender offer not approved by the board of directors. This in effect threatens the dilution of the firm value by distributing the company's assets directly to the shareholders before control of the company is surrendered, and thus is another anti-takeover provision.

1.12.5 Compensation Plans With Changes in Control Provisions

These provisions allow "participants in incentive bonus plans to cash out options or accelerate the payout of bonuses if there should be a change in control. The details may be a written part of the compensation agreement, or discretion may be given to the compensation committee." (GIM)

1.12.6 Cumulative Voting

These provisions allow "a shareholder to allocate his total votes in any manner desired, where the total number of votes is the product of the number of shares owned and the number of directors to be elected. By allowing them to concentrate their votes, this practice helps minority shareholders to elect directors." (GIM) These are seen to increase shareholder rights. They grant the minority more power to be represented, in that they can focus on electing at least some of the directors. These provisions can also be made contingent on there being a substantial shareholder.

1.12.7 Director Indemnification Contracts

These are contracts between the company and particular officers and directors, using the bylaws, charter, or both, indemnifying them from certain legal expenses and judgments resulting from lawsuits pertaining to their conduct. Some firms have both "Indemnification" in their bylaws or charter and these additional indemnification "Contracts"... "The cost of such protection can be used as a market measure of the quality of corporate governance [Core 1997, 2000]."" (GIM)

1.12.8 Dual Class Common Stock

Dual Class Common Stock attaches more or less voting power to different shares. Thus allowing for a disproportionate amount of control to be put in the hands of those holding these preferred shares.

1.12.9 Golden Parachutes

"Golden Parachutes are severance agreements that provide cash and non-cash compensation to senior executives upon an event such as termination, demotion, or resignation following a change in control. They do not require shareholder approval. While such payments would appear to deter takeovers by increasing their costs, one could argue that these parachutes also ease the passage of mergers through contractual compensation to the managers of the target company [Lambert and Larcker 1985]. While the net impact on managerial entrenchment and shareholder wealth is ambiguous, the more important effect is the clear decrease in shareholder rights. In this case, the "right" is the ability of a controlling shareholder to fire management without incurring an additional cost. Golden Parachutes are highly correlated with all the other takeover defenses."(GIM) GIM treat these as restrictions of shareholder rights.

1.12.10 Limitations on Director Liability

"Limitations on director Liability are charter amendments that limit directors' personal liability to the extent allowed by state law. They often eliminate personal liability for breaches of the duty of care, but not for breaches of the duty of loyalty or for acts of intentional misconduct or knowing violation of the law."(GIM)

1.12.11 Pension Parachutes

"Pension Parachutes prevent an acquirer from using surplus cash in the pension fund of the target to finance an acquisition. Surplus funds are required to remain the property of

the pension fund and to be used for plan participants' benefits." (GIM)

1.12.12 Poison Pills

"Poison Pills provide their holders with special rights in the case of a triggering event such as a hostile takeover bid. If a deal is approved by the board of directors, the poison pill can be revoked, but if the deal is not approved and the bidder proceeds, the pill is triggered. Typical poison pills give the holders of the target's stock, other than the bidder, the right to purchase stock in the target or the bidder's company at a steep discount, making the target unattractive or diluting the acquirer's voting power. Poison pills are a crucial component of the "delay" strategy at the core of modern defensive tactics."(GIM)

1.12.13 Secret Ballot

"Under a Secret Ballot (also called confidential voting), either an independent third party or employees sworn to secrecy are used to count proxy votes, and the management usually agrees not to look at individual proxy cards. This can help eliminate potential conflicts of interest for fiduciaries voting shares on behalf of others, and can reduce pressure by management on shareholder-employees or shareholder-partners."(GIM) GIM see the inclusion of this provision in by-laws as increasing shareholder rights.

1.12.14 Severance Agreements

"Executive Severance agreements assure high-level executives of their positions or some compensation and are not contingent upon a change in control (unlike Golden or Silver Parachutes)."(GIM)

1.12.15 Silver Parachutes

"Silver Parachutes are similar to Golden Parachutes in that they provide severance payments upon a change in corporate control, but differ in that a large number of a firm's

employees are eligible for these benefits."(GIM). These do not protect key decision makers in a merger, and are thus potentially just an expression of the power of management. They are classified by GIM in the "Other" group rather than in the "Protection" group.

1.12.16 Special Meeting Limitations

"Special Meeting limitations either increase the level of shareholder support required to call a special meeting beyond that specified by state law or eliminate the ability to call one entirely. Such provisions add extra time to proxy fights, since bidders must wait until the regularly scheduled annual meeting to replace board members or dismantle takeover defenses. This delay is especially potent when combined with limitations on actions by written consent."(GIM)

1.12.17 Supermajority Requirements for Merger Approvals

"Supermajority requirements for approval of mergers are charter provisions that establish voting requirements for mergers or other business combinations that are higher than the threshold requirements of state law. They are typically 66.7, 75, or 85 percent, and often exceed attendance at the annual meeting."(GIM) Note that the state laws (see above) are not uniform on this issue. These provisions are similar to Control-Share Acquisition Laws defined above.

1.12.18 Unequal Voting Rights

"Unequal Voting rights limit the voting rights of some shareholders and expand those of others. Under time-phased voting, shareholders who have held the stock for a given period of time are given more votes per share than recent purchasers. Another variety is the **substantial-shareholder provision**, which limits the voting power of shareholders who have exceeded a certain threshold of ownership." (GIM)

1.12.19 Written Consent Limitations

"Limitations on action by Written Consent can take the form of the establishment of majority thresholds beyond the level of state law, the requirement of unanimous consent, or the elimination of the right to take action by written consent. Such requirements add extra time to many proxy fights, since bidders must wait until the regularly scheduled annual meeting to replace board members or dismantle takeover defenses. This delay is especially potent when combined with limitations for calling special meetings (see above)." (GIM)

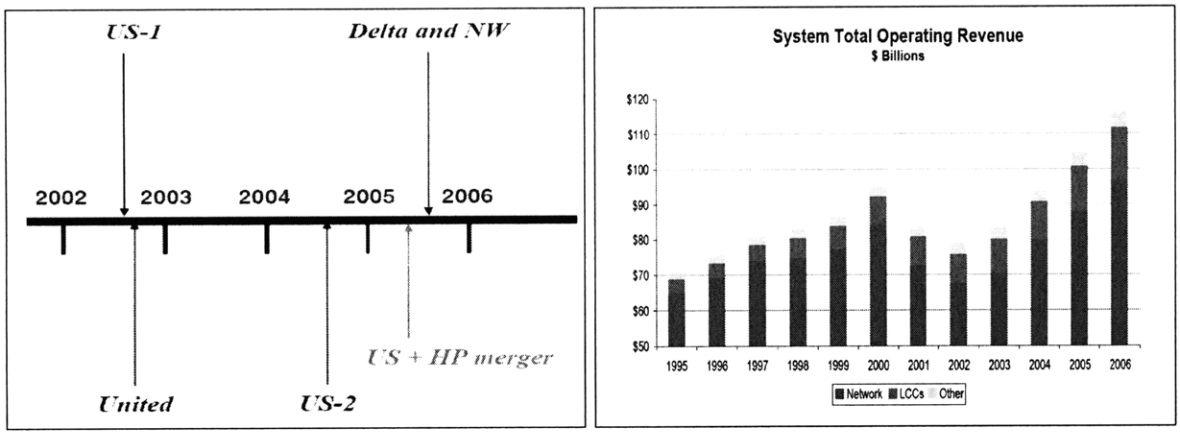
Chapter 2

Estimation of A Dynamic Oligopoly Entry Game in the US Airline Industry: Hubs, and LCC

2.1 Introduction

The recovery and analysis of airline profits and their determinants have long been elusive. Airlines, by and large, have lost money since the invention of the first planes by the Wright brothers in 1903. As shown in figure 1(a), several of the large legacy carriers have been in and out of Chapter 11 in recent years (compare Borenstein and Rose 2007). However, the number of passengers has increased steadily over this time period[1], and as shown in figure 1(b) the operating revenues in the industry have risen significantly as well (following the post September 11th decline). In this paper I develop and estimate a dynamic model of airline competition, with entry and exit, and recover the supply and demand sides of the domestic US airline market. The model exploits the information

[1]The number of domestic passengers has gone from 551M in 2002, to 679M in 2007. There is some decline in 2008 to under 650M.



(a) Recent Bankruptcies

(b) Operating Revenues

Figure 2-1: Recent Trends

embedded in the decisions of consumers in choosing between the many airline product offerings in each quarter, regarding consumer preferences, demand, and marginal costs, as well as the information embedded in the quarterly decisions of entry and exit, which are assumed to reveal each airline's belief of the (expected present discounted) profitability of the markets it chooses to serve.

As is common in the dynamic game literature, this paper makes the simplifying assumptions of airlines maximizing the profits from each market separately and not taking into account the added benefits to the entire network; of the transitions between states following a first order Markov process where the payoff relevant variables are only the market specific variables; of the individual airline transition probabilities being independent conditional on the state space; and while the model can have multiple equilibria, this paper assumes that the data is generated by one (and only one) of them. It also specifies a relatively simple nested logit demand model over the quarterly tickets sales, which is used as an input to the dynamic game. However, this paper extends previous applied work on dynamic games by allowing for firm identities to matter, while accounting for the impact this has on the size of the state space by structuring the model in a manner which

facilitates the use of state-of-the-art mathematical solvers. This allows for the exploration and exploitation of the highest level of richness in airline costs and profits possible under the current computational optimization technology.

The recovery of the costs of serving the heterogenous US markets - and consequently a better understanding of the profit structure - allows for an analysis of many of the key questions that have been at the focus of the vast airline literature both within and outside of economics. This paper addresses two of them: The desirability of hub networks, and the impact of the increased presence of low cost carriers (LCC) on the costs and profits of incumbent airlines, which I detail below:

2.1.1 The Case for Hubs

Following the deregulation of airlines in the late 1970s, many of the "legacy carriers" chose to concentrate a large portion of their operations in certain airports (the "hubs") and to connect the other cities served (the "spokes") to these hubs by non stop flights. Figure 2 shows the distribution of hubs across the US. The rationale for hubs is that there are significant benefits or returns to scale from having a large presence in the hub airport that outweigh, in certain situations, the additional costs (and inconvenience to passengers) of having the many additional connecting flights, and travelling larger distances when serving two spoke end points (compare Hendricks, Piccione and Tan, 1995) . Hubs facilitate the use of larger planes which reduce the cost per passenger and allow for a reduction in the number of direct connections. They also allow for economies of scope in having a concentration of manpower in the hub, and may also allow for more bargaining power with the airports. They have been claimed to also be attractive to consumers since they offer more variety and frequency of flights (at the hub), and more expertise[2]. There are thus potentially both supply (cost savings) and demand (revenue increases) advantages to hubs. Finally, there

[2]Separating variety and frequency is difficult since both measures, but frequency in particular, are largely demand driven.

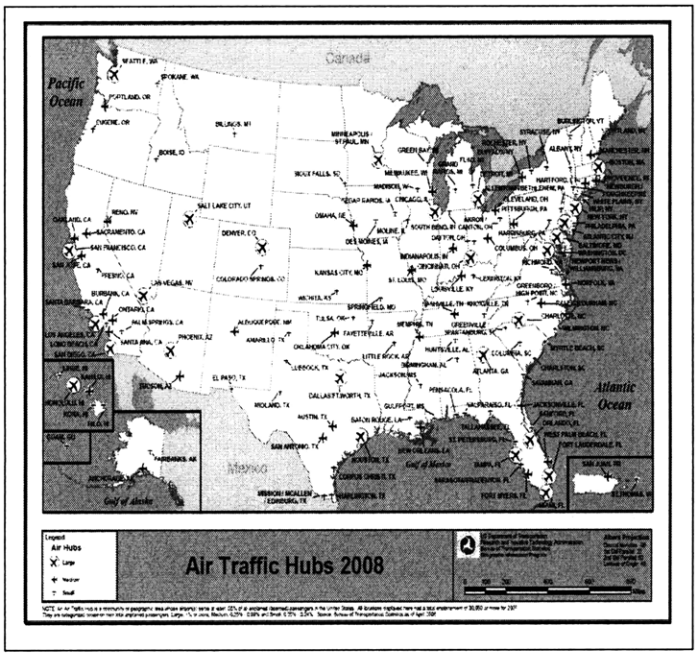


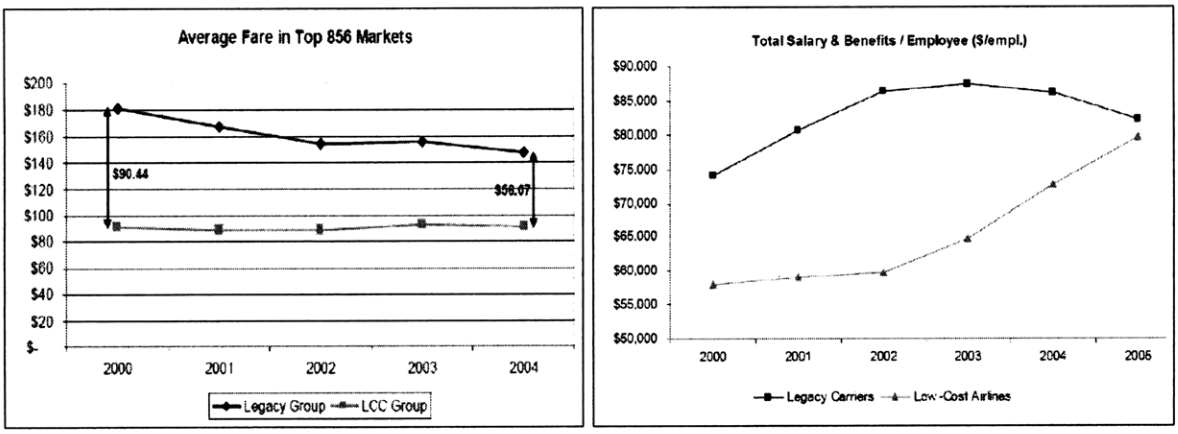
Figure 2-2: Hubs in 2008

are claims that hubs serve as an entry deterrent, given the complementarities in profits between different routes (any city added connects the entire network to that city).

Given the many puzzles surrounding the airline industry, these claims require empirical support, by examining the determinants of profits in general and fixed costs in particular. I find that hub carriers have higher variable profits that their competitors in their own hub markets but lower than average profits in non-hub markets. Consumers prefer to travel with the hub carrier in hub markets and have a significant distaste for flying with other carriers in these markets. They also prefer to travel with airlines that have more destinations from the origin, or more flexibility. However, the static game estimates further reveal a distaste for connecting flights, or a preference for nonstop flights, which can offset the preference for using the hub airlines. Furthermore, the profits garnered by carriers in their hub markets are significantly reduced when including fixed costs and they do not increase the costs of entry for other carriers. These estimates further question the desirability and overall profitability of the hub structure, even in the hub markets themselves.

2.1.2 The Impact of Low Cost Carriers (LCC) on Rival Costs and Profits

The **second** question analyzed is the heterogenous effects of the different carriers on their rivals. There is a natural interaction between the six legacy carriers. However, a key feature has been the growing market share of low cost carriers, leader amongst which is Southwest Airlines, followed, more recently, by JetBlue. The differing products offered by these carriers induce a response by the actual or potential other players in the market. There seems to be a convergence in operating procedures (such as less food on flights), and as figure 2 shows, a convergence in fares and labor costs as well. Previous work, including Berry, 1992, and, more recently, Ciliberto and Tamer 2009 examined the effects of firm heterogeneity on entry into airline markets in a static framework. This paper extends this work to a dynamic framework, allowing for firm identities, and thus for heterogeneous



(a) Fares (b) Salaries

Figure 2-3: Convergence in Operating Procedures: Legacy and LCC

effects of the legacy carriers and low cost carriers on the profits of their rivals, accounting for the heterogeneity in markets. The interactions between players are decomposed into their differing effects on rival variable profits and marginal costs (both part of a static pricing game), as well as on the costs of entry and the fixed operating costs. I find variable profits to generally be higher for the LCC. Consumers have a significant preference for these airlines in general (accounting for observable features of the products offered), but especially for Southwest and JetBlue, upon which I therefore focus. This likely reflects some unobservable features of the services provided (or the characteristics of their products). Marginal costs are also significantly lower for these airlines, suggesting that they are doing something better on the cost side. Furthermore, there are large strategic effects between airlines, but the (negative) impact of the presence of LCC on the profits of legacy competitors is most pronounced. The preliminary results for the entry costs and fixed costs suggest that there too there are significant strategic effects, where the presence of all major airlines increases the costs associated with entry and the fixed costs, but the LCC play a special role.

The paper is organized as follows: in section 2 I briefly survey the relevant literature.

Section 3 discusses the datasets used and the construction of the sample. Section 4 sets up the general framework and the model. Section 5 then discusses the estimation, separating the variable profits and marginal costs recovered from a static pricing game, and the entry and fixed cost parameters recovered from the dynamic game. Section 6 collects and discusses the results. Finally, section 7 concludes and suggests some of the many extensions and future work that can be done using the framework in the paper.

2.2 Related Literature

As to the methodology, there are a number of dynamic game applications using one of two methods. The first is a simulation of moment inequalities approach, where the value functions are simulated forward using policy functions and transition probabilities estimated from the data, and the estimates are those under which value functions of the policies chosen are greater than the value functions from any alternative policies (as developed by BBL 2007). The second is a maximum likelihood with best response equality constraints approach (as developed by AM 2007), which will be discussed at greater length below. Given the complexity of these methods, previous applied work, including Ryan 2009 (cement, BBL), Collard-Wexler 2005 (ready-mix concrete, AM), Sweeting 2007 (radio, BBL), Beresteanu and Ellickson 2006 (supermarket products, BBL) and Macieira 2006 (supercomputers, BBL), generally assumes symmetric equilibria. This is a difficult assumption to justify in most settings. This paper is closer to the AM approach, but allows for individual airline heterogeneity and a state space visible to all players which transitions consistent with the equilibrium probabilities thus allowing for the treatment of the effect of the airlines' own and rival market features (such as hubs), by exploiting state-of-the-art optimization techniques.

The application, the study of airlines, has received much treatment in a voluminous literature spread across many fields, for which there are now many useful surveys (see for example Borenstein and Rose 2007 and 2008). I briefly sample that most relevant to the

analysis here.

For entry into airlines markets, Berry (1992) which builds on Bresnahan and Reiss (1991) analyses entry as a static game of discrete choice. The profit function is restricted to ensure the uniqueness of the number of players, by assuming firm characteristics only affect the fixed costs, and a symmetric post entry game. Ciliberto and Tamer (2009) use a similar framework, but rather than requiring a unique number of entrants (and restricting the profit function to ensure this), they consider multiple equilibria and allow for a different number of entrants and different selection mechanisms in different markets. Their estimation builds upon the set estimation procedures in Chernozhukov Hong and Tamer (2007), and thus they allow for multiple equilibria within a set constrained by the requirement of airlines earning positive profits in markets. The identified estimated parameters are those for which there is a selection function such that the predicted choice probabilities in the model match the empirical choice probabilities in the data. Both papers find heterogeneity in the manner in which competitors' profits are affected by the presence of their rivals and an important role played by airport presence. This paper is similar in its identifying assumption of airlines operating in city pair markets when they produce positive profits, however rather than looking at static single period profits I require the difference between the value function of operating and of not operating in the period to be positive. The dynamic framework relies on the assumption that the equilibrium estimated is the one most likely given the data. This framework allows for an estimation of fixed and entry costs from the moments of entry and exit, and for the study of both the strategic effects airlines impose on one another and the effects of hubs. However, it requires a full specification of the state space today and in all possible future periods and of the transitions to and from all elements of the state space.

Aguirregabiria and Ho (2009) analyze hubs in a dynamic framework as well, and as such are closer to this paper. Their markets are city pairs, but their incumbency in a market is defined by operating non-stop flights in the market. The analysis is restricted to 2004.

The restrictions they impose on the state space require binning all states with four or more incumbents together and discretizing the variable profits to a grid of 11 points. They also do not include the characteristics[3] and identity of rival airlines in the state space observed by each airline. Their paper thus focuses on the effect of hubs on airlines' own profits. They find much lower fixed costs, entry costs and variable profits than those found here. In contrast, this paper looks at a larger time period and specifically includes the identity and characteristics of all incumbent airlines, which, in the dynamic model treats all 6 legacy carriers separately and bins the low cost carriers together. Equilibrium transition probabilities are estimated to and from all elements of the state space. This allows for the study of the effects of airline characteristics - most notably hubs - and identities - such as being a low cost carrier - on own and rival profits and costs.

Finally, in preliminary work, motivated by the BBL approach, Benkard, Bodoh-Creed and Lazarev (2008) estimate and project simple probit entry probability functions to simulate the effect of mergers. This paper estimates similar activity probabilities, but these are used as initial values in the search for activity probabilities which represent a MPE of the full dynamic model.

For the demand side, Berry, Carnall and Spiller (2006), followed by Berry and Jia (2008) use a characteristic based model of demand, which is a simplification of the now canonical BLP (1995) framework to a bimodal distribution of tastes, to estimate (variable) costs and markups, defining products as unique combinations of airline-fare-itineraries. This paper specifies a pricing game at the quarterly level - consistent with the data driven time periods in the dynamic model - and thus uses a simple nested logit to estimate demand, which is then projected on the state space. The demand model used here is closer to that used by Peters (2006), who looked at data from 1985 and found static demand models to not predict post merger prices well. There are also a host of reduced form studies. For example, Borenstein (1989, 1991) finds that flights on airlines with hubs at end points

[3] They have a measure of the "mean value of hub size for the incumbents".

command higher prices. However, recently (Borenstein 2005) he finds these premiums to have declined. Goolsbee and Syverson (2008) finds preemptive price cuts in expectation of Southwest entry (which is generally into markets in which there is a Southwest presence at one of the endpoints). However, are but a few of the many studies focusing the importance of hubs and the effects of LCC[4].

2.3 Data Construction

The main datasets used are two of the three datasets (merged by ticket id) from the Origin and Destination Survey (DB1B, hereinafter "the survey"), which is a 10% random sample of all domestic US tickets aggregated up to the quarter and the aggregate information in the T-100. These are public and commonly used (for example Ciliberto and Tamer (2009), Berry, 1992, and Borenstein, 1989). I use the 20 quarters from $2002-2006$[5]. The Coupon dataset has coupon specific information for each domestic itinerary in the survey, including the operating carrier, number of coupons, origin and destination airports, trip break code, number of passengers, fare class, and distance. Each coupon is a separate observation and represents a city pair trip (these may be pieces of the same itinerary). An itinerary is the entire trip and may contain many coupons (a round-trip contains at least 2). The Ticket dataset has the number of coupons, the origin airport, round-trip indicator, reporting carrier, a credibility indicator, the itinerary fare, the number of passengers, and distance and miles flown[6]. These are merged (by operating carrier) with the T-100 Domestic Segment Dataset, which includes all (100% of the data rather than just a sample) of domestic market data by air carriers, and origin and destination airports for passengers enplaned, including

[4] See, for example, Borenstein and Rose (2007,2008) for more.

[5] Future versions of the paper will examine the effect of dropping 2002 which following September 11th, was an atypical year.

[6] The third dataset, the Market dataset, has directional market characteristics of each domestic itinerary in the survey, with a seperate observation for each market (defined as an airport pair), in the itinerary. It is not used due to inconsistencies in the market definition.

load factors, number of passengers and flights, etc.[7]. Tickets not in the T-100, or that are not provided on a regular basis (at least once a week) are dropped.

For the entry decisions, a market is a (nondirectional) city pair. Airlines decide which cities to connect and in doing so are "in the market" for itineraries involving both cities as both origins and destinations. On the demand side, the products are the tickets sold based on the origin and destinations of consumers (compare Aguirregabiria and Ho 2009 that look at non-stop itineraries). The numbers of stops are a product characteristic. Time periods are a quarter as dictated by the data. I think of the airlines as supplying these products in different ways: some with more direct connections, and some with complex hub structures. The effective seller is the ticketing carrier[8].

This data has many dimensions and its reliability is not perfect. Accordingly, following the previous airline literature using this data, the sample is reduced by dropping tickets with more than two stops, multiple ticket carriers (per directional trip)[9], credibility questioned by the Department of Transportation, segments of international trips or non-contiguous domestic travel with Hawaii, Alaska and Territories; less than 120 passengers per quarter, and particularly high (over 2000 dollars) prices[10]. I keep all classes of tickets, including one-way tickets since the objective is to determine the total profitability of the route. Airports in the same MSA are joined (to reflect the competition induced by the multiple airports in a given city), and the size of the market is seen as the geometric mean of the population of the endpoint cities[11].

Airports commonly seen as hubs are coded as such. I thus have the following cities as

[7] There is also a T-100 Market database, which again has inconsistent definitions of a market and thus is not used. For example, a carrier change is defined as serving a different market.

[8] The ticketing carrier sells the tickets. The operating carrier is determined by the airlines (not necessarily the owner of the plane or in any other fixed definition). The reporting carrier is seen as pretty meaningless by the DOT.

[9] This represents less that one percent of the data.

[10] These tickets are dropped due to suspected reporting error. In further versions I plan to explore the impact of this cut.

[11] Data were available at http://www.census.gov/popest/metro/CBSA-est2006-annual.html. Note that I only looked at markets that had at least one ticket carrier at some point in the sample.

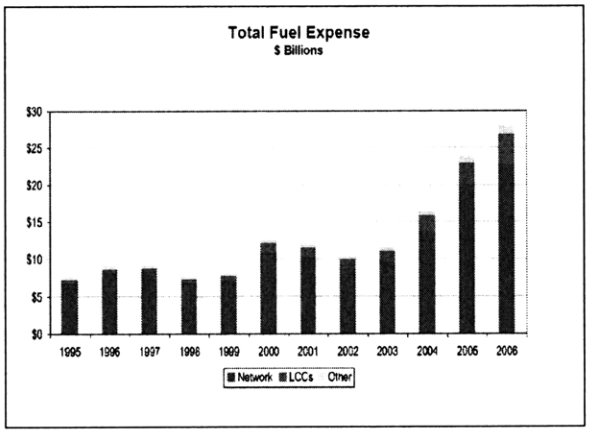


Figure 2-4: Spending on Fuel

hubs: Atlanta (Delta), Chicago (American, United), Charlotte (US), Cincinnati (Delta), Dallas (American), Denver (United), Detroit (Northwest), Houston (Continental), Memphis (Northwest), Minneapolis (Northwest), Philadelphia (US), Pittsburgh (US), Salt Lake City (Delta).

This data is also merged with a time series of jet fuel prices from the United States Department of Energy's Energy information Administration (to be used as cost shifters), aggregated to the quarterly level. As can be seen in figure 4 the spending on fuel has increased dramatically. Airlines explicitly cite fuel costs as a reason for the increase in prices and these prices do indeed work well at explaining ticket prices, as discussed below.

2.4 General Framework

I adapt the general structure proposed by AM 2007, which is amenable to the use of the computational techniques I employ. To set notation, assume there are N airlines $i \in I = \{1, 2...N\}$, potentially operating in M markets, where markets are combinations of the D different US cities[12]. These markets are not directional in that we assume that for

[12]More precisely, these are metropolitan statistical area combinations, as I will explain below.

the LA-Boston combination, for example, airlines in the market sell (one way and round-trip) tickets originating and culminating in both cities. Thus for D origin and destination cities there are $M = D(D-1)/2$ markets. Airline choice sets have to be made to both reflect the data limitations, as well as to provide a tractable framework for the dynamic analysis. We assume the following timing for airlines decisions, where each quarter is a time period:

2.4.1 Timing

1. Each airline observes the state space at the beginning of the period (which is determined by the actions of the previous period).

2. Airlines observe their private productivity shock.

3. Airlines choose an action $a_{imt} \in \{0, 1\}$, where a 1 corresponds to being active in the market. This choice, of course, may require entry or exit, depending on the state in the previous period. Airlines know whether they enter or exit in the period, but form expectations over which of the other airlines will be in the market, given the state space of the previous period. Based on these beliefs airlines choose capacity for the market and the characteristics of the tickets that they offer, and play a price-competition game with the other airlines that chose to be in the market for the period. These latter choices will not be modeled but rather will be assumed to shape the variable profits which we estimate as a function of the evolving state space, which ensues from the entry and exit decisions.

4. Consumers choose the ticket with the characteristics that maximize their utility. Airline payoffs (the sum of the prices they collected for the tickets sold) are realized.

2.4.2 State Space

The state space is driven by both data limitation and the feasibility of the computation. The structure of this problem - the study of hubs and the heterogeneous effects of competition - requires treating markets and airlines heterogeneously. However, many of the market level variables do not evolve and so the state space can be made relatively rich by allowing for one market specific state. Markets include the following variables (which remain fixed in the estimation[13]): hub variables, the nonstop distance between cities, the size of the market, the density of passengers in the market (taken from the first quarter of 2002 and held fixed throughout the sample) and whether the market is a tourist market. The variables that evolve are the number (and identity) of the incumbents in the market[14].

Given that identities matter, there are 2^N states associated with N players in a given market. There will also be N different value functions for each element of the state space in each market. The analysis of hubs requires including all 6 legacy carriers: American Airlines, United Airlines, Continental Airlines, US Airways, Delta and Northwest Airlines. The analysis of the effects of LCC requires having at least one more player. I thus chose the most parsimonious player state space with $N = 7$. Given the prominence of Southwest Airlines and, more recently, JetBlue, the seventh player is either one of these two carriers.

Following previous work, and particularly CT (2009), I order markets by the geometric mean of the city populations. I begin with all markets between the 50 largest cities (and show below that these are not a bad approximation to all US markets). This gives me a total of 1225 markets. I further eliminate 14 markets between cities that are very close geographically, leaving me with a total of 1211 markets.

As discussed below, I am thus left with $1211 \cdot 2^7 = 155,008$ states. This represents the

[13] This assumption simply implies that the airlines *beliefs* are such that they do not expect an evolution, on average, in the future.

[14] In future verisons I plan to include the sum of the number of destinations flown by the carrier from the two connecting cities which is a deterministic function of the airlines in the market (assuming we limit it to the select number of markets).

richness (and computational burden) of the model.

2.4.3 Profit function

Airlines' per period profit function from all markets is:

$$\Pi_{it}(a_{it}, s_t, \varepsilon_{it}) = \sum_m \pi_{im}(s_t) - \sum_m a_{im} C_{im}(s_t, \varepsilon_{imt})$$

where π_{im} is the variable profits from each market, and $C_{im}(.)$ are the fixed and entry costs incurred by serving the market:

$$C_{im}(s_{mt}, \varepsilon_{imt}) = FC_{imt} + \varepsilon_{imt} + (1 - a_{imt-1}) EC_{imt}.$$

Note that given the timing assumption, the state space at time t, s_{mt}, represents the identity and number of firms that were in the market at time $t-1$. a_{imt} is the action taken at time t. Simple put, airlines incur an operational fixed cost if they are active, and an additional cost for entering the market (i.e. becoming active after a quarter in which they were inactive)[15]. We can think of the payoff from not being in the market as $\mu_i + \varepsilon_{it}(0)$ but since this payoff is not separately identified from the fixed cost we redefine fixed costs as net of this opportunity cost. Exit costs can also not be separately identified, since identification comes from the (two) moments of entry and exit. The structural fixed and entry cost parameters are modeled as:

$$\begin{aligned} FC_{imt}(\boldsymbol{\gamma}) &= \gamma_1 + \gamma_{2i} + \gamma_3 X_m + \gamma_4 MyHUBS_{im} + \gamma_5 HUBS_m + \gamma_6 Legacies_i + \gamma_7 LCC_i \\ EC_{imt}(\boldsymbol{\varphi}) &= \varphi_1 + \varphi_{2i} + \varphi_3 X_m + \varphi_4 MyHUBS_{im} + \varphi_5 HUBS_m + \varphi_6 Legacies_i + \varphi_7 LCC_i. \end{aligned}$$

[15] Natually, other definitions of inactivity could be used, exploring the possibility of seasonality in the service of some markets, etc. For entry I require an airline to active for at least two consecutive periods, following inactivity. Similarly, for exit, I require an airline to be inactive for at least two periods following activity. This is expected to alleviate some of the possible errors in the data collection, which results from an (imperfect) 10% survey.

The costs are decomposed into a constant and carrier specific fixed effects, X_m - market level variables (which I begin by having as just the distance between cities), $MyHUBS$, the number of cities which are the airline's hub, $HUBS$, the number of hubs for other carriers[16], *Legacies* - the number of (other) legacy carriers in the market, and LCC - whether a low cost carrier (here Southwest or JetBlue) is in the market (other than the airline itself, and so this is zero for the LCCs).

Finally, $\varepsilon_{it} = \{\varepsilon_{imt} : m = 1, 2...M\}$ are the private information idiosyncratic shocks incurred by each airline in each market m. We assume they are i.i.d. over airlines, markets and time[17] with an extreme value CDF G_ε.

2.4.4 The Dynamic Entry Game

This game has the standard markov-structure: Airlines maximize the expected present discounted value of profits, taking into account all payoff relevant variables. Denote the strategy functions by $\sigma = \{\sigma_i(s_t, \varepsilon_i), i \in I\}$. This gives a value function for each airline i over the states:

$$V_i^\sigma(s_t, \varepsilon_{it}) = \max_{a_{it}} \{\Pi_{it}(a_{it}, s_t, \varepsilon_{it}) + \beta E[V_i^\sigma(s_{t+1}, \varepsilon_{it+1})|s_t, a_{it}]\}$$

which takes as given the strategies of the other airlines (belonging to σ) and chooses a_{it} as a best response - maximizing the expected discounted profits. The MPE (markov perfect equilibrium) implies that: $\forall \sigma_i \in \sigma$ we have:

$$\sigma_i(s_t, \varepsilon_{it}) = \arg\max_{a_{it}} \{\Pi_{it}(a_{it}, s_t, \varepsilon_{it}) + \beta E[V_i^\sigma(s_{t+1}, \varepsilon_{it+1})|s_t, a_{it}]\}$$

[16] I chose to count up all hubs for other carriers to allow for a difference between having one and more than one carrier with a hub in a given city.

[17] This assumption, which is common in the dynamic game literature, may be a strong one. It is possible for certain markets or airline-market combinations to have serially correlated shocks. Such would be the case if airlines were reluctant to exit certain markets even if they were unprofitable (for example due to network benefits not captured in the model). I plan to explore this important extension in future work.

or that all airlines are best responding to each other. As in the AM framework, players' strategies depend on one another only through the conditional choice probabilities, i.e. the probabilities that airlines choose a_{imt} given the state space. These integrate the strategy functions over the private information shocks.

$$P_i^{\sigma}(a_i|s) \equiv \Pr(\sigma_i(s,\varepsilon_i) = a_i) = \int I\{\sigma_i(s,\varepsilon_i) = a_i\} g_i(\varepsilon_i) d\varepsilon_i.$$

This gives the equilibrium condition in terms of probabilities, which essentially will form the constraints in the dynamic optimization problem. In order for the conditional choice probabilities to represent an MPE they must satisfy:

$$P_i^{\sigma}(a_i|s) = \int I\{\sigma_i(s_t,\varepsilon_{it}) = \arg\max_{a_{it}}\{\Pi_{it}(a_{it},s_t,\varepsilon_{it}) + \beta E[V_i^{\sigma}(s_{t+1},\varepsilon_{it+1})|s_t,a_{it}]\}\} dG_{\varepsilon}(\varepsilon_{it}).$$

Equilibrium existence follows the proofs in AM and Doraszelski and Satterthwaite (2007) for any absolutely continuous (with respect to the Lebesgue measure) density function.

Airlines can and should jointly optimize their entire network. However, for tractability we make some simplifying assumptions:

First, for simplicity we treat each market separately. We assume that each regional airline manager maximizes the expected present discounted profits from each market and does not consider the private shocks or decisions that the airline makes regarding other markets. In other words, although there is some commonality in the matter in which consumers respond to airlines (for example through the fixed effect or brand effect and through the total number of destination served), and entry into one market may now enable consumers to fly between other cities that become connected in the airline's network, airline managers do not take this into account.

Second, we simplify the structure of the transition probabilities. We assume a first

order markov process, where:

$$\Pr(s_{mt+1}|s_{mt}, a_{imt}, s_t; P_{-im}) = \Pr(s_{mt+1}|s_{mt,}a_{imt}; P_{-im}).$$

In other words, rather than considering the state space of all markets considered (including the airline's own state in these other markets) the payoff relevant variables are the market specific state variables. This assumption extends the previous one in that airlines do not consider the entire state space for all markets even with regards to the profits in their own market. Thus, airlines consider how the this being a hub market for themselves and for their competitors affects profits in the market, but not how the profits from being in this market are affected by the identity and characteristics of the airlines in related markets. For example, serving both LA-Boston and LA-NY may be something valued by consumers in adding to the flexibility the airline offers, and thus the profitability of entering LA-Boston may depend on whether the airline is in LA-NY, but here we assume the airline does not consider this. The conditional transition probability is thus assumed to be independent of the state space in other markets (which is the assumption we relied on above in specifying the size of the state space).

Third, following what is standard in this literature, we assume that the individual probabilities are independent conditional on the state space and so:

$$\Pr(s_{mt+1}|s_{mt,}) = \prod_{i=1}^{N} \Pr_i(a_{imt}|s_{mt}).$$

Fourth, while the model may have multiple equilibria, we assume the data are generated by one MPE, which players expect to be played into the future.

These assumptions allow for a redefinition of the equilibrium (note the addition of m subscripts), where an airline chooses to be active depending on the value function from each market iff:

$$V_{imt}(s_{mt}, 1) - V_{imt}(s_{mt}, 0) \geq 0$$

$$\Longleftrightarrow$$

$$E[\Pi_{imt}(a_{imt}, s_{mt}, \varepsilon_{it})] + \beta[E[V^P_{im,t+1}|s_{mt}, 1] - E[V^P_{im,t+1}|s_{mt}, 0]] \geq \varepsilon_{imt}.$$

The first expectation is the expected profits in the market given the state space at the beginning of the period,

$$E[\Pi_{imt}(\cdot)] = -a_{imt}(1-a_{imt-1})EC_{im}(s_{mt}) + \sum_{s'_m \in S_m} [\pi_{im}(s'_m) - FC_{im}(s'_m)]FP(s'^*_m|\sigma_{im}(s_{mt}, \varepsilon_{it}) = a_{imt}, s_{imt})]$$

where entry costs are incurred when an inactive firm becomes active, and the variable period profits depend on which firms decide to be active in the period. Thus, there is an expectation taken using the transition probability matrix, $FP(\cdot)$, which is a function of the true equilibrium probabilities. The second is the expectation of the value functions from next period onward (once again using the transition probability matrix).

This implies equilibrium probabilities of firm i being in market m at time t, of the form:

$$P_{im}(a_{it} = 1|s_{mt}) = G_\varepsilon(E[\Pi_{it}(a_{it}, s_t, \varepsilon_{it})|s_{m,t-1}] + \beta[E[V^P_{im,t+1}|s_{mt}, 1] - E[V^P_{im,t+1}|s_{mt}, 0]]).$$

Now, we can think of the per period profit function as:

$$\Pi_{imt} = (1 - a_{imt})[z_{imt}(0)'\theta] + a_{imt}[z_{imt}(1)'\theta + \varepsilon_{imt}]$$

where: $z_{imt}(0, s_{mt})$ is a vector of zeros, $z_{imt}(1, s_{mt}) \equiv (E[\hat{\pi}_i(s_m)|s_{mt}], E[FC_i(s_m)|s_{mt}], a_{imt}(1-a_{imt-1})EC_i(s_{mt})$ and $\theta \equiv \{1, \gamma, \varphi\}$.

Denote the variance of ε by σ_ε. We then have a MPE being a vector of $P = \{P_{im}(s)\}$ such that for all (i, m, s_{imt}):

$$P_{im}(a_{imt} = 1|s_{mt}) = \Psi(\tilde{z}^P_{imt}\frac{\theta}{\sigma_\varepsilon} + \tilde{e}^P_{imt}) \qquad ((\text{EQ}))$$

where $\Psi(\cdot)$ is the extreme value CDF (and θ is all of the parameters in the model), and we have the infinite sums of:

$$\tilde{z}^P_{imt} = z_{imt}(1, s_{mt}) - z_{imt}(0, s_{mt}) +$$
$$\sum_{j=1}^{\infty} \beta^j [$$
$$\sum_{s' \in S} \{E\{(P_{im}(a_{im,t+j} = 0|s'_{m,t+j}))z_{im,t+j}(0, s'_{m,t+j}) + P_{im}(a_{im,t+j} = 1|s'_{m,t+j})z_{im,t+j}(1, s'_{m,t+j})\}|a_{imt} = 1, s_{im,t+j-1}\}\} -$$
$$\sum_{s' \in S} \{E\{(P_{im}(a_{im,t+j} = 0|s'_{m,t+j}))z_{im,t+j}(0, s'_{m,t+j}) + P_{im}(a_{im,t+j} = 1|s'_{m,t+j})z_{im,t+j}(1, s'_{m,t+j})\}|a_{imt} = 0, s_{im,t+j-1}\}]$$

$$\tilde{e}^P_{imt} = \sum_{j=1}^{\infty} \beta^j [\sum_{s' \in S} E\{P_{im}(a_{im,t+j} = 1|s'_{m,t+j})(Euler - \ln P_{im}(a_{im,t+j} = 1|s'_{m,t+j}))\}|a_{imt} = 1, s_{m,t+j-1}\} -$$
$$\sum_{s' \in S} \{\{\{P_{im}(a_{im,t+j} = 1|s_{m,t+j})(Euler - \ln P_{im}(a_{im,t+j} = 1|s_{m,t+j}))|a_{imt} = 0, s_{m,t+j-1}\}].$$

Thus airlines compare the value of being active with that of not being active in terms of the payoff today and in all future period paths beginning at being active or not being active today respectively. Thus includes the effect of the activity status today on future entry and exit through the equilibrium probabilities. Once again, expectations are taken using the transition probability matrix which provides a probability for every possible transition from a given state. The derivation of $e^P_{imt} \equiv E(\varepsilon_{imt}(a_{imt})|s_{mt}, \sigma^*_i(s_t, \varepsilon_i) = a_i)$, the expectation of the error conditional on it rendering a_i the optimal action, using the Euler-Mascheroni constant was shown by Hotz-Miller (1987?). As is standard, the infinite sums are solved by solving a system of linear equations.

Equation (EQ) represents the conditions required for their to be an equilibrium in the model, namely that each firm is best responding to the actions of all other firms. Given the assumptions above, an equilibrium exists (see AM 2007). There are as many best response constraints as the dimension of the state space, multiplied by the number of carriers.

2.4.5 The Transition Probability Matrix

Note above that the transitions are the product of the individual choice probabilities. This still however leads to a prohibitively large number of transition, since we have for each market $(7 \cdot 2^7)^2$ possible transitions. After looking at the data, I find that in 99% of the data, there are no more than 2 movements (changes of status - entry or exit) per period. I thus constrain all transitions involving more than two transitions to be zero (and scale the permissible transitions accordingly). To give a sense of the data, table 2 lists the transitions (by number of players, not identities).

Table 2

Transitions

	0	1	2	3	4	5	6	7	**Total**
0	**74**	10	0	0	0	0	0	0	84
1	10	**608**	91	12	1	0	0	0	722
2	2	63	**896**	194	40	6	4	0	1205
3	0	12	128	**1275**	363	57	14	12	1851
4	0	0	6	242	**1830**	545	97	10	2730
5	0	0	0	21	345	**2848**	735	62	4011
6	0	0	0	1	25	425	**5158**	613	6222
7	0	0	0	0	1	24	336	**5823**	6184
Total	86	693	1211	1745	2605	3905	6344	6510	23009

2.5 Estimation

2.5.1 General Strategy

The estimation strategy is a result of both the complexity inherent in the airlines' optimization problem, as well as the feasibility constraints imposed by the state-of-the-art solvers available today. Over time, as solvers improve and computational power increases

larger problems (and richer state spaces) will become feasible.

As mentioned, airlines are interested in maximizing profits, which are the difference between the revenues they can garner from ticket purchasing consumers and their costs. Airlines make many choices, including the choice and allocation of their fleet, the structure of their networks, and so on. For simplicity, I think of the airlines as solving this complex profit optimization problem using backwards induction. This implies that when playing the pricing game at the ticket level, airlines take capacity (as well as product characteristics and their fleet and network) as given. Consequently, at this level, the marginal costs of serving an additional customer are just the costs of filling a seat, conditional on capacity, as well as some probability of incurring a cost for adjusting capacity. These latter costs likely involve compensating consumers for overbooked flights (adding more capacity at the "last minute" is much less common) and are not likely to be very large in the overall calculus of firm costs. The difference between the revenues from an optimally chosen pricing structure and these marginal costs represent the variable profits from the route. These actual (or potential) variable profits then go into the calculus of the decision regarding the airline's activity status in each route, in forming the expected present discounted profits.

The model allows for an estimation of variable profits from the dynamic game as well. However, this would add many more parameters to be fit, and ignore the abundance of ticket level data which is available. Consequently, my approach is to estimate all parameters that can be estimated from the static game and then construct projections of these variable profits, for each market, on the state space and use them as inputs into the value function and the dynamic model. Specifically, I specify a static pricing game to obtain variable profits, and estimate them with all of the cross sectional data described above. Note that, when estimating the static game markets are directional: consumer choose from all available tickets beginning at their desired origin. I thus combine the profits from each direction in computing the total *market* profits. I then move to estimate the structural parameters of the entry and fixed costs specified above.

2.5.2 Variable Profits

Approach

In modeling the demand, choices have to be made to both reflect the data limitations, as well as to provide a tractable framework for the dynamic analysis. On the data side, the time of purchase (and, consequently, the choice set available to each consumer) is not observed; many of the ticket restrictions, which are key determinants of the price, are unobservable as well. Indeed, a disaggregate analysis, like a BLP 1995 approach (treating each ticket separately), requires identifying products as any tickets with a different price (or any other characteristic); and thus, products are rarely repeated. Each quarter will have tens of thousands of products, yielding a huge amount of product shares. In my model, entry is into a market in a particular quarter, and so, in the model, airlines consider the quarterly profits. Thus, instead of estimating the per product variable profit and aggregating to the quarterly level, I simplify the analysis by thinking of a (reduced form) game over the entire quarterly traffic. This eliminates the ability to have consumer specific coefficients (or even the simple two-type bimodal distribution in BCS and Berry and Jia 2009). It does, however, simplify the instrumental variable methodology. I define the price of a product as the average price paid by each passenger to the carrier in the market. If p_{km} is the price of each ticket sold by airline k in market m, redefine p_j as:

$$p_j = \frac{\sum_{k \in Km} p_{km}}{s_j MS},$$

where MS is the size of the market, the geometric mean of the MSAs of the endpoint cities.

Static Pricing Game and Demand

I use the common discrete choice framework (for example Berry 1994), where consumers have the following utility function:

$$u_{ijt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt} + \zeta_{igt}(\sigma) + (1-\sigma)\varepsilon_{ijt}.$$

x_j is a vector of characteristics of the product and market characteristics. p_j is the product price, β and α are the vector of tastes for the product characteristics and price respectively; ξ_j are the unobserved ticket features; ε_{ijt} is an i.i.d. (across all products and consumers) logit error, ζ is common (for consumer i) to all products in group g and has a distribution function that depends on σ, $\sigma \in [0,1]$,where $\zeta_{ig}(\sigma) + (1-\sigma)\varepsilon_{ijt}$ is still distributed logit. As σ goes to 1 there is no i.i.d error (within group correlation of utility levels goes to one). With $\sigma \in (0,1)$ product shares have the common nested logit form:

$$s^i_{j/g} = \frac{\exp[(x_{jt}\beta - \alpha p_j + \xi_{jt})/(1-\sigma)]}{D_g} \tag{1}$$

where

$$D_g = \sum_{j \in J_g} \exp[(x_{jt}\beta - \alpha p_{jt} + \xi_{jt})/(1-\sigma)].$$

Define

$$\delta_{jt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$$

as the mean utility of product j. Assume the outside good has a zero mean utility, and so the share of consumers who decide to fly (and purchase a ticket) follows the same logit form (substituting in for the outside good) :

$$\bar{s} = s^i_{j \in J} = \frac{D_g^{(1-\sigma)}}{1 + D_g^{(1-\sigma)}}. \tag{2}$$

The (unconditional) share of product j is the product of (1) and (2):

$$s_j^i = s_{j/g}^i \cdot s_{j\in J}^i = \frac{\exp[(x_{jt}\beta - \alpha p_j + \xi_{jt})/(1-\sigma)]}{D_g^{\sigma}(1 + D_g^{(1-\sigma)})} \tag{3}$$

Specifically, the share of the outside good is:

$$s_0^i = \frac{1}{1 + D_g^{(1-\sigma)}}.$$

This framework is convenient. Taking logs (and suppressing the index for each individual), we have

$$\ln(s_j) - \ln(s_0) = \frac{1}{1-\sigma}\delta_j - \sigma \ln(D_g)$$

and taking logs of (2) we have $\frac{\ln(\bar{s}) - \ln(s_0)}{1-\sigma} = \ln(D_g)$ and so:

$$\ln(s_j) - \ln(s_0) = \frac{1}{1-\sigma}\delta_j - \sigma\frac{\ln(\bar{s}) - \ln(s_0)}{1-\sigma}$$

and rearranging we have:

$$\begin{aligned}
\delta_j &= (1-\sigma)\ln(s_j) - \ln(s_0) + \sigma\ln(\bar{s}) \\
&= (1-\sigma)\ln(s_j) - \ln(s_0) + \sigma[\ln(s_j) - \ln(\bar{s})] \\
&= \ln(s_{j)} - \ln(s_0) - \sigma\ln(\bar{s})
\end{aligned}$$

where the second equality comes from substituting in (3). This yields the following:

$$\ln(s_{j)}^i) - \ln(s_0^i) = x_{jt}\beta - \alpha p_{jt} + \sigma\ln(\bar{s}) + \xi_{jt}. \tag{4}$$

Using this relationship, we can obtain the coefficients β, α and σ using a linear instrumental variable approach, we where know that, at minimum, p_j and $\bar{s}$ are endogenous with respect to ξ_j. Finally for the supply side of this last stage game, we can assume a standard Bertrand-

Nash game and then have:

$$p_j = c_j + \frac{s_j}{\frac{\partial s_j}{\partial p_j}}$$

and so differentiating (3) we have that:

$$p_j = c_j + [\frac{\frac{(1-\sigma)}{\alpha}}{1 - \sigma\bar{s} - (1-\sigma)s_j}]$$

and so we can form

$$\hat{c}_j = p_j - [\frac{\frac{(1-\hat{\sigma})}{\hat{\alpha}}}{1 - \hat{\sigma}\bar{s} - (1-\hat{\sigma})s_j}]$$

and obtain measures of the variable profits in the market. $p_{jm} - \hat{c}_{jm}$.

We can project the marginal cost on product characteristics and estimate:

$$\hat{c}_j = w_j\gamma + \omega_j$$

where w_j are the product characteristics affecting the marginal cost, which, once again simplifies to this linear form.

Instrumental Variables are needed, as mentioned above to account for the simultaneity of the determinations of prices and quantities. A good cost shifter for the average prices are the fuel costs and so I use a 2SLS strategy of instrumenting for the price and internal share with fuel costs and, what are commonly known as the BLP instruments, the sum of the characteristics of the other products - which are the sum of the average characteristics of the other airlines in the market[18]. The latter group of instruments stem from the assumption that firms play a pricing game where the characteristics of all other products affect the prices they can charge and the overall share of consumers choosing to fly, but that airlines are not adjusting the product characteristics (and specifically the ξ_{jt}) jointly

[18]These instruments provide reasonably large first stage R-squares, and highly significant parameters.

with the other players[19].

For comparison[20] I also replace these BLP instruments with the set of instruments used by Hausman (1996). The identifying assumption behind these instruments - which exploit the panel structure of the data - is that, given the controls, market-specific valuations are independent across markets (but potentially correlated within a market). This allows for the use of ticket prices in other markets as valid IVs. These prices are thus assumed to be correlated across markets due to common marginal costs but, given the mentioned assumption, not due to market specific valuations. All prices in all markets and all quarters could potentially be used as instruments. Following Nevo 2001, I use the average price in all markets (excluding the market being instrumented for) in a given quarter[21].

2.5.3 Estimation of the Dynamic Game

The Optimization Problem

Given the estimation of the bottom node of the game, we estimate the parameters of entry and fixed costs from the entry and exit decisions/moments. We therefore track the activity status of each airline and construct the following constrained maximum[22] pseudo likelihood estimator[23]. We maximize ($\theta \in \Theta$):

$$\tilde{L}(\Theta, P) = \sum_{i=1}^{N}\sum_{m=1}^{M}\sum_{t=1}^{T} \mathbf{1}(a_{imt}=1)\ln(\Psi(a_{imt}|s_{mt};P,\Theta)+(1-\mathbf{1}(a_{imt}=1))\ln(1-\Psi(a_{imt}|s_{mt};P,\Theta))$$

[19] This assumption, which is questionable in many industries, is questionable here as well. It may be unreasonable to assume that airlines are not adjusting their ticket features, although this is difficult within a time period given the complexity in coordinating the entire network.

[20] In this version of the paper, I use the first version of estimates for as inputs in the dynamic game.

[21] There is no claim being made here regarding the optimality of this choice of instruments. I also include the hub characterization of the other produces here, since these are, by construction, not adjustable.

[22] More precisely, it is the supremum of the maximization of the pseudo likelihood.

[23] The term "pseudo" comes from these probabilities not necessary representing the equilibrium probabilities, but rather best responses to an arbitrary vector P. See below and in AM for more on this.

subject to the equilibrium condition (EQ) above of:

$$P_{im}(a_{imt} = 1|s^*_{imt}) = \Psi(\tilde{z}^P_{imt}\frac{\theta}{\sigma_\varepsilon} + \tilde{e}^P_{imt}).$$

Simply put, we are maximizing the likelihood of observing the activity patterns in the data, given the model, subject to all actions representing best responses to P, or being consistent with the most likely equilibrium given the data, in the game specified above. We will call the solution to this problem the **"Full Maximum Likelihood Estimator"**.

Feasibility and Computational Methodology

This problem is computationally challenging. Thus, while the objective function is smooth, there is a (non-linear) equilibrium constraint for each element of the state space, for each player. There are thus $7 \cdot 1211 \cdot 2^7 = 1,085,056$ probabilities, for each of which there are all the transitions described above. The value function has to be solved for each player and each element of the state space as well, by inverting a system of linear equations, where:

$$V_{im}(s) = \sum_{a \in A} P_{im}(a) E(\Pi(s) \cdot \beta \sum_{s' \in S} V_{im}(s') TP(s, s').$$

This estimator is consistent, and efficient (see AM). To ease this computation burden, previous work has simplified the state space by assuming that players are symmetric and by discretizing the state space. AM note that "this estimator can be impractical if the dimension of P is relatively large....this is the case in models with heterogeneous players...even when the number of players is not too large". It is not possible to assume symmetry here, since the purpose of this study is to understand the differential effects that the differing network features have on the players themselves, and the heterogenous effects they have on their actual and potential competition.

When the population probabilities P^0 are known, the equilibrium constraints are not needed and the estimator is root-M consistent. When a $\sqrt{M}$ non-parametric estimator

of $P^0, \hat{P}^0$ is available (as is the case for example with a frequency estimator or a kernel method, when there are no unobservable market characteristics), estimates of θ, resulting from the maximization of the likelihood: $\tilde{L}(\Theta, \hat{P}^0)$, or from what we will call the **"Two Step Estimator"**, are consistent as well (see full details in AM 2007). It is difficult to determine and establish consistency of the estimators of the probability methods in most applications. The use of market fixed effects when feasible, is helpful, but there still could of course be market-time specific unobservables corrupting the estimates. AM propose a Nested Pseudo Likelihood method: The prescription set forth by this methodology is that (potentially non-consistent) estimates of $\hat{P}_k^0$ are formed, $\hat{\theta}_k$s are obtained from the maximization of $\tilde{L}(\Theta, \hat{P}_k^0)$, $\hat{P}_{k+1}^0(\hat{\theta}_k)$ are formed, using the equilibrium constraints with $\hat{\theta}_k$, new $\hat{\theta}_{k+1}$ are formed from the maximization of $\tilde{L}(\Theta, \hat{P}_{k+1}^0)$ and so on until:

$$(\hat{P}_{k+t}^0 - \hat{P}_{k+t-1}^0) \leq r,$$

where r represents the stopping rule. This sequence is well defined when there is a unique value of θ that maximizes the pseudo-likelihood function for each value of P, which is assumed in all applications using this method. When this sequence of $\{\hat{P}_k^0, \hat{\theta}_k\}$ converges (if it converges), its limit represents the maximum of the constrained problem. This is what they call the "**Nested Fixed Point**" estimator. In the Monte-Carlo examples presented in the AM paper convergence is achieved. Interestingly, the "two-step" estimates provided very similar results, suggesting that in any case, the estimated probabilities should be used in initializing the solution algorithm for this problem.

To ease the computational burden, I follow the MPEC (Mathematical Programming with Equilibrium Constraints) approach, advocated by Judd and Su (2008) and Dubé, Fox and Su (2009). The MPEC structure of the problem essentially relies on the "augmented likelihood function", $\mathcal{L}(\theta; \sigma, X)$, presented above, which explicitly expresses the dependence of the likelihood on σ. θ and σ do not need to be consistent with the conditions of the

equilibrium of the model[24]; however, when adding the equilibrium conditions as constraints, the solution to the problem will be a θ which represents the most likely equilibrium. Given this formulation, I can use solvers which rely on quadratically convergent constrained optimization methods, based on Newton's method (see Schmedders 2008 for a review of optimization methods)[25]. The solvers do not solve a fixed point, or require the specification of an algorithm for solving the equilibrium conditions, and the augmented likelihood uses single valued functions. Furthermore, the constraints need only to be solved at the point of the optimal solution; an LU decomposition is computed, and backsolving is used (rather than inverting matrices); derivatives are computed using automatic differentiation[26], which "eliminate[s] this as a serious problem"[27]; and the sparsity in the Jacobians and Hessians is exploited.

I begin by estimating smooth functions for the activity probabilities. For each player, I estimate the logit probability:

$$P_i(a, s_m) = G(X\beta)$$

where $G(\cdot)$ is the standard logistic CDF. X includes the activity status of each of the players in the previous period, a quadratic function of the distance, the geometric mean of the population, the passenger density, whether it is a tourist market, the number of hubs for the airline itself in the market, and the number of hubs for the other legacies. I cluster by market to flexibly allow for serial correlation[28]. The results are presented in table 5 below. As can be seen, the strongest predictor of being active in a market is last period's activity status, reflecting the stability of decisions over time and potentially the

[24] Compare to the discussion of the "pseudo" maximum likelihood above.

[25] This as opposed to Guass-Siedel methods, commonly used in past work, which have at best linear convergence, and (even local) convergence for which is difficult to prove with nonlinear equations.

[26] This refers to methods that computed analytic gradients and Hessians efficiently and use the chain rule of differentiation to build a sequence of simple operations. Languages such as AMPL and GAMS aid in computing these gradients and Hessians using insights incorporated in symbolic software.

[27] See Judd and Su 2008.

[28] In future versions I plan to explore more flexible specifications of these activity probabilities.

large role for entry and fixed costs. The other coefficients are consistent with the variable profit results presented below. Note that these are merely activity probabilities and not structural parameters. However, these results can be compared to Benkard et al (2008) who use similar probit probabilities to simulate the effects of Delta-Northwest merger and find them to have much predictive power[29].

I use the predicted probabilities generated by these regression to form the initial probabilities for all elements of the state space in every market. In other words, this initializes the values of all of the equilibrium constraint probabilities discussed above. In addition, as is standard, β is treated as a parameter. Given that these are quarters, I chose $\beta = 0.98$.

Finally, I chose the knitro solver, which is one of the most powerful solvers today, designed to handle linear and nonlinear problems with dimensions running into the hundreds of thousands. It has the versatility of three different algorithms which it can choose between, including direct and conjugate gradient interior point methods, as well an active set algorithm to rapidly solve binding constraints using linear programming. The main advantage of the AMPL language is that, once the problem is transformed into the form above, the code is straightforward and the communication with a multitude of available solvers is made easy. AMPL prepares the problem for the solver, and, when the presolve option is used, it transforms the problem into an equivalent smaller problem which is easier to solve. It removes unnecessary constraints and applies useful transformations. However, AMPL, can require a large amount of memory and even the most advanced solvers are not without limitations. In this application, the richness of the model is largely dependent on the size of the state space that can be estimated, where each element of the state space essentially adds an equilibrium constraint for each player. For the full model specified here, there are over a million constraints. This is generally above the limit of what can be done with the best solvers and so the problem has to be estimated in parts (or, equivalently, the

[29] I note that one of the key predictors, especially for Southwest entry was the amount of passenger traffic that could be added to the network. This variable could be tracked in further versions of the paper, along with the total number of destinations from the cities discussed above.

number of markets that can be used in the estimation has to be reduced). Furthermore, both AMPL and the solver demand a lot of RAM memory. A 64 bit operating system (and AMPL version) is required to process a problem using over 4 GB of RAM, which in my experience is the case for most non-trivial applications. The largest server I was able to use has 65 GB of (shared RAM), and approximately another 30 GB of (slower) swap memory. These represent the computational constraints with which I was faced[30].

The difference between the full maximum likelihood and the two step estimator, given this formulation, hinges on whether the initial estimates are treated as starting values, or are held fixed in the estimation[31]. Note that even for the two step estimator, each player's value functions have to be solved for, using the transition probabilities (as functions of the initial values). I have found that 50 market can solved for in a matter of minutes, while 300 markets requires about a day. For the full model, even 50 markets can take a matter of days. To get initial estimates I estimated the markets in groups of 50 and in groups of 300 (while averaging the scaled coefficients). The results vary between groups and between the averages of the groups.

2.6 Results

This section collects the results of all stages of estimation. We begin by discussing the results from the static demand: the demand parameters and the marginal costs. We then discuss the projection of these parameters on the state space, which feed into the value function. Finally we move to discuss the parameters of the entry costs and fixed costs from the estimation of the full dynamic game.

[30] These were the limits when writing the first draft in June of 2009. I hope and expect these limit to soon be seen as laughable.

[31] In the code, the difference between the two-step and the full maximum likelihoos estimates is two comment characters.

2.6.1 Demand Parameters

In order to track the impact of the necessary simplifications of the state space, I present several specifications, using all relevant (and available) variables. My approach is to specify as rich a demand model as possible and then project the resulting profits onto as rich a state space as possible. Summary statistics for all filters considered (all markets, all markets between the 100 largest MSAs, and all markets between the 50 largest MSAs) are presented in table 1. The results are presented in tables 3(a) and 3(b). Once again I present the results for the three filters considered, which suggest that the limits made on the subset of markets are reasonable.

I cluster by market to allow for serial correlation, but, as can be seen, all coefficients are highly significant and of the expected sign. I include fixed effects for all major carriers (the six legacy carriers, and six of the biggest LCC including Southwest, Jetblue, Frontier, AirTran, Spirit Airlines and ATA Airlines), and for each time period.

The demand is downward sloping in price. Distance (measured in miles) is positive but non monotone, reflecting an inverted U-like relationship where air travel becomes more attractive as it crowds out other modes of travel, but at further distances travel is needed less and is consumed less. The number of destination cities may reflect more convenient gate access and expertise as well as flexibility in rerouting passengers and thus is positive. The tourist dummy captures travel to or from Las Vegas or Florida and fits the high level of traffic to these cities not captured by the other variables in the model. Travellers prefer less stops and more direct flights. They also prefer travel to and from the hub of the ticket carrier they are using, and have a negative preference for airlines other than the hub airline at hub airports. Within this model, the airline dummies are positive and large for the main LCC: JetBlue and Southwest ("B6" and "WN" respectively), but not for the legacies, capturing the features of the LCC service not captured by the relatively parsimonious specifications possible with the data. Finally, as expected, travelers prefer round-trip tickets. To get a sense of the monetary value of the characteristics it is useful

to use the value of a marginal dollar of price as a scale. Doing this suggests for example that on average, passengers would pay $143 dollars to travel an extra 1000 miles, and $241 less for a connecting flight. They are willing to pay $53 extra to travel with a carrier that has a hub at their destination airport, and about $55 more for a tourist spot. In comparison, when using the Hausman instruments in table 3(b), the magnitudes change: on average, passengers would pay $500 to travel another 1000 miles, and $957 less for a connecting flight. The hub premium increases as well to $93.5 dollars, as does that for tourist destinations.

2.6.2 Marginal Costs

For the marginal costs recall that these are those in the pricing equation, where price is defined as the average price for a given airline's product in a given market and a given quarter. Here, we find the tourist variable coefficient to be extremely low. This likely reflects much lower prices for these markets, which is likely due to the higher elasticity of travel to these destinations. In other words this is a control for the parsimonious model which assumes the same Bertrand-Nash pricing game in all markets. The other coefficients are as expected. The more round-trip tickets, the higher the cost. Similarly, a larger number of connections increases the cost (per ticket), because there is more travel. Note that this is the final level costs. The choice of more connections is helpful in the aggregate analysis of the airline, since it eases constraints on the rest of the network. The more speciality in terms of destinations from the origin, the lower the cost. Costs follow an inverted U shape with respect to distance where there are savings in costs at much higher distances. The large coefficients on the hub variables are surprising and may, once again, caution the simplicity of this analysis, but suggest that costs are higher at hub airports to the hub carriers themselves. As expected, all LCC have much lower marginal costs. In other specifications, the number of markets is reduced. This helps reduce the size of the data, but also, potentially, allows for a more homogenous group of markets. This

homogeneity has advantages in fitting the (parsimonious) model, but omits some of the information embedded in markets in which at least one of the endpoints is in a small MSA. Popular vacation spots would be an obvious example of this. However the second and third specifications, corresponding to CT and excluding all markets not between cities in the largest 100 MSA (55% of the markets) and 50 MSAs (15% of the markets, for a total of 1211) respectively, have roughly similar results. Similarly, the results using the alternative set of instruments in table 3(b) are qualitatively similar[32].

The average number of connections and the percent of round-trips is likely to be endogenously determined, and thus the assumption made in these specifications is that product characteristics are fixed at the beginning of the period. However, specifications without these variables yield similar results. These variables are not part of the state space for the dynamic game and are used to get the best fit for the variable profits where are used below. Taken as a whole, these results can be related to the two main questions of the paper, hubs, and LCC. Regarding hubs, the results from the static estimation suggest that while consumers prefer flying with hubs or with airlines offering more connections from the origin, more generally, they have a very high distaste for flights with connections (stops). Thus, the benefits from having a large hub presence comes at a high cost if indeed this requires many more connecting flights. It is difficult to comment on the magnitude without accounting for the benefits of the whole network, but these results do suggest the potential for segregation by different airlines in offering different products that meet either the flexibility features, or the nonstop features respectively. As to the LCC, it is clear that consumers have a high preference for the LCC brand and especially for Jetblue and Southwest. This suggest that there are some unobservable characteristics of their products which consumers like. Marginal costs are also lower for these airlines and overall variable profits are thus higher. The increased number of nonstop flights offered by these carriers

[32] As mentioned, there are some differences in magnitudes which I plan to explore. There is also a difference in sign in the effect of the HubDest parameter on the marginal cost, but these estimates were not of the same sign in table 3(a) either.

increases consumers' willingness to pay as well.

2.6.3 Structural Profit Parameters

Now, for the purposes of the dynamic model we need a simple projection of the variable profits (revenues minus marginal costs), estimated above, on the state space. The results are presented in table 4 below. As can be seen the results from this parsimonious regression reflect those from the full (static) variable profit estimation. The fixed effects are negative for all airlines, relative to the omitted category of Southwest and JetBlue. The distance follows a U shape, the population, passenger densities and additional profits earned from tourist markets are all positive. Variable profits are higher in the airlines' own hub markets, but lower in markets in which their competitors have hubs. The LCC, Delta and Northwest all have significant negative effects on their rivals, when they are in the market. Interestingly, the strategic effects of the players on rival profits suggest that Southwest and Jetblue significantly and to a large degree, reduce the profits of their competitors[33].

The constant is 6.7 million dollars of quarterly variable profits. This reflects the baseline profit of the omitted category (the LCC). Baseline profits (fixed effects) are lower for all other carriers (relative to the omitted category) and are generally three to four million dollars. The variables distance, passenger density and population are scaled down (and so coefficients are per 1000 miles, passengers or residents respectively).

As mentioned above, these are seen to be the structural parameters of variable profits. Thus, in the estimation of the dynamic game we project these parameters on the state space of every market in each time period and input these projected profits into the value function. This both exploits the DB1B and T-100 data available, and reduces the burden on the estimation of the dynamic game below[34].

[33] This projection, although clearly much more parsimonious than the estimation of variable profits above, has a relatively high R-square of 40%.

[34] In future versions I plan to explore more flexible specifications of these projections, given the large

2.6.4 Entry Costs and Fixed Costs

I present the results (from the estimation of 300 market blocks) in table 6 below. Column 1 represents the average from all groups, while columns 2 and 3 represent the averages from the first two and second two blocks of 600 markets respectively. As can be seen, the estimates are of the expected magnitude, but vary between blocks of markets. Markets are ordered by the geometric mean of the city populations (not by the passenger traffic or profitability), but this likely still introduces some systematic differences (and so perhaps choosing the subset of markets randomly may be better). The preliminary results here suggest that entry costs are the equivalent (for the LCC omitted category) of slightly less than 3 years of variable profits, and so can be quite high, but not surprising. Fixed costs are about 60% of the variable profits, and so are considerable as well. Both costs vary by player, where American, Northwest and United seem to have lower costs, while the costs for Continental, Delta and US Airways are higher. Distance (measured here and in the profit projections in thousands of miles) is negative for some of the blocks, which is unlikely. This result seems to be sensitive (it was not the case when I averaged over market blocks of fifty markets). Thus, this issue may be alleviated by added a quadratic in distance, as in the static results, as well as increasing the size of the blocks (both of which I plan to explore). Entry costs are much lower in markets in which the carrier has hubs, and interestingly (although to a smaller degree) for hub markets more generally. Entry costs are higher the more legacy carriers incumbent in the market, and, to a smaller degree when there are LCC incumbents. Fixed costs, are higher in hub markets. This echoes the results of the marginal costs being higher in airlines' own hub markets as well. If this is indeed the case this results suggest that the added benefits of hubs are smaller than what is commonly perceived. In the specification presented here legacy carriers and especially

---

sample size.

LCC, being in the market, raises fixed costs[35]. These results were also not stable across blocks and are puzzling. I will revisit these as well, as the scale of the estimation increases.

Returning the our motivation, we find evidence that while hubs increase consumers willingness to pay, as do an increased number of destinations from the origin airport, they come with considerable added marginal and fixed costs which may outweigh the benefits even in the hub markets themselves. The increase in fixed costs is higher for the hub airline in its own hub airports than for competitors. The additional connections required are both more costly to the airlines and also can garner less from consumers, especially when competitors offer nonstop flights. Similarly, I do not find that hub airports are more expensive for competitors to enter. Taken as a whole, these results do not paint an optimistic picture for the use of major hub airports.

As mentioned, low cost carriers are found to be more profitable in the static demand and to significantly reduce the variable profits of the legacy airlines. In the estimates in table 4[36] we find that on average the presence of a LCC in the market is associated with a reduction of over $\$660,000$. These effects are much smaller than the effects of legacy carriers on themselves and on the LCC. These airlines are more appealing to consumers and to have lower marginal costs. However, surprisingly, LCC impose less of a deterrent to the entry of legacy carriers in the airports they serve, as compared to that which the legacy carriers impose on other legacy carriers and on the LCC. In addition, the results presented here do not show an impact of LCC presence in a given market on the fixed costs of the other incumbents. Thus, while there are huge benefits to LCC in both desirability and costs, future work is needed to understand the source of the convergence in operating procedures and fares shown above, and the to explore the potential for the LCC benefits to be emulated by other carriers.

There are many caveats to these preliminary estimates. Firstly, the structure of the

[35] This may represent, for example, higher advertising and spending in these markets to combat the LCC presence.

[36] There results are from the projections of the variable profits on the state space as discussed above.

model is such that, conditional on the parameters of the model, all randomness comes from the extreme value errors. This is a serious limitation with which these types of dynamic models are faced. From a computational point of view the solver has to fit the σ and the entry and fixed costs. Entry and exit can come from either the draws of the error or these costs. This is likely a reason for the high sensitivity of the results. It may (and does for some of the blocks) even produce negative costs. Furthermore, as mentioned, we cannot identify exit costs separately and thus the fixed costs are essentially net of exit costs (which are not estimated), as well. Secondly, larger blocks need to be estimated. I am currently estimating the markets in two blocks, but ideally they should be estimated together. Thirdly, the full maximum likelihood estimates (for the full number of markets) should be used. Unfortunately, these take time, and will likely only be feasible for smaller blocks of markets. In their absence, the accuracy of the results hinges on the consistency of the initial probability estimates and their small sample bias, and so richer specifications of these initial probabilities should be explored. Fourthly, correct standard errors need to be obtained. Given the complexity in deriving analytical standard error for this multi-step process, a bootstrap methodology is more feasible. This methodology requires drawing from the data (with replacement), accounting for market clusters, and completing all stages of the estimation for each draw[37]. This is time intensive and somewhat difficult (and thus often not done in the applied dynamic game papers), given that solvers can, in some instances fail, and may need to be run several times to find an optimum. These caveats will hopefully be addressed in future versions of the paper.

2.7 Conclusion and Future Work

In this paper I applied a dynamic entry game model to the complex airline industry, in an effort to recover market specific profitability and its determinants. Specifically, I focused

[37] There are many distributions to think about. For example, when projecting the profit parameters we take the point estimates, even though these can lie in large regions and are sometimes insignificant.

on the desirability of hubs and the strategic effects of the heterogeneous players. The model used allowed for the exploitation of the benefits of state-of-the-art mathematical solvers and optimization software, which have recently been strongly promoted.

The results in this paper have important implications: I found that while hubs offer benefits that consumers desire and higher profits to hub airlines in their own hub markets, this network may also come at a cost in only offering many more flights with connections which are much less desirable to consumers. In addition, the preliminary estimates of fixed costs suggest that costs may be higher in the hubs as well further dampening their desirability and that they do not deter entry. As expected, Southwest and JetBlue are more profitable than legacy carriers, and their brands offer benefits not captured by the limited product characteristics in the data. Their presence in the market imposes particularly large strategic effects for the other carriers, in lowering variable profits and in raising fixed costs. More work is needed to understand the nature of the impact of LCC on their rivals and the potential for their advantages to be emulated.

As mentioned throughout, while this paper takes a further step in employing and enriching what can be done with dynamic models, there are important limitations both to the overall use of dynamic entry games in this application and in the preliminary results presented above. The estimation can be improved by increasing the number of markets in each block and in the full estimation, but memory limits and the computational capacity of solvers will inevitably constrain the size of the state space - and, consequently, the richness of the model that can be estimated, and thus final judgement will have to be made in the interpretation of the results. First and foremost, the simplifications which essentially abstract out of the joint network optimization made by airlines may indeed be deemed unreasonable.

The bridge between applied econometric work and state-of-the-art computational software is an important one. Economic models in general and dynamic models in particular, test the boundary of these tools and greatly benefit from increases in computational power.

Any flexibility granted by using better methods, offers more room to develop better representative economic models.

The extensions to this paper are immediate. More has to be done to ensure that we indeed at the limit in capturing the richest profit function possible. As more of the ticket transactions move online, better data is also becoming available. Then, with this profit function at hand, many counterfactual experiments may be estimated. For example, the preliminary work by Benkard et al, regarding the simulation of mergers can be extended to exploit the full structural model. Given that identity specific value functions are estimated, specific mergers can be explored, once choices are made regarding how we view the new merged entity. I plan to explore some of these extensions in future work.

Bibliography

[1] Aguirregabiria, V., and P. Mira (2007), "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75, 1 – 53.

[2] Aguirregabiria, Victor; Ho, Chun-Yu (2009), "A Dynamic Oligopoly Game of the US Airline Industry: Estimation and Policy Experiments, mimeo.

[3] Bajari, P., L. Benkard and J. Levin (2007), "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75, 1331 – 1370.

[4] Benkard, C. Lanier, Aaron Bodoh-Creed and John Lazarev (2008), "Simulating the Dynamic Effects of Horizontal Mergers: U.S. Airlines", mimeo.

[5] Beresteanu, Arie and Paul B. Ellickson (2006), "The Dynamics of Retail Oligopoly", mimeo.

[6] Berry, S. (1992), "Estimation of a model of entry in the airline industry", *Econometrica*, 60(4), 889 – 917.

[7] Berry, S. (1994), "Estimating Discrete-choice Models of Product Differentiation" *Rand Journal of Economics*, 25, 242 – 262.

[8] Berry, S, J. Levinsohn, and A. Pakes (1995) (BLP), "Automobile Prices in Market Equilibrium." *Econometrica*, 63, 841 – 890.

[9] Berry, S., M. Carnall, and P. Spiller (2006), "Airline Hubs: Costs, Markups and the Implications of Customer Heterogeneity," in *Advances in Airline Economics, Vol. 1: Competition Policy and Antitrust*, (D. Lee, ed.) Elsevier Press. 183 – 214.

[10] Berry, Steven, and Panle Jia (2009), "Tracing the Woes: An Empirical Analysis of the Airline Industry", mimeo.

[11] Boguslaski, C. H. Ito, and D. Lee (2004), "Entry Patterns in the Southwest Airlines Route System", *Review of Industrial Organization*, 25(3), 317 – 350.

[12] Borenstein, S. (1989), "Hubs and high fares: Dominance and market power in the US Airline industry", *Rand Journal of Economics*, 20(3), 344 – 365.

[13] Borenstein, S. (1991), "The Dominant Firm-Advantage in Multiproduct Industries: Evidence from the U.S. Airlines", *Quarterly Journal of Economics*, 106(4), 1237 – 1266.

[14] Borenstein, S. (1992), "The Evolution of U.S. Airline Competition." Journal of Economic Perspectives, Vol. 6.

[15] Borenstein, Severin (2005), "U.S. Domestic Airline Pricing, 1995 – 2004", *U.C. Berkeley Competition Policy Center Working Paper*, CPC05 – 48.

[16] Borenstein, Severin, and Nancy Rose (2007), National Bureau of Economic Research working paper #13452, forthcoming in N. Rose ed., *Economic Regulation and Its Reform: What Have We Learned?*, University of Chicago Press.

[17] Borenstein, Severin, and Nancy Rose (2008), forthcoming in N. Rose ed., *Economic Regulation and Its Reform: What Have We Learned? (chapter 2)*, University of Chicago Press.

[18] Bresnahan, T., and P. Reiss (1993), "Measuring the importance of sunk costs", *Annales D'Economie Et De Statistique*, 31, 181 – 217.

[19] Chernozhukov, V. H. Hong and E. Tamer (2007), "Estimation and Confidence Regions for Parameter Sets in Econometric Models", *Econometrica,* 75(5), 1243 – 1284.

[20] Ciliberto, F. and E. Tamer (2009) (CF), "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, forthcoming.

[21] Collard-Wexler, Allan, (2005), "Plant Turnover and Demand Fluctuations in the Ready-Mix Concrete Industry", mimeo.

[22] Doraszelski, U., and Satterthwaite, M. (2007), "Computable Markov-Perfect Industry Dynamics: Existence, Purification, and Multiplicity," manuscript. Department of Economics. Harvard University.

[23] Dube, Jean-Pierre H., Jeremy T. Fox, and Che-Lin Su, (2009), "Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficient Demand Estimation", mimeo.

[24] Evens, W.N. and I.N. Kessides (1993), "Localized Market Power in the U.S. Airline Industry", *The Review of Economic Studies*, 62, 53 – 82.

[25] Fourer, R. D.M. Gay, and B.W. Kernighan (2003), *AMPL: A Modeling Language for Mathematical Programming*, Brooks/Cole - Thomson Learning, 2nd Edition.

[26] Gayle, K. (2004), "Does Price Matter? Price or Non-Price Competition in the Airline Industry, mimeo, Department of Economics, Kansas State University.

[27] Goolsbee, Austan and Chad Syverson, "How do Incumbents Respond to the Threat of Entry? Evidence from the Major Airlines", *Quarterly Journal of Economics* , 123(4), 1611 – 1633.

[28] Hausman, Jerry (1996), "Valuation for New Goods Under Perfect and Imperfect Competition", in *The Economics of New Goods*, Studies in Income and Wealth Vol. 58, ed. by T. Bresnahan and R. Gordon. Chicago: National Bureau of Economic Research.

[29] Hendricks, K. M. Piccione, and G. Tan (1995), "The Economics of Hubs: The Case of Monopoly", *The Review of Economics and Statistics,* 62(1), 83 – 99.

[30] Hendricks, K., M. Piccione, and G. Tan (1997), "Entry and exit in hub-spoke networks," *Rand Journal of Economics*, 28(2), 291 – 303.

[31] Hendricks, K., M. Piccione, and G. Tan (1999), "Equilibria in Networks," *Econometrica*, 67, 1407 – 1434.

[32] Hotz, V. Joseph, and Robert A. Miller, (1993), "Conditional Choice Probabilities and the Estimation of Dynamic Models", *Review of Economic Studies*, 61(2), 265 – 289.

[33] Judd, Kenneth L., and Che-Lin Su, (2008), "Constrained Optimization Approaches to Estimation of Structural Models", mimeo.

[34] Januszewski, J. and M. Lederman (2003), "Entry Patterns of Low-Cost Airlines ," manuscript, Rotman School of Management, University of Toronto.

[35] Kasahara, H., and K. Shimotsu (2008), "Pseudo-likelihood Estimation and Bootstrap Inference for Structural Discrete Markov Decision Models," *Journal of Econometrics*, 146, 92 – 106.

[36] Lederman, M. (2008), "Are Frequent-Flyer Programs a Cause of the Hub Premium?" *Journal of Economics and Management Strategy*, 17, 35 – 66.

[37] Nevo, Aviv, (2000), "A Practitioner's Guide to Estimation of Random Coefficient Logit Models of Demand", *Journal of Economics and Management Strategy,* 9(4),513 – 548.

[38] Nevo, Aviv, (2001), "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69(2), 307 – 342.

[39] Macieira, Joao, (2006), "Measuring the Effects of Innovation Subsidies: The Case of Supercomputers", mimeo.

[40] Pakes, A., M. Ostrovsky, and S. Berry (2007): "Simple estimators for the parameters of discrete dynamic games (with entry/exit examples)," *Rand Journal of Economics*, 38(2), 373 – 399.

[41] Peters, Craig, (2006), "Evaluating the Performance of Merger Simulation: Evidence from the U.S. Airline Industry", *Journal of Law and Economics, XLIX,* 627 – 649.

[42] Reiss, P.C. and P.T. Spiller, (1989), "Competition and Entry in Small Airline Markets", *Journal of Law and Economics*, 32, 179 – 202.

[43] Rust, J. (1987), "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher", *Econometrica*, 55, 999 – 1033.

[44] Ryan, Stephen, (2009), "The Costs of environmental Regulation in a Concentrated Industry", mimeo.

[45] Schmedders, K. (2008), "Numerical Optimization Methods in Economics. In S.N. Durlauf and L.E. Blume, eds., *The New Palgrave Dictionary of Economics*, 2nd Edition, Palgrave Macmillian.

[46] Sweeting, Andrew, (2007), "Dynamic Product Repositioning in Differentiated Product Markets: The Case of Format Switching in the Commercial Radio Industry", *NBER Working Paper* 13522.

2.8 Tables:

Table 2.1: Summary statistics

Variable	Mean	Std.	N	Mean	Std.	N	Mean	Std.	N
Price	362.442	108.73	814066	352.05	105.176	612319	331.017	95.827	282724
Fuel	126.411	43.375	814066	125.89	43.38	612319	124.98	43.419	282724
Roundtrips	0.891	0.112	814066	0.885	0.115	612319	0.872	0.121	282724
Connections	0.911	0.315	814066	0.888	0.334	612319	0.806	0.391	282724
Destinations	63.077	26.504	814066	64.107	26.129	612319	65.579	26.485	282724
Frommyhub	0.055	0.228	814066	0.058	0.233	612319	0.076	0.265	282724
Tomyhub	0.053	0.224	814066	0.057	0.233	612319	0.076	0.265	282724
Distance	1148.841	621.705	813925	1181.412	630.388	612302	1271.623	666.848	282724

These are the main variables used in the demand estimation. The first three columns are for all markets, the second three is for the first filter which keeps the 100 largest MSAs and the third are for the main sample used which keeps the markets between the 50 largest MSAs. As can be seen, most of the variables have roughly the same summary statistics.

Table 3(a)

Utility Function and Marginal Cost Parameters (BLP Instruments)

	allmkts	filter-1	filter-2	mc-1	mc-2	mc-3
	(1)	(2)	(3)	(1)	(2)	(3)
Price	-.007***	-.007***	-.006***	-	-	-
	(.0002)	(.0002)	(.0003)			
Lwshare	.279***	.274***	.249***	-	-	-
	(.011)	(.013)	(.022)			
Tourist	.390***	.397***	.406***	-53.043***	-53.712***	-55.712***
	(.023)	(.025)	(.032)	(2.369)	(2.587)	(3.263)
HubOrigin	-.190***	-.195***	-.148***	5.511***	3.784***	9.523***
	(.015)	(.018)	(.026)	(1.305)	(1.425)	(1.894)
HubDest	-.066***	-.059***	-.093***	-2.997**	-1.223	8.373***
	(.014)	(.016)	(.024)	(1.259)	(1.386)	(1.880)
Roundtrips	1.300***	1.436***	1.799***	21.871***	25.323***	-.681
	(.045)	(.055)	(.103)	(2.965)	(3.374)	(5.805)
Connections	-1.687***	-1.783***	-1.958***	35.050***	33.224***	23.240***
	(.024)	(.029)	(.049)	(1.611)	(1.745)	(2.166)
Destinations	.012***	.013***	.016***	-.382***	-.395***	-.226***
	(.0002)	(.0003)	(.0005)	(.023)	(.029)	(.049)
Frommyhub	.368***	.382***	.356***	57.569***	58.998***	53.962***
	(.019)	(.022)	(.031)	(1.593)	(1.780)	(2.331)
Tomyhub	.719***	.754***	.792***	50.874***	52.347***	52.465***
	(.019)	(.021)	(.031)	(1.553)	(1.700)	(2.174)
Distance	.001***	.001***	.001***	.073***	.066***	.078***
	(.00004)	(.00004)	(.00007)	(.004)	(.005)	(.007)
Distancesq	-3.35e-07***	-3.45e-07***	-3.32e-07***	-6.59e-06***	-3.49e-06**	-8.74e-06***
	(1.30e-08)	(1.49e-08)	(2.19e-08)	(1.49e-06)	(1.70e-06)	(2.39e-06)

Table 3(a) Cont'd

Utility Function and Marginal Cost Parameters (BLP Instruments)

	allmkts	filter-1	filter-2	mc-1	mc-2	mc-3
	(1)	(2)	(3)	(1)	(2)	(3)
AA	-.273***	-.306***	-.342***	70.115***	62.321***	44.618***
	(.030)	(.033)	(.051)	(2.108)	(2.412)	(3.970)
CO	-.430***	-.516***	-.604***	67.590***	59.979***	45.512***
	(.030)	(.033)	(.049)	(2.041)	(2.312)	(3.735)
DL	-.281***	-.344***	-.534***	92.407***	85.830***	52.335***
	(.031)	(.035)	(.054)	(2.386)	(2.875)	(4.922)
NW	-.524***	-.610***	-.803***	63.022***	55.912***	39.528***
	(.029)	(.033)	(.050)	(2.105)	(2.459)	(4.130)
US	-.184***	-.221***	-.241***	50.863***	46.671***	31.646***
	(.029)	(.032)	(.049)	(2.059)	(2.359)	(3.884)
UA	-.298***	-.359***	-.464***	74.510***	72.253***	56.097***
	(.030)	(.034)	(.050)	(2.069)	(2.414)	(3.979)
B6	.475***	.516***	.656***	-37.678***	-38.238***	-36.205***
	(.066)	(.067)	(.085)	(3.453)	(3.574)	(4.881)
F9	-.134***	-.106***	.129**	-18.519***	-16.865***	-18.183***
	(.034)	(.036)	(.050)	(1.994)	(2.144)	(3.061)
FL	.180***	.308***	.603***	-48.095***	-46.653***	-51.839***
	(.037)	(.041)	(.067)	(1.961)	(2.188)	(3.365)
NK	.699***	.533***	.622***	-74.194***	-69.661***	-79.750***
	(.102)	(.098)	(.109)	(3.667)	(3.689)	(4.604)
TZ	-.052	.024	.387***	-27.351***	-27.376***	-26.417***
	(.037)	(.040)	(.058)	(1.950)	(2.163)	(3.287)
WN	.323***	.292***	.398***	-14.173***	-13.392***	-17.443***
	(.030)	(.033)	(.052)	(1.895)	(2.110)	(3.249)
Obs.	644516	485915	224977	644516	485915	224977

Table 3(b)

Utility Function and Marginal Cost Parameters (Hausman Instruments)

	all-h	f-1-h	f-2-h	mc-a_all	mc-1	mc-2
	(1)	(2)	(3)	(1)	(2)	(3)
price	$-.002^{***}$	$-.002^{***}$	$-.003^{***}$	–	–	–
	(.0002)	(.0002)	(.0003)			
Lwshare	$.176^{***}$	$.210^{***}$	$.295^{***}$	–	–	–
	(.012)	(.016)	(.034)			
Tourist	$.615^{***}$	$.611^{***}$	$.593^{***}$	-53.840^{***}	-54.015^{***}	-54.923^{***}
	(.026)	(.030)	(.041)	(2.308)	(2.499)	(3.059)
HubOrigin	$-.287^{***}$	$-.253^{***}$	$-.130^{***}$	10.545^{***}	9.976^{***}	14.486^{***}
	(.016)	(.021)	(.032)	(1.270)	(1.385)	(1.839)
HubgDest	$-.130^{***}$	$-.095^{***}$	$-.074^{**}$	2.566^{**}	4.880^{***}	13.047^{***}
	(.016)	(.020)	(.031)	(1.235)	(1.354)	(1.830)
Roundtrips	1.361^{***}	1.418^{***}	1.639^{***}	11.933^{***}	13.520^{***}	-13.102^{**}
	(.046)	(.061)	(.124)	(2.894)	(3.298)	(5.703)
Connections	-1.914^{***}	-1.963^{***}	-1.912^{***}	43.043^{***}	45.085^{***}	42.256^{***}
	(.024)	(.029)	(.064)	(1.585)	(1.706)	(2.132)
Destinations	$.013^{***}$	$.014^{***}$	$.016^{***}$	$-.344^{***}$	$-.397^{***}$	$-.281^{***}$
	(.0002)	(.0002)	(.0004)	(.022)	(.027)	(.047)
Frommyhub	$.187^{***}$	$.173^{***}$	$.128^{**}$	50.633^{***}	51.221^{***}	44.369^{***}
	(.025)	(.031)	(.050)	(1.543)	(1.714)	(2.230)
Tomyhub	$.564^{***}$	$.573^{***}$	$.559^{***}$	45.080^{***}	44.653^{***}	41.707^{***}
	(.023)	(.029)	(.053)	(1.506)	(1.639)	(2.078)
Distance	$.001^{***}$	$.001^{***}$	$.001^{***}$	$.080^{***}$	$.073^{***}$	$.083^{***}$
	(.00004)	(.00005)	(.00008)	(.004)	(.004)	(.007)
Distancesq	$-3.11e-07^{***}$	$-3.38e-07^{***}$	$-3.05e-07^{***}$	$-8.30e-06^{***}$	$-5.59e-06^{***}$	$-1.00e-05^{***}$
	(1.34e-08)	(1.56e-08)	(2.39e-08)	(1.44e-06)	(1.63e-06)	(2.31e-06)

Table 3(b) Cont'd

Utility Function and Marginal Cost Parameters (Hausman Instruments)

	all-h	f-1-h	f-2-h	mc-a_all	mc-1	mc-2
	(1)	(2)	(3)	(1)	(2)	(3)
AA	−.520***	−.539***	−.511***	65.537***	60.223***	45.528***
	(.039)	(.042)	(.057)	(2.100)	(2.394)	(3.934)
CO	−.712***	−.765***	−.748***	65.945***	61.072***	48.187***
	(.038)	(.040)	(.050)	(2.044)	(2.311)	(3.735)
DL	−.608***	−.671***	−.733***	82.527***	77.059***	51.221***
	(.043)	(.048)	(.062)	(2.355)	(2.813)	(4.814)
NW	−.758***	−.829***	−.934***	58.774***	53.782***	40.776***
	(.038)	(.040)	(.051)	(2.096)	(2.435)	(4.081)
US	−.355***	−.391***	−.372***	47.578***	44.840***	31.465***
	(.036)	(.039)	(.054)	(2.057)	(2.347)	(3.835)
UA	−.585***	−.642***	−.646***	72.230***	72.443***	60.410***
	(.039)	(.042)	(.053)	(2.065)	(2.394)	(3.938)
B6	.713***	.718***	.688***	−39.982***	−41.295***	−39.302***
	(.075)	(.074)	(.092)	(3.535)	(3.674)	(4.974)
F9	−.067*	−.041	.153***	−17.870***	−16.637***	−19.340***
	(.035)	(.035)	(.044)	(1.988)	(2.149)	(3.098)
FL	.478***	.551***	.676***	−50.835***	−49.898***	−55.232***
	(.034)	(.037)	(.063)	(1.987)	(2.220)	(3.386)
NK	1.106***	.880***	.815***	−79.987***	−69.467***	−74.902***
	(.098)	(.093)	(.105)	(4.447)	(3.567)	(4.511)
TZ	.112***	.162***	.423***	−28.098***	−28.795***	−28.003***
	(.036)	(.038)	(.051)	(1.969)	(2.189)	(3.301)
WN	.470***	.391***	.360***	−19.491***	−18.594***	−25.226***
	(.031)	(.034)	(.061)	(1.943)	(2.170)	(3.348)
Obs.	644476	485875	224937	644516	485915	224977

Table 4

Projection of the Variable Profits on the State Space

Variable Profit Projection	
Constant	6694214.000***
	(855010.000)
AA	-3630095.000***
	(508058.800)
CO	-4367321.000***
	(446053.300)
DL	-3546622.000***
	(471656.600)
NW	-4471127.000***
	(467761.500)
US	-4486935.000***
	(549021.800)
UA	-4848132.000***
	(519141.500)

Table 4 Cont'd

Projection of the Variable Profits on the State Space

Variable Profit Projection	
Distance	-2409303.000***
	(443088.200)
Distancesq	743346.000***
	(139236.800)
Population	454.471***
	(160.204)
Passenger Density	53355.020***
	(6386.904)
Tourist	1128930.000**
	(527532.800)
OwnHubs	6107603.000***
	(385589.100)
OtherHubs	-1627747.000***
	(212550.800)
AA in Market	163744.400
	(265317.300)
CO in Market	-156505.000
	(207198.200)
DL in Market	-1336960.000***
	(293430.900)
NW in Market	-798468.300***
	(188964.000)
US in Market	212201.300*
	(125126.900)
UA in Market	206435.800
	(173316.800)
SWJB in Market	-662852.400***
	(136145.000)
Obs.	127800

Table 5

Estimates of Predicted Activity Probabilities

	AA	CO	DL	NW	US	UA	SWJB
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	−5.192***	−4.017***	−3.747***	−3.410***	−3.769***	−3.798***	−4.085***
	(.228)	(.192)	(.157)	(.160)	(.162)	(.167)	(.170)
LastAA	**4.271*****	.609***	.086	.409***	−1.166***	.398***	.197*
	(.153)	(.092)	(.095)	(.094)	(.107)	(.105)	(.120)
LastCO	.659***	**3.351*****	−.009	.436***	.069	−.241***	.059
	(.100)	(.103)	(.087)	(.086)	(.089)	(.081)	(.097)
LastDL	.140	−.088	**4.184*****	.575***	.072	−.460***	.169
	(.113)	(.106)	(.110)	(.093)	(.105)	(.100)	(.114)
LastNW	.325***	.471***	.590***	**4.074*****	.142	.226***	−.449***
	(.109)	(.095)	(.084)	(.107)	(.092)	(.085)	(.111)
LastUS	−1.101***	.345***	.208**	.232***	**5.466*****	.940***	.282***
	(.109)	(.082)	(.091)	(.086)	(.125)	(.090)	(.094)
LastUA	.414***	−.231***	−.421***	.052	.632***	**4.244*****	.263**
	(.097)	(.082)	(.089)	(.082)	(.087)	(.104)	(.102)
LastSWJB	−.075	.261***	.133	−.161**	.017	−.251***	**7.489*****
	(.100)	(.071)	(.084)	(.070)	(.067)	(.081)	(.168)
Distance	4.691***	2.970***	3.674***	1.492***	1.543***	3.124***	.430
	(.372)	(.307)	(.285)	(.291)	(.285)	(.315)	(.268)
Distancesq	−1.213***	−.834***	−1.057***	−.384***	−.413***	−.770***	−.008
	(.150)	(.110)	(.100)	(.104)	(.098)	(.114)	(.089)
Population	.0004***	.0002***	.00009*	−.00003	.0004***	.0001**	.0001***
	(.00005)	(.00008)	(.00005)	(.00004)	(.00005)	(.00005)	(.00004)
Pdensity	.002	−.003***	−.0006	−.0001	.004***	.0007	.004***
	(.001)	(.001)	(.0006)	(.0006)	(.001)	(.001)	(.001)
Tourist	1.347***	.406*	−.120	−.402*	.500*	.297	.484**
	(.338)	(.220)	(.219)	(.223)	(.301)	(.273)	(.208)

Table 5 Cont'd

Estimates of Predicted Activity Probabilities

	AA	CO	DL	NW	US	UA	SWJB
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Hubs-AA	1.540***	–	–	–	–	–	–
	(.258)						
Hubs-CO	–	2.545***	–	–	–	–	–
		(.334)					
Hubs-DL	–	–	1.416***	–	–	–	–
			(.180)				
Hubs-NW	–	–	–	2.776***	–	–	–
				(.409)			
Hubs-US	–	–	–	–	.927***	–	–
					(.158)		
Hubs-UA	–	–	–	–	–	1.095***	–
						(.129)	
Totalhubs-oAA	−.215***	–	–	–	–	–	–
	(.083)						
Totalhubs-oCO	–	−.497***	–	–	–	–	–
		(.074)					
Totalhubs-oDL	–	–	.075	–	–	–	–
			(.077)				
Totalhubs-oNW	–	–	–	.039	–	–	–
				(.074)			
Totalhubs-oUS	–	–	–	–	−.187***	–	–
					(.067)		
Totalhubs-oUA	–	–	–	–	–	.116*	–
						(.069)	
Total Hubs							−.434***
							(.058)
e(N)	23009	23009	23009	23009	23009	23009	23009

Table 6

Estimates of Fixed and Entry Costs

Type of Costs:	Entry (1)	Entry (2)	Entry (3)	Fixed (1)	Fixed (2)	Fixed (3)
Constant	81.5	140	22.7	4.17	8.13	.222
AA	−30.8	−32.8	−28.9	−9.48	−20.3	1.29
CO	3.87	28.4	−20.6	−10.2	−21.1	.636
DL	.441	17.2	−16.3	−9.43	−19.5	.668
NW	−87.2	−164	−10	−7.08	−14.6	.398
US	41.3	116	−33.3	−11.3	−23.9	1.35
UA	−37.0	−54	−19.9	−10.4	−21.4	.624
Distance	−22.5	−46.8	1.8	−.05	−1.05	.047
My Hubs	−33.7	−70.1	2.78	8.75	17.9	−.37
Other Hubs	−15.1	−2.27	−7.51	.095	.415	−.225
Legacies	4.15	8.09	.219	.445	.942	−.052
LCC	2.12	2.62	1.61	4.46	9.88	−.952

This table represents the parameters of the fixed and entry costs. Column 1 is an average over all four groups of markets. Column 2 is an average over the first two groups and column 3 is an average over the last two groups. Estimates are in the millions of dollars.

Chapter 3

Federal Appeals: Selection, Outcome and Predictability

3.1 Introduction and Motivation

As appeals are an integral part of most judicial systems, their existence is often taken for granted. However, legal systems have experienced many difficulties with their appellate structure. There is much criticism and concern regarding the undue burdens on the courts and their increasing caseloads, which many see to be related to the process. Judges do not have a sufficient amount of time to devote to each case, and many public dollars are spent considering cases that should have never been filed. There are also concerns regarding access to the courts and the availability of legal redress, given the high attorney fees and the length of the process. Naturally, this reality can be exploited by powerful bodies with much legal counsel at their disposal, who may target individuals for whom the legal remedies and process are too costly to employ. From the other end, there are those voicing concerns regarding the quality of the decisions. Some see the American system, particularly the federal judges who are appointed directly by the president, to be biased in a variety of ways.

In this paper, I empirically explore whether certain types of cases are more likely to be appealed, and whether there are certain types of cases that are more likely to prevail on appeal. Using data on all federal civil law cases (trial and appeal) I find that there are clear situations in which cases have a higher probability of success on appeal. Notably, appellate courts are more likely to rule in favor of the US as a litigant, and to overturn decisions for the plaintiffs, especially when the trial was by jury. There are also (sometimes different) systemic tendencies in the propensity to appeal.

These findings are interesting in evaluating the role for the appellate level in the judicial system. The appellate procedure chosen in the US is not the only one. There are alternate appellate schemes employed in other countries and more novel ones considered in the literature (see for example Shavell, 1995). Once systematic biases are identified, there are many potential deterrents to frivolous or patently unworthy appeals, which could prevent such appeals from ever being filed. The existence of such appeals are one of the chief sources for the heavy caseloads imposed on judges, which, quite naturally, could lead to a variety of negative consequences[1]. There are also ways to reduce the resources spent on such appeals, and to allow for the judges' time to be spent on the cases requiring their expertise.

However, litigation in general, and litigation at the appellate level in particular, is the result of a violation of one or more of the assumptions underlying the Coase Theorem. Consequently, while my focus is not on testing a particular model, any systematic tendencies leading to appeal or to success on appeal (of the selected cases that are appealed) are likely to be related to these violations and any proposed reform will affect the out-of-court settlement process as well. The form of selection impacts the interpretation of my findings and my findings serve as stylistic facts models of settlement breakdown should aim to explain. The breakdown of settlement has been explained in three main ways (discussed at

[1] Of course there are positive consequences as well, such as having many repeat observations for similar cases to let the common law system function with more plurality of opinions, etc. These seem less likely, however, when the cases have obvious outcomes.

greater length below): First, there is a large bargaining literature (focusing on the pre-trial rather than appeal stage) with a variety of predictions regarding the probability of going to trial and efficiency. These, generally, either employ an inconsistent priors approach or one of many asymmetric information models. Second, there are models of the client and attorney relationship, focusing on the potential agency problems arising out of the various compensations schemes. Third, there are models which deviate from rationality assumptions, where even under symmetric information there can be optimism which causes settlement breakdown. Empirical work has largely focused on the pre-trial stage and has related various data to some of the theoretical models (with the inconsistent priors approach receiving special treatment). In this paper, I focus on appeals where the legal and factual realities are different; and rather than focusing on one particular model, I informally discuss the implications of my findings for a range of possible models of settlement breakdowns and the incentives to appeal and briefly address the implications for potential reform.

This paper is organized as follows: In section 2, I lay out the legal foundation and background on the functions of the federal appellate level and the underlying framework for the breakdown of settlement, and explain the various suspected biases in appellate decisions. In section 3, I survey the relevant literature. In section 4, I present the data. In section 5, I discuss the methodology and present the results, and finally, in section 6, I conclude and discuss future work.

3.2 Legal Foundation and Background

The choice of the federal data is largely one of convenience: federal data is meticulously collected and is available. For the analysis of predictability, I chose appeals since the operative factual reality at the trial level is much less clear. Parties are not always sure what has really happened, or what they can prove or verify that happened, and it is up to the court, after hearing much testimony and evidence, to establish the relevant facts

of the case - the "legal reality". There is much more uncertainty at the trial level and much more maneuverability for the lawyers, who can attempt to shape the determination of the facts (by judge or jury), using the sometime huge body of evidence. This uncertainty leads to much less predictability, and the breakdown of settlements is thus often related to disagreements over this uncertainty. In contrast, at the civil appellate level, courts are supposed to interfere with a trial court decision only if it involves a mistake of law (as opposed to a mistake of fact). The factual reality determined by the trial court is to serve as a foundation for the appellate court, unless there is a "clearly erroneous" decision. This reality is particularly acute when the trial is by jury, as the decision, and the factual reality upon which it was based, are not detailed in formal written opinions. Thus, the uncertainties should mostly concern the legal ramifications of this factual reality. However, statistically, there are very few cases in which "new law" is required; cases are generally very similar to cases that have been heard before and thus the legal implications of a given factual reality should be clear[2].

Civil proceedings are a better testing ground for appellate predictability than criminal proceedings, since in the latter type of proceedings the facts established by the trial level are more frequently reviewed. The perception is that the stigma imposed on a guilty defendant by the criminal process is so severe, that it can be reexamined by the appellate levels. The defendant is thus often given at least two shots at proving his innocence. In addition, these stigmas may be valued very highly by defendants - leading to a frequent asymmetry in stakes. In other words, while in a case regarding monetary compensation decisions can be assumed to largely depend on their expected monetary value, the considerations are much

[2]There is an intermediate standard of appellate review between that of legal determinations (reviewed de novo) and that of factual determinations (reviewed for clear error): decisions that fall within the discretion of the trial judge. These include many cases applying settled legal standards to changing facts, which are reviewed for "abuse of discretion". The standard is supposed to be more restrictive than "de novo" and less restrictive than "clear error".

Another factor is that there are some appeals that are heard without an actual trial having been conducted. These are appeals from decisions on summary judgment, motion to dismiss, etc. In all those cases the appellate court faces the exact same record the trial court did when making the decision under appeal. For some circuits (the second circuit for example), summary judgments are reviewed de novo.

less clear in criminal cases. In addition, criminal proceedings, including criminal appeals, are considerably cheaper than civil proceedings.

The cases that are appealed may not be random. Appealed cases are "survivors"; cases that could not be resolved otherwise, for which it is privately optimal for parties to go to trial. To set ideas, each one of the parties will litigate iff:

$$V_i(litigation) > \max\{V_i(settlement), 0\},$$

where $i = \{a, d\}$, the appellant and defendant. To simplify, assuming risk neutrality, and that parties are motivated strictly by the case at hand (they do not, for example, have reputational concerns), litigation should only occur if there is no settlement amount that both the appellant and the defendant would prefer to going to trial and the expected value of litigation is positive. The appellant would accept any offer greater than $V_a(trial)+SC_a$, where

$$V_a(trial) = E_a[W] - E_a[LC_a].$$

In other words, the appellant would accept any offer greater than the expected value (using his beliefs regarding the probability distribution) of appeal awards (winnings, W, minus his litigation costs LC), plus the costs to him of securing such a deal SC_a.

The defendant would be willing to offer anything lower than $-V_d(trial) - SC_d$ where

$$V_d(trial) = -E_d[W] - E_d[LC_d].$$

In other words, the defendant would be willing to pay her expected losses in trial plus her expected litigation costs in trial, minus the costs of settling. Litigation will happen when

$$V_a(trial) + E[SC_a] > -V_d(trial) - E[SC_d]. \qquad \text{(Litigation)}$$

Under public symmetric information regarding the expected award to the appellant,

this implies:

$$E_d[LC_d] + E_a[LC_a] < E_d[SC_d] + E_a[SC_a],$$

or that litigation happens when the costs of litigation are smaller than the expected costs of settlement. However, even assuming a simple bilateral one shot bargaining game, absent the strong assumption of both parties observing the same information and forming the same expectations, there are many other factors that could explain litigation (where the litigation condition holds): Thus for example, asymmetric information, or asymmetric beliefs, regarding the probability of winning or the stakes, affect the minimum offer that would be accepted by the plaintiff and the maximum offer which would be made by the defendant. This asymmetry is possible due to actual private informational advantages one of the sides may enjoy (such as a superior knowledge of the law, a familiarity with the judge, etc.), or for behavioral reasons - such as overconfidence. It also may result from an agency problem where the lawyers may *convince* their clients that such is the case. Furthermore, even with symmetric information, different priors could lead to appeals.

The particulars of the asymmetric information can lead to different outcomes. For example, the one sided asymmetric information model in Bebchuk (1984) leads to a fairly robust prediction (confirmed below) of the likelihood of settlement decreasing with the size of the stakes, as the optimism is magnified with the increase in the expected award. However, the findings regarding the expected plaintiff win rates at trial (shared by Reinganum and Wilde (1986)) depend heavily on the structure of the information asymmetry and as shown in Shavell (1996), under general assumptions, any plaintiff win rate is possible.

Given the complexity and multiplicity of potential explanations for settlement breakdown, I will begin by examining the potential factors determining the predictability of appeals, ignoring the option for out-of-court settlement. I will then move to informally discuss the interpretation of my findings for a complete model of the incentives for appeal and the causes of settlement breakdown. I chose the factors that may influence predictability in two ways: The first is by examining the role and function of appellate courts, as

set forth in the law (I will call these "legal biases"). The second is by examining the various claims of systematic biases in the decision making of the courts (these will be termed "systemic biases").

3.2.1 Legal Biases – General Deference to the Trial Court

As mentioned above, generally, in civil proceedings, appeals are not viewed as being a "fresh new start" with a higher judicial level. The trial is seen as ending at the district court level. It is the district court that hears all the witnesses and gains a "hands on" impression based upon which it determines "what happened". An appeal is merely a check on this legal decision, made based upon the factual reality determined by the district level. The factual reality is not interfered with, even if it is seemingly mistaken, unless there is something flagrantly wrong. The appellate court mostly examines whether the law was generally determined and applied correctly. The main factors determining the outcome of an appeal are thus whether the case even merits the review of the appellate court: whether the district court decision involved a mistake (of the nature described above), and whether the district court decision was largely based on fact, as opposed to on an elaborate legal determination. If appealed cases do not meet these criteria, there is a high probability that they will be rejected and less room for disagreement about the outcome. However, unfortunately, variables capturing these factors are not (and can not) be made available, as they require a subjective reading of each trial court decision.

There are two tests that can proxy for this legal reality. Given the anecdotal evidence and the common perception of most cases being decided based on existing law applied to the facts of the case, and not involving flagrant errors, we can examine first, whether there is an overwhelming rejection of appeals. This would be the case if appealed cases are representative of all cases tried (that do not merit a second review). A second related test is what percentage of appeals has a concurring or dissenting opinion or is reheard en

banc[3]. This provides an indication of the uncertainty that the case generates. Judges have incentives to write dissenting and concurring opinions, not only in order to try to establish a majority, but also to signal their legal philosophy and affect the future development of the law, which also gives them prestige, recognition, respect and promotion opportunities. This measure is clearly imperfect. On the one hand, predictable cases are sometimes treated seriously for political or public opinion reasons. On the other hand, even complex cases may not engender a separate opinion being written, as judges balance a host of strategic considerations, and face severe burdens on their time. Many cases have only a majority opinion, despite disagreement amongst members of the panel.

Note that the findings here naturally impact the explanation for appeals: A finding of appellate cases being complex cases which often are accepted and over which there is often disagreement lends itself to one of the bargaining failure models for appeals. Parties may find it hard to determine and agree upon the expected outcome of appeals. In contrast, a finding of appeals being representative of the general universe of cases may require other explanations for the violation of the Coase theorem and the breakdown of settlement.

3.2.2 Systemic Biases

Jury Trials

Either party in federal court has a constitutional right to unilaterally demand a trial by jury. The right to a jury has been entrenched in the Seventh Amendment. Thus we should expect the trial to be by jury whenever a jury brings with it an ex-ante advantage to either of the parties (appellant-defendant). As mentioned above, what distinguishes trials by jury is that the decision is not lengthy and detailed. Historically, such trials are seen as reflecting the views of the litigants' peers, perhaps benefiting from a perspective not always available to the judges. The jury is guided by a judge who is to ensure that

[3]Most cases require a decision by the appellate panel to reach this stage.

the jury is accurately weighing the proper questions and considerations and who has the power to interfere with the decision if it is grossly mistaken (especially in matters of law). Accordingly, conventional wisdom is indeed that appeals against jury trials rarely succeed (see Clermont and Eisenberg, 2001, and references therein). However, there have also been claims of juries exhibiting a strong plaintiff-bias; i.e. of them being more sympathetic and empathetic towards the supposed victim-plaintiff than the "colder" judges especially regarding the size of the awards (see for example Helland and Tabarrok 2000). This bias may indeed lead to a need for corrections to be made at the appellate level. There have also been further claims regarding a recent trend of appeal courts intervening even in matters of fact, and looking at the wisdom of the decision as a whole (from my personal experience, this claim is reasonable). Below, I look at the effect of the trial being by-jury on the appeal rate and appeal success rate, both in general and also when the jury ruled for the plaintiff (where the claimed jury bias towards plaintiffs may be more relevant).

US Appellants and US Respondents

Federal judges are appointed by the president. Their allegiance to the federal government and the US as a whole is likely to be an important factor in their appointment[4]. Settlement with the US as a party may also be more complex, given that the US is a repeated player with reputational concerns. This raises the following questions: Are federal courts more likely to defer to the discretion of the Federal Government? Is the US as a party, treated differently in appeals? Can we distinguish between cases in which the US is the appellant and cases in which the US won in the first round and is "targeted" on appeal? These questions will be addressed below. I note however, that these questions are asked with a high level of generality given the aggregate nature of the inquiry here. I expect that looking more closely at individual judges (such as those with promotion ambitions), would

[4]This, of course, despite the fact that, once appointed, judges are - at least formally - insulated from the political sphere.

yield interesting results.

Who won? Plaintiffs vs. Defendants

As mentioned above, there have been claims as to a potential bias towards plaintiffs in jury trials. This raises the following questions: Is there a systematic tendency in either or both of the trial and appeal levels for either party that are not purged by the potential for out-of-court settlement? These tendencies can be related to behavioral models, including for example overconfidence or sunk cost effects, as discussed below.

Procedural Progress

Cases can end at the district level at different stages. I distinguish between cases dismissed at various stages of the trial. I have explored specifications with all 13 levels of dismissal, as well as with coarser categories. There are opposing considerations for the effects of the procedural stage. On the one hand, cases dismissed early may be thought of as being resolved, and thus as not requiring an appeal. On the other hand, litigants may feel that when the case was dismissed "too early" they will not have had "their day in court" and thus will attempt to seek another hearing at the appellate level.

Life of the Case

I distinguish between three time periods: The time that elapses from the date the case was filed in the district court to the time there was a decision at the district court (the first time category); the time that elapses from the filing of a notice of appeal until the final decision on appeal is rendered (the second time category); and the time from the date the case was filed in the district court to the time a final decision was reached on appeal (the third time category). The length of the first and second time periods can result from several factors such as each court's workload, the stage at which the decision is given, the litigants dragging out the case, and the complexity of the case. The third time period is also

influenced by the length of time until an appeal is filed (which may be due to settlement attempts and also to tardiness). The length of these time periods can have several effects. The more complex cases may be cases that courts are interested in accepting (ruling for the plaintiff, or granting the appeal), and the courts may be reluctant to reject cases that have dragged on for so long. However, the longer cases may also have been dragged out by undeserving parties[5].

Self Representation

Conventional wisdom is that self representation is a very bad idea, especially in the "adversarial" system, where it is up to the litigants to raise all the claims. The court generally evaluates the legal claims posed by the litigants and does not help them, especially in civil cases. Thus it is particularly interesting to see how appellate courts view decisions made when the plaintiff represented himself at the district level, and also when the appellant represent himself. It is also interesting to see whether plaintiffs that represent themselves are more likely to appeal. In different specifications, I have examined self-representation by the plaintiff at the district level when the plaintiff won, self representation at the district level when the plaintiff lost, self representation by the defendant when the defendant won, self representation by the defendant when the defendant lost, and self representation by both of the parties. For the appellate outcome I also examine the effect of self representation at the appeal, of the appellant, of the defendant, and of both the appellant and the defendant. Note that appellants that represent themselves do not risk agency problems with their lawyers. However, given their general lack of legal education this effect may be of second order.

[5] If the length of time from the point of the district court decisions to a decision on appeal can be reasonably expected by the parties, the effect may be to make litigation less likely since it reduces the present value of the appellant award and the defendant's costs, thereby creating value (the value of time until recovery) that the parties can use to make settlement more attractive. Delay will of course have a stronger effect on the reservation value of the party with higher discount rates. It also depends on whether the awards can be expected to account for the time that elapsed. However, unfortunately we do not observe the expected length of the case at the point of the decision to appeal.

3.3 Existing Literature

There is a sizeable Law and Economics literature on the selection of cases for the trial level. There seems to be a consensus that tried cases are unrepresentative of the universe of filed cases. However, there are many different models of bargaining failure and settlement. As mentioned, these models generally follow one of two approaches. The first follows the inconsistent priors approach of Priest and Klein (1984), according to which both parties estimate case quality with error, and cases then proceed to trial when, randomly, the plaintiff is sufficiently more optimistic than the defendant. This approach predicts that, as the fraction of tried cases declines[6], the surviving cases will have a 50% likelihood of winning, as disagreement is more likely when the cases are of a quality near the decision standard for winning. As Waldfogel (1995, 1998, 1999) explains, this model implies that plaintiff victories and trial rates are functions of the decision standard (the fraction of cases that would yield plaintiff wins if tried), the uncertainty of the parties regarding the quality of their case, and the degree of stake asymmetry. These measures are unobservable and thus empirical papers testing this model have used proxies for them. Thus for example Kessler, Meites and Miller (1996) look at cases where *they* expect defendants to have higher stakes and Eisenberg and Farber (1997) look at cases where *they* expect more uncertainty. The second approach is one of many asymmetric information models. One of the leading models is Bebchuk's (1984) screening model which assumes that the defendant has private information and thus implies that informed defendants reject settlement offers when they have stronger cases, thus selecting cases where plaintiffs are less likely to win (as compared to the general universe of cases). Of course this selection would be reversed if, in the model, the plaintiff had the private information and the defendant made the offer. Reinganum and Wilde (1986) show that in a signalling model where the informed party makes the offer there is a fully separating equilibrium where the plaintiff mixes between accepting

[6] For example as clients become better at predicting case quality.

the offer and proceeding to trial (and thus receives a settlement offer equal to the payoff in trial). Shavell (1996) shows that any plaintiff win rate in trial is possible under more general assumptions. These models have thus received less empirical treatment and it is difficult to distinguish between the classes of models (compare Waldfogel 1998 that does this using an uninformed plaintiff-informed defendant model and the criticism of this in Daughety and Reinganum 2008). Finally, there are more stylized models of bargaining, examining the role of multiple litigants and specific procedural rules (such as discovery, most favored nations clauses, and joint and several liability, see Daughety and Reinganum 2005, 2008).

There are also several papers model modeling the agency problem between the client and lawyer, where the lawyer pays all of the litigation costs and receives a fraction of the award. In these models lawyers have an incentive to bargain harder at the settlement phase (see Kahan and Tuckman (1995), Polinsky and Rubinfeld (2002), and Chen and Wang (2006)). Alternatively, when lawyers are paid by the hour, their incentives are to prolong the litigation (see Spier 1998, Miller 1987). There are some empirical studies that focus on the relationship between lawyer compensation schemes and the likelihood of settlement (see for example Brickman 2003, and Kritzer, 1998, 2002).

Finally, there are papers that deviate from the assumptions of rationality and assume (based on anecdotal or experimental evidence) optimism (of the litigants and/or their lawyers) to the point of there not being any mutual acceptable settlement amounts (see Loewenstein et al (1993) and Mnookin (1993)).

As to trials at the appellate level, there is some work examining the role of the appellate level and the function of appeal courts (see for example Shavell 1995, and Daughety and Reinganum 2000). However, there has been much less empirical work on predictability and selection in appeals. Of most relevance is the Clermont and Eisenberg (2001) paper that looks at an earlier smaller sample than that considered here and finds large differences between appellate courts and trial juries which they claim to result from appellate judges

"sizeable misperceptions regarding the jury" and some advantages to defendant-appellants which they see as misperceptions regarding pro-plaintiff biases in trial courts. They also find an affirmance rate of about 80% of appeals which they conclude is evidence that case selection has a limited effect at the appellate level and can be ignored in the interpretation of the results.

In sum, many of the models do not generate predictions amenable to their empirical validation, especially across the many dimensions of the potential biases explored here. Furthermore, the appellate environment changes some of the predictions. For example, Waldfogel's (1999) claim that there is more uncertainty with juries refers to the trial level and says nothing about the appeal where, as mentioned, we may actually expect more certainty. The approach taken here is thus to estimate the significant tendencies in the data for the selection of cases for appeal and the appeal outcomes and then to informally discuss the implications for the main classes of models. This work can thus be seen as preliminary work in preparation for the estimation of the magnitudes of the particular form of settlement breakdown, such as informational asymmetry or agency costs from a more general structural model, which I plan to explore in the future.

3.4 Data

I used the data on civil terminations and on appellate terminations from the Federal Integrated Data Base. This data is gathered by the administrative office of the United States Courts, assembled by the Federal Judicial Center, and disseminated by the Inter-University Consortium for Political and Social Research (ICPSR). It includes information on every federal case terminated since 1970. The procedure is that whenever a federal case is terminated a court clerk submits a form with information about the case, including (but not limited to) the names of the parties, the category of the subject manner (there are 98 categories), the jurisdictional basis for the case (i.e. why federal?), the amount demanded, the amount received, the dates of filing and termination, the procedural method of disposition,

and which party prevailed.

I chose the years 1992 – 2003. This, since in my dataset the median amount of time in the third time category (of the time categories discussed above) is 32.4 months and I wanted to capture a sufficiently long period. I chose to go back to 1992 since there were some differences in the coding of variables in earlier years. The civil datasets are linked to each other and to the appeal datasets, using the docket number that each case receives when it is filed. I also used the year, circuit and district in which the cases were filed, and the category for the area of law which they regard. For duplicates (which largely result from several decisions being made on a given case), I chose the latest appellate decision and the last civil decision, which is generally the most substantial one. This leaves out about 10% of the appeal decisions and approximately 0.65% of the district court cases. Next I chose to limit the categories to cases largely involving monetary claims[7]. I drop prisoner petitions, and require that each case have a nonzero monetary demand (at the district level). The categories with which I remained are detailed in a chart 1 in the appendix.

I am left with 17,679 cases which terminated between the years 1992 – 2003 for which there is both civil and appellate data, and 469,764 district cases in total (from which appeals are made).

3.5 Methodology and Results

3.5.1 Legal Biases

As mentioned, given the deference to the district level, absent settlement, we expect a high level of rejected appeals. And indeed, the percentage of appeals that are rejected is overwhelming: In my dataset, out of my 17,680 observations, only 12.42% of the cases

[7] It may be possible to refine the categories further in order to isolate, from within the cases that are financially motivated, the cases that are primarily concerned with the specific case at hand (rather than the effects on other future cases), such as by looking at the identity of the parties. I plan to return to this in future work.

achieved some goal, either reversing, entirely or in part, or remanding the case to the district court. Only 7.94% of the cases were considered to have been reversed. The percentage of cases in which there were differing opinions in the panel is also exceedingly small. Of the cases for which this was tracked[8], in 91.86% of the cases there was only one opinion of the court and the court did not sit en banc (i.e. did not expand the panel[9]). This may be seen as evidence of the simplicity of most appeals. Thus, a simple and robust coarse prediction for appellate outcomes is that they will be rejected. Both findings therefore, are facts to be explained by a model of appeal incentives, although, as mentioned, the second findings may be explained in other ways including the time constraints and strategic interactions between judges as repeated players. There is therefore no evidence that appealed cases are particularly complex cases.

Appeals as a Low Probability - High Return Gamble

This overwhelming large percentage of rejections naturally leads to the question of whether appeals are seen as a costly low probability gamble with high expected returns. The data does not include the amounts received on appeal, is top-coded at 10 million dollars at the civil level, and the ability to follow through cases that are remanded to the district court is limited. In addition, it is hard to get a reliable estimate of the actual attorney costs expended on a given appeal. Thus, the strategy I employed was to compute a generous proxy for the amount awarded in the appeal (as captured by the "stakes" variable described above). The amount is calculated as the amount *awarded* at the trial level - if an amount was awarded at the trial level and the appeal was accepted - and at the amount *demanded* at the trial level - if the plaintiff lost at the trial level and won on appeal. This measure is particularly generous, since the awards are generally only a very small fraction of the amount demanded, and, as mentioned, my definition of winning if very broad. The mean

[8] This is slightly over 40% of the cases.

[9] There may be some issues with the coding of this last variable.

of the amount received on appeal is $366,000 while the mean of the amount demanded is $2,014,000. Now, adding up these inflated wins and dividing them by the total number of appeals, yields an expected gain of $258,640 dollars per appeal. The average length of an appeal (the time from when the appeal is filed until the final judgment is given) is 9.7 months and the average attorney fee is 250 dollars per hour, and so it is hard to conclude from this generous measure whether the appeal is, on average, wasteful. Chart 2 details the average success rates, winnings, and duration of the case, by case category. The results are mixed. Thus for example for category 355, "Motor Vehicle Product Liability" the expected winnings are approximately 550 thousand dollars, where for category 152, which deals with student loans, they are under 700 dollars. It is clear that there is much dispersion in the expected winnings, and that, for some categories, appeals may indeed be profitable, but for others they are clearly not. There is likely much dispersion in the costs of legal representation for the various categories. More detailed work is required to model and to understand the motivations in the specific categories. While the precise objective function for appellants, and the manner in which they form beliefs are not known, Fournier and Zuehlke (1989) find, for example, that movements in the awards in completed trials in the same category, influence the litigants' settlement decisions. This suggests that such calculations are being made to some degree and that these expected measures should be considered. Cases with a negative expected value for appeals require an explanation for why they are filed in the first place. Here, the motivation for the lack of settlement by the defendant is clear: there is no reason for her to pay when the outside option of the appellant can be reasonably expected to be negative. The motivation for the appellant is more difficult to rationalize. Such cases may be more amenable to an explanation regarding the agency problems with the attorneys (rather than sustainable optimistic beliefs in an environment where appeals are, on average, wasteful to file). Attorneys may not have the incentives to provide accurate information to the clients (for example in one shot relationships with clients with little ramifications on future business, where the attorney is

compensated based on the amount of time worked and not as a percentage of winnings). The validity of such rationales can be explored by looking more closely at these classes of cases. There also may be room for behavioral biases to explain such cases, where the appellant wrongly believes that even though, on expectation, appeals are a bad idea in this class of cases, his case is special. Once again, such explanations require more structure as to the exact nature of the behavioral bias in the given case.

3.5.2 Systemic Biases

I use simple probability models for both the selection of cases for appeal (from the cases tried), and the propensity of the appellant to succeed. The right-hand-side variables are similar in both specifications, with the obvious exception of the variables that are not available at the time of appeal (such as some of the time categories). Naturally, for the selection equation I used the entire merged database of district and appellate cases, while for the outcome equation I only used the cases which could successfully be linked. I control for the case categories, circuit fixed effects, and time fixed effects (for the year the trial case and appeals were decided), and for district fixed effects when estimating the selection equation for the trial districts. I compare average marginal effects under a logit and probit model with a linear probability model. Note that my data is a panel of repeat cross sections of cases (which sometimes take several periods to be decided). Despite the many fixed effects, the large number of observations (on average there are over 1000 observations per fixed effect) in this static model should compensate for them (as the bias is proportional to the variance of the parameter estimate) especially when average marginal effects are used, or in other words the bias arising from the "incidental parameter problem" is likely to be small (compare Hahn and Newey, 2004; Fernandez-Val 2009). Standard errors are clustered by districts to account for serial correlation.

As mentioned, the key omitted variable is whether the trial level decision resulted in

a mistake that merits appellate intervention[10]. However, there is no obvious endogeneity problem (and thus no obvious instruments). Similarly, while the sample of cases ending up in an appeal is likely not random, this paper focuses on the determinants of the cases which have been appealed (rather that all cases).

Specifications

Tables 3 and 4 present the results for the two specifications below respectively:

$$appealed = appealed(category, district, circuit, time, beforetrial, districttime, demanded\ jury, plaintiffwon, plaintiffwonjury, selfrep, selfred)$$

$$won = won(category, circuit, time, beforetrial, usappellant, us\_respondent, totaltime, appealtime, districtime, jury, plaintiffwon, plaintiffwonjury, selfrep, selfrepa)$$

Summary statistics for the variables are detailed in tables 1 and 2 below. The binary variable "**Appealed**" indicates whether the trial case appears in the appeal database, during the time frame. Only a small percentage of cases are appealed. As mentioned above, only about 4% of the cases show up as appeals, which is consistent with the percentages which can be inferred from the US Courts publications. The binary variable "**Won**" is a dummy variable denoting a successful appeal. An appeal was deemed to have been successful (or "won"), when the trial decision was reversed, entirely or in part, and when

[10] As discussed above, this is when there is a mistake in law, or an extreme patenly obvious error in determining the factual basis of the case. See also note 2.

the case was remanded to the district court. Cases were not followed to their final outcome when they are sent back to the lower level.

"**Before Trial**" is a dummy for cases that were dismissed at early stages of the trial, including before the issue was joined, and before or during pretrial. "**Jury**" is a dummy variable denoting whether the district trial was by jury. "**US Appellate**", and "**US Respondent**" are dummy variables indicating whether the US initiated or was a respondent to the appeal. "**Total Time**" is the total time the case was alive from the time it was filed until a final decision was reached on appeal; "**District Time**" is the length of time from the time the case was filed until a final decision was reached at the district court; "**Appeal Time**" is the total amount of time that elapsed from the time the appeal was filed until a final decision was reached on appeal. "**Plaintiff Won**" is a dummy variable indicating whether the plaintiff won at the district level. "**Plaintiff Won-Jury**" is an interaction dummy variable indicating whether the plaintiff won and it was a jury trial. "**Amount Demanded**" indicates the amount demanded at the district level.

The self representation dummy variables include: "**Self Rep**" indicating whether the plaintiff represented himself at the district level; "**Self Rep (defendant)**" indicating whether the defendant represented himself at the district level; and "**Self Rep (appeal)**" indicating whether the appellant represented himself on appeal"[11]

Finally, fixed effects for case categories, circuit, district, and time period (for the year the trial case and appeal were decided) are included.

[11] As mentioned above, in other specifications (not reported I have also looked at whether the plaintiff represented himself at the district level and won; whether the plaintiff represented himself at the district level and lost; whether the defendant represented himself at the district level and won; indicating whether the defendant represented himself at the district level and lost; whether both the defendant and plaintiff represented themselves at the district level; whether the defendant represented himself on appeal; and whether both the plaintiff and the defendant represented themselves on appeal. The variables presented are representative for the effects of self representation.

Summary of Findings

Given the fewer observations for self representation, I present the specifications with and without these variables. As can be seen, there are several systemic factors that significantly influence the probability of winning the appeal:

As a rule, jury trials are less likely to be overturned. However, when the plaintiff wins at the trial level, appeals are more likely to be successful. The US seems to do better, as it is more likely to win when it is the appellant, and less likely to lose when it is a defendant. Note that this does not imply favoritism, since cases can not be assumed to be selected at random. Cases that took longer either at the district level or at the appeal level, have a slightly better chance at being successful; and the time it takes for the decision to be rendered is a positive predictor of success. Once again, this result should be interpreted with caution. It is likely that many cases are thrown out immediately, and so, conditional on surviving for a small period of time, the length of time it takes to get a decision may not be as predictive. An appeal when the plaintiff won at the district level is more likely to be successful, but, as mentioned, the added likelihood of success increases dramatically when the lower level trial was by jury. This suggests that appeal courts are comfortable trusting juries when they throw out the case, but not when they make awards for the plaintiff. Finally, as expected, self representation is not a good idea. Parties that represent themselves at either the trial or appellate level are less likely to win an appeal.

Looking at the selection equation results, we find, once again, that the length of time is positively related to the decision to appeal (although the magnitude is very small). As the stakes are higher, and the amounts demanded increase, parties are more likely to appeal as well. Appeals are made more frequently when the plaintiff wins, which is consistent with the outcome equation, however jury decisions are more commonly appealed, even though these appeals are less likely to succeed. Furthermore, there seems to be a tendency NOT to appeal jury trials when the plaintiff is successful in a jury trial, which is inconsistent with the higher tendency for such appeals to be overturned.

It is on this last point where the three types of probability models most differ (and even sometimes have opposing signs). Recall from the summary statistics that the plaintiffs winning in jury trials is a pretty infrequent event in the data, which happens less than 1% of the time. Interestingly, other than this difference, the probability models generally behave in the same way, despite less than 4% of the cases going to appeal. In parallel work with Whitney Newey I am exploring a more general categorization of cases under which these models differ.

Discussion

As mentioned, these findings both are influenced by and inform the modeling of the incentives to appeal. I illustrate some of the implications of the findings below[12]:

Jury trials are more likely to be appealed. However, consistent with the cases surviving to appeal being representative of the general body of cases, jury trials are deferred to at the appellate level. Similarly, while the results regarding the propensity to appeal are not consistent across specifications, in the two (preferred) average marginal effect specifications, jury trials where the plaintiff won at the district level are slightly less likely to survive to appeal. However, those that do survive are more likely to succeed, which may be explained by the common perception of juries being plaintiff biased. Subject to the further caveat of there being very few jury trials[13], these findings challenge the explanations for settlement breakdown.

For the bargaining breakdown explanations, we would expect jury trials to be more likely to settle and less likely to survive, since they generally present less grounds for intervention, and thus there should be more agreement, based on an interpretation of the law and the district court decision. A model would thus have to produce results where

[12] Given the small impact of the life of the case, and the insignificant impact of the procedural stage, these are omitted from the discussion.

[13] A separate interesting question is the selection of cases for jury. As mentioned, either party can request this unilaterally.

there is more private information (perhaps stemming from other sources) in these cases, or some reason to have more inconsistent priors. Similarly, we may expect there to be more asymmetric information when the plaintiff wins in jury trials since there are the two (potentially opposing) effects: the general deference to juries (due to the nature of their decision making process and short judgments) and the various perceptions regarding the juries potentially being too friendly to the plaintiff. Parties may also differ in their beliefs regarding how fair the awards were (especially in the absence of written formal decisions). However, in the probit and logit specifications we find the opposite tendency of these cases being less likely to show up as appeals. Once again, the model would have to explain why the surviving cases exhibit similar tendencies, regarding appeal success rates, to those which could have been predicted absent selection, despite the cases surviving to appeal not being random with respect to these variables. Here structural estimation which could produce estimates of magnitudes would be useful. Informally, the findings could result from the tendencies speculated (the nature of the jury decision), but, potentially, also from these cases being more (or less) likely to survive to appeal, where even if the entire universe of decided jury cases are similar in the validity of the judgment to all cases, the selected sample of those appealed is less (or more) likely to be worthy of intervention.

An agency story would have to explain why clients may be more easily fooled (into proceeding to trial) in jury trials. For example it could be claimed that juries can be portrayed as being less professional and thereby "wrong" and clients may believe they can indeed "remedy" this wrong on appeal. Similarly a behavioral bias would give some foundations to litigants being less likely to agree when the trial was by jury, perhaps feeling that a "true understanding" of their case requires expert judges.

Regarding cases involving the US as a party, we do not have information on the propensity to appeal. However, the results on the US being more likely to succeed - or a deference to the federal government - which is consistent with the perceptions discussed above, still requires an explanation as to why these tendencies are not purged by settlement. In other

words, why are cases with the US as a party not settled more frequently leading to a selected sample of cases that go to trial in which, on average, the US has a weaker case, where the outcome would be less expected? Once again returning to the potential explanations, an asymmetric information story would have to claim perhaps that there is more private information in cases in which the US is a party, where certain judges are more friendly to the government or more variance in the noise for the inconsistent prior explanation. Such claims require a closer look at cases involving the US. Similarly, agency or behavioral stories would have to argue that clients do not internalize the US tendencies. It is of course also possible that the reason for the finding is that selection works the other way, and cases involving the US settle in proportions leaving the appealed cases to be those where the US has a stronger case. Once again such explanations would have to incorporated in the model.

Cases in which plaintiffs won are more likely to reach appeal and are more likely to be successful on appeal, although the latter magnitude is significantly smaller. Appellants seem to be acting based upon the perception of some plaintiff bias and the selected cases tried at the appellate level seem to still be more stronger. This can be explained if there is more uncertainty in such cases or more private information, but seems more likely to be related to a behavioral biases. These can include an effect of plaintiffs winning in court reinforcing their belief regarding the merits of their case and causing them to refuse to settle, or alternatively a sunk cost effect, where defendants refusing to accept the loss in trial (compare for example Genesove and Mayer, 2001). However, a full analysis of magnitudes (and the interpretation of the difference in magnitudes and the predicted features of the surviving cases) likely requires the structural estimation of a particular model.

The size of the stakes makes appeals more likely. As mentioned, this result naturally feeds into models of bargaining failure. When there are large amounts at stake even small disagreements (in terms of the deviation in subjective probabilities) may make the high

costs of litigation worthwhile. It may, however, also feed into behavioral models regarding the refusal to accept large losses (if the defendant loses), or large incentives for the lawyers to push forward when their compensation scheme justifies it (such as when their hourly compensation is higher).

Finally self representation leading to more appeals is most likely to be the result of "crazy" litigants. Recall that these are generally unskilled litigants that "believe" in their own skills to the degree that they do not feel it worthwhile to hire (or accept) representation. It is thus easy to think of similar reasons for them rejecting reasonable offers in the settlement phase and for them to still perform poorly at the appeal.

Taken as a whole, the results of this preliminary work can be seen as stylistic facts that structural models of appeal motivation and selection must explain. On the face of it, it seems highly unlikely for one model to explain all of the results. However, there is clearly room for much theoretical work and more precise structural estimation of such models and the testing of competing predictions that may differentiate them in more specific scenarios.

3.6 Conclusion and Future Work

In the above I used a relative new and unexplored dataset to examine the issue of the predictability of judicial decisions. I found that first and foremost appeals are predictably rejected. I cannot conclude from this inspection of the data whether it would be possible to form rational predictions when more features of the case are taken into account, which would make the decision to appeal justified in terms of the expected profits the appeal generates. It seems that this is true for some types of cases but not for others. There is no evidence of settlements having a selection effect whereby cases that survive to appeal are complex and have uncertain outcomes. The predictability of appeals can be explained, to some degree, as though there were no outside settlement option at all. I have very informally discussed how different models of settlement failure and appellate incentives may go about explaining some of the main findings.

A more complete understanding of the realities at the appellate level then feeds into the consideration of reform. If indeed frivolous litigation is to be deterred, an understanding of the motivation of negative expected value appeals is needed. Similarly, a finding concluding that there are inexplicable breakdowns of bargaining, questions whether the number of cases tried could not be reduced by simply improving the settlement process, such as by offering venues for the parties to discuss settlement and more indications of the expected reality at the appellate level.

The extremely preliminary nature of this work makes the extensions straightforward. As mentioned, first and foremost, a formal model of appeal incentives which can yield predictions consistent with these findings should be specified. Such a model should be amenable to the structural estimation of the magnitude of the underlying causes for appeals (such as agency costs or asymmetric information), which would feed into the policy recommendations of the sort sketched above. I plan to return to these in future work.

Bibliography

[1] Alicke, Mark D., M.L. Klotz, David L. Breitenbecher, Tricia J. Yurak and Debbie S. Vredenburg (1995), "Personal Contract, Individuation, and the Better-Than-Average Effect, *Journal of Personality and Social Psychology* 6, 804.

[2] Angrist, Joshua D. (2001), "Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice", *Journal of Business and Economic Statistics* 19(1) .

[3] Bailis, Daniel S. & Robert J. MacCoun (1996), Estimating Liability Risks with the Media as Your Guide: A Content Analysis of Media Coverage of Tort Litigation, 20 *Law & Hum. Behavior.* 419, 426 .

[4] Bebchuk, Lucian Arye, (1984), "Litigation and Settlement under Imperfect Information", *The RAND Journal of Economics*, 15(3), 404 – 415.

[5] Bertrand, Marianne, Esther Duflo and Sendhil Mullainathan (2004), "How Much Should we Trust Differences-In-Differences Estimates", *Quarterly Journal of Economics,* 119(1) 249 – 275.

[6] Brickman, Lester (2003), "Effective Hourly Rates of Contingency-Fee Lawyers: Competing Data and Non-Competitive Fees", 81 *Wash. U. L. Q.* 653.

[7] Camerer, Colin and Dan Lovallo.(1999), "Overconfidence and excess entry: an experimental approach," *American Economic Review* 89(1), 306-318.

[8] Chen, Kong-Pin, and Jue-Shyan Wang, (2006), "Fee Shifting in Litigation with Contingency Fees", *Journal of Law, Economics and Organizations*, 23, 519 – 546.

[9] Clermont, Kevin M. and Theodore Eisenberg (2001), "Appeal from Jury or Judge Trial: Defendants' Advantage", *American Law and Economics Review:* 3(1), 125.

[10] Clermont, Kevin M., and Theodore Eisenberg,(2002), "Litigation Realities", *Cornell Law Review*, 88, 119 – 154.

[11] Cooter, Robert D. and Daniel L. Rubinfeld (1989), "Economic Analysis of Legal Disputes and Their Resolution", *Journal of Economic Literature,* 27, 1067.

[12] Cross, Frank B., Decisionmaking in the U.S. Circuit Courts of Appeals, *California Law Review*, 91(6), 1457 – 1515.

[13] Daughety, Andrew F. and Jennifer F. Reinganum,(2000), "Appealing Judgments", *The RAND Journal of Economics*, 31(3), 502 – 525.

[14] Daughety, Andrew D. and Jennifer F. Reinganum, (2005), "Economic Theories of Settlement Bargaining", *Annual Review of Law and Social Science*, 1, 35 – 59.

[15] Daughety, Andrew F. and Jennifer F. Reinganum, (2008), "Settlement", in *The Encyclopedia of Law and Economics*, vol. 10, Procedural Law and Economics. Chris William Sanchirico ed.

[16] Eisenberg, Theodore, and Henry Farber, (1997), "The Litigious Plaintiff Hypothesis", *RAND Journal of Economics*, 28, S92–S112.

[17] Eisenberg, Theodore, Paula L. Hannaford-Agor, Michael Heise, Neil LaFountain, G. Thomas Munsterman, Brian Ostrom and Martin T. Wells (2006), "Juries, Judges, and Punitive Damages: Empirical Analyses Using the Civil Justice Survey of State Courts 1992, 1996 and 2001 Data", *Journal of Empirical Legal Studies*, 3(2), 263 – 295.

[18] Fischman, Joshua B. (2007), "Decision-Making Under a Norm of Consensus: A Structural Analysis of Three-Judge Panels", mimeo.

[19] Folkes, Valerie S. (1988), "The Availability Heuristic and Perceived Risk", *Journal of Consumer Research* 15.

[20] Fournier, Gary M. and Thomas W. Zuehlke (1989), "Litigation and Settlement: An Empirical Approach", *The Review of Economics and Statistics*, 189.

[21] Fernandez-Val, Ivan (2009), "Estimation of Structural Parameters and Marginal Effects in Binary Choice Panel Data Models with Fixed Effects", mimeo.

[22] Genesove, David and Christopher Mayer, (2001), "Loss Aversion and Seller Behavior: Evidence from the Housing Market", *Quarterly Journal of Economics*, 116, 1233 – 1260.

[23] Hahn, Jinyong and Whitney Newey (2004), "Jackknife and Analytical Bias Reduction for NonLinear Panel Models", *Econometrica* 72(4), 1295.

[24] Helland, Eric, and Alexander Tabarrock (2000), "Runaway Judges? Selection Effects and the Jury", *Journal of Law, Economics and Organization*, 16(2), 306 – 333.

[25] Hersch, Joni and W. Kip Viscusi, "Punitive Damages: How Judges and Juries Perform" (2004), 33 *Journal of Legal Studies* 1.

[26] Jughes, James W. and Edward A. Snyder (1995), "Litigation and Settlement under the English and American Rules: Theory and Evidence", *Journal of Law and Economics*, 38(1) 225.

[27] Kahan, Marcel, and Bruce Tuckman, (1995), "Special Levies on Punitive Damages: Decoupling Agency Problems and Litigation Expenditures", *International Review of Law and Economics*, 15, 175 – 185.

[28] Kahneman Daniel and Amos Tversky (1982), *Judgment under Uncertainty: Heuristics and Biases*, New York, Cambridge University Press.

[29] Kessler, Daniel, Thomas Meites, and Geoffrey Miller (1996), "Explaining Deviations from the Fifty Percent Rule: A Multimodal Approach to the Selection of Cases for Litigation", *The Journal of Legal Studies*, 25, 233.

[30] Kritzer, Herbert M.(1998), "The Wages of Risk: The Returns of Contingency Fee Legal Practice", 47 *DePaul L. Rev.* 267.

[31] Kritzer, Herbert M. (2002), "What we Know and Do Not Know about the Impact of Civil Justice on the American Economy and Policy: Lawyer Fees and Lawyer Behavior in Litigation: What does the Empirical Literature Really Say?, 80 *Tex. L. Rev.* 1943.

[32] Kruger, Justin. (1999), "Lake Wobegon Be Gone! The "Below-Average Effect" and the Egocentric Nature of Comparative Ability Judgments," *Journal of Personality and Social Psychology,* 77(2), 221-232.

[33] Kuran, Timur and Cass R. Sunstein, (1999), "Availability Cascades and Risk Regulation", 51 *Stan. L. Rev.* 683.

[34] Loewenstein, George, Samuel Issacharoff, Colin Camerer and Linda Babcock, (1993), "Self Serving Assessments of Fairness and Pretrial Bargaining", *Journal of Legal Studies*, 22, 135 – 158.

[35] Malmendier Ulrike and Geoffrey Tate (2008), "Who Makes Acquisitions? CEO Overconfidence and the Market's Reaction, *Journal of Financial Economics*, $89(1), 20 - 43$.

[36] Malmendier Ulrike and Goeffrey Tate (2005a) "Does Overconfidence Affect Corporate Investment? CEO Overconfidence measures Revisited", *European Financial Management*, $11(5), 649 - 659$.

[37] Malmendier Ulrike and Goeffrey Tate (2005b), "CEO Overconfidence and Corporate Investment", *Journal of Finance,* 60(6), 2661 – 2700.

[38] Miller, Dale T. and Michael Ross (1975), "Self-Serving Biases in the Attribution of Causality: Fact or Fiction", *Psychological Bulletin* 85(2) 213.

[39] Miller, G. (1987), "Some Agency Problems in Settlement", *Journal of Legal Studies*, 16, 189 – 215.

[40] Mnookin, R. (1993), "Why Negotiations Fail: An Exploration of the Barriers to the Resolution of Conflict", *Ohio State Journal on Dispute Resolution*, 8, 235 – 256.

[41] Polinsky, A. Mitchell, and Daniel L. Rubinfeld, (2003), "Aligning the Interests of Lawyers and Clients", *American Law and Economics Review*, 5, 165 – 188.

[42] Posner, Richard A. (1973), "An Economic Approach to Legal Procedure and Judicial Administration", *The Journal of Legal Studies*, 2(2), 399 – 458.

[43] Priest, George L., and Benjamin Klein, (1984), "The Selection of Disputes for Litigation", *The Journal of Legal Studies*, 1 – 55.

[44] Oppenheimer, Daniel M., (2004), "Spontaneous Discounting of Availability in Frequency Judgment Tasks", *Psychological Science*, 15(2) 100.

[45] Reinganum, Jennifer F. and Louis L. Wilde (1986), "Settlement, litigation, and the allocation of litigation costs", *Rand Journal of Economics* 17(4), 557.

[46] Ross, Michael, and Fiore Sicoly (1979), "Egocentric biases in availability and attribution", *The Journal of Personality and Social Psychology* 37 322. (Also in Judgment Under Uncertainty: Heuristics and biases (Daniel Kahneman, Paul Slovic and Amos Tversky edi. (1982)).

[47] Rhode, Debora, L. (2004) "Essay: Frivolous Litigation And Civil Justice Reform: Miscasting The Problem, Recasting The Solution", 54 *Duke L.J.* 447.

[48] Shavell, Steven (1995), "The Appeals Process as a Means of Error Correction", *Journal of Legal Studies* 24(2) 379.

[49] Shavell, Steven (1996), "Any Probability of Plaintiff Victory at Trial is Possible", Journal of Legal Studies, 25, 493 – 501.

[50] Shavell, Steven (2006), "The Appeals Process and Adjudicator Incentives", *Journal of Legal Studies*, 35, 1.

[51] Siegelman, Peter and Joel Waldfogel (1999), "Toward a Taxonomy of Disputes: New Evidence Through the Prism of the Priest/Klein Model, *The Journal of Legal Studies*, 28(1), 101 – 130.

[52] Spier, Kathryn and Bruce Hay, (1998), "Settlement of Litigation", in Peter Newman ed., *The New Palgrave Dictionary of Economics and the Law*, Macmillian Reference Limited, 442 – 451.

[53] Spier, Kathryn, (2007), "Litigation", in A. Mitchell Polinsky eds., *The Handbook of Law and Economics*, North Holland (2007).

[54] Spier, Kathryn, (2008), "Economics of Litigation", in Blume and Durlauf, eds., *The New Palgrave Dictionary of Economics*, 2nd Edition, Palgrave Mcmillian.

[55] Taylor, Shelly E.(1982), "The Availability bias in Social Perception and Interaction", in *Judgment Under Uncertainty: Heuristics and biases* (Daniel Kahneman, Paul Slovic and Amos Tversky edi. (1982)) 190.

[56] Thomas, Robert E. (1998), "Psychological Impact of Scrutiny on Contingent Fee Attorney Effort", 101 *W. Va. L. Rev.* 327.

[57] Tversky Amos, and Daniel Kahneman (1973), "Availability: A Heuristic for Judging Frequency and Probability", *Cognitive Psychology*, 5 (September), 207-232.

[58] Tversky Amos, and Daniel Kahneman (1992), "Advances in Prospect Theory: Cumulative Representation of Uncertainty", 5 *Journal of Risk and Uncertainty*, 297.

[59] Waldfogel, Joel (1995), "The Selection Hypothesis and the Relationship between Trial and Plaintiff Victory", *The Journal of Political Economy*, 103(2), 229 – 260.

[60] Waldfogel, Joel (1998) "Reconciling Asymmetric Information and Divergent Expectation Theories of Litigation", *Journal of Law and Economics*, XLI, 451 – 476.

[61] Viscusi, W. Kip (1997), "Alarmist Decisions with Divergent Risk Information", *Econ. J.*, 107, 1665 – 70.

[62] Weinstein, Neil D.(1980), "Unrealistic Optimism About Future Life Events," *Journal of Personality and Social Psychology*, 39, 806-820.

[63] Wooldridge, Jeffrey M. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press.

Chart 1

Description of Case Categories

Contracts	Real Property	Torts – Personal Injury
Indemnity on admiralty cases	Land Condemnation	Airplane
Insurance	Foreclosure	Airplane product liability
Marine	Rent, lease, ejectment	Assault, libel and slander
Miller Act	Torts to land	Federal Employers' liability
Negotiable Instruments	Tort - product liability	Marine
Judgments overpayments, enforcements	Other real property actions	Marine - product liability
Medicare Act: Overpayments		Motor vehicle
Student loans		Motor vehicle - product liability
Veteran benefits		Other personal injury
Stockholders suits		Workman's Compensation(Accidents)
Other contract actions		Personal Injury - medical malpractice
Employee relations		Personal Injury - product liability
State Fair Trade Act		Asbestos personal injury - product liability
Hospital Care Act		
Contract product liability		

Torts – Personal Property Damage	Property Right	Misc.
Other Fraud	Copyright	Antitrust
Truth In Lending	Patent	Securities, Commodities, Exchange
Other personal property damage	Trademark	Tax Suits
Property damage - product liability		

Chart 2

Category	Average Winnings	Average Duration	Success Rate	Nonconsensus	Observations
110	232.5783	10.05249	.1465084	.0859564	2191
120	145.1517	9.196581	.1324786	.0320513	468
130	79.58156	9.29078	.106383	.0714286	141
140	123.6514	8.637019	.0889423	.0323887	416
150	26.07813	10.48438	.078125	.0714286	64
151	563.0833	10.33333	.1666667	.1666667	36
152	.6984925	4.743719	.0251256	.030303	199
160	837.6515	9.712121	.1969697	.1	66
190	196.1671	9.379337	.1238612	.0528455	5159
195	339.7209	9.883721	.127907	.0833333	86
210	78.11765	11.94118	.1176471	.2	17
220	83.65086	8.538794	.0517241	0	232
230	180.9828	9.172414	.0862069	.047619	58
240	371.217	10.07752	.1782946	.1132075	129
245	224.6522	13.17391	.173913	.1	23
290	201.5667	9.6	.1190476	.091954	210
310	279.5347	11.0297	.1386139	.175	101
315	390.8333	13.2	.1	.2	30
320	314.0583	9.220834	.0979167	.0869565	480
330	295.6583	8.967626	.1366906	.0763359	278
340	114.5632	9.296492	.1	.0357143	570
345	447.4211	9.631579	.1052632	.1111111	19

Category	Average Winnings	Average Duration	Success Rate	Nonconsensus	Observations
345	447.4211	9.631579	.1052632	.1111111	19
350	257.4851	8.483734	.1074965	.0462046	707
355	549.7806	10.70323	.1290323	.0784314	155
360	235.3898	9.044063	.1023152	.0715726	2678
362	375.8875	9.05726	.1042945	.0722222	489
365	310.8146	9.741617	.122288	.0941176	1014
368	1429.74	25.13842	.3531073	.361194	354
410	666.6558	11.21311	.1229508	.1157895	244
850	502.1284	10.34925	.1432836	.0980392	335
870	130.552	9.40411	.1383562	.0939597	730

This table is computed for the purposes of examining the legal biases in section 5.1. Average winnings are in thousands of dollars. Average duration is in months. Nonconsensus represents the percentage of cases for which the court did Not sit en banc and there were no concurring or dissenting opinions.

Table 3.1: Summary Statistics (District Court Decisions)

Variable	Mean	Std. Dev.	N
Very Early	0.419	0.493	479092
Early	0.34	0.474	479092
During Pretrial	0.104	0.306	479092
During Trial	0.006	0.077	479092
After Trial	0.027	0.161	479092
After Arbitration	0.001	0.036	479092
Jury	0.017	0.13	479092
US Appellate	0.002	0.044	479092
US Respondent	0.007	0.086	479092
Plaintiff Won	0.221	0.415	479092
Self Rep	0.024	0.154	251977
Plaintiff Won - Jury	0.008	0.092	479092
District Time	11.803	14.786	469764

Table 3.2: Summary Statistics (Appeals)

Variable	Mean	Std. Dev.	N
Very Early	0.175	0.38	17679
Early	0.485	0.5	17679
During Pretrial	0.125	0.33	17679
During Trial	0.024	0.153	17679
After Trial	0.216	0.412	17679
After Arbitration	0.002	0.046	17679
Jury	0.14	0.347	17679
US Appellate	0.031	0.173	17679
US Respondent	0.125	0.331	17679
Plaintiff Won	0.259	0.438	17679
Self Rep	0.109	0.312	7375
Self Rep (Appeal)	0.068	0.252	17679
Plaintiff Won - Jury	0.084	0.277	17679
Total Time	32.423	21.643	17679
Appeal Time	9.700	8.178	17679

These summary statistics refer to the variables used in the selection and outcome models.

Table 3

Outcome Equation

	Probit	Logit	LPM	Probit	Logit	LPM
	(1)	(2)	(3)	(1)	(2)	(3)
Before Trial	-.001	-.002	-.003	.0005	.0005	.0006
	(.006)	(.006)	(.007)	(.004)	(.005)	(.005)
Jury	-.031**	-.029*	-.030*	-.031***	-.032***	-.033***
	(.015)	(.015)	(.016)	(.009)	(.009)	(.010)
US Appellate	.035	.039	.046*	.026	.032*	.041***
	(.028)	(.029)	(.026)	(.017)	(.017)	(.015)
US Respondent	-.007	-.005	-.007	-.024***	-.026***	-.026***
	(.014)	(.015)	(.013)	(.009)	(.010)	(.009)
Total Time	-.0006	-.0006	-.0007	.00004	-1.00e-05	-.0001
	(.0004)	(.0004)	(.0005)	(.0002)	(.0002)	(.0002)
Appeal Time	.013***	.012***	.016***	.011***	.011***	.015***
	(.0009)	(.0009)	(.001)	(.0006)	(.0006)	(.0007)
District Time	.001**	.001**	.001**	.0005**	.0006**	.0008***
	(.0004)	(.0004)	(.0005)	(.0003)	(.0003)	(.0003)
Plaintiff Won	.0003	.0005	.003	.014*	.015**	.017**
	(.012)	(.013)	(.013)	(.007)	(.007)	(.008)
Plaintiff Won-Jury	.075**	.072**	.069***	.076***	.081***	.081***
	(.030)	(.031)	(.023)	(.019)	(.021)	(.015)
Self Rep.	-.023*	-.032**	-.011	-	-	-
	(.014)	(.014)	(.010)			
Self Rep (Appeal)	-.064***	-.071***	-.050***	-	-	-
	(.009)	(.009)	(.010)			
Circuit Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
District Fixed Effects	No	No	No	No	No	No
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7359	7359	7375	17679	17679	17679

Table 4

Selection Equation

	Probit	Logit	LPM	probit	logit	lpm
	(1)	(2)	(3)	(1)	(2)	(3)
Before Trial	.002	.002	.002	-.0005	-.0005	-.003
	(.002)	(.002)	(.002)	(.002)	(.002)	(.003)
District Time	.0007***	.0006***	.001***	.0009***	.0008***	.001***
	(.0001)	(.0001)	(.0003)	(.0001)	(.0001)	(.0002)
Amount Demanded	1.51e-06***	1.49e-06***	2.69e-06***	2.66e-06***	2.60e-06***	3.92e-06***
	(2.54e-07)	(2.39e-07)	(4.77e-07)	(1.97e-07)	(1.87e-07)	(4.26e-07)
plaintiffwon	.032***	.035***	.024***	.035***	.038***	.029***
	(.003)	(.003)	(.004)	(.003)	(.003)	(.004)
Jury	.102***	.095***	.161***	.112***	.106***	.172***
	(.011)	(.011)	(.013)	(.009)	(.010)	(.010)
Plaintiff Won-Jury	-.004	-.007***	.085***	.002	-.004	.102***
	(.003)	(.002)	(.014)	(.006)	(.004)	(.024)
Self Rep.	.061***	.058***	.084***	-	-	-
	(.006)	(.006)	(.012)			
Self Rep. (defendent)	.007*	.005	.003	-	-	-
	(.004)	(.003)	(.004)			
Circuit Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
e(N)	251964	251964	251977	469764	469764	469764