Essays on Industrial Organization and Innovation

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th>Citable link</th>
<th><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:40050046">http://nrs.harvard.edu/urn-3:HUL.InstRepos:40050046</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Essays on Industrial Organization and Innovation

A dissertation presented
by

Jiashuo Feng

to

The Department of Economics

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

in the subject of
Economics

Harvard University
Cambridge, Massachusetts
May 2018
Essays on Industrial Organization and Innovation

Abstract

In my dissertation, I explore policy-relevant topics in the pharmaceutical industry and in the intellectual property system.

In Chapter 1, I begin by studying prescription drug demand, and argue that standard cross-sectional demand estimates can lead to misleading substitution patterns and policy implications. The motivating empirical fact behind my research is that patients on a chronic drug the previous year will keep taking the same drug 80-90% of the time, with most of the remaining patients taking no drug rather than another substitutable drug. I show through quasi-experiments that history-dependence plays a significant role in generating the data, and that most of the continuing patients would pick another drug if their current drug were not available. I conclude by constructing an empirical model of demand around these findings, which can provide more accurate counterfactuals for analyzing firm behavior and policy questions surrounding pricing, formulary design, and the value of drug innovation.

In Chapter 2, I argue that existing models of prescription drug pricing miss out on key market features, including the importance of pricing intermediaries, known as pharmacy benefit managers (PBMs), and the history-dependent nature of demand found in Chapter 1. In the chapter, I look to understand the impact of PBMs on pricing, drug company profits, and overall spending. First, I contribute to the literature by constructing average negotiated prices from financial filings, which reflects the actual amount paid to drug companies per drug. These prices are quite different to the list prices commonly used in the literature. Then, I build a model to explain the negotiated prices, which incorporates the PBMs role in negotiating prices using their insurance design power. Incorporating PBMs is crucial for
explaining several features of the price data, and my model suggests that PBMs reduce drug company profits by 25%, but capture up to 40% of the savings. I conclude by applying the model to evaluate vertical integration of PBMs with drug companies, as well as the pros and cons of potential government-led price negotiation. Vertical integration leads PBMs to favor drugs owned by their parent company in formulary decisions, and the government’s effectiveness in negotiating prices would depend on its willingness to exclude drugs, which may be politically challenging in the US.

In Chapter 3, I turn to studying the nature of patents, a key component of the innovation system. More specifically, my co-author and I find a strong link between how patent documents are crafted and its subsequent value and usage. To do this, we exploit variation generated by patent examiners at the US Patent Office, and show that the activities of Patent Assertion Entities (PAEs), commonly referred to as patent trolls, and the probability of litigation are quite sensitive to the language in the final patent document. These results show the pitfalls of the influential “rational ignorance” view of the patent office, which argues that there is little at stake in the patent examination process.
## Contents

Abstract ................................................................. iii
Acknowledgments ..................................................... xi

Introduction ............................................................ 1

1 History-Dependent Demand in Chronic Drug Markets: Evidence from Drug Introductions ............................. 4
  1.1 Introduction .......................................................... 4
  1.2 Background: The Mechanics of Prescription Drug Demand and Motivating Empirical Fact ................................. 7
    1.2.1 Drug Choice Mechanics - Patients, Doctors, and Pharmacists ........................................ 7
    1.2.2 Chronic Drugs and History-Dependence ................................................................. 8
    1.2.3 Features of the Anti-Cholesterol Drug Market ...................................................... 10
    1.2.4 Demand Data - MarketScan and Medical Expenditure Panel Survey ............................. 12
    1.2.5 Motivating Summary Statistics: Conditional Market Shares ....................................... 14
  1.3 Quasi-Experimental Evidence: History-Dependence and Second Choices ............................................. 16
    1.3.1 Previous Literature - History-Dependence in Other Markets and Identification Strategies .................. 17
    1.3.2 Data Setup - Quarterly Panel, Treatment and Control Groups ....................................... 18
    1.3.3 Methodology and Identification Tests ........................................................................... 21
    1.3.4 Estimation Results - Inertia ....................................................................................... 23
    1.3.5 Establishing Design Validity - Inpatient Starters, Smoothness, and Balance on Observables .................. 26
    1.3.6 History-Dependence More Broadly - Generics and Incremental Drugs ............................... 28
    1.3.7 The Role of Doctors ................................................................................................. 35
    1.3.8 Second Choice for Continuing Patients - Evidence from Baycol Removal ......................... 39
  1.4 An Empirical Model of Demand for Chronic Drugs .................................................................................. 41
    1.4.1 Why a Choice Model? .............................................................................................. 41
    1.4.2 A Choice Model for Chronic Drugs ........................................................................... 42
    1.4.3 Potential Implications for Existing Research .............................................................. 44
### 1.5 Conclusion ................................. 45

#### 2 Pricing Intermediaries in Prescription Drug Markets: Impact and Implications for Government Negotiation 47

- 2.1 Introduction .................................. 47
- 2.2 Background: Drug Pricing Literature, PBMs, and the Anti-Cholesterol Market 51
  - 2.2.1 Drug Pricing - Background, Literature, and Policy .......................... 51
  - 2.2.2 PBMs and Industry Pricing Structure .......................................... 53
  - 2.2.3 Anti-Cholesterol Market - Institutional Details .............................. 57
  - 2.2.4 Net Price Data .................................. 60
- 2.3 Pricing Model - Finite-Period Dynamic Game .............................. 60
  - 2.3.1 Overview of Model ........................................ 60
  - 2.3.2 Assumptions and Limitations of the Model ................................. 62
  - 2.3.3 Demand System - Switching Costs ........................................ 67
  - 2.3.4 Drug Company Pricing - Take-It-Or-Leave-It Offers and Dynamics .... 71
  - 2.3.5 PBM Behavior - Formulary-Setting Behavior .............................. 74
- 2.4 Results - Estimation Approach and Model Estimates ......................... 78
  - 2.4.1 Demand System - Switching Costs and Copay Sensitivity ............... 78
  - 2.4.2 Model Computation and Estimation ....................................... 81
  - 2.4.3 Parameter Estimates and Model Fit ....................................... 83
- 2.5 Counterfactuals - Impact of PBMs and Policy Implications ................ 88
  - 2.5.1 Do PBMs Reduce Drug Expenditures? ....................................... 88
  - 2.5.2 The Effects of Vertical Integration .................................... 92
  - 2.5.3 Implications for Medicare Part D Price Negotiation .................... 93
  - 2.5.4 Quantifying the Effects of History-Dependent Demand on Pricing ... 94
- 2.6 Conclusion ....................................... 96


- 3.1 Introduction ...................................... 98
- 3.2 Data .............................................. 106
  - 3.2.1 Data Sources, Samples and Variable Definitions ......................... 106
  - 3.2.2 Summary Statistics ........................................ 109
  - 3.2.3 Illustration of Main Findings ........................................ 112
- 3.3 Estimating Examiner Effects on Patent Outcomes .............................. 115
  - 3.3.1 Research Design ........................................ 115
  - 3.3.2 Baseline Estimates of Examiner Effects .................................. 118
  - 3.3.3 Validation of Baseline Design: Addressing Non-Random Assignment and Selection ............................. 121
3.3.4 Robustness Checks ............................................. 134
3.4 Implications for Patent Assertion Entities ...................... 135
  3.4.1 Research Design ............................................. 136
  3.4.2 PAEs and Examiner Behavior .............................. 137
  3.4.3 PAEs and Patent Invalidity ................................. 145
  3.4.4 Robustness Checks and Additional Results ................. 150
3.5 Conclusion ..................................................... 154

References .......................... 155

Appendix A Appendix to Chapter 1 .......................... 163
  A.1 Suggestive Evidence of History-Dependent Demand Based on Firm Behavior 163
  A.2 Additional Data Background ................................ 165
  A.3 Robustness - Methodology, Other Events, and Other Chronic Drug Markets 165
    A.3.1 Heterogeneity by User Characteristics ................ 167
  A.4 Using the Framework to Assess Long-Term Health and Spending Impacts of
    New Drugs .................................................. 168
  A.5 Micro-Foundations - A Detailed Discussion .................. 170
  A.6 Standard Claims Datasets and Multi-Agent Models ............ 172
  A.7 Additional Tables and Figures ............................... 173

Appendix B Appendix to Chapter 2 .......................... 183
  B.1 Constructing Net Price Estimates ............................ 183
  B.2 Illustrative Model - Inertia and Copay Tiers ................. 186
    B.2.1 Model Outline ............................................. 186
    B.2.2 PBM Decision ............................................. 187
    B.2.3 Entering Firm Pricing .................................... 187
    B.2.4 Summary .................................................. 189
  B.3 Additional Graphs ............................................. 190

Appendix C Appendix to Chapter 3 .......................... 193
  C.1 Examiner Effect Estimation Methods ........................ 193
  C.3 Additional Tables and Graphs ................................ 201
List of Tables

1.1 2002 Anti-Cholesterol Market Shares ................................. 15
1.2 Major Anti-Cholesterol Drugs - Approval Date and Ownership ........ 20
1.3 Estimates of the Impact of Inertia ..................................... 25
1.4 Estimates of the Impact of Inertia - Inpatient Starters Panel .......... 27
1.5 Conditional Market Shares - Baycol User in Previous Year .......... 40
2.1 Demand System Estimates .............................................. 80
2.2 Parameter Estimates .................................................. 84
2.3 Summary of Counterfactual Market Outcomes ....................... 91
3.1 Summary Statistics ..................................................... 110
3.2 Signal Standard Deviations of Examiner Causal Effects on Patent Outcomes . 119
3.3 Validation of Baseline Estimates of Examiner Effects ................. 126
3.4 Robustness Checks on Examiner Causal Effects on Patent Outcomes .... 135
3.5 Patent Acquisition and Examiner Behavior ........................... 140
3.6 Patent Litigation and Examiner Behavior ................................ 143
3.7 Summary Statistics on Proxies for Patent Invalidity .................. 147
3.8 Examiner Behavior and Likelihood of Patent Invalidity ................ 148
3.9 Heterogeneity in Examiner Causal Effects on Patent Acquisition by PAEs . 151
A.1 Estimates of the Impact of Inertia - 60-Day Window .................. 173
A.2 Heterogeneity in Inertia .............................................. 174
C.1 Raw Standard Deviations of Patent Outcomes across Examiners ...... 201
C.2 Signal Standard Deviations of Examiner Causal Effects by Technology Categories 202
C.3 Signal Standard Deviations of Examiner Prosecution Behaviors ...... 203
C.4 PAE Patent Portfolios and Litigated Patents across Technology Categories . 204
C.5 Patent Acquisition and Examiner Behavior, Full Sample .............. 205
C.6 Patent Litigation and Examiner Behavior, Full Sample ............... 206
C.7 Examiner Behavior and Likelihood of Patent Invalidity, Full Sample ... 207
C.8 Examiner Behavior and Other Patent Outcomes ....................... 208
C.9 Patent Acquisition by PAEs and Grant Decision at the European Patent Office 209
## List of Figures

1.1 Graphical Evidence of Inertia from Quasi-Experiment .................................. 24
1.2 Smoothness of Flow of New Users ................................................................. 28
1.3 History-Dependence in Generic Drug Choice .................................................. 31
1.4 Adoption of Brand Drug by Previous Choice .................................................. 32
1.5 The Impact of Vytorin on Generic Zocor Adoption .......................................... 33
1.6 Adoption of Combination Drugs by Previous Choice ....................................... 34
1.7 Impact of Crestor Choice on Vytorin Adoption .............................................. 35
1.8 Number of Distinct Drugs Prescribed by Physicians in 2014 ........................... 37
1.9 Market Shares by Doctor Medical School Graduation Year ............................ 38
1.10 Decision Structure with Three Drug Options ............................................... 44

2.1 Anti-Cholesterol Market Summary - Demand ............................................... 58
2.2 Anti-Cholesterol Market Summary - Prices .................................................... 59
2.3 Model of Market Structure ................................................................................. 61
2.4 Model Fit - Predicted vs. Actual Prices ......................................................... 85
2.5 No PBM - Coinsurance ...................................................................................... 90
2.6 No Merck-Medco relationship \( (b_1 = 0) \) ...................................................... 93
2.7 No Drug-Specific Inertia ................................................................................... 95

3.1 The Effect of Examiners on Patent Acquisition and Litigation ....................... 113
3.2 The Allocation of Patent Applications to Examiners by Application’s Last Digit 124
3.3 A Busyness Instrument for the Effect of Examiners on Patent Acquisition by PAEs ................................................................. 129
3.4 Out-of-Sample Tests of Baseline Estimates of Examiner Effects .................... 134
3.5 Heterogeneity in Patent Acquisition Behavior across Groups of PAEs ........... 152

A.1 MarketScan Sample Size Over Time ............................................................... 173
A.2 Impact of Inertia By Period .............................................................................. 175
A.3 Smoothness of Flow of New Users (Unbalanced Panel) .................................. 176
A.4 Event study results broken down for major anti-cholesterol entry events .......... 177
A.5 Market Share - All Cholesterol Drugs ............................................................. 178
Acknowledgments

I would like to thank my advisers Ariel Pakes, David Cutler, and Josh Lerner. I also want to thank my classmate and co-author Xavier Jaravel.

In addition, I would also like to acknowledge Philippe Aghion, Pierre Azoulay, Ernst Berndt, Raj Chetty, Lauren Cohen, Edward Glaeser, Stephen Haber, Nathan Hendren, Nathan Hipsman, Larry Katz, Bill Kerr, Jay Kesan, Scott Kominers, Robin Lee, Ross Levine, Alex Mackay, Luca Maini, Alan Marco, Markus Mobius, Anders Munk-Nielsen, Ben Roin, Scott Stern, and Heidi Williams for providing detailed comments and suggestions on my dissertation chapters.

I would also like to thank all of my classmates and especially my officemates over the years in Littauer: Kirill Borusyak, Aubrey Clark, Max Eber, Andy Garin, Luca Maini, Jann Spiess, Edoardo Teso, and Wentao Xiong.

Next, I would like to thank various academics and industry contacts for providing valuable background information and data: Ryan Baum and SSR Health for providing data and suggestions on net price construction; Bob Nease and Edmund Pezalla for insights into PBM formulary decisionmaking. In addition, I would like to thank Michael Frakes and Melissa Wasserman, RPX Corporation, LexMachina, and Juristat for providing data on patent examination, litigation, and patent assertion entities.

Finally, my research also benefited from comments from participants in the Harvard Industrial Organization Lunch, the Harvard Labor/Public Lunch, the Harvard Innovation and Entrepreneurship Brownbag, the NBER Productivity Seminar and Summer Institute, Boston University, Duke, Hoover Institute, UBC Sauder, and the US Patent Office Visiting Speaker Series.
To my parents, my grandfather, and my love.
Introduction

In my dissertation, I look to contribute to our understanding of the US innovation system, focusing on two important factors. The first is the impact of downstream market structure on innovation, going back to the theories of Schumpeter. The second is the role of intellectual property, another literature that stretches back many decades. In the first two chapters of the dissertation, I will focus on two first-order issues in the structure of prescription drug markets that are under-explored in the literature, namely history-dependent demand and the impact of pricing intermediaries. Having a better grasp of how the downstream market works can lead to a better understanding of innovation in the pharmaceutical industry. In the final chapter, I explore the importance of the language in a patent document on how the patent is subsequently used, relating it to the recent increase in patent litigation, much of it driven by so-called patent trolls.

The pharmaceutical industry plays an important part in the US innovation system. First, the pharmaceutical industry directly contributes a significant fraction of total R&D spending in the US. According to the National Science Foundation, the pharmaceutical industry spent $58 billion on R&D in 2015, about one-sixth of all industry R&D spending in the US. Furthermore, the pharmaceutical industry often builds on and attempts to commercialize research funded by government agencies such as the National Institute of Health (NIH), which alone invested $30.4 billion in medical research in 2015.

In addition, pharmaceutical companies make almost half of its total revenues in the US, likely making policies and market features in the US important factors in driving R&D decisions. According to data from IMS Health, the industry garners 46% of its global
revenues in the US. In addition, the US market is unique in that the government does not directly control prices. Instead, much of the market is covered by private insurance, and intermediaries, known as pharmacy benefit managers (PBMs), are responsible for negotiating prices on behalf of insurers. This system has been criticized due to rising drug prices, highlighted by recent cases of hyperinflation involving Turing Pharmaceuticals, Valeant, and Mylan. Congress has blamed both drug companies and PBMs for their role in forming the high prices, and has discussed various policies including government negotiation of prices. However, drug companies have argued that greater pricing pressure reduces incentives to innovate. Put together, all of this suggests that a better grasp of US drug markets can help improve our understanding of R&D decisions in the pharmaceutical industry.

In Chapters 1 and 2, I contribute to our understanding of US prescription drug markets in two ways. First, in Chapter 1, I contribute to the literature on prescription drug demand by analyzing quasi-experiments that show strong history-dependence in demand and uncover second choices. I then use these features as a guide to construct an empirical model of demand, which can lead to different inferences about welfare and policy relative to models in the literature.

In Chapter 2, I study the impact of the aforementioned PBMs on market outcomes, with an eye towards understanding the impacts of vertical integration and the potential impact of government-led negotiations. I start by collecting data on average negotiated prices, which look very different from the list prices typically used in the literature. I then model the negotiated price as the outcome of a dynamic game played by drug companies and PBMs, with both types of agents taking into account the dynamic nature of demand. Finally, I use the model to assess counterfactual prices and spending under scenarios without vertical integration and with the government replacing PBMs as price negotiators.

Another important aspect of the US innovation system is the patent system, which has drawn scrutiny in recent years due to the rapid growth in patent litigation. Although firms and inventors use a variety of ways to protect their intellectual property (Cohen et al, 2000),
one important and observable method is to file for a patent at the US Patent Office (USPTO).
The patent system has come under fire in recent years, due to the growth in the number of
patent lawsuits and the emergence of patent assertion entities (PAEs), which are commonly
and pejoratively referred to as patent trolls. According to analysis by RPX, the number of
defendants in patent lawsuits increased from about 1,000 in 2000 to an average of about
6,000 per year between 2010 and 2015. Most of that growth has come from PAE assertions.
PAEs are unlike typical R&D firms in that they buy up patents and make all of their revenue
through enforcing those patents, either from settlements or from damages. Critics of PAEs,
including many large tech companies, argue that they are damaging the innovation system
by indiscriminately enforcing patents that contain vague language.

In Chapter 3, my co-author and I study how the language in a patent document affects
its value and subsequent usage, with a focus on the purchasing activities of patent assertion
entities (PAEs), commonly referred to as patent trolls, and patent litigation more broadly. To
do this, we exploit the quasi-random assignment of patent applications to examiners within
art units at the USPTO. This helps us generate variation in the language, as examiners
exhibit significant variation in the amount of editing they force applicants to make in their
patent documents. We find that patent outcomes differ significantly across examiners, with
one standard deviation in examiner effects (computed using a empirical Bayes approach
often used in the teacher value-added literature) equal to $3 million in private value and
64% of the baseline probability that a patent is litigated. In addition, we find that PAEs
overwhelmingly purchase and assert patents that were grant by “lenient” examiners; these
examiners issue patents that are much more likely to be litigated but also to be legally
invalid.
Chapter 1

History-Dependent Demand in Chronic Drug Markets: Evidence from Drug Introductions

1.1 Introduction

In this chapter, I argue that there is strong history-dependence in prescription drug demand. I begin by documenting the empirical fact that patients who took an anti-cholesterol drug in the previous year continue taking the same drug at very high rates (80-90%), with many of the remaining patients choosing not to take any drug. Using quasi-experiments, I show that strong history-dependence in demand contributes significantly to this pattern, and that most of the continuing patients would choose another drug if their current option were removed. Then, I build a model of demand that can capture these features, which standard demand models in the literature fail to incorporate. To conclude, I discuss how the model can be used to understand the value of drugs, drug company strategy, and formulary design, which in turn affects the analysis of key policies in the industry.

The nature of prescription drug demand likely plays an outsized role in influencing drug market outcomes, including pricing, investment, and other strategic actions such as
advertising. Unlike other industries, production of prescription drugs, with the exception of biologics, is relatively straightforward, which makes classic supply-side issues such as scale and learning-by-doing less relevant. Instead, it is likely that the nature of demand influences a whole range of decisions and market outcomes. Observations from the industry suggest that demand influences a wide range of drug company decisions, including pricing, R&D direction, advertising, copay assistance programs, and FDA priority review voucher purchases. In addition, from conversations with executives at pricing intermediaries and insurance companies, substitution patterns strongly influence the design of drug insurance.¹

In this chapter, I provide quasi-experimental evidence on two important demand features and then construct a model of demand around these facts that can be used to study firm decisions. To begin, I establish evidence of history-dependent demand by using a strategy based on discontinuities around drug launches. As a unit, patients, doctors, and pharmacists generate demand that exhibits inertia, a tendency to choose the same drug, and a more general history-dependence when it comes to choosing generic or incremental drugs.² I then provide additional evidence that suggests that this is due to doctor risk aversion when prescribing to continuing patients. Second, I show that most patients would switch to another drug rather than the outside option if their preferred drug is removed. This feature is important to account for in assessing the impact of removing a drug from insurance coverage. Finally, I embed these patterns into a parsimonious model of prescription drug demand in chronic markets, which can then be used to analyze various phenomena in these markets.

This chapter adds to the literature on consumer inertia and also the literature on prescription drug demand. There is a sizable literature on consumer inertia and brand loyalty, which has used random assignment, dominated choices, and discounts to identify

¹Pricing intermediaries in drug markets are known as pharmacy benefit managers, which are discussed in greater detail in Chapter 2.

²Generic drugs contain the same active ingredient (molecule) as the corresponding branded version, and enter after the patent expires on the branded drug. Incremental drugs refer to reformulations or combinations involving the molecule in an existing drug.
inertia in insurance and consumer goods markets. I contribute to this literature by offering an identification strategy in settings with new products and consumer inflexibility in decision timing. In terms of prescription drug markets, my paper is the first to provide causal evidence of history-dependence in chronic drug demand, a possibly important factor in many drug company decisions.

Previous studies on learning and advertising, including Crawford and Shum (2005), Dickstein (2014), Sinkinson and Starc (2018), and Lee (2016) either implicitly or explicitly find history-dependence in their demand analysis, but do so by imposing modeling assumptions. They also analyze effects over a few weeks or months, and neglect to draw out the implications for pricing.

I also build on and contribute to the long literature on prescription drug demand. The simplest approach to modeling drug demand is to use a static discrete choice or random coefficients logit framework, an example of which can be found in Dunn (2012), who assesses price indices in the anti-cholesterol market. More complicated approaches include nested logit modeling for drug markets with either different classes of drugs or with generics, examples of which include Arcidiacono \textit{et al.} (2013) and Bokhari and Fournier (2013), who study the value of new drug introductions using demand. Several papers touch on the multi-agent nature of drug demand, starting with Ellison \textit{et al.} (1997). They make use of a multi-stage budgeting structure to study molecule and generic choice. Several papers also touch on the issue of demand dynamics, including short-run learning models estimated by Crawford and Shum (2005) and Dickstein (2014), and Lee (2016), as well as differential responses to advertising based on user history explored by Sinkinson and Starc (2018). Most of the papers mentioned here use list price directly in the demand system, but recent papers including Einav \textit{et al.} (2016) explore the role of copays in drug choice, using the Medicare

\footnote{I discuss the literature in detail in Section 1.3.1.}

\footnote{The strategy is similar in spirit to Handel (2013), who compares the insurance plan choice of new and continuing workers.}

\footnote{See Appendix A.1 for a detailed discussion.}

\footnote{An exception is Ching (2010), which looks at the impact of global learning about generic quality on generic pricing and competition. Researchers have investigated pricing and inertia in other health care markets, mainly in Medicare Part D plan choice. These include Ericson (2014), Wu (2015), Fleitas (2016), and Ho \textit{et al.} (2017).}
Part D donut hole. My model incorporates history-dependence in the form of switching costs, which I identify using the quasi-experimental structure.

1.2 Background: The Mechanics of Prescription Drug Demand and Motivating Empirical Fact

To begin, I provide a brief overview of the institutional details surrounding demand in chronic drug markets. These details are conducive for my analysis of history-dependent demand and also provide some context for my findings.

1.2.1 Drug Choice Mechanics - Patients, Doctors, and Pharmacists

In this paper, I treat observed prescription drug choice as the result of a joint decision made by patient, doctor, and pharmacist. Doctors are responsible for diagnosing health problems in patients, and then make a decision on whether or not to use medication to treat the problem, as well as which drug to take.\(^7\) For each of the possible drugs, patients with insurance will face various copay amounts, which often come in two tiers for branded drugs and a separate tier for generics, rather than the full price. However, the doctor is often not aware of the details of the patient’s insurance, possibly dampening price sensitivity. Forces that push back against this include insurers calling doctors to make them aware of the drugs with lowest copays on their patient’s plan and doctors generally knowing that generic drugs have much lower copays.

Pharmacists also play a role in drug choice, by alerting doctors to insurance coverage issues and by substituting generic medication. Once a patient obtains a prescription, they usually go to the pharmacy to obtain their medication. The pharmacist they see can view the formulary on a patient’s drug insurance, which usually lists copayments for various drugs. The pharmacist can choose to contact the doctor to change the prescription to a

\(^7\)For example, in the case of a high cholesterol diagnosis, doctors can suggest changes in diet and exercise or prescribe one of many anti-cholesterol drugs.
cheaper alternative, a practice known as “therapeutic substitution.”\(^8\) A more common form of pharmacist behavior is to dispense generic drugs that contain the same active ingredient, or molecule, as the branded one prescribed by the doctor. This guides patients on a branded drug to switch to the generic version, which I see in my analysis. Each state in the US has generic substitution laws that govern pharmacist behavior, ranging from mandatory to optional substitution.\(^9\)

Finally, the patient is responsible for obtaining and refilling prescriptions, which can lead them to pick the outside option. As mentioned above, the patient needs to go to the pharmacy\(^10\) to obtain his or her prescriptions. If they do not go, then they will have no record of prescriptions in the claims data, and will appear to be diagnosed but taking no drug.

### 1.2.2 Chronic Drugs and History-Dependence

In this paper, I focus on demand in chronic drug markets, which represent some of the largest drug markets and allow me to observe patient-level prescriptions over time. Chronic drug markets have the feature that once a patient starts taking medication, they keep taking medication indefinitely.\(^11\) This feature allows me to track repeated choice over time for a large number of people.\(^12\) Examples of chronic markets include treatments for high cholesterol, diabetes, multiple sclerosis, asthma, hypertension, and HIV, all of which are large markets in terms of revenue and/or number of patients both in the US and globally.\(^13\)

One aspect of prescriptions in chronic drug markets is that they are valid for up to 8 months.\(^8\) Therapeutic substitution without notifying the doctor is prohibited in many states.


\(^10\) An alternative that has become more popular is ordering prescriptions through the mail.

\(^11\) Allowing for occasional gaps due to non-compliance.

\(^12\) As I note later, one can also track the choices of doctors over time, but the sample is much smaller.

one year, creating the possibility of significant history-dependence. A written prescription from a doctor is valid for an amount of time depending on a drug’s schedule. The Drug Enforcement Administration (DEA) classifies drugs into six schedules, with Schedule II drugs having the highest potential for abuse, and Schedule VI having the lowest. Painkillers such as OxyContin and attention deficit hyperactivity disorder drugs such as Adderall are examples of Schedule II drugs, while cholesterol and diabetes medication are in Schedule VI. Each state then sets its own laws on the length of prescription validity, with the shortest for Schedule II drugs and the longest for Schedule VI. For example, Massachusetts sets a 30-day supply limit on Schedule II drugs, but a one year limit on Schedule VI drugs. Therefore, patients on Schedule VI drugs can choose to refill their medication several times without having to visit their doctors.

Additional suggestive evidence of history-dependent demand comes from drug company decisions and commentary. As I detail in Appendix A.1, drug company decisions surrounding incremental drug entry and FDA priority review vouchers suggest that there is significant history dependence in demand. For example, they value FDA priority review vouchers, which takes four months off review time, at a significantly higher amount than what the drugs earn in those four months. In addition, drug companies regularly discuss breakdowns such as new user market share or the source of switchers to their new combination drugs.14

Finally, anecdotal evidence suggests that doctors take an “if it ain’t broke, don’t fix it” approach to prescribing drugs, which can generate history-dependence at the patient level. From speaking with doctors and industry analysts, the conventional wisdom is that doctors are reluctant to switch patients away from an effective drug, even if the doctor thinks a different drug is superior or cheaper. This behavior can still generate patient-level history-dependence, even if the patient never switches doctor. New drugs can enter after a

14For example, Merck asserted in their earnings calls that Vytorin (a combination of Zocor and Zetia) was intended to attract users beyond just those already on Zocor. AstraZeneca also routinely discussed new user market share in the years after Crestor’s launch. Gilead recently launched a successful HIV combination drug, Genvoya, but were pressed into acknowledging that only 8% of prescriptions were switches from people not already on some components of the combination.
patient begins treatment or existing drugs may change copay tiers over time, changes that patients and their doctors do not respond to.

### 1.2.3 Features of the Anti-Cholesterol Drug Market

Here, I provide some background on the anti-cholesterol drug market, which I focus on in my analysis. The market is large in terms of revenue and number of patients, has major entry events during my analysis period, and relative homogeneity in the medical characteristics of the major drugs.

Anti-cholesterol drugs help patients manage high cholesterol, an increasingly common condition, and represent a significant share of total prescriptions and spending in the US. A recent CDC report on the market estimates that 27.9% of US adults over the age of 40 are currently taking anti-cholesterol medication.\(^{15}\) These drugs help reduce LDL cholesterol levels, which have been linked in the medical literature to increased rates of cardiovascular disease. Given the prevalence of the condition, it is unsurprising that cholesterol drugs represent a large share of spending and prescriptions. At its peak in 2004 and 2005, cholesterol drugs represented 10% of all drug spending.\(^{16}\) Even now, cholesterol drugs represent about 7% of all drug claims.\(^{17}\)

There are several classes of anti-cholesterol drugs, although I will later focus on the statin class due to its market dominance and homogeneity. Prescription drugs are typically classified into classes by their biological mechanism of action, and there exist several classes within the set of anti-cholesterol drugs. The biggest sellers by number of prescriptions and revenue are Lipitor (sold by Pfizer), Zocor (sold by Merck), and Crestor (sold by AstraZeneca), which are in the statin class. Statins are typically the only drugs taken for

\(^{15}\)National Center for Health Statistics Data Brief No. 177, December 2014

\(^{16}\)Based on calculations from the Medical Expenditure Panel Survey (MEPS). The denominator includes all drug spending, including non-prescription drugs. The first major generic entry occurs in 2006, which has led to a reduction in spending.

moderate or high reductions in LDL levels, as they are more potent than alternatives in other classes. Other drugs such as Zetia provide an alternative mechanisms for lowering LDL. After my data sample ends, a new class of drugs, PCSK9, has entered the market with increased ability to reduce LDL levels but surprisingly limited commercial success.\textsuperscript{18} While physicians debate the differences between statins, they generally have similar efficacy and side effect profiles, with muscle pain as the most common complaint \textit{Pedersen and Gaw (2001)}. My identification strategy for showing history-dependence deals with unobserved patient-drug match quality, but the scientific evidence provides additional re-assurance.

A first helpful feature of the anti-cholesterol market is the number of entry events during the period in which I have data. In Table 1.2, I have displayed the major drug entry events during the time period covered by the MarketScan data. Before 1996, Zocor, and to a lesser extent Pravachol and Mevacor, were the predominant cholesterol drugs on the market. The biggest brand entry events are Lipitor in 1997, Zetia in 2002, Crestor in 2003, and Vytorin in 2004.\textsuperscript{19} The first generic to enter was Mevacor in 2001, followed by Zocor and Pravachol in 2006, and finally Lipitor in 2011. There are also entries of incremental drugs, such as Vytorin (combination) and Lescol XR (extended-release).

I leverage a second type of event that occurs in the market, the removal of a drug from the market, to identify second choices of continuing patients. The FDA removed a less popular statin, Baycol, from the market in August 2001, after finding thirty-one instances of fatal side-effects caused by the drug. Baycol was less popular than the other statins mentioned above, as it was a latecomer to the market in 1999, reaching a peak of 6\% of all anti-cholesterol prescriptions in 2000. The usefulness of the removal event is that I can see how all existing users on Baycol behave after the drug is no longer in their choice set, which allows me to observe the second choices of all patients. I discuss the intuition in Section 1.2.5 below.

\textsuperscript{18}This may have to do with the reluctance of patients to switch away from generic statins, a fact document in Section 1.3.6.

\textsuperscript{19}Zetia is a drug in a different class that is less potent, and Vytorin is a fixed dosage combination of Zocor and Zetia, which was popular for a brief period in the high intensity treatment class.
The large number of cholesterol users is another desirable feature, making the market both important and easier to study. In my MarketScan data, about 10.3 million people have a prescription for anti-cholesterol medication at some point, out of 150 million unique users. This is slightly below but roughly in line with CDC estimates of usage rate in the adult population. The number of people on cholesterol medication allows me to set a very fine-grained empirical strategy for isolating history-dependence.

1.2.4 Demand Data - MarketScan and Medical Expenditure Panel Survey

Here, I provide some background on my main data sources, with a focus on the Truven MarketScan dataset, which is crucial for both my quasi-experimental analysis and for my demand estimation. I also make use of data from the Medical Expenditure Panel Survey (MEPS) for representativeness in demand.

The Truven MarketScan dataset is based on medical claims data collected by Truven from large US companies. Truven offers various services to help companies manage their medical spending, and anonymizes the data for researcher use. People remain in the dataset as long as their companies still use Truven’s services.

The Truven MarketScan dataset provides better patient tracking and coverage period relative to other commonly used datasets in the literature, which is crucial for my quasi-experimental setup. The MarketScan dataset tracks a large number of working-age users over the period from 1996-2013, ideal for setting up my empirical strategy and measuring long-term outcomes. Figure A.1 shows the number of enrollees in the entire dataset, which has grown from about 2 million pre-2000 to almost 60 million in the late 2000s. In comparison, MEPS surveys overlapping waves of about 30,000 users, but only tracks each user for two years, precluding the possibility of measuring long-run patterns in drug choices.

---

20 According to a 2014 CDC report, general usage rate of anti-cholesterol medication has risen from 19.9% to 27.9% between 2003 and 2012, with about a 15% usage rate in the 40-59 population. The latest census estimates in 2010 put the 45-64 age category at around 80 million people. Our data covers working adults up to age 65.

A similar dataset in nature is Medicare Part D claims data, which covers most Americans over age 65. However, many patients start on anti-cholesterol medication well before 65, so we would not have as many first-time users to analyze. In addition, Medicare Part D has only been in existence since 2006, which is after many of the major drug launches documented in Table 1.2.

Another key advantage of the MarketScan is the ability to infer formulary data, which is important for my demand estimation. MarketScan data comes with plan identifiers for each patient, and I infer the anti-cholesterol drug formulary for each plan from realized user choices. The size of the cholesterol market helps in that even plans covering a small number of patients have many realizations of each option, which helps rule out any selection issues based on copay and makes it easier to confidently identify formulary exclusions. I then merge the constructed formulary onto user choice data in order to estimate a demand model.

One major disadvantage for the MarketScan relative to Medicare Part D data is its lack of doctor and pharmacy identifiers, which, as I mentioned earlier, forces me to consider the patient, doctor, and pharmacist as a single decision unit. This limits my ability to delve more deeply into the importance of joint decisionmaking on demand patterns, which means I am less able to speak to policy interventions targeted specifically at patient or doctor behavior. Therefore, throughout the paper, patient choice implicitly refers to a joint decision by the patient and his or her doctor, although I explore how much we can learn from the data about doctor behavior in Section 1.3.7.

A second disadvantage is the nature of the population covered by MarketScan. As alluded to before, the dataset covers families with someone working at firms that use MarketScan to help them manage their health plans, which misses out on older and poorer individuals covered by Medicare and Medicaid, respectively, and uninsured individuals also taking anti-cholesterol medication. In addition, as shown in Table 1, the population

---

22 Ellison et al. (1997) explore the role of doctors and pharmacies in a multi-stage budgeting model of demand for cephalosporins, which they use to speculate about the role of policies targeting pharmacies.
skews male.

A final disadvantage is the significant attrition rate in users, which I address by constructing both balanced and unbalanced panels. Attrition rates from year-to-year is generally in the 20-25\% range. This comes from individuals leaving firms and firms leaving the MarketScan sample. Firms come in and out of the data, based on whether they need help managing their health insurance plans, which do not suggest an obvious selection problem, but may still lead to biases in the analysis of long-term effects. I construct both unbalanced and balanced panels to look for consistency in results.\footnote{See Appendix A.2 for a detailed discussion of panel selection.}

To address representativeness, I make use of claims data from MEPS. MEPS surveys waves of respondents on their health-related spending and outcomes, including prescription drug claims, with each wave surveyed for a two-year period. The benefit of MEPS is that it provides weights on each respondent in order to arrive at a sample that is representative of the US market. In particular, this covers older patients not in the MarketScan dataset, and addresses other representativeness issues in MarketScan such as selection on employment. I use MEPS to construct quantity data by drug and strength, in order to arrive at more accurate market shares for some of my analysis.

### 1.2.5 Motivating Summary Statistics: Conditional Market Shares

To conclude this section, I provide summary statistics in the form of conditional market shares, which motivate my quasi-experimental investigation into history-dependence. In Table 1.1, I compute 2002 market shares in the anti-cholesterol market, conditional on the patient’s 2001 prescription.\footnote{To simplify the numbers, I include only the two most popular drugs in 2002, Zocor and Lipitor.} This approach similar to the approach taken in Sinkinson and Starc (2018) to build their model of the impacts of advertising on demand. Patterns look similar in other years, but I choose the 2001-2002 period because Baycol was taken off the market in August of 2001, as discussed in Section 1.2.3 Lipitor is the more popular of the two drugs in this period, possibly due to their lower price, differences in doctor beliefs, and
heavy advertising.

Table 1.1: 2002 Anti-Cholesterol Market Shares

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Unconditional</td>
<td>0.22</td>
<td>0.44</td>
<td>0.34</td>
</tr>
<tr>
<td>Diagnosed but no drug</td>
<td>0.13</td>
<td>0.25</td>
<td>0.62</td>
</tr>
<tr>
<td>Lipitor</td>
<td>0.02</td>
<td>0.83</td>
<td>0.15</td>
</tr>
<tr>
<td>Zocor</td>
<td>0.78</td>
<td>0.04</td>
<td>0.18</td>
</tr>
<tr>
<td>Baycol</td>
<td>0.24</td>
<td>0.35</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Notes: Market shares are computed based on data from the Medical Expenditure Panel Survey, and are computed at the patient level. Patients who have multiple drugs on record are assigned the drug they take the most in a given year. The market size is defined as all patients who are diagnosed with high cholesterol or taking anti-cholesterol medication.

Several key features stand out in Table 1.1, the first of which is the consistency in drug choice for patients taking medication in both years. Conditional on taking Lipitor in 2001, 83% of patients continue taking it in 2002, with only 2% switching to Zocor and 15% not taking any drug.25 A similar pattern holds for 2001 Zocor users. The difference could be driven by heterogeneity in match quality between patient and drug, but as I show later, there is strong history-dependence that accounts for part of the consistency.

A second key feature touched on above is the popularity of the outside option, especially relative to potential substitutes for continuing patients. Here, the outside option refers to the patient not taking any cholesterol drug, and includes all patients who have been diagnosed with high cholesterol. First, the patients diagnosed with high cholesterol in 2001 but who did not take a drug continue to have high rates of choosing the outside option. The shares of Lipitor and Zocor in this group are lower than the overall rates. This is unsurprising as many of these “outside option” patients did not have a severe enough problem to warrant the need for drugs, and continue not to need them. Second and more surprisingly, for

25In results not reported here, I find that most users who start on a drug but stop taking it will return to the same drug at a later date.
patients who took Zocor in 2001, the outside option is much more popular than Lipitor. The same holds in reverse. The observational data can again be interpreted in two ways: either the other drug is a very poor fit for these patients or there are unobserved intensive margin shocks, which one can think of as adherence or general side-effects. From a medical perspective, the latter is more likely.

To get at the issue of why the outside option has such high share, I make use of the aforementioned removal of Baycol from the market. In the final row, I show the 2002 market shares for patients who chose Baycol in 2001. Baycol is removed from their choice set, and based on a naive prediction assuming an independence of irrelevant alternatives and the popularity of the outside option for previous Lipitor and Zocor users, one would expect about 10-20% of Baycol users to turn other statins. Instead, we see that about 60% of the users move to either Zocor or Lipitor. This suggests that other statins are not particularly bad matches, and that in a counterfactual without Lipitor, most of the 83% of continuing Lipitor patients would move to Zocor. I explore this more formally in Section 1.3.8.

1.3 Quasi-Experimental Evidence: History-Dependence and Second Choices

In this section, I test for the presence of history-dependence in the anti-cholesterol market by analyzing whether a patient starting on a drug causally impacts his or her choices in later years. I also use the removal of Baycol from the market to understand the second choices of continuing patients.

My strategy for demonstrating this causal relationship involves identifying patients who start treatment right before and after a new drug launches, and then showing persistently large differences in their subsequent drug choices. To verify the validity of the design, I provide evidence that patients in the groups are similar and also re-run the analysis in a sub-sample of users that start treatment after a hospitalization.

Using this framework, I also generate evidence on how history-dependence varies in the
important cases of generics and incremental drugs, two aspects unique to pharmaceuticals. The fact that switching to a generic drug is easy for patients on the branded counterpart but hard for patients taking other drugs is particularly relevant for my structural analysis in Section 2.3.

I conclude with a discussion on the role doctors play in generating history-dependence. If they are the ones carrying effects across periods, then new patient choice will be influenced by previous market outcomes.

I relegate to Appendix A.7 my results on history-dependence in other chronic drug markets and how I can use the framework to measure the long-term health impact of taking a given drug using observational data, and additional discussion of the underlying micro-foundations of the observed patterns.

1.3.1 Previous Literature - History-Dependence in Other Markets and Identification Strategies

Before documenting my empirical strategy, I provide a brief overview of the extensive literature on consumer inertia, and highlight similarities and differences present in my setting. Inertia typically refers to consumers staying on exactly the same product, whereas in my research, I establish inertia and additional effects on related drugs.

In prior literature, inertia has been explored in plan choice and consumer goods markets, using random initial assignment, dominated options, and discounts to gain identification. Handel (2013) explores the role of inertia in negating adverse selection in health insurance markets, and uses dominated plans and new user choice to identify and quantify inertia. Another paper in the health insurance area is Ericson (2014), who demonstrates inertia in plan choice by taking advantage of random plan assignment for the Medicare Part D Low-Income Subsidy Program. Other papers documenting inertia in plan choice include Honka (2014) in auto insurance and Shcherbakov (2016) in cable TV plans. Outside of plan choice, Dubé et al. (2009) explore switching costs in the supermarket setting, looking at inertia after exposure to supermarket discounts. Other papers in this area include Shum
(2004), who analyzes the market for breakfast cereals.

My specific setting allows me to construct an identification strategy based on a temporal discontinuity, one that may be suitable for demand in innovative markets. Temporary discounts, dominated options, and random assignment are hard to find in this setting, so I construct a strategy in the spirit of new user choice used in Handel (2013). I leverage drug launches, which creates differences in product availability when patients make their first choice of a drug.

### 1.3.2 Data Setup - Quarterly Panel, Treatment and Control Groups

To begin my analysis, I construct a quarterly panel of patient drug choices and classify new patients into treatment and control groups based on when they begin treatment. My simplifying choices gloss over some details, which I discuss and account for later in robustness checks.

First, to make my analysis tractable, I convert raw prescription drug claims data into a patient by quarter dataset, recording the most frequent prescription choice of the patient in that quarter. The chosen option can be summarized by the molecule, an indicator for brand vs. generic, and an indicator for extended-release, a common form of incremental innovation.

This simplification of the claims to a patient by quarter dataset raises two issues that turn out to be minor: some patients use a combination of drugs and they may also switch in the middle of a quarter. Combination regimens are common in other markets such as oncology drugs, but the data suggests that they are less of an issue in the anti-cholesterol market, as only 3% of my patient-quarters contain prescription claims for different drugs in the same week. To assign one choice to these patient-quarters, I give each option a random

---

26 One could in theory construct an quasi-experiment based on variation in exposure to free sample or copay coupon programs.

27 Extended release drugs are less popular in the anti-cholesterol market, but much more prevalent in other markets such as diabetes treatments.

28 This imperfectly weeds out quarters with claims on multiple drugs where patients are actually switching.
ID, and take the option with the lowest ID, which should be consistent across treatment and control groups. Missing out on true switches is less of a worry, as patients usually will still be on the drug in the next quarter. In robustness tests shown in Appendix A.7, I deal with this issue by generating indicators from the raw data surrounding whether patients are ever prescribed a given drug, which will pick up on isolated prescriptions unaccounted for in the quarterly data.

Second, I construct a list of when patients begin treatment. For each patient, I assign them a start date based on the earliest date they have a recorded anti-cholesterol drug claim. One issue especially relevant in the unbalanced panel is that I don’t see a patient’s history before they enter the MarketScan panel. To mitigate the risk of mis-labeling entering users who were already using anti-cholesterol medication, I filter out users who have cholesterol claims in the first six months they are in the dataset. Mostly users have either monthly or quarterly prescription entries, but there still may be some noise introduced here. This is less of a worry in the balanced panel, as many new patients have several years of claims history.

Using this list, I sort new patients into treatment and control groups for each drug entry event. To do this, I compare each patient’s start date to the list of entry event dates documented in Table 1.2. For my core analysis, a patient is in the control group if they start in the 180 days before a drug enters and the treatment group if they start in the 180 days after. I later test the robustness of these results by using 60-day windows, as having narrower windows helps rule out changes in environment over time, including any changes in medical guidelines and clinical evidence on drugs.29 The events themselves are fairly spaced out, but 40% of new patients qualify for two groups (e.g. treatment group for Zetia and control group for Crestor). For these patients, I assign them to one event randomly. Some patients also fall into no groups.

29For example, guidelines for statin use might change in a given year, leading to a new patient cohort with different characteristics than their predecessors.
Table 1.2: Major Anti-Cholesterol Drugs - Approval Date and Ownership

<table>
<thead>
<tr>
<th>Drug</th>
<th>Approval Date</th>
<th>First Date in Claims</th>
<th>Drug Company</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mevacor</td>
<td>8/31/1987</td>
<td>1/1/1996</td>
<td>Merck</td>
</tr>
<tr>
<td>Pravachol</td>
<td>10/31/1991</td>
<td>1/1/1996</td>
<td>BMS</td>
</tr>
<tr>
<td>Lescol</td>
<td>12/31/1993</td>
<td>1/1/1996</td>
<td>Novartis</td>
</tr>
<tr>
<td><strong>Lipitor</strong></td>
<td><strong>12/17/1996</strong></td>
<td><strong>1/31/1997</strong></td>
<td><strong>Pfizer</strong></td>
</tr>
<tr>
<td>Mevacor (G)</td>
<td>12/17/2001</td>
<td>12/20/2001</td>
<td>-</td>
</tr>
<tr>
<td>Mevacor (XR)</td>
<td>6/26/2002</td>
<td>7/30/2002</td>
<td>Covis</td>
</tr>
<tr>
<td><strong>Crestor</strong></td>
<td><strong>8/12/2003</strong></td>
<td><strong>8/20/2003</strong></td>
<td><strong>AstraZeneca</strong></td>
</tr>
<tr>
<td>Pravachol (G)</td>
<td>4/24/2006</td>
<td>4/24/2006</td>
<td>-</td>
</tr>
<tr>
<td><strong>Zocor (G)</strong></td>
<td><strong>6/23/2006</strong></td>
<td><strong>6/23/2006</strong></td>
<td>-</td>
</tr>
<tr>
<td>Simcor [Zocor/Niacin]</td>
<td>2/15/2008</td>
<td>3/12/2008</td>
<td>Abbvie</td>
</tr>
<tr>
<td><strong>Lipitor (G)</strong></td>
<td><strong>11/30/2011</strong></td>
<td><strong>11/30/2011</strong></td>
<td>-</td>
</tr>
<tr>
<td>Lescol (G)</td>
<td>4/11/2012</td>
<td>4/16/2012</td>
<td>-</td>
</tr>
</tbody>
</table>

Notes: Summary of entry events during the 1996-2013 period, with the first date in which the drug appears in MarketScan data and the drug company that owns the drug also reported. There is often a small delay between approval and subsequent availability to patients, due to logistical issues. The major events in the sample period (1996-2013) involve Zocor, Lipitor, and Crestor, and are highlighted in bold. (G) refers to generic and (XR) refers to extended release. Drugs with bracketed information are fixed-dosage combination drugs. Baycol was pulled from the market in 2001 due to safety issues. BMS refers to Bristol Myers Squibb. Sources: Drugs@FDA and drugs.com
1.3.3 Methodology and Identification Tests

My methodology for isolating history-dependence is based on comparing patients who start medication just before and just after a new branded drug launches. It relies on the assumption that patients do not actively choose their treatment start date, which I test using smoothness and balance checks and by analyzing a subsample of patients who are presumably inflexible in their start dates. The structure provided by the methodology also allows for an alternative way to evaluate the long-term health impacts of new chronic drugs.

Using the quarterly choice data and new patient classification from Section 1.3.2, I track the choices of patients in the two groups to infer the causal effect of being assigned the new drug as initial treatment on the probability a patient chooses that drug in later periods. More formally, the regression specification is:

\[ Y_{it} = \alpha + \beta_t Y_{i0} + \epsilon_{it} \]  

(1.1)

where \( Y \) represents whether the patient chose the entering drug for the group they are in, \( i \) indexes the user, and \( t \) the quarters since launch. \( \beta_t \) represents the effect of initially choosing the new option on whether the patient is on the new option \( t \) quarters later.

To analyze the quasi-experiment in an intent-to-treat manner, I instrument for \( Y_{i0} \) using \( Z_i \), an indicator for whether the patient is in the treatment group. Being in the treatment group opens up the possibility that the patient starts on the new drug, and the instrumental variables framework helps recover a treatment-on-treated estimate (\( \beta_t \)). The framework essentially scales any gap in outcome between the groups in period \( t \) by the initial gap in outcome.

The identification assumption necessary to interpret \( \beta_t \) causally is that patients do not actively choose the start date of their treatment, which is reasonable given institutional details but possibly problematic. This is intuitively reasonable in a health setting, as patients probably start treatment based on medical need, on the advice of their doctors. However, there are still mechanisms that may create unobservable differences between the treatment and control groups. One possibility is that patients who end up in the treatment group
anticipate future entry, and wait to start on the new drug for either poor match quality with existing treatments or financial reasons. For example, in the case of Hepatitis C, many patients with mild symptoms waited for the second entrant in the latest generation, Viekira Pak, because of the high price tag on the first entrant, Sovaldi. Another possibility is that firms often advertise heavily after launching a drug, and therefore the treatment group may be selected based on general responsiveness to advertising.

Therefore, I test this passive timing identification assumption by checking smoothness in the number of patients starting treatment, balance on observables across the two groups, and by re-running the analysis on a subset of users with concurrent inpatient events. The first two would present suggestive evidence in favor of the assumption, but still may miss out on unobservables. The third would provide more definitive evidence, as users almost certainly do not choose their inpatient admission dates based on the availability of new cholesterol medication.

For the core regressions, I set the time index \( t \) to be quarters since drug launch, but also provide robustness checks by setting it to quarters since the individual started treatment. One problem with the 180-day window described above is that patients begin treatment in different quarters. My main approach will be to compare treatment and control patients at any given point in time, indexed by quarters after drug launch, with the null hypothesis being that the two groups should make the same choices given the same environment. I also try an alternative approach where \( t \) indexes quarters since the individual started on treatment. The advantage of this approach is that it lines up all patients in the analysis sample, so that we are tracking them at the same age and experience. This change only accentuates the effects that I find, because treatment patients enter later than their control counterparts, so for any given relative period \( t \), the new drug will be more popular for the treatment group.\(^{30}\)

\(^{30}\)For example, at \( t = 1 \), the drug has not even launched for some control patients, whereas the new drug is already popular by the time some treatment patients enter.
1.3.4 Estimation Results - Inertia

I present estimates of the causal impact of starting on a new drug on the probability of choosing the same drug in later periods, what I term inertia. Assigning a patient to a new drug generates a persistently higher probability of choosing that drug in the future. To check the validity of the estimates, I repeat the analysis on patients who start treatment after an inpatient episode.

In Figure 1.1, I present two pieces of graphical evidence on inertia. The left graph in Figure 1.1 compares the market share on the new drug between the “before” and “after” groups, pooling together all entry events. The graph shows that there is a large initial gap between the two groups during the launch quarter, with the control group having negligible share on the new drug, and that the gap between the groups persists over time. If my aforementioned identification condition holds, then there is an equal number of patients in the control group versus the treatment group who are suitable for the new drug, but there remains a gap due to differences in initial choice conditions. Another way to view this is that the new drug becomes adopted over time at some natural rate, but switching costs are preventing suitable patients from taking the drug. The right graph shows a case study of 2007 Crestor market share by starting cohort quarter, with the vertical line representing the launch date of Crestor in late August 2003. There is a clear discontinuity in the market share, with post-launch cohorts more likely to choose Crestor, even four years later.
More formally, I present estimates of Equation (1.1) in Panel A of Table 1.3, which show large effects of inertia over several years. As explained earlier, the $\beta_t$ coefficient can be interpreted as a treatment-on-treated estimate of inertia. The final column of Panel A in Table 1.3 shows that the causal effect of starting on a new branded drug on probability of subsequent usage is still 55 percent at 15 quarters after drug launch, where 100 percent would represent absolute persistence and 0 percent no inertia.\footnote{See Figure A.2 for a full set of $\beta_t$ estimates.} As mentioned earlier, previous studies of prescription drug demand had touched on dynamic effects, but none documented large effects over several years.

\footnote{See Figure A.2 for a full set of $\beta_t$ estimates.}
Table 1.3: Estimates of the Impact of Inertia

Panel A: Unbalanced Panel

<table>
<thead>
<tr>
<th></th>
<th>t+3</th>
<th>t+6</th>
<th>t+9</th>
<th>t+12</th>
<th>t+15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initially Chose New Drug</td>
<td>0.687***</td>
<td>0.632***</td>
<td>0.547***</td>
<td>0.580***</td>
<td>0.545***</td>
</tr>
<tr>
<td>(0.00922)</td>
<td>(0.0142)</td>
<td>(0.0171)</td>
<td>(0.0196)</td>
<td>(0.0197)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>1354497</td>
<td>1005048</td>
<td>774485</td>
<td>611098</td>
<td>419943</td>
</tr>
<tr>
<td>R²</td>
<td>0.424</td>
<td>0.230</td>
<td>0.159</td>
<td>0.126</td>
<td>0.095</td>
</tr>
</tbody>
</table>

Panel B: Balanced Panel

<table>
<thead>
<tr>
<th></th>
<th>t+3</th>
<th>t+6</th>
<th>t+9</th>
<th>t+12</th>
<th>t+15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initially Chose New Drug</td>
<td>0.685***</td>
<td>0.542***</td>
<td>0.473***</td>
<td>0.408***</td>
<td>0.380***</td>
</tr>
<tr>
<td>(0.0218)</td>
<td>(0.0317)</td>
<td>(0.0344)</td>
<td>(0.0368)</td>
<td>(0.0363)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>140133</td>
<td>127515</td>
<td>121390</td>
<td>117554</td>
<td>101802</td>
</tr>
<tr>
<td>R²</td>
<td>0.420</td>
<td>0.205</td>
<td>0.141</td>
<td>0.107</td>
<td>0.088</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

* p < 0.05, ** p < 0.01, *** p < 0.001

Notes: Estimates of Equation (1.1) for 3, 6, 9, 12, and 15 quarters after drug launch, using the unbalanced and balanced panels of MarketScan enrollees. The coefficients reflect the effect of randomly assigning a user to a drug on the probability that the user chooses the same drug in later quarters.

The coefficient estimates shown in Table 1.3 are not completely monotonic, which has to do with the pooled nature of the analysis combined with differential attrition in the unbalanced panel. As noted earlier, due to issues of attrition in the panel, some events have less representation in later quarters. For example, as noted above, there was high attrition in 2005. Therefore, the Crestor entry event will have differential weights in early versus late period results. I show in Panel B of Table 1.3 that the coefficients do become monotonic once I account for attrition by using a balanced panel for my analysis.

Even accounting for attrition, the coefficients reflect a non-uniform hazard rate. As shown in Panel B of Table 1.3, $\beta_t$ drops significantly in earlier periods, but then the trajectory becomes flatter in later years. This suggests that switching becomes less frequent as a patient takes it more, either because of learning or habit formation. This result is consistent with
comments by PBM decisionmakers, who suggested that there is a higher rate of switching in the first six months to a year, after which switching is much less frequent. For tractability reasons, I later simplify demand to a first-order Markov process, but the actual patterns are richer.

The qualitative nature of the pooled results also holds across each entry event during the period. In Figure A.4 in the Appendix, I display graphs representing the analysis for major entry events in the period. The results are generally very consistent across entry events, with an initial gap between treatment and control groups that persists over time.

1.3.5 Establishing Design Validity - Inpatient Starters, Smoothness, and Balance on Observables

My main test of experimental validity is to re-run the same analysis on a subsample of patients who start on anti-cholesterol drugs after a hospitalization. I also show smoothness in the rate of patients starting on anti-cholesterol drugs over time, as well as balance on observables.

The key threat to validity is that some patients wait for new drugs before starting treatment, so I pick a sub-sample of patients starting treatment after hospitalization, who probably are inflexible in start timing. As described earlier, my design is no longer valid if the two groups are unobservably different in preferences, which could be the case if patients have flexibility in choosing their start time. One group unlikely to have such flexibility is hospitalized users. Those hospitalized for heart-related health events, such as heart attacks, typically start on anti-cholesterol medication right after, in order to reduce the probability of future events. Therefore, they would have little control over their start time.

I generally find similar but noisier results in this sub-sample. The sub-sample is about 6% of the full sample of patients, making the results much noisier. Table 1.4 reports the estimates, which are slightly smaller in magnitude than the ones estimated using the full sample, but generally similar. This bolsters the case that the estimates from the larger sample are reliable.
### Table 1.4: Estimates of the Impact of Inertia - Inpatient Starters Panel

<table>
<thead>
<tr>
<th></th>
<th>t+3</th>
<th>t+6</th>
<th>t+9</th>
<th>t+12</th>
<th>t+15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initially Chose Drug</td>
<td>0.543***</td>
<td>0.571***</td>
<td>0.600***</td>
<td>0.375**</td>
<td>0.463***</td>
</tr>
<tr>
<td></td>
<td>(0.0593)</td>
<td>(0.0836)</td>
<td>(0.101)</td>
<td>(0.123)</td>
<td>(0.121)</td>
</tr>
<tr>
<td>Observations</td>
<td>81599</td>
<td>59061</td>
<td>45697</td>
<td>35076</td>
<td>23573</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.257</td>
<td>0.138</td>
<td>0.101</td>
<td>0.065</td>
<td>0.064</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Estimates of Equation (1.1) for 3, 6, 9, 12, and 15 quarters after drug launch, in a sub-sample of patients who start treatment after a hospitalization. The coefficients are comparable but noisier relative to those in Table 1.3.

As additional evidence, I also look for smoothness in new users and balance on observables. Figure 1.2 plots the number of new patients starting on anti-cholesterol medication by month, which shows that there is little evidence of spikes near key entry events marked by the vertical lines, which start dates were sensitive to entry events.\(^3\) In terms of balance on observables, MarketScan does not have a very rich set of characteristics, but I find that age, gender, hospitalization in the starting quarter, and number of other prescriptions are balanced across the two groups. Finally, I re-run the core analysis with age, gender, and other observable health measures as controls, and find identical estimates.

\(^3\)I show the balanced panel results here because it mitigates issues surrounding user turnover in the unbalanced panel. Figure A.3 in the Appendix has the corresponding plot for the unbalanced panel, and also shows no evidence of discontinuities, apart from trends within a given year, driven by users who join the dataset.
Notes: A plot of the number of new patients starting on cholesterol medication during each month in the balanced 10-year MarketScan panel. The vertical lines represent the entry dates of Crestor, Vytorin, generic Zocor, and generic Lipitor. The graph generally shows a smooth trend, with no noticeable spikes around entry events.

1.3.6 History-Dependence More Broadly - Generics and Incremental Drugs

Modifying the previous framework slightly, I provide additional evidence on how consumers behave with respect to generics and incremental drugs, two important and distinctive aspects of the pharmaceutical industry, with generics particularly relevant for my later modeling choices. Generics appear to be difficult to switch from, but easy to switch to for consumers on the corresponding brand drug. Switches to incremental drugs appear to be easier for users on a drug that contains a common molecule.

As mentioned in Section 1.2, generics and incremental drugs play a key role in many drug markets, making it important to account for them in our demand analysis, especially with an eye towards firm pricing behavior around entry of these drugs. As shown in Figure A.5 in the Appendix, Zocor and Lipitor generics begin to dominate the market once they enter. In addition, Vytorin, a combination of Zocor and Zetia, enjoys a period of popularity between
its launch in 2004 and the aforementioned ENHANCE trial announcement in January 2008.\textsuperscript{33}

The shortcoming of using the same framework above to analyze generic entry events is that it misses out on causal effects across types of drugs, which are also relevant for drug company behavior. The results from using the same quasi-experimental setup as above can be interpreted as the effect of starting on generic Zocor on the probability a patient chooses generic Zocor in later periods. This is alone is interesting, as a strong effect would suggest that branded drugs that enter after generic entry would struggle to convert users.\textsuperscript{34} However, it misses out on other relevant mechanism for market analysis. This includes the impact of choosing various branded drugs on later generic adoption. For example, if Merck knows that patients on branded Zocor will switch immediately to generic Zocor, they have little incentive to build market share in the period before generic entry.

I explore additional mechanisms in two complementary ways: by breaking down the control group for generic entry events and by analyzing branded launch events that occur a little before generic entry events. The first approach is to look at the different types of choices made by patients starting just before brand and generic entry events, and how they appear to affect subsequent generic choice. For example, for the generic Zocor entry event, I can look at patients who start on Lipitor, and analyze their later adoption rates of generic Zocor. The drawback of this analysis is that the initial choice reflects unobserved heterogeneity in preferences, which will carry over into whether the patient later chooses the generic. A second, complementary approach is to analyze branded entry events that occur in the few years before the generic entry of interest, substituting in generic choice as the outcome. As discussed above, the branded entry causes treatment users to take the entering drug at a higher rate. I can then trace out the impact of starting on Crestor on later adoption of generic Zocor.

\textsuperscript{33}For my demand estimation, I leave out Vytorin, as it is almost exclusively used for the highest treatment intensity level, and complicates both the state space and introduces the issue of multi-product firms.

\textsuperscript{34}This accentuates any price differences that already put late brand entrants at a static disadvantage, which is the focus of Gilchrist (2016). The limited commercial success of the recently launched PCSK-9 cholesterol drugs is consistent with this story.
The results from the first approach show strong inertia in terms of staying on generic drugs, a high adoption rate of generics for patients who start on the same molecule, a much lower one for patients starting on other branded drugs. Panel (a) of Figure 1.3 shows the basic result surrounding inertia for generic Zocor. Treatment users exhibit a persistently higher market share on generic Zocor. A different way to see this is to break down the control group for branded entry events, as I do in Figure 1.4, which shows generic almost never switching to new branded drugs.

35 Results for generic Lipitor entry are very similar.

36 This is accentuated by the fact that for entries in this period, the only previously available generic is Mevacor, which is generally considered to be inferior to newer drugs. As mentioned above, the caveat here is there could be persistent unobserved heterogeneity in tastes or price sensitivity.
Figure 1.3: History-Dependence in Generic Drug Choice

Notes: Shows a breakdown of the fraction of patients who choose generic Zocor by (a) treatment and control groups defined using the 180-day window used for Figure 1.1 and (b) treatment and several control groups, with the control groups broken down by type of initial drug choices. The first graph shows a similar pattern to those for branded drug entries, with those in the treatment group more likely to adopt generics. The second graph breaks the controls into patients who started on branded Zocor (orange), patients who started on a different drug (green; typically Lipitor and Crestor), and the few patients who start on generic Mevacor (red). The graph shows that over 80% of users who start on branded Zocor switch to the generic version, whereas the adoption rate of Zocor generics is much lower for the other groups. This is suggestive of greater switching costs across molecules.
**Figure 1.4: Adoption of Brand Drug by Previous Choice**

*Notes:* Entry event analysis with control group breakdown for branded drug entries. Users already on generic drugs, mostly Mevacor for branded entries in this period, are much less likely to adopt the new branded drug relative to control users who picked an existing branded drug.

Panel (b) of Figure 1.3 then provides suggestive evidence of other forms of history-dependence surrounding generics, by breaking down the control group by the drug users start on before generic Zocor enters. The first noticeable feature in the graph is that over 80% of control group patients who start on branded Zocor switch over to generic Zocor within a year, suggestive of almost automatic switching. This is not surprising, given the previous discussion of generic substitution laws in Section 1.2.1. Another result of note is that patients who start on a different branded drug switch at much lower rates, barely reaching 20% after four years. This provides suggestive evidence that previous choice influences subsequent choice, which drug companies can use to form strategies around generic entry.

The complementary approach based on analyzing branded drug entries with generic choice as outcome confirms the existence of a causal mechanism linking brand and generic drug choices. The specific approach is to use the entries of Crestor (2003) and Vytorin (2004) to analyze generic Zocor adoption in later years. The treatment group in each case is pushed
towards choosing the new entering brand. I then trace out the difference between treatment and control groups in later adoption of generic Zocor. In Figure 1.5, I analyze theVytorin entry event by plotting the causal coefficient from a modified version of Equation (1.1), with generic Zocor choice as the outcome instead of Vytorin choice.\(^{37}\) The graph shows that starting on Vytorin has a negative causal effect on the later adoption of generic Zocor, a form of history-dependence different from inertia. One interpretation of this is that Vytorin takes such patients away from branded Zocor, which later results in a more difficult switch, absent the effect of generic substitution laws.

![Figure 1.5: The Impact of Vytorin on Generic Zocor Adoption](image)

*Notes:* A plot of the estimated coefficients from Equation (1.1), using the Vytorin entry event but with generic Zocor usage as the outcome instead of Vytorin usage as the outcome. The graph shows that the treatment group, those pushed towards starting on Vytorin, are less likely to adopt generic Zocor, until the announcement of the ENHANCE trial around quarter 15, at which point patients on Vytorin start to switch to generic Zocor.

Finally, I show using the same two approaches that incremental drugs, such as combinations, also exhibit various forms of history dependence. Here, I focus on combination

\(^{37}\)Similar but noisier results are present for the Crestor entry event.
drugs, including Vytorin, a combination of Zocor and a non-statin Zetia. First, I analyze the Vytorin entry event by breaking down the control group. The patterns in Figure 1.6 show that control group patients who start on a drug containing a molecule also present in the launched combination adopt at the highest rates, although still far lower than the rates observed for generics. I also take the second approach, by analyzing the impact of starting on Crestor on later usage of Vytorin. In Figure 1.7, I present estimates showing that there is a significant negative causal impact of starting on Crestor on later usage of Vytorin.

![Figure 1.6: Adoption of Combination Drugs by Previous Choice](image)

**Notes**: Entry event analysis with control group breakdown for combination therapies, primarily Vytorin. Patients on medication that overlap with the components of the combination are significantly more likely to adopt the combination, whereas patients already on generics rarely adopt the new drug.

---

38 Other, much less popular combinations include Simcor (Zocor plus Niacin).

39 As mentioned earlier, Merck officials in earnings calls had to explicitly state that they wanted Vytorin to be adopted by all patients, not just those already on Zocor, possibly contrary to the prevailing view among investors.
Notes: A plot of the estimated coefficients from Equation (1.1), using the Crestor entry event but Vytorin usage as the outcome instead of Crestor usage. The graph shows that the treatment group, those pushed towards starting on Crestor, are less likely to adopt Vytorin.

1.3.7 The Role of Doctors

So far, I have ignored discussion of the micro-foundations driving the observed statistical patterns. Here, I focus on the role of doctors, as they can generate differences in the dynamic incentives of drug companies with respect to market share. I relegate discussion of traditional demand mechanisms to the Appendix A.5, as I will be implicitly assuming for the rest of the paper that history-dependence is a structural feature of the anti-cholesterol market.

Here, I investigate whether doctors generate additional history-dependence effects by prescribing to new patients in a way that is affected by their previous prescriptions, which would affect drug company incentives to invest in market share and hint at whether effects exist in acute markets. In my analysis far, I have focused on doctors and patients as a single decisionmaking unit. However, doctors themselves may exhibit additional inertia, in
the sense that their previous prescriptions influence their current choices for new patients. In effect, doctor inertia would accentuate the impact of previous market share on current outcomes. It also matters for understanding whether similar forces may exist in acute markets, where only doctors are making repeated choices over time, each time for a new patient.  

I find that doctors generally prescribe many different drugs and exhibit little variation in market shares by medical school graduation cohort, suggesting that they exhibit little additional inertia beyond the effects generated through interacting with existing patients. To gain insight into doctor behavior, I bring in publicly available Medicare Part D files, covering the 2013-2014 period. For each year, the data provides aggregate prescriptions for each doctor. As shown in Figure 1.8, 75% of doctors prescribe at least three different cholesterol drugs in a given year, with an average Herfindahl index of 0.41. In addition, there does not appear to be a noticeable difference in prescription behavior by medical school graduation year. Under the same conceptual framework as my empirical methodology, one might expect that doctors may stick with the most popular prescription at the time of their medical school graduation, based on how they were taught. However, Figure 1.9 suggests that there is little difference across cohorts. For example, Crestor prescription rates for doctors graduating before and after 2003 is relatively similar. The only noticeable difference is that recent graduates are more likely to prescribe generic medications.  

Overall, the signs point to doctors updating current preferences based on medical guidelines, while not changing the prescriptions of existing patients.

40A leading example is the case of Hepatitis C drugs, where patients are cured after a 12-24 week drug regimen, and therefore do not make repeated choices across multiple years.

41Anecdotally, there has been a shift to the usage generic names of drugs in medical school training to to prescribe generics when available.
Figure 1.8: Number of Distinct Drugs Prescribed by Physicians in 2014

Notes: A histogram showing the variety in number of distinct molecules prescribed (counting generic and brand of the same molecule as one drug), across the 210k doctors that prescribe any cholesterol medication to Medicare Part D patients in 2014. About 75% of doctors prescribe at least three drugs with different molecules.
Notes: A plot of market share (prescriptions) by medical school graduation cohort in the Medicare Part D data. The graph shows little variation across graduation cohorts, suggesting that doctors are only minimally influenced by their own past choices.

The evidence is consistent with the story that doctors take an “if it ain’t broke, don’t fix it” to prescribing. Physicians are generally afraid of unpredictable side-effects and adverse drug interactions that emerge when patients switch, and therefore are leery of changing a patient’s prescription, even under pressure from payers. Therefore, they will only consider switching in cases where the existing drug causes side effects or if the patient has a major medical event. The data does indicate that patients are more likely to switch drugs when they need a change in intensity of treatment or experience a hospitalization.

There are important caveats to note here relative to the literature on doctor prescription behavior, as the findings are not necessarily inconsistent with doctor-driven heterogeneity in prescribing patterns or effects from detailing. For my analysis, I am only interested in

---

42 See discussion article “Drug switching: lowering costs vs. adverse interactions, potential error” in the July/August 2000 issue of the ACP-ASIM Observer.

43 This story is in the spirit as the breakdown model of demand in Smallwood and Conlisk (1979).
whether doctors causally carry previous prescription choices over to new patients, which would create a stronger gradient by medical school graduation year. Doctors can still have consistent preferences over time that leads to persistence in prescribing behavior, but this variation can average out over a cohort. Detailing can also have an effect on future prescriptions, as doctors leave returning patients on the same drug.

A final caveat is the limitations created by the data, which future research can address using more detailed data. First, I am observing behavior several years after the last big entry event, so there may have been differences across graduation cohorts in earlier years. Second, medical school graduation year may not provide a sharp instrument for studying preferences, as doctors may learn about prescribing both during medical school and later in residency. Having doctor-level micro data may be able to help identify lasting effects of initial prescribing behavior on later prescriptions, using a similar framework to the one presented here.

### 1.3.8 Second Choice for Continuing Patients - Evidence from Baycol Removal

Having established robust evidence of inertia, I now assess why the outside option is the second most popular option (behind the incumbent drug) for patients previously taking medication. This could mean that the second choice of most continuing patients is the outside option, for example due to bad matches with other drugs, or it could reflect unobserved heterogeneity in preference for the outside option, such as an adherence shock. The distinction is crucial for assessing counterfactuals under which one of the drugs is removed, which is often done to assess drug value or to understand drug insurance design.

Ideally, a researcher would want to analyze cases where one of the drugs is randomly removed from a patient’s choice set. After drug A is removed, if patients previously taking A mostly move to the outside option, it would suggest that other drugs are a bad match for most patients taking A. On the other hand, if most patients move to another drug, it would suggest that adherence shocks or a common shock for all drugs is driving the popularity of the outside option.
My approach for approximating the ideal experiment is to analyze patient responses to the removal of Baycol from the market, which show that the second choice of a majority of continuing patients is another drug. As mentioned above, Baycol was withdrawn from the market in August 2001 due to a number of severe side-effects. The conditional shares in Table 1.5 show that about 60% of Baycol users switch to one of the two most popular statins after it is withdrawn. Table 1.5 also shows Baycol market share in 2000 for conditional on previous use in 1999. The 60% of users who switch to another drug is higher than both an IIA projection using year 2000 Baycol conditional shares (50% based on the first row in Table 1.5) and the share for patients not previously using a drug (38% based on the second row in Table 1.1). This suggests that there is significant unobserved heterogeneity in preference for the outside option. In fact, the difference is likely to be a lower bound of second choice probabilities for drugs without safety concerns, as the severe side-effects from Baycol could deter existing users from taking other drug.

Table 1.5: Conditional Market Shares - Baycol User in Previous Year

<table>
<thead>
<tr>
<th>Conditional Market Share</th>
<th>Zocor Share</th>
<th>Lipitor Share</th>
<th>Baycol</th>
<th>No Drug</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>0.03</td>
<td>0.07</td>
<td>0.80</td>
<td>0.09</td>
</tr>
<tr>
<td>2001</td>
<td>0.07</td>
<td>0.20</td>
<td>0.61</td>
<td>0.12</td>
</tr>
<tr>
<td>2002</td>
<td>0.24</td>
<td>0.35</td>
<td>0</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Notes: Market shares are computed based on data from the Medical Expenditure Panel Survey, and are computed at the patient level. Patients who have multiple drugs on record are assigned the drug they take the most in a given year. The market size is defined as all patients who are diagnosed with high cholesterol or taking anti-cholesterol medication.

The possibility that this Baycol quasi-experiment is biased due to scarring effects can be addressed by finding and analyzing formulary exclusions. As noted, Baycol was withdrawn due to severe side-effects, which could lead to additional aversion to all statins among existing Baycol users. I am in the process of identifying plans in the MarketScan data that exclude a drug for reasons not related to safety, and seeing how existing users of the excluded drug respond.
1.4 An Empirical Model of Demand for Chronic Drugs

My goal in this section is to construct an empirical model of chronic drug demand that incorporates previous choice and insurance design. I start by discussing the motivation for estimating a demand model. Next, I construct a model that can match the features of demand isolated in Section 1.3. To conclude, I discuss how incorporating my demand system can be used to reassess some of the key questions explored in previous research.

1.4.1 Why a Choice Model?

The main advantage of a choice model is that it can help us understand substitution patterns on the demand side. In Section 1.3, I isolated one particular dimension of demand, namely the causal impact $\beta_t$ of choosing a drug in period 0 on the probability of choosing the same drug $t$ quarters later. However, other factors governed by drug companies and insurance designers are also important in governing demand, particularly insurance design and advertising. Therefore, to understand the impact of demand patterns on firm behavior and market outcomes, one needs to a model that incorporates the impact of all of these factors on consumer choice.

In this paper, I will focus on the impact of history-dependence and insurance design, incorporating both factors in my model of demand. Traditionally, demand models have been functions of prices and product characteristics, which can be used to understand the impact of new goods and policies that affect pricing such as mergers. The goals here are similar, but consumers in drug markets face insurance plans rather than prices, so the demand system should be a function of insurance design. In addition, given the evidence presented in Section 1.3, history-dependence is a first-order feature that should be included in demand modeling. Other factors such as advertising and coupons will be accounted for in drug-by-period fixed effects. Overall, demand in the model, $D_t(p, D_{t-1})$, will be a function of a vector of cost sharing $p$ and demand in the previous period.\textsuperscript{44} Cost sharing

\textsuperscript{44}A more general demand function would include more lags. This structure is similar to the one used in Sinkinson and Starc (2018), but I will separate out history-dependence from unobserved heterogeneity.
can be infinite to reflect the lack of availability of a given drug.

Estimating such a demand system will be the first step towards understanding insurance coverage policies, pricing, the value of new drugs, and innovation policy. The most direct application of the demand model will be for understanding the effects of insurance design on market shares. Given a particular insurance design, the demand system will output a set of predicted market shares. In order to understand pricing and evaluate policies, one needs to go a step further and embed demand in a model of competition in drug markets. In Chapter 2, I use a simplified version of the demand model to understand drug pricing, by modeling the competitive process for setting insurance parameters. One can also explore policies that affect insurance design, such as Medicare Part D regulations on the minimum number of drugs that need to be covered in a given drug market. Another potential use is to estimate the value of new drugs, by evaluating counterfactuals where a drug is removed. Finally, the presence of previous choices in the demand function allows us to quantify first-mover advantage as well as the returns to developing incremental drugs, both of which factor into innovation policy in the pharmaceutical industry.\textsuperscript{45}

1.4.2 A Choice Model for Chronic Drugs

Here, I lay out my model of demand, which satisfies the general framework laid out in Section 1.4.1, while being flexible enough to capture the patterns found in Section 1.3. The model builds on the existing literature, especially the state-dependent demand model constructed by Sinkinson and Starc (2018).

Given the results above and internal consistency considerations, the main features that the model needs to satisfy are:

1. High conditional market shares for incumbent drugs (defined as the drug chosen in the previous year by a patient)

2. Very low conditional market shares for non-incumbent drugs relative to the outside

\textsuperscript{45}The effectiveness of FDA priority review vouchers rely on first mover advantage and both FDA exclusivity rules and patent policy affect the returns to incremental drugs.
3. Significant switching rates to other drugs if the incumbent drug is unavailable

4. Small but non-zero extensive margin elasticity with respect to cost sharing

My model will generate these features by having a cost for existing users to obtain a new prescription. As shown in Table 1.1, market shares conditional on choosing drug A in the previous year are mostly split between drug A and the outside option. This suggests some doctor and/or patient unwillingness to switch prescriptions. When drug A is removed from the market, a sizable fraction of people do indeed switch to another drug. Given these patterns, I set up a model with a simple nested structure, which is shown in Figure 1.10. At the top level, a patient decides whether or not to take any drug. This is similar to the structure in Sinkinson and Starc (2018), and can account for the significant fraction of continuing patients who choose the outside option. At the second level, patients choose between different drugs, and gain a boost from taking the incumbent drug. Intuitively, the shocks at the lower level should be highly correlated, and therefore the incumbent drug becomes the clear winner most of the time. If the incumbent drug is removed, then it will reduce the value of the “inside” nest, increasing the share of patients picking the outside option. However, I allow for unobserved heterogeneity in the value of taking a drug, so many continuing patients will have high unobserved value for any drug, and thus switch to another drug.

Formally, each drug $j$ will generate an average utility for a patient $i$ in period $t$:

$$v_{ijt} = \delta_{jt} - \alpha p_{jt} + v_i I_{j \neq 0} + \psi_{ij}$$

(1.2)

where $v_i$ is a taste shock for any drug, and $\psi_{ij}$ is a match-specific shock for each drug.
In ongoing work, I aim to estimate the model by matching moments from the individual-level claims data. In Chapter 2, I estimate a simpler version of the model using maximum likelihood in order to have a simple demand system for dynamic game estimation. However, the control function approach for dealing with unobserved heterogeneity will fail to satisfy standard assumptions. Therefore, my alternative approach here is to match conditional market shares, as well as the difference in market shares between the treatment and control groups around drug launches. In addition, I will also incorporate moments found in the literature surrounding cost sharing variation to identify the $\alpha$ parameter.

1.4.3 Potential Implications for Existing Research

Here, I discuss the potential implications for existing research on prescription drug markets. My goal is to explore these implications in greater depth in future research.

One area in which the demand model could help is in understanding insurance design. Existing research, such as Einav et al. (2016), have used within-patient cost sharing variation to estimate elasticities. Accounting for history-dependence can help separate out the elasticities generated by new patients versus those generated by existing patients. This distinction may be important for understanding across-drug or across-market patterns in cost sharing elasticities and insurance design.

---

46 The main problem is that the indicator for incumbent drug is binary, and any predictive regression will form residuals that are not normal, a typical requirement for the control function approach to be exactly right.
Another way in which the model can help is in understanding pricing in drug markets, which is relevant for pricing policy and assessing the value of new drugs. As I explore in Chapter 2, incorporating history-dependent demand can help us better understand pricing in drug markets. This in turn is useful for assessing the price and welfare impacts of new drugs, which is the goal of several existing papers in the literature, including Dunn (2012), Arcidiacono et al. (2013), and Bokhari and Fournier (2013).

A final, potentially fruitful application of the model is towards assessing the impacts of innovation policy. As discussed earlier, several important drug innovation policies hinge on demand dynamics. First, FDA priority review vouchers may depend on history-dependence for their effectiveness, but may also have additional dynamic effects. History-dependence will lead to a first-mover advantage for the redeemer of the voucher, making it more valuable than just the extra four month’s worth of sales. However, the voucher will also have unknown dynamic effects on future pricing in the market. Second, both patent policy and FDA exclusivity policy affect the development of incremental drugs, such as reformulations and combinations. The incentives to develop incremental drugs and the optimal timing of their introduction will depend on demand.

1.5 Conclusion

In this chapter, I have provided quasi-experimental on history-dependence and unobserved second choices, and then built an empirical model of chronic drug demand that can be used for industry and policy analysis.

My main contribution is to provide a quasi-experimental framework for showing the importance of inertia and other forms of history-dependence in chronic drug demand, which should be accounted for in most analyses of the industry. My estimates suggest that inertia has significant effects on choice probabilities over several years, which I then use to analyze pricing outcomes. However, as I detail in Appendix A.1, there are many firm investment and entry decisions that also appear to be affected by history-dependence in demand, suggesting that most research into the industry should take these effects into
consideration. It also points to future research into the effect of demand on innovation, which can help guide market and innovation policies related to the pharmaceutical industry.

My other contribution is to provide a model of demand that incorporates history-dependence and can generate the key demand features in the anti-cholesterol market. Existing models in the literature do not explicitly separate out history-dependence and heterogeneity, which prevents analyses of dynamic issues in the industry. My model flexibly allows for the data to separate out these two channels. In ongoing work, I aim to estimate the model and apply it to problems analyzed the literature discussed in Section 1.4.3.

In terms of other applications, the quasi-experimental framework provides insight into dynamic linkages in choices by leveraging a temporal discontinuity, an approach that can be applied in other prescription drug markets and other industries. Here, I focus on variation generated by drug launches to alter initial conditions, under the assumption that patients do not actively choose the time they begin treatment, a more likely and verifiable assumption in prescription drugs. As discussed in the context of the ENHANCE trial, other significant events can also be analyzed using this framework. Beyond pharmaceuticals, it is a little harder to find settings where start timing is exogenous, but institutional details that limit choices by age serve as a prime candidate for analyzing other dynamic linkages in choices and outcomes.

In the context of prescription drugs, my framework can be used to analyze the health and spending impacts of new drugs. As briefly explored in the Appendix A.4, the framework can be used to assess the long-term health and spending impacts of a new drug relative to the existing standard of care. In the case of anti-cholesterol drugs, this analysis yields little health differences, consistent with clinical trial evidence. However, the methodology may help in other contexts in using observation health data to evaluate causal treatment effects for drugs.
Chapter 2

Pricing Intermediaries in Prescription Drug Markets: Impact and Implications for Government Negotiation

2.1 Introduction

In this chapter, I build a model of prescription drug pricing in chronic drug markets that is one of the first to include a formulary designer, known in the industry as a pharmacy benefit manager (PBM). Relative to models in the literature, my model is better at explaining negotiated price data in the anti-cholesterol market without resorting to high marginal costs, and allows for an evaluation of the impact of the PBM industry. Using the model, I find that the presence of PBMs cut drug company profits by 25%, but that PBMs capture up to 40% of the savings. I conclude by showing that vertical integration between drug companies and PBMs can lead to higher spending, and that government-led negotiation would need a strong exclusion threat to achieve current levels of spending.

There is significant interest in the determinants of drug pricing in the United States,
highlighted by discussions in Congress and reports in the media on recent cases of drug inflation. These cases include drugs with no close substitutes such as Daraprim, which increased from $13.50 per tablet to $750 within a few weeks, and the EpiPen, which increased from $100 to $600 over a decade, but also include cases in markets with close substitutes, such as the insulin market. All of this has led to significant Congressional and judicial scrutiny of agents involved in the industry’s pricing structure, including drug manufacturers and PBMs, with accusations of collusion and proposed centralized price negotiation for Medicare Part D. In order to make progress in understanding how these agents impact drug pricing, one needs to include these agents in a competitive model of drug markets.

In my paper, I highlight the role played by PBMs by including them in a model of drug pricing in the anti-cholesterol market. Specifically, I model the pricing structure as a finite-period dynamic game between drug companies, in which they make simultaneous price offers to a representative PBM in every period, and the PBM, in turn, sets a formulary that impacts demand for each drug. Both drug companies and PBMs account for the history-dependent nature of demand in the market, which I capture in a demand system based on the findings in Chapter 1. Taking the model to data on average negotiated prices in the anti-cholesterol market allows me to recover the structural parameters of the game, in particular the objectives of PBMs. I conclude by applying the model to understand the impact of PBMs on drug company profits and overall spending, and by using the model to highlight the pros and cons of potential government-led price negotiation.

To begin, I construct a model of drug pricing that solves some of the problematic aspects of previous drug pricing models while also allowing for an evaluation of the impact of the

---

1See NY Times articles “Drug Goes From $13.50 a Tablet to $750, Overnight” and “EpiPen Price Rise Sparks Concern for Allergy Sufferers.”

2See NBC News article “Is Insulin the New EpiPen? Families Facing Sticker Shock Over 400 Percent Price Hike.”

3As I detail in Section 2.2, PBMs negotiate prices with drug companies, using their ability to set formularies as a tool for extracting discounts.

4See articles “Drug Markets Accused of Fixing Prices on Insulin” (NY Times), “Drugmakers Point Finger at Middlemen for Rising Drug Prices” (Wall Street Journal), and “Medicare Should Leverage Buying Power to Pull Down Drug Prices, White House Says” (NPR).
PBM industry. Previous models in the literature have typically modeled drug companies as directly setting prices for consumers, with consumer demand a function of prices. This leads to two problems. First, direct pricing models lead to high estimates of marginal cost, even though small molecule drugs are known to have negligible marginal cost. For example, Arcidiacono et al. (2013) use a standard Nash pricing framework to study the anti-ulcer market, and find implied marginal costs that fluctuate significantly and rise as high as $120 for a $150 small molecule drug. Second, demand systems in the literature are forced to generate price or cost sharing elasticities greater than one, which is implausible if taken at face value, especially given recent research by Einav et al. (2016), which finds very inelastic demand with respect to cost sharing. My model solves these issues by adding the PBM as a profit maximizing intermediary, which negotiates prices with drug companies by leveraging their formulary setting power. PBMs serve the function of mapping prices to cost sharing, avoiding the need for high marginal costs or price sensitive consumers while also allowing us to better understand the role of the PBM in drug markets.

More specifically, my model ends up taking the form of a dynamic pricing game with drug companies bidding for formulary position in each period. As detailed in Chapter 1, demand in several large chronic drug markets, including the anti-cholesterol market, is strongly history-dependent. This is likely to have significant impacts on the decisions of both drug companies and PBMs. As detailed in Section 2.2.2, this is corroborated by decisionmakers in the industry. As such, I model prices as generated by a dynamic pricing game played by drug companies and PBMs, with previous market share serving as the state variable. In each period, drug companies bid for formulary position by submitting price offers to a representative PBM. The PBM then picks the formulary that maximizes its objective function. Negotiated prices are a result of equilibrium play in this game.

In order to have the right negotiated prices to feed into the model, I make a minor

---

5See Figures 3 and 6 in their paper.

6As I detail later, the representative PBM in my model represents the overall behavior of an imperfectly competitive industry.
contribution by measuring the average prices paid to drug companies, known as net prices. The existing literature on drug pricing generally works with data on list price, or sticker price, given the lack of data on discounts offered by drug companies to PBMs. Using drug-specific revenue data from financial filings, I construct novel data on net prices for anti-cholesterol drugs, finding significant differences relative to list prices. Unlike list prices, net prices do exhibit declines, particularly after generic entry of rival drugs. This drop is preceded by an increase in net prices before rival generic entry.

I then estimate the model using the average net price data, recovering parameters of the pricing game, in particular the PBM profit function. First, I estimate a switching-cost model of demand on individual-level prescription drug choice and formulary data, based on my results in Chapter 1. I then embed demand estimates in my model of pricing structure, and estimate the model using the aforementioned net price data. To do this, I play out the dynamic game involving firms and PBMs, with the former making take-it-or-leave it offers in each period to the latter and both types of agents accounting for consumer behavior. By matching to data on net prices, I recover key parameters surrounding the profit function of PBMs. The estimates suggest that PBMs generally aim to maximize cost-effectiveness to please their customers, but are limited in their ability to exclude popular drugs and are also influenced by other factors such as vertical integration.

The estimated model yields significant insight into the impact of PBMs, potential government price negotiation, and issues with vertical integration in the market. First, I find that PBMs are effective at counteracting the significant inflationary pressure from consumer inertia, because they are somewhat willing to exclude drugs from the insurance plans they design. This result stems from the lack of consumer sensitivity to cost sharing, which means that PBMs cannot be effective by just threatening to set higher copays. Overall, I find that PBMs cut significantly into drug company profits, but keep about 20-40% of the savings as their own profits (based on their financial filings), passing the rest along to payers. These results suggest that Medicare Part D bargaining would have an advantage in terms of saving the fees currently captured by PBMs, but would need a credible exclusion
threat to achieve current negotiated prices in the face of consumer inertia. Finally, prices in the anti-cholesterol market point to distortions created by vertical integration of drug companies and PBMs that are not competed away. This has implications for recent vertical merger cases involving pharmacies, insurers, and PBMs, such as ones between CVS and Caremark, CVS and Aetna, Rite Aid and Albertsons, and Cigna and Express Scripts.

The rest of the Chapter is organized as follows. In Section 2.2, I provide a more detailed overview of prescription drug markets. In Section 2.3, I lay out my model of drug pricing. In Section 2.4, I discuss estimating a demand system and then estimating my model of pricing. In Section 2.5, I run counterfactuals to study the impact of PBMs and potential government-led price negotiation. Section 2.6 concludes.

### 2.2 Background: Drug Pricing Literature, PBMs, and the Anti-Cholesterol Market

In this section, I provide necessary details on the drug pricing literature, the institutional details surrounding PBMs and pricing structure in the industry, and the specific market I study using my model.

#### 2.2.1 Drug Pricing - Background, Literature, and Policy

In this part, I provide some background on the key issues surrounding drug pricing. This includes the literature on dynamic pricing patterns, the problems with using list price data that is common in the literature, the anecdotal evidence on the pricing structure in the industry, and the most salient policy debates and discussions.

Pharmaceutical pricing has long drawn the interest of academic researchers, who have documented dynamic pricing patterns in various drug markets, particularly near patent expiration. An early paper that presents facts on pricing in the pharmaceutical industry is Caves et al. (1991).\(^7\) One key finding in the paper is significant price increases in the

---

\(^7\)A comment by Ariel Pakes on the paper advocates the use of micro panel data to understand observed
two years leading up to patent expiration. Subsequent papers, such as Dunn (2012) on the
anti-cholesterol market, have noted increasing trends in prices over time, despite competitor
and generic entry. Recent work by Aitken et al. (2016) provides a broad overview of pricing
trends in recent years, providing a breakdown that shows price increases on incumbent
branded drugs generate a sizable fraction of the growth in drug expenditures.

A shortcoming in the literature that I address here is the use of list prices in most
analyses, an especially large problem in the US market. List prices, which are known as
Average Wholesale Price (AWP) or Wholesale Acquisition Cost (WAC), are akin to sticker
prices for cars. Most payers, including public ones such as Medicaid, end up receiving an
unobservable but significant discount (or rebate) off the list price or even the price paid
at the pharmacy, in effect paying a net price to drug companies plus some payment to
intermediaries. This creates a problem when analyzing pricing dynamics, because discounts
do not necessarily remain constant or move in lockstep with list prices. For example, Gilead,
makers of the new generation of Hepatitis C drugs, set similar list prices in 2013 and 2014,
but reported average discounts off the list price of 46% in 2014, after only offering 22% in
2013, possibly due to competitor entry and PBM exclusion threats. A broader analysis by
Bloomberg and SSR Health suggest that net prices paid to drug companies is increasing,
albeit at a slower rate than list prices. Therefore, using list prices can lead to incorrect
inferences, an issue addressed here that is not accounted for in many existing analyses of
prescription drug markets.

In terms of policies surrounding drug pricing, recent controversies surrounding hy-

---

8See Berndt and Newhouse (2012) for a detailed overview of the institutional details behind these prices
and the pricing system more generally.

9Pharmacy prices are often recorded in prescription claims data.

10Source: conference call transcripts available through the NASDAQ website. Also discussed in


12Exceptions in the health literature include the aforementioned paper by Aitken et al. (2016).
perinflation and price increases more generally have drawn the attention of policymakers, with some proposals to centralize Medicare Part D bargaining and bring more scrutiny to PBM activity. Controversial cases of hyperinflation involving Martin Shkreli’s Turing Pharmaceuticals and Mylan’s EpiPen have drawn Congressional scrutiny, but even areas with several competitors such as insulin markets have exhibited large price increases. As a result, some congressional members have levied collusion accusations against companies selling diabetes medication.\textsuperscript{13} In addition, there have been calls for centralized Medicare Part D price bargaining, which is currently banned under federal law.\textsuperscript{14} Currently, PBMs bargain on behalf of Part D plans, and they have drawn significant congressional and legal scrutiny in the past 15 years, which I detail in the following section.

\subsection*{2.2.2 PBMs and Industry Pricing Structure}

Next, I turn to discussing PBMs, outlining their importance in prescription drug markets and the controversies surrounding them in policy debates.

The Role Played by PBMs in Drug Markets

The PBM industry is highly concentrated, and large PBMs are quite profitable and often owned by other entities. According to the Pharmacy Benefit Management Institute, the top three PBMs in terms of total prescription claims in 2015 were Express Scripts, CVS/Caremark, and OptumRx, totaling 73\% of all prescription claims in the US.\textsuperscript{15} There have been several large mergers in recent years, including mergers between Express Scripts and Medco in 2012 and OptumRx and Catamaran in 2015. They generate significant annual

\footnotetext[13]{See Washington Post article detailing the accusations by Senators Bernie Sanders and Elijah Cummins: \url{http://wapo.st/2fhhFRb}}


net revenue, as Express Scripts made $2.5 billion in 2015.\footnote{Express Scripts garnered 26% market share in 2015, and is the clearest to analyze, as it is not part of a larger company.} Finally, throughout the past twenty years, both drug companies and pharmacy chains have periodically owned large PBMs. This includes the 1993 acquisition of Medco by Merck that lasted until 2003, which is relevant for my analysis of the anti-cholesterol market, as Merck owned one of the best-selling drugs, Zocor.\footnote{See Quartz article “Big pharmacies are dismantling the industry that keeps US drug costs even sort-of under control” for a detailed overview of major PBM industry events. \url{https://qz.com/636823/big-pharmacies-are-dismantling-the-industry-that-keeps-us-drug-costs-even-sort-of-under-control/}} Currently, CVS, a pharmacy chain, owns Caremark, the second largest PBM by market share.

The key role played by PBMs in the pricing system is to construct a prescription drug benefit for their customers, a key part of which is to design a formulary by assigning drugs to copay tiers or excluding them from the plan entirely. PBMs design drug plans for large health insurance companies (including Medicare Part D plans), large employers that self-insure, and other entities such as unions. Their customers usually set copay tier amounts, but the PBM is in charge of assigning drugs to different tiers. For example, a customer might set the copay tiers to $30 and $50, and then the PBM would assign drug A to the cheaper tier and drug B to the more expensive tier. PBMs also have the ability to severely limit access to drugs by setting stringent criteria or to exclude them from the benefit entirely, a practice that has gained traction in recent years.

PBMs profit by leveraging this benefit design power to negotiate discounts off list prices, which helps them attract more fee-paying customers. Using their power to set copay tiers for a large number of patients, PBMs negotiate with drug manufacturers to obtain discounts off the listed (gross) price. They then pass some of the savings onto customers, which makes their service more attractive. PBMs generally collect a fee per patient from their customers, but also pocket some fraction of the discounts. In addition, they profit by running mail order services, making especially high margins on generic drugs.\footnote{See Barron’s 2005 article “Pfizer’s New Headache”, available at: \url{http://www.barrons.com/articles/}}
Based on anecdotal evidence, PBMs are hampered by consumer inertia both in its ability to move drugs to a higher copay tier and to exclude them, forces I allow for in my later model. In a market with several substitutable competitors, PBMs would theoretically be able to extract large savings by making drug companies bid for one favorable position.\textsuperscript{19} However, multiple decisionmakers in the industry suggested that consumer inertia curbs their ability to extract discounts. First, moving a drug to a higher copay tier can backfire in the form of higher spending if there is significant inertia, as the same patients will still pick the drug, but the drug company may offer no discounts. Second, excluding a drug entirely may lead to complaints from PBM customers, as patients on the excluded drug are forced to switch to a different drug. This may reduce the demand for the PBM’s services.

A final aspect of PBM behavior relevant to my analysis is that they are somewhat forward-looking, based on the length of their contracts and their behavior around Zocor generic entry. Based on publicly available information, PBMs sign contracts of varying lengths with customers. In particular, they sign long-term contracts with large insurers, sometimes up to 10 years.\textsuperscript{20} PBM behavior around generic Zocor entry provides more direct evidence of dynamic incentives. As detailed in Aitken \textit{et al.} (2009) and media coverage,\textsuperscript{21} Express Scripts anticipated the launch of generic Zocor by favoring branded Zocor over Lipitor in its formulary placement, even though Zocor was significantly more expensive. Their motivation was to facilitate generic adoption, which would lead to long-run cost savings.

\textsuperscript{19}This actually played out in the aforementioned Hepatitis C market.

\textsuperscript{20}A recent Bloomberg report highlights the 10-year contracts between Express Scripts and Anthem, as well as Catamaran and Cigna. “Express Scripts’ Anthem Loss Goes Deeper Than Numbers” available at: https://www.bloomberg.com/gadfly/articles/2017-04-25/express-scripts-anthem-loss-cuts-deep

\textsuperscript{21}See Chicago Tribune article “Generic Zocor won’t be a market healer” available at http://articles.chicagotribune.com/2005-12-29/business/0512290220_1_generic-zocor-lipitor-generic-version
PBMs and Regulators

Throughout the last two decades, PBMs have drawn significant scrutiny from policymakers for their role in increasing drug prices, in part due to the opacity of their contracts. PBM contracts with both drug companies and their customers are proprietary, making it difficult to discern the net price paid to drug companies and the share of the discounts kept by PBMs.\textsuperscript{22} This has led some policymakers and drug companies to blame PBMs for increasing drug costs, including the recent controversy surrounding the price of EpiPens. For example, they claim that PBMs extract a lower net price from drug companies, but then pocket all of the discounts, which then means that insurance companies pay close to the full gross price, leading to higher premiums for patients.

In addition, regulators have repeatedly investigated PBMs for violating anti-kickback and anti-fraud statutes, with at least one lawsuit asserting foul play specifically in the anti-cholesterol market. Standard volume-based discounts are legal, but any misleading advertising by PBMs to promote drugs on which they receive the largest discounts would violate the law.\textsuperscript{23} In addition, as mentioned earlier, drug companies once owned large PBMs, leading to accusations of favoritism towards their own drugs in formulary design. A prominent legal case involved the Justice Department’s 2003 lawsuit against Medco for favoring drugs owned by Merck, its parent company.\textsuperscript{24} The complaint, later settled, included the specific accusation that Medco favored Merck’s Zocor over Pfizer’s Lipitor, with the former costing more than the latter.

In recent years, PBMs have moved away from integrating with drug companies, but have moved towards integrating with pharmacies and insurance companies. In this paper, I

\textsuperscript{22}Express Scripts claimed in a 2017 conference call that its clients receive 89% of rebates on average. Others argue that these statistics can be misleading, as PBMs often shroud discounts in other fees, which creates an inflated pass-through number.

\textsuperscript{23}See \texttt{http://www.pbmwatch.com/pbm-litigation-overview.html} for a list of major lawsuits involving PBMs and kickbacks.

\textsuperscript{24}See articles “U.S. Is Joining Lawsuit That Says Medco Put Profits Before Patients” in the Wall Street Journal and “Medco to Pay $29.3 Million to Settle Complaints of Drug Switching” in the New York Times for more details. Other large PBMs such as AdvancePCS faced similar accusations at around the same time.
abstract away from the distribution side of the market, but several large pharmacies have significant market power in the US. In the last decade or so, there have been actual and proposed vertical mergers between PBMs (Caremark, Express Scripts, EnvisionRx) and several large insurers and pharmacy chains. These include Caremark and CVS (2006), CVS/Caremark and Aetna (2017, proposed), Rite Aid/EnvisionRx and Albertsons (2017, proposed), and Cigna and Express Scripts (2017, proposed). Although I do not directly address the effects of PBMs on the pharmacy industry, my evidence on PBMs favoring their parent company in insurance design suggests that PBMs may also strongly favor specific pharmacy chains when integrated, lowering competition in the pharmacy market.

2.2.3 Anti-Cholesterol Market - Institutional Details

To conclude, I cover some of the key summary statistics and institutional details relevant for understanding market outcomes in the anti-cholesterol market, the focus of the empirical part of my paper. For more in-depth background on anti-cholesterol drugs, see Section 1.2.3.

In terms of demand, the market size grows at a large but slowing rate in the period, and market share is also significantly affected by entry events. Figure 2.1 summarizes market demand based on MEPS data. I compute market size in number of individuals, counting those on medium intensity dosage levels plus plus those diagnosed with high cholesterol but not taking medication. The market size grows steadily over time, but the growth rate slows down starting around 2003. The share of users on each drug generally remains pretty steady, outside of the initial entry of Lipitor and Crestor, and the impact of Zocor generics on Lipitor market share. Branded market shares drop very quickly after generics enter: Zocor drops from 20% to 4% a year either side of generic entry and Lipitor drops from 12% to a little over 1% from 2011 to 2013.

25 This assumption is based on my analysis of the dosage that new patients start on, which is predominantly in the medium-intensity class from 1998 onwards.

Figure 2.1: Anti-Cholesterol Market Summary - Demand

Notes: Graphs summarizing demand in the medium-intensity treatment level anti-cholesterol market. **Top**: number of patients who are taking cholesterol medication and who are diagnosed with high cholesterol but not taking medication, based on MEPS data. The sum represents the market size, which grows at a slowing rate over time. **Bottom**: given the market size, a plot of the fraction of the market on each branded medication. The red sums together people taking the three main drugs plus people taking generics and other minor drugs. Generic entry significantly reduced market shares of branded Zocor (June 2006 generic entry) and branded Lipitor (December 2011 generic entry).
The net price data shows significantly different gross vs. net price patterns, with significant net price changes around brand and generic entry events. The inflation-adjusted gross and net price data for the three major statins is summarized in Figure 2.2. The list price graph, as is usual in most drug markets, shows a steady growth over time for all three drugs. However, the net price graph suggests that this is misleading, as net prices does exhibit decreases.

![Figure 2.2: Anti-Cholesterol Market Summary - Prices](image)

**Figure 2.2: Anti-Cholesterol Market Summary - Prices**

*Notes:* A comparison of average pharmacy and net prices in each year for the three major drugs over the analysis period, showing very different patterns when accounting for rebates. Prices are weighted across medium-intensity treatment level dosages for each drug, with the pharmacy prices coming from median prices in the MarketScan claims data. Smoothed net prices incorporate estimated discounts described in Appendix B.1 and are smoothed out to account for unevenness in wholesale purchasing across years. All prices are adjusted to Year 2000 US dollars.

The three most striking features of the net price data, which my model will look to explain, are the gap in price between Zocor and Lipitor but small difference between Lipitor and Crestor, the responses in anticipation of generic Zocor entry, and the drop in prices after Zocor generics enter. Similar to the list price data, Zocor appears to have a significantly higher price relative to Lipitor and Crestor, despite being less popular than Lipitor. Merck does begin to offer increasingly large discounts starting in 2002, but Zocor remains significantly more expensive. Unlike the list price data, Crestor and
Lipitor prices drop after Zocor generics enter midway through 2006, which may reflect a response to increased competition. Traditional models had justified list price increases after competitor drugs go generic as signs of price discrimination, where firms target the most price insensitive users. The drop in price post-generic entry follows an increase in prices on all three drugs before generic Zocor enters.

2.2.4 Net Price Data

I make use of SSR Health data, company financial filings and earnings call transcripts, and IMS Health Top-Line data in order to construct estimates of average net price for each drug in a given year. As mentioned earlier, it is important to construct net prices in order to gain insight into actual market outcomes. I obtain net price estimates from SSR Health from 2007 onwards, and then replicate their analysis and extend the series backwards by collecting net revenue data from financial filings, available through SEC EDGAR, and earnings call transcripts, available through NASDAQ and Factiva. I take the revenue data and divide by sales data from MEPS in order to arrive at net price estimates.27

2.3 Pricing Model - Finite-Period Dynamic Game

Here, I begin with an overview of my model and then proceed to discuss the choices made in the model and potential limitations. I then describe the behavior of each part of the pricing system (demand, drug companies, and the PBM industry) in greater detail.

2.3.1 Overview of Model

I summarize my model of market structure in Figure 2.3. In the model, drug companies and a representative PBM face consumers with behavior captured by a switching-cost demand system. Drug companies maximize dynamic profits by making take-it-or-leave-it net price offers to the PBM in each year. The PBM then pick a formulary arrangement, namely a

27See Appendix B.1 for more details on the construction of prices and minor adjustments.
copay tier for each drug, that maximizes its own profits. Patients then make drug choices based on the formulary they face. The system generates equilibrium prices in each period. In equilibrium, no drug company or PBM wants to deviate from their choices.

In terms of characteristics of the dynamic game, I model the game as a finite-period one, which ends for each drug company once its drug goes generic. I assume perfect foresight, which helps ease computation burden, but may miss out on the impacts of unexpected changes in the environment during the period.

Figure 2.3: Model of Market Structure
2.3.2 Assumptions and Limitations of the Model

As I mentioned in the introduction, the goal of the model is to help us better understand the impact of PBMs on drug prices and drug spending. While some of the advantages of the model will become clear as I go through the estimates and counterfactuals, I wanted to begin by discussing the reasons behind the assumptions I make in my model, as well as its limitations and generalizability to other markets.

The Nature of Price Negotiation

The most important assumption in the model is the nature of the price negotiation between drug companies and PBMs. My structure is a simplified version of the bidding structure used by Express Scripts, currently the largest PBM, but potentially simplifies away from more complex structures that may be used by other PBMs.

My model is a simplified version of the negotiation structure used by Express Scripts, which involves annual bidding by drug companies for formulary arrangements. Based on conversations with former executives at Express Scripts, their approach is to set up annual auctions for formulary positions in each drug market.²⁸ In the auction, Express Scripts lays out a set of formulary arrangements, with each arrangement placing each drug in the low copay tier, high copay tier, or excluded tier. Companies submit a price bid for each arrangement, and Express Scripts then picks the arrangement that maximizes its profits. My model simplifies the bid from a high-dimensional vector to a single price. This potentially misses out on more complex equilibrium play by drug companies, but captures the essence of the trade-offs they face. If they bid too high, the PBM will pick an arrangement that excludes them. If they bid too low, they would unnecessarily sacrifice price without obtaining too much more quantity.

The model is unlike other papers studying negotiations in the health care sector, as it does not involve a Nash-in-Nash framework, mainly due to data limitations. A standard

²⁸Markets can be as narrow as a specific genotype of a disease, as was the case with Express Scripts and Hepatitis C. It excluded Sovaldi for genotype 1, but included it for genotypes 2 and 3. http://www.modernhealthcare.com/article/20141222/NEWS/312229943
The negotiation framework in the health care literature is a Nash-in-Nash framework. In this context, price setting would involve several drug companies and the five large PBMs, with final prices splitting surplus relative to threat points. Apart from the institutional details mentioned above, one reason I avoid such a framework is that demand for PBM services and the competition in the PBM industry is not well understood. A second reason is that I do not have PBM-specific negotiated prices for each drug. Without these components, it would not be possible to estimate a Nash-in-Nash bargaining model, so I stick with the simplified Express Scripts structure.

The absence of the Nash-in-Nash framework also precludes the ability to study negotiation over a set of drugs. One main conjecture raised in the industry is that larger companies with a large portfolio of drugs may have a more favorable negotiation position. This is raised as a potential concern in recent anti-trust cases, and there is anecdotal evidence that PBMs negotiate with drug companies across multiple markets. However, in the anti-cholesterol market, there do not appear to be any large changes in prices as a result of the introduction of new drugs in other markets by Merck, Pfizer, or AstraZeneca.

In addition, the model does not allow for long-term contracts, which were used by PBMs during the analysis period. From discussions with decisionmakers at a different PBM, I learned that some PBMs do sign multi-year contracts with drug companies in mature drug markets. Having long-term contracting would allow for richer bargaining dynamics, as parties can agree to a price path rather than just a price for the current period. Depending on their ability to commit, PBMs could control prices without as strong an exclusion threat, as they could use the threat of a persistently high copay tier. For tractability reasons and based on institutional details, I abstract away from this possibility. Multi-year contracting greatly complicates the nature of the game, as drug companies and PBMs would have a much richer action space. In addition, in recent years, the largest PBMs have likely employed annual contracting, as they announce formularies annually, with many changes from year-to-year.

---

29See 2002 Wall Street Journal Article “Pharmacy-Benefit Managers At Times Toil for Drug Firms.”
within each drug market.\footnote{Industry analysts closely follow annual formulary announcements from Express Scripts, CVS, and OptumRx.}

Finally, in my policy application, I assume that government negotiation of prices would follow a similar bidding format to the one in my model, which excludes other potential structures. One potential issue with my framework is that governments may have additional tools when negotiating prices. For example, the National Institute for Health and Care Excellence (NICE) essentially sets a price cap, based on the medical value delivered by a given drug. Another possibility is that governments can make take-it-or-leave it offers to drug companies, giving it a stronger bargaining position. However, in the US system, these options seem less likely, as there is no agreed upon organization to evaluate drugs and less political support for strong government involvement. Therefore, it’s likely that the government will play the role of PBMs, setting formularies based on the price offers it receives. Another likely possibility not evaluated here is an annual price increase cap.\footnote{Evaluating this in my model would require including previous prices as additional state variables, which would enter as constraints on bidding.}

While this may reduce price growth, it may lead to higher introductory prices.

\textbf{Simplifications - Other Firm Controls, Exclusion, List Prices, and Volume Discounts}

My model also simplifies away from other important institutional details, including other firm strategies such as advertising and copay vouchers, weaker forms of formulary exclusion, and the distinction between rebates and discounts.

The demand system in my model rules out interactions between pricing and other firm controls such as advertising and copay vouchers. There is a long literature on the impacts of advertising, both direct-to-consumer and doctor detailing, on drug demand. In addition, there has also been growing attention paid to copay vouchers, which help cover high copays faced by consumers. Both actions represent actions available to firms, and may interact with pricing strategy. For simplicity, I rule out interactions, and have any effects from other actions absorbed into period specific drug fixed effects.
Another potential issue is that my model assumes there are three possible formulary positions for each drug, which abstracts away from more complicated formulary-setting tools. In recent years, insurance plans have typically had two copay tiers for branded drugs with the possibility of exclusion, which forms the basis of formulary choice in my model. However, in addition to excluding a drug outright, a PBM can also put softer restrictions on a drug. For example, many plans restrict drugs by requiring prior authorization, where a doctor has to write a detailed explanation for why a patient needs that drug, or step therapy, where the patient has to start on a different drug before switching to the restricted drug. Adding these options would greatly complicate the formulary options, without adding much from a qualitative perspective. My model allows for a probabilistic distribution across potential formularies, which means a predicted exclusion rate of 20% can partly capture step therapy or prior authorization restrictions.

Next, my model does not use or address list prices, an important quantity in the literature and policy debates. The model captures the data generating process for the net price, but does not base any objective functions on or make predictions about the list price. The list price is important for the uninsured segment of the population, which generally pays out of pocket at list price minus a small discount offered by pharmacies. It is possible that the list price reflects outcomes absent a negotiating PBM, which my results roughly reflect. Another common conjecture in the industry is that PBM profits depend on list price, which then pushes drug companies to set higher list prices to create greater joint surplus. This is likely to be the case if PBM customers sign contracts that pay fees to the PBM based on the difference between net price and list price for each drug, but it is difficult to understand why a large insurer or even private employer cares about anything but expected spending.

Finally, a key distinction made in the industry is between discounts and rebates, a distinction that I also abstract away from. Discounts in the industry can be thought of as the difference between list price and the per unit payment made to the drug company. This data is now available for Medicare Part D plans. In addition, drug companies pay PBMs

---

32 About 25% of the anti-cholesterol market.
rebates based on volume of sales. My net price data would include both of these discounts. I abstract away from this distinction, as it’s likely both drug companies and PBMs can estimate what demand will be given the formulary arrangement, and therefore know what the true net price will be even if bids are in the form of a discount plus rebates. Therefore, volume discounts likely serve more as a way to obfuscate prices than to encourage effort on the part of PBMs. I simplify away from these volume discount issues in my model.

Model Limitations - PBM Competition Issues

The nature of my model and data limitations also preclude analysis of important issues within the PBM industry, most importantly mergers. As discussed in Section 2.2.2, there have been several horizontal and vertical mergers in the PBM industry in recent years, especially after the end of my sample period. My model treats the PBM industry as a black box, albeit a competitive one. Without an objective function grounded in demand for PBM services, the model cannot be used to assess the impacts of PBM mergers. Addressing the nature of PBM demand and profit function would be a fruitful area of future research.

Generalizability - Coinsurance, Demand

Finally, my model can be generalized to evaluate drug markets with slightly different institutional features, including the form of cost sharing and the nature of demand. My model can be easily adapted to studying drug markets where cost sharing takes the form of coinsurance, which is often the case for expensive biologic drugs. Although plans rarely have multiple coinsurance rates, the PBM still possesses the ability to exclude a drug, as Express Scripts famously did in the case of Hepatitis C drug Sovaldi. Therefore, instead of

33 There are minor actions in addition to formulary setting that PBMs can take to affect demand, such as sending letters to doctors encouraging them to prescribe a preferred tier drug.

34 There is also an argument that this is a way to charge higher cost sharing for drugs covered under coinsurance.

35 Anecdotally, insurers and private employers do switch PBM services from time to time. This includes a recent case involving Anthem and Express Scripts. Harvard University also switched from OptumRx to Express Scripts in 2017.
picking copay tiers, the PBMs decision is just a binary exclusion decision.

My model can also be used for studying markets with different demand structures, although it will be difficult to incorporate demand heterogeneity. The simplest demand structure is one with no dynamic linkage, which my model will nest. Other demand structures will also work as long as it exhibits a Markovian structure. For example, Hepatitis C markets have the opposite dynamic linkage to the ones in chronic drug markets. The more users previously treated, the less future demand there will be. The structure of my model will be able to handle this. One weakness of my model is the inability to handle heterogeneity in parameters such as price sensitivity. Once heterogeneity is introduced, the state space would then need to keep track of additional parameters capturing the type of patient currently on a given drug.

2.3.3 Demand System - Switching Costs

My demand model is a simplified version of the model I use in Chapter 1. The key aspects in the model are that consumers face copays rather than list prices and that I identify switching costs using variation in cohort means, motivated by my identification strategy in Chapter 1.

I capture patient behavior in a switching cost model of demand. The key aspects in the model are that consumers face copays rather than list prices and that I identify switching costs using variation in cohort means, motivated by my identification strategy in Chapter 1.

In the demand model, consumers are myopic, and chose either a drug or the outside option\textsuperscript{36} in every quarter based on three factors: the quality of the options, the formulary they face, and the molecule of the drug they chose in the previous period. The first factor is represented by a quality parameter assigned to to each drug and year, in order to reflect changes in medical knowledge and account for advertising campaigns. This can be thought of as the revealed-preference drug quality. Another factor is the plan formulary, which is just a set of copays associated with each option. A final factor is consumer history-dependence, which is represented by an indicator of whether the given option contains the same molecule.

\textsuperscript{36}In this context, one can think about the outside option as diet and exercise.
as the option chosen by the patient in the previous period.

The assumption that only the previous period matters helps simplify the dynamic model, but loses out on some of the richness captured in the reduced form estimates. The quasi-experimental estimates shown in Section 1.3.4 allowed for a non-parametric inertia effect over time, and offered evidence that experience reduces switching rates. For my structural analysis, I need to simplify to a first-order Markov process for demand. Without such an assumption, I would have to keep track of market shares for each cohort, which dramatically increases the state space.

More formally, consumers have utility functions

\[
u_{ijkt} = \delta_{jkt} - ap_{j(f)t} + \gamma I_{j=m_{i,t-1}} + \nu I_{m_{i,t-1} \neq 0} + \zeta_{ij} + \psi_{ij} + \epsilon_{ijkt}(2.1)\]

where \(i\) indexes the patient, \(j\) the molecule (0 is the outside option), \(k\) the form of the option (brand/generic), and \(t\) the year.

The first two terms, capturing product quality and copay sensitivity are standard in demand systems. \(\delta\) is a fixed effect specific to drug, form, and year. This is allowed to fluctuate to pick up variation in advertising, changes in medical evidence, and other factors that affect the outside option such as macroeconomic conditions. \(p\) is the copay faced by the user for the given option, and implicitly depends on the formulary \(f\), which I discuss in Section 2.3.5.

There are two switching-related terms in the equation, one for switching between drugs and one for switching from any drug to the outside option. First, I include a binary indicator, \(I_{j=m_{i,t-1}}\), for whether the choice of molecule in period \(t-1\), \(m_{i,t}\), matches option \(j\) being considered in the current period. The coefficient \(\gamma\) will capture the switching cost by boosting the incumbent molecule, and reflect the patterns found in the quasi-experiment.

Second, I include an indicator for whether the patient chose any drug in the previous period: \(I_{m_{i,t-1} \neq 0}\). The coefficient \(\nu\) will capture this intensive margin effect. The importance of the second term comes from the fact that while some patients do switch from a drug to the outside option, they are far less likely to pick it than a newly diagnosed patient. Including
the intensive margin term helps the demand system better reflect the prevailing substitution patterns if a patient is forced to switch from their current drug, possibly due to formulary exclusions.

In terms of errors, $z_{ij}$ captures any individual-specific match quality, $v_{ijt}$ represents any serial correlation in shocks that evolves according to $v_{ijt} = \rho v_{ij,t-1} + z_{ijt}$, and $e_{ijkt}$ are idiosyncratic errors drawn from a type-I extreme value distribution (logit errors).

I deal with these unobserved heterogeneity issues by instrumenting using a predicted value based on the starting-cohort mean in my estimation. Given the problem outlined above, what I require is a variable that is uncorrelated with both $z_{ij}$ and $v_{ijt}$, but still captures variation in the actual previous choice indicator $I_{j=m_{j,t-1}}$. My proposed solution, motivated by my quasi-experimental approach, is to the leave-one-out mean of the indicator within the same starting cohort, for each molecule $j$ and time period $t$:

$$I_{y,i,j,t} = \frac{1}{|C_y| - 1} \sum_{k \in [C_y \setminus i]} I_{j=m_{k,j,t-1}}$$

$$I_{j=m_{j,t-1}} = \tau_1 + \tau_2 I_{y,i,j,t} + v_{ijt}$$

$$\hat{I}_{j=m_{j,t-1}} = \hat{\tau}_1 + \hat{\tau}_2 I_{y,i,j,t}$$

where patient $i$ is in cohort $C_y$ if they started on treatment in year $y$. $I_{j=m_{j,t-1}}$ is the value I substitute in. Since we use leave-one-out means, then this quantity will be uncorrelated with any individual error terms, as it will only reflect errors for other individuals.

I can then use this instrument in two ways: substituting in a predicted value from the instrument for the endogenous regressor or adding the residual $v_{ijt}$ as an additional regressor. As described in Terza et al. (2008) and outlined in work starting with Newey (1987), two common approaches for instrumenting for endogenous regressors in nonlinear models are two-stage predictor substitution (2SPS) and two-stage residual inclusion (2SRI). The 2SPS provides an analogous approach to the standard two-stage least squares routine.
in linear regression. The equivalent here is to substitute in $I_{j=m_{i,t-1}}$ for $I_{j=m_{i,t-1}}$ in the logit model. The 2SRI approach, advocated for by Terza et al. (2008), involves keeping the basic specification, but adding in the residual to absorb the serially correlated error terms. I will subsequently report results using both approaches.

The requirement for the instrument to work is that there is significant variation in market shares across cohorts, a fact established in Chapter 1. This requirement for cohort-level variation is akin to the standard first stage relevance condition on instrumental variables. Crucially, as I showed in Figures 1.1b and A.7, there are major difference across cohorts in choice over time, stemming from variation in drug availability and announcements of clinical trial results. Other sources of variation include direct-to-consumer advertising campaigns, which Sinkinson and Starc (2018) show have a greater effect on new users by leveraging variation in advertising supply driven by election ads.

I also take a similar approach to instrumenting for the previous period intensive margin indicator $I_{m_{i,t-1}} \neq 0$. The purpose of the term is to capture substitution patterns in the event one of the drugs becomes unavailable due to formulary exclusion. Based on raw patterns, patients already taking drugs are more likely to switch to another drug rather than the outside option. An instrument is necessary here as well, as unobservable match quality will drive patients to continually take some drug, and I use a leave-one-out mean of the rate of any drug usage in a given starting year cohort.

One simplification in my model is the homogeneity across users, which is necessary for tractability in my dynamic game estimation. In standard demand models, parameters such as $a$ are modeled as patient-specific coefficients, in order to reflect heterogeneity in price sensitivity. However, as I describe in Section 2.3.1, the state variable reflects previous period market shares. Adding heterogeneity would mean that one has to keep track of types of individuals on each drug, multiplying the dimension of the state space. The lack of heterogeneity here is mitigated by the fact that switching costs and PBM behavior also play a large role in generating effective price elasticities, whereas in previous models, heterogeneity in $a$ alone generated variation in price elasticities.
Finally, I use raw copay data to identify copay elasticities, and later verify that the results are comparable to those found in quasi-experimental studies. The standard concern in the literature is that using raw prices in estimating demand systems can bias the price sensitivity coefficient, as products with unobservably higher quality are priced higher. This is less of a concern here, as I identify a using variation across plans in copay differences between drugs. Although copays could reflect plan-specific unobservable differences in tastes, this is probably less likely in a prescription drug market setting, as tiers are set by PBMs covering a diverse set of customers. To check that using raw copay data yields reasonable copay elasticities, I will later compare estimates to those recovered by Einav et al. (2016), who use copay variation from the Medicare Part D donut hole to estimate copay elasticities. They generally find inelastic copay sensitivity.\footnote{One concern in light of the results here is that they are looking for within patient and year changes in choice, which may be diminished by the effects of inertia.}

\subsection*{2.3.4 Drug Company Pricing - Take-It-Or-Leave-It Offers and Dynamics}

Next, I formalize a dynamic game between drug companies, who make net price offers every period in order to maximize dynamic profits.

In the model, drug companies submit one take-it-or-leave-it net price offers to PBMs in each period, which simplifies away from more generalized offers while still capturing the essential forces. In selecting a net price offer, drug companies take into account competitor, PBM, and patient behavior. Realized prices then represent an equilibrium outcome of the game. For tractability reasons, I choose a simpler bidding behavior, which should capture the key trade-offs, but may miss out on richer equilibria.\footnote{I discuss this in detail in Section 2.3.2.}

The two core components of the dynamic game are the state variable, previous market share, and the control, which is the net price submitted by drug companies to PBMs. Having previous market share as the state variable allows me to capture the history-dependent nature of demand. Drug companies trade off prices and market share today, as lower prices

\footnotesize{37}
will lead to lower profits today but better market position tomorrow, in the form of

More formally, drug company $j$ chooses net price offer $P_{jt}$ in period $t$ to maximize dynamic profits represented by the following value function:

$$V_{jt}(x_t; N_t) = \max_{P_t} P_t N_t \sum_{k=0}^n x_{kt} Pr(m_t = j|m_{t-1} = k, f(P))$$

$$+ \beta^{dc} V_{j,t+1}(x_{t+1}(P), N_{t+1})$$

$P_t$ is a vector of net price offers and $x_t$ is a vector representing the state variable, as drug companies make simultaneous offers each period and affect each other’s states. I assume that each drug company owns one drug. $N_t$ represents the market size in period $t$, which I assume to be exogenous to the game.

The key functions here are the mapping from price offers to formularies and the mapping from a formulary to choice probabilities. $f(P)$ represents the mapping from the vector of price offers to a formulary, which is based on PBM considerations. I discuss this in detail in the following section. $Pr(m_t = j|m_{t-1} = k, f(P))$ represents the realized market share for company $j$ in this period for the set of patients with previous choice $k$ facing a formulary $f$. I assume the outside option is represented by $k = 0$, and there are $n$ drugs each owned by a different drug company. $\beta^{dc}$ is the drug company discount rate.

The state variable for each drug updates according to the equation:

$$x_{j,t+1} = [1 - \mu_{t+1}(1 - \kappa)] Pr(m_t = j), \forall j > 0 \quad (2.2)$$

so the state variable just reflects previous period market shares diluted by new patients in the subsequent period $\mu_{t+1}$. Since I cannot pin down the role of doctors in carrying over previous market outcomes to new patients, I flexibly allow for a carryover term $\kappa$, where $\kappa = 0$ means there is no carryover, and $\kappa = 1$ means new patients behave as if they are probabilistically assigned a previous choice based on previous period market shares.
The value function contains no expectations, as I assume perfect foresight. This is a reasonable starting point, as Lipitor is almost launched at the start of the period, and Crestor is in clinical trials as of 1998. Furthermore, there are no large changes in medical evidence in the statin class. The main benefit of this assumption is that eases the computation burden, as having shocks to the system would mean computing later periods several times, once for each realization of shocks. This will miss out on effects generated by unpredictable changes in medical evidence, market size growth, and macroeconomic conditions such as the great recession.

**Generic Drugs**

One key element of the anti-cholesterol market during this period is the role of generic entry. I deal with by having generics replace their branded versions in the dynamic game once they enter, and having them be passive but still influential in the game.

One key assumption to make the model work in a non-stationary environment is that drug companies play a finite period game that ends when generics enter. Generics typically enter when a branded drug’s patents expire, although some drugs are also granted FDA marketing exclusivity that could run longer. Here, I assume companies become inactive players once the patent on their drug expires at $T_j$, and that the terminal payoff $V_{j,T_j}$ is a constant that I normalize to zero. I test the robustness of this assumption by also making terminal payoffs a function of market share in the last period before patent expiration.

Another key assumption I make is that existing market shares transfer from branded drugs to generics, which creates anticipatory effects. As documented in Section 1.3.6, over 80% of patients on a branded drug switch over to the generic once it becomes available. I make the assumption that the generic inherits the branded drug’s state in its first period, and replaces the branded drug in the game. This reflects the history-dependence found in Section 1.3.6, and the mechanics will influence play in earlier periods, as competitors of the expiring drug will want to avoid giving the generic a large base. On the other hand, PBMs may want to move patients to the expiring branded drug to save costs in the future,
as anecdotal evidence suggests Express Scripts did in anticipation of the entry of generic Zocor.

Once generics enter, I assume they only play a passive role in the pricing game. Typically, PBMs automatically place generics in a separate generic copay tier, which has a lower associated copay than the preferred tier copay. Therefore, I assume generics do not play an active role in the pricing game, and are assigned a price equal to the pharmacy price in the data, which is significantly lower than branded net prices. However, generics still affect the competitive environment for remaining branded drugs, as generics represent an option that will automatically be cheaper for patients and more attractive to PBMs.

2.3.5 PBM Behavior - Formulary-Setting Behavior

Finally, I model the behavior of PBMs, who maximize profits by leveraging their formulary-setting power to negotiate discounts off list price and then taking a cut of the discounts. A PBM’s ability to extract large discounts is limited by its ability to move volume and the aversion of their customers to exclusions.

Given data limitations, my model simplifies the pricing structure to include a representative PBM. As alluded to earlier, the PBM industry is opaque, as contracts between drug companies and PBMs, as well as PBMs and its customers, are not publicly available. Ideally, my model would contain an oligopoly structure for the PBM industry as well, where they compete for customers and each obtain their own price offers from drug companies. This would allow for a richer set of outcomes, and allow for a full bargaining model akin to Ho et al. (2017). Instead, I include a representative PBM in my model, which one can think of as a black box representing the PBM industry. The PBM takes one price offer from each drug company, and chooses a probabilistic distribution of formularies, which looks to capture the heterogeneity in formularies observed in the MarketScan data. An equivalent interpretation is that I am modeling a continuum of non-interacting PBMs, each with some idiosyncratic profits from selecting a given formulary.

The PBM makes formulary decisions to maximize its profits, which is based on attracting
customers as well as generating other sources of revenue. As mentioned in Section 2.2.2, PBMs profit by charging customers for their services. PBMs are typically paid a fee per patient covered, and therefore would want to maximize demand. In addition, PBMs make some margin on selling generic medication through its mail order programs, and some were owned by drug companies in the 1990s and early 2000s. This gives them additional incentives to select a formulary that guides patients towards generics or towards drugs owned by their parent company. One challenge here is that in a more competitive environment, PBMs would not be able to favor drugs owned by their parent company. However, given the lawsuits surrounding the Merck-Medco relationship, this behavior does occur.

The key trade-off faced by the PBM in maximizing demand is to offer large discounts while still providing a generous drug benefit. In order to obtain large discounts, PBMs need to be aggressive in excluding or moving expensive drugs to higher copay tiers. However, their willingness to move drugs to higher copay tiers hinges on demand being responsive, and their willingness to exclude drugs depends on customer preferences. Their customers are typically large employers and insurance companies, who are acting on behalf of employees or members who may strongly dislike plans that exclude a drug they have been taking.

Given these forces, I now formalize the PBM profit function. Given net prices \( P \) offered by pharmaceutical firms, PBMs maximize profits by choosing a probabilistic distribution over formulary arrangements \( f_t \in \mathcal{F}_t \), where \( \mathcal{F}_t = \{ p_t, \bar{p}_t, \infty \}^n \). \( n_t \) is the number of branded drug competitors active in period \( t \), with generic drugs automatically assigned to a lower copay tier \( p_g \). The value of \( p_t \) and \( \bar{p}_t \) are exogenous to the game.

PBMs choose the optimal formulary distribution in every period based on the following profit function:

\[ \text{PBMs choose the optimal formulary distribution in every period based on the following profit function:} \]

\[ ^{39} \text{This means that } |\mathcal{F}_t| \text{ is smaller after generics enter.} \]
\[ K_t(x_t; N_t) = \max_{f_t} N_t \left[ W(f_t|x_t) - \theta \sum_j P_j s_{jt}(f_t|x_t) - \chi \sum_j I(p_j = \infty)x_{jt} \right. \]

\[ \left. + M(s_t(f_t|x_t)) + \sigma_{\omega, \omega_{f_t}} + \beta^{PBM} K_{t+1}(x_{t+1}(f); N_{t+1}) \right] \]

The first three terms represent demand for PBM services, which factor directly into PBM profits under the assumption that they charge a fixed fee per person they serve. As discussed informally above, PBMs want to pick a formulary that will attract the most customers.\(^{40}\) The first factor is welfare or consumer surplus generated by the formulary, \(W(f|x_t)\), which is derived from the demand system. Higher copays and exclusions will lead to lower welfare for all patients. The second term represents expected spending on anti-cholesterol drugs, with \(s_{jt}(f|x_t)\) representing the realized market share for drug \(j\) in period \(t\) as a function of the formulary chosen and the current state variable. Together, these two terms form a cost-effectiveness measure of the services offered by PBMs, with \(\theta\) serving as the weight. The third term contains an indicator for exclusion, \(I(p_j = \infty)\), and the parameter \(\chi\) looks to capture any additional aversion to formularies with exclusions on the part of PBM customers.

The fourth term, \(M(s_t(f))\), represents a catch-all other motivations that drive PBM profits. As mentioned in Section 2.2.2, several large PBMs were owned by drug companies, including the Merck-Medco relationship relevant for my current study. This may create an incentive for Medco to favor formularies that generate higher Zocor market share. The other highly-discussed aspect of PBMs is their profit margin on running mail-order services for generic drugs. Therefore, they may also favor formularies that lead to more generic market share. I add in flexible terms for these factors in the following way:

\[ M(s_t(f)) = b_1 s_{t1}(f) I(t < 2004) + b_2 \sum_{j \in G_t} s_{jt}(f) \]

\(^{40}\)In reality, PBMs often sign contracts with customers first, but still have reputational concerns when choosing a formulary.
where $b_1$ represents a boost for Zocor (indexed to be $j = 1$) in the years during which it owned Medco, and $b_g$ represents any margins PBMs make on selling generic drugs. $b_g$ may also capture any mis-measurement in the price of generic drugs. $G_t$ is the set of drugs that have gone generic by period $t$.

The fifth term, the error term $\omega_{ft}$, captures idiosyncratic noise in the profit function of PBMs. The error term is not known to drug companies when making offers, and helps ensure the existence of an equilibrium. Without the smoothness introduced by the error, the discrete copay tier structure can generate best response functions may never intersect, as one drug company may want to undercut another by a small amount to gain a large bump in demand. Effectively, the resulting realization will be a probabilistic distribution across potential formularies, although only small values of $\sigma_\omega$ can generate the pricing patterns observed in the data.\footnote{With large $\sigma_\omega$, drug companies can rely on the error term to guarantee some volume, which leads them to offer high prices.}

Finally, the final term captures the dynamic incentives for PBMs, who prefer certain states to others. In the model, PBMs are limited in their ability to move volume by history-dependent demand, as highlighted by the presence of $x_t$ in the first four terms of their objective function. Therefore, some states may be more favorable for them. For example, if many patients are already a given generic drug coming into the period, then PBMs will have an easy time generating high generic demand and pressing for higher discounts from remaining branded competitors. I flexibly allow for PBMs to have some dynamic motives through the parameter $\beta_{PBM}$.

One element missing here is the role of list prices in PBM profits, which helps simplify my model, but does not allow me to assess the common critique that PBMs prefer higher list prices. I assume in my model that PBMs are paid a fee per patient covered, but policy and industry discussion often points to PBMs receiving a cut of rebates. Rebates are the difference between list price and net price, leading to accusations that PBMs prefer very high list prices in order to inflate rebates. The challenge with including list prices is that it...
adds another decision to the drug company problem. Another reason why rebates based on list price makes less sense is that PBM customers should only care about net price or net spending on drug benefits, and therefore competition between PBMs should eliminate the role of list price in payments.

2.4 Results - Estimation Approach and Model Estimates

In this section, I present the estimates from the model and discuss features of the model that help fit the data. I first estimate the demand system separately, recovering switching cost coefficients and revealed-preference drug qualities. Then, I plug the demand system into the dynamic game, and recover parameters of the game by matching model predictions to my net price estimates.

2.4.1 Demand System - Switching Costs and Copay Sensitivity

I begin by estimating the switching cost demand system separately from the rest of the dynamic game. The estimates capture the history-dependence documented in Chapter 1, and highlight the importance of instrumenting for previous choice. I also find some sensitivity to copays and some variation over time in implied drug quality.

My approach proceeds in two steps: first I estimate switching cost and copay sensitivity parameters using the MarketScan dataset and then calibrate drug quality parameters to match the representative market shares from MEPS. As detailed earlier, one advantage of the MarketScan data is the ability to infer the formulary facing each patient. This then allows me to estimate copay sensitivity and switching cost parameters. However, the MarketScan data is not a representative sample of the US market, so I separately calibrate

---

42 One could also assume that list prices are exogenous, or based on demand from uninsured patients for drugs.

43 Consultants working for PBM customers argue that some employers do make the mistake of making payments to PBMs based on savings off list price.

44 These data sources are detailed in Chapter 1.
the $\delta_{jkt}$ parameters to match the market shares from the representative MEPS dataset, fixing
the other parameters to be equal to the estimates from the first step.

Estimates of switching costs are very large relative to copay differences, are sensitive to
the inclusion of instruments and the previous intensive indicator, and are in line with the
quasi-experimental estimates. Table 2.1 displays the parameter estimates from MarketScan.
The switching cost in dollar terms will be $\frac{\gamma}{\delta}$, which captures the equivalent increase in copay
that will negate the inertia. Column (2) of Table 2.1 display the estimates for $\gamma$ without any
instruments and without the previous intensive indicator $I_{m_{i,t-1}} = 0$, which equates to about
$4.60 on a daily basis. As a comparison, gaps between copay tiers are typically less than $1.
Column 2SPS-2 uses the 2SPS approach described earlier to instrument for the two previous
choice indicators, including the intensive margin term. This reduces the between-drug
switching cost to about $2.2.45 The instrument helps absorb the additional individual-level
unobservables, and the inclusion of the previous intensive indicator helps pick up on the
general lack of switching to the outside option, rather than just loyalty to one particular
drug. Finally, the final estimate suggests that, all else equal, the annual switching rate is
around 80%, which is generally in line with the quasi-experimental coefficients.46

45 The table also contains estimates from a 2SRI approach, which yields similar results, but contains some
differences in the impact of the intensive margin indicator.

46 This is just based on the CDF of the extreme value distribution, evaluated at $\gamma$. 
Table 2.1: Demand System Estimates

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>2SPS-1</th>
<th>2SRI-1</th>
<th>2SPS-2</th>
<th>2SRI-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copay ($\alpha$)</td>
<td>0.662***</td>
<td>0.719***</td>
<td>0.633***</td>
<td>0.724***</td>
<td>0.668***</td>
<td>0.739***</td>
</tr>
<tr>
<td></td>
<td>(0.00225)</td>
<td>(0.00253)</td>
<td>(0.101)</td>
<td>(0.141)</td>
<td>(0.111)</td>
<td>(0.149)</td>
</tr>
<tr>
<td>Same Drug ($\gamma$)</td>
<td>3.341***</td>
<td>2.674***</td>
<td>2.502**</td>
<td>1.460***</td>
<td>2.070***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00235)</td>
<td>(0.567)</td>
<td>(0.778)</td>
<td>(0.204)</td>
<td>(0.277)</td>
<td></td>
</tr>
<tr>
<td>Prev. Drug ($v$)</td>
<td>1.385**</td>
<td>0.330</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.494)</td>
<td>(0.494)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Drug FE</th>
<th>x</th>
<th>x</th>
<th>x</th>
<th>x</th>
<th>x</th>
<th>x</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>49,413,573</td>
<td>49,413,573</td>
<td>49,413,446</td>
<td>49,413,446</td>
<td>49,413,446</td>
<td>49,413,446</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Estimates of Equation (2.1). The results show an inelastic response to cost sharing and significant inertia. Incorporating the instrument and the term for previous intensive margin choice diminishes some of the raw estimates. The instrumental variables specifications (2SPS and 2SRI) have standard errors clustered at the cohort-level to reflect the source of the exogenous variation used in the estimates.

The switching cost estimates are also large relative to those found in the literature, levels that would suggest large impacts on price in regular markets through invest-then-harvest type strategies. Dubé et al. (2009) find switching costs on the order of 15-19% of the purchase price, in the orange juice and margarine markets. In their simulations (under a stylized oligopoly model), they find that when switching costs reach 3-4 times the list price, average prices over a product’s lifecycle will be double relative to the case with no inertia, and will exceed the no inertia price even in the case of forward-looking consumers. This is due to a standard invest-then-harvest strategy, where firms build market share by setting low prices initially, but then raise prices to cash in on captured consumers. The nature of the market structure in my model limits this type of behavior, as I discuss in the results and counterfactuals.

The implied copay elasticities are quite small, consistent with evidence in the literature, suggesting that PBMs play a role in generating higher price elasticities. In a recent paper,
Einav et al. (2016) estimate copay elasticities for various drugs, using the Medicare Part D donut hole to generate variation in copays within a patient. They find copay elasticities for Lipitor of 0.33, significantly less than 1. My estimate of $\alpha$, shown in Table 2.1, are all between 0.6 and 0.7. Under a copay of $1 per pill, this translates to a copay elasticity of around 0.25 for Lipitor in the years before 2007. As patients do not face prices directly, this suggests that PBMs play a role in creating a high effective price elasticity through its formulary setting behavior.

Finally, the estimation also provides drug quality estimates for each drug and year, which suggest differences in revealed preference drug quality across drugs and over time. Figure B.1 in the Appendix provides the estimates of $\delta_{jkt}$ from Equation (2.1). In terms of relative quality, Lipitor has a significantly higher quality at the start of the period relative to Zocor, which lasts until generic Zocor enters. Crestor has lower implied quality than the other two. In terms of trends over time, Zocor stays relatively flat, while Lipitor and Crestor drop in quality after the generic entry of Zocor in 2006. This is due to the fact that generic Zocor steals market share rather than expanding the fraction of patients choosing a drug. Other fluctuations in implied quality may be driven by advertising campaigns, macroeconomic conditions\(^47\), and changes in medical evidence.

\subsection{Model Computation and Estimation}

Before turning to the estimation results, I describe my approach to computation and estimation for the dynamic game laid out in Section 2.3. The inner loop of the computation plays out the pricing game under a given parameter guess, and the outer loop finds parameters that best fit the net price data, under a maximum likelihood framework.

For the inner loop, I compute the final-period value functions on a finite grid of points in the state space and then iterate back to the first period. As is standard, I construct a set of points on which to evaluate the value function. In practice, Crestor is the only active

\(^{47}\) As shown in Figure 2.1, the fraction of patients choosing the outside option (one minus the “All Drugs” number) increases during the Great Recession.
participant in the final period (2013), and therefore the only computation is a single-agent profit maximization one. Then, I take the value function from the last period in order to solve for a pricing equilibrium in the penultimate period. I proceed backwards until I reach the first period (1996).

One key aspect of the model is that I allow for a continuous control, by interpolating value functions. In a given period, drug companies play a simultaneous pricing game, arriving at an equilibrium where no one wants to deviate in pricing. In practice, I start with a guess for the equilibrium price and then update the price one drug company at a time by finding their optimal price given the other prices. The continuous control means that I need to be able to compute $V_{t+1}$ under any value of $x_{t+1}$, which I do by constructing a smooth scattered interpolant using the output from computing the value function on a discrete grid. I can then efficiently evaluate this interpolant to evaluate the value function every time I need to compute profits.

Given the inner loop structure, I then estimate the model by minimizing squared prediction error in an outer loop. For the outer loop, I search over parameters to minimize the squared error between predicted and observed prices

$$\min_{\theta, \chi, \beta^{PBM}, b_1, b_2, \kappa} \sum_{t} \left( \sum_{j \in G_t} (\hat{P}_{jt}(\theta, \chi, \beta^{PBM}, b_1, b_2, \kappa) - P_{jt})^2 \right)$$

where $\hat{P}_{jt}$ is the model’s predicted equilibrium price. This is equivalent to an assumption of uncorrelated measurement errors in the prices. The basic structure of the likelihood function does not allow for serial correlation or period-specific errors, which I look to relax in robustness checks.

I use MEPS data to estimate the market size $N_t$ and the fraction of new patients $\mu_t$ in a given year. The market size is based on the number of people taking anti-cholesterol medication or diagnosed with high cholesterol. The $\mu_t$ is based on assuming a 5% patient

---

48I find the initial guess through a grid search, to avoid starting in a region where the company has no variation in incentives. This is present in my model because there are price ranges under which exclusion is essentially certain.
attrition from year-to-year, possibly from improving medical condition, and assuming that
the difference \( N_t - 0.95N_{t-1} \) represents the count of new patients, so \( \mu_t = \frac{N_t - 0.95N_{t-1}}{N_t} \). \(^{49}\)

Next, I use the available MarketScan formulary data to construct \( \tilde{p}_t, \hat{p}_t \) in every period. I
do this by taking the copay tier dollar values for each plan, and then taking the median
value for the preferred and non-preferred tiers, weighting by the number of users on each
plan. The use of MarketScan is necessary, as MEPS does not provide plan identifiers, and
therefore I am unable to construct copay tiers. Finally, I also use median generic prices and
copays from MarketScan to set \( P_g = 1 \) and \( p_g = 0.33 \).

2.4.3 Parameter Estimates and Model Fit

In this part, I provide an informal discussion of identification, discuss the results of the
estimation, and highlight important components of the model for fitting the data, including
the PBM exclusion threat and the discrete copay structure.

Each parameter is roughly derived from a key moment in the net price data. Although
the model is non-linear, the PBM equation is simple enough to generate a rough mapping
from data to parameters. \( \theta \), the weight on expected cost, is identified by the average price
levels in the net price data. The less the PBM’s customer cares about costs relative to welfare,
the more likely drug companies will set higher prices, as they know PBMs will still put
them in a favorable position. \( \chi \) captures the exclusion penalty, which reflects whether
higher previous market share leads to higher prices in the current period, as the threshold
for exclusion is pushed up. \( \beta^{PBM} \) is identified based on actions around generic entry, as
discussed earlier in the case of their generic Zocor strategy. \( b_1 \) is identified off any abnormal
gap between Lipitor and Zocor prices in the pre-2003 period. Similarly, \( b_g \) is identified off
lower brand drug price levels after generic entry, as they would face PBMs looking to switch
people to generics. Finally, \( \kappa \) is derived from differences in pricing versus future new patient
fraction \( \mu_{t+1} \). If there is no relationship, it would suggest significant carryover of previous
market shares through any doctor channels.

\(^{49}\)Attrition is required to avoid negative numbers of new patients in a given year.
The best fit parameter values, shown in Table 2.2, provide some insights into the PBMs profit function, suggesting that they generally look to design cost-effective formularies to gain customers and earn fees, but also factor in other sources of revenue. Most of the estimated parameters aim to capture PBM behavior, as we know the least about their business model and profit function. The estimation yields a estimate of $\theta = 0.973$, which is the relative weight of cost against welfare. It is statistically indistinguishable from 1, suggesting something close to a cost-effectiveness criteria. Another key parameter $\chi$ is large and significant, suggesting that PBMs have trouble excluding popular drugs. The other parameter with statistical significance is $b_1$, which suggests that Medco may have had incentives to increase the market share of Zocor based on ownership.

<table>
<thead>
<tr>
<th>Parameter Estimate</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta$</td>
<td>0.973</td>
</tr>
<tr>
<td></td>
<td>(0.083)</td>
</tr>
<tr>
<td>$\chi$</td>
<td>0.826</td>
</tr>
<tr>
<td></td>
<td>(0.110)</td>
</tr>
<tr>
<td>$\beta_{PBM}$</td>
<td>0.100</td>
</tr>
<tr>
<td></td>
<td>(0.109)</td>
</tr>
<tr>
<td>Merck Boost ($b_1$)</td>
<td>1.155</td>
</tr>
<tr>
<td></td>
<td>(0.127)</td>
</tr>
<tr>
<td>Generic Boost ($b_g$)</td>
<td>0.449</td>
</tr>
<tr>
<td></td>
<td>(0.360)</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>0.086</td>
</tr>
<tr>
<td></td>
<td>(0.130)</td>
</tr>
</tbody>
</table>

Notes: Standard errors are computed using the standard maximum likelihood approach.

The statistically insignificant parameter estimates suggest weaker effects from PBM dynamic incentives and not much carryover of history-dependence through doctors. The estimates show a small and insignificant value for $\beta_{PBM}$. The lack of a dynamic effect may come from the fact that PBMs may only care about this around generic entry, and do not factor in future savings on branded drugs. It is also possible that many of their contracts are short-term, and little incentive to manipulate market shares for their current customers.
The carryover parameter, $\kappa$, is also small and insignificant, suggesting that drug companies dampen dynamic strategies if they know there is about to be an influx of new patients. This is consistent with the limited evidence on doctors presented in Section 1.3.7.

In terms of fit, the model generally captures the patterns in the net price data, including the relative prices between the drugs and the trends in prices around generic entry. Figure 2.4 presents a comparison of the price dynamics predicted by the model at the best-fit parameters versus the net price data presented earlier. The predictions accurately capture the absolute and relative price levels of the three main drugs, in particular the higher price of Zocor relative to Lipitor in earlier years. In addition, it captures the trends in drug prices before and after generic entry. Prices tend to increase for all drugs before generic entry, except for Lipitor in 2006, reflecting a desire to “harvest” existing market share before competition intensifies.

![Figure 2.4: Model Fit - Predicted vs. Actual Prices](image)

**Notes:** A comparison of model predictions and data on net prices. The model captures the general dynamics in the net price data, including the key features laid out in Section 2.2.3. The model rationalizes the gap between Zocor and Lipitor prices in early years, the decrease in net prices after generic Zocor enters the market, and the general increase in price in anticipation of generic Zocor and Lipitor entry. Aspects not matched are due to inflexible features of the model discussed in greater detail in Section 2.4.3. This includes the size of Zocor’s drop in net prices, Lipitor’s pricing in 2006 relative to 2005, and pricing in 1996 and 1997.

The discrete copay structure in the model helps to explain the higher price of Zocor
in the earlier period. As noted before, a key aspect of the net price data is that Zocor maintains a significantly higher price relative to Lipitor in the early part of the sample period. This is unusual, as Lipitor is the more popular drug and quickly leads in market share. Although having $b_1$ in the model helps to match the quantitative difference, I show in the counterfactuals that Zocor is still priced above Lipitor in equilibrium even with $b_1 = 0$. This is because Lipitor has a stronger dynamic incentive to gain market share, but can only do so by undercutting Zocor significantly to push it out of the preferred tier, as PBMs have difficulty satisfying existing Zocor patients. I provide an illustrative model highlighting this effect in Appendix B.2.

Another key aspect of the model is the PBMs ability to exclude drugs, which limits the range over which drug companies can price, in turn curbing more extreme dynamic strategies. As shown in the net price data, there are no significant price increases, contrary to the existing literature on the market impacts of consumer switching costs. This is mostly due to the threat of exclusion in the model, which limits the range over which drug companies can price, even if they have significant market share. This then has an effect on previous periods, as drug companies have less leeway to execute an “invest-then-harvest” type strategy.

As an additional check on model fit, I show that the model yields reasonable predictions for formulary outcomes and exclusion probabilities. In addition to the pricing predictions, the model also outputs the equilibrium formulary choices made by the PBM. I do not use data to discipline these predictions, as representative data on formularies is difficult to obtain, as MEPS does not include insurance plan indicators, unlike the MarketScan data. However, the results presented in Figure B.2 in the Appendix are reflective of formulary trends in MarketScan and industry reports. First, there is essentially no exclusion in the early years, consistent with MarketScan data, while Crestor faces some exclusion when it

---

50MEPS does have average realized copays from patient choice, which my model does make predictions on, but the reliability of the data may present problems.
First enters. Second, Zocor has a much higher probability of ending up in the high copay tier in the 1996-2006 period. A 2001 report by the California HealthCare Foundation showed that all the major PBMs put Lipitor in the preferred tier, whereas Zocor was only in the preferred tier of the formularies of Medco and AdvancePCS. Finally, after generic Zocor enters, the model predicts that PBMs will move Lipitor and Crestor to higher copay tiers and exclude them. In reality, PBMs do appear to move remaining branded drugs to higher copay tiers and exclude and apply step therapy and prior authorization to them. Although the model does not make these actions available to PBMs, they essentially make it very cumbersome for patients and doctors to obtain insurance coverage for the drugs, akin to a partial exclusion.

There are key limitations to the model in terms of matching pricing dynamics, which result from issues in market size computation and model specification. In chronological order, the first issue with the model is the price predictions in the 1996-1998 period. The model outputs lower quality estimates, as shown in Figure B.1 of the Appendix, as the outside option is more popular in early years. This problem may be due to an imprecise calculation of market size, as it’s possible that many diagnosed patients did not even consider statin treatment. A second issue surrounds the pricing level of Zocor. The data suggests a more gradual decline in Zocor net price relative to the model. The source of the discrepancy is that I assume that the Merck-Medco relationship ends entirely after 2003, when in fact it maintained some volume guarantee contracts as part of the spinoff. This is

---


54 Based on step therapy and prior authorization data from Medicare Part D formulary data.

55 Medical guidelines generally become more encouraging of statin use over this period.

56 See NY Times article “With Ties Lingering, Medco Leaves Merck” available at http:
solvable by adding another flexible parameter to capture the post-2003 relationship, but for my core analysis, I wanted to limit the parameter space. The third issue surrounds pricing in 2006, the year during which generic Zocor enters. In the data, all three drugs exhibit increased net prices, but my model predicts that Lipitor will set a much lower price, and raise prices in the following year. This is the result of the weaker dynamic incentives of PBMs estimated in the model, as Lipitor has more of an incentive to acquire market share than the PBM has of moving market share to Zocor. As I mentioned earlier, in reality, PBMs may only pursue dynamic strategies at specific times.

2.5 Counterfactuals - Impact of PBMs and Policy Implications

In this final section, I run counterfactuals to understand the roles played by PBMs and demand in generating equilibrium prices, with the goal of understanding the pros and cons of potential centralized Medicare Part D price negotiation. I also run counterfactuals turning off history-dependence in demand to better understand price patterns in other drug markets.

2.5.1 Do PBMs Reduce Drug Expenditures?

First, I look to analyze the impacts of PBMs on prices and total expenditures in the anti-cholesterol market. My main counterfactual will be to look at the overall impact of PBMs on drug expenditures. I also run additional counterfactuals to highlight the importance of exclusion relative to copay tiers in curbing expenditures.

For my primary counterfactual here, I remove the PBM from the model and have patients face a fixed fraction of prices set by drug companies. This structure is akin to a coinsurance structure, where insurance companies still provide some insurance, but PBMs are not involved to provide an exclusion threat.\(^{57}\) For the exercise, I fix a coinsurance rate \( r = 0.33, \)

\(^{57}\)The motivation for moving from a copay structure to a coinsurance one is that without the threat of

//www.nytimes.com/2003/08/20/business/with-ties-lingering-medco-leaves-merck.html. The article details the penalties Medco would face if Zocor market share dropped below a national target.
which is larger than the observed net price to copay ratio in the data and even typical coinsurance rates. A smaller ratio would lead to higher equilibrium prices, and lead to a larger estimate of PBM impact.

The results from the exercise show much larger swings in prices and a generally higher price level. Figure 2.5 shows the counterfactual price dynamics under the modified model. In general, price levels are significantly higher, as there is no longer an exclusion threat from PBMs. In addition, the prices exhibit larger swings, reflective of the larger pricing range available to drug companies. For example, Lipitor sets an extremely low entry price to capture market share initially, and then raises prices from under $1 to $4 within two years. In a copay tier structure, this strategy would be less appealing, as low prices may not yield much additional demand as the drug remains in the low copay tier, and high prices would lead to exclusion rather than profitable “harvesting” of market share.\textsuperscript{58} The model also predicts bumps in prices before the Lipitor, Crestor, generic Zocor, and generic Lipitor entry events and subsequent declines after. This also captures some degree of “harvesting” by all companies.

\textsuperscript{58}From the standard invest-then-harvest strategy discussed earlier.
Notes: Counterfactual prices under a scenario where patients face a fixed fraction of the prices set by drug companies.

More specifically, the change in structure alters the relative prices of the drugs, particularly Lipitor and Zocor. As mentioned earlier, the sustained higher price of Zocor relative to Lipitor is one of the most interesting features of equilibrium prices in the market, and, as discussed in Section 2.4.3, the structure in my model helps to explain this qualitative result. Without the structure, Lipitor is priced more in accordance with its dominant market position and edge in static demand.

The results of the counterfactual also provide an estimate on whether PBMs increase or decrease overall drug expenditures. Table 2.3 shows that net revenue earned by drug companies over the 18 year period would be $195 billion under the structure without PBMs, as opposed to $151 billion in the data. This number does not include payments made to PBMs, which is the other relevant component in assessing PBM impact on drug expenditures. As noted earlier, Express Scripts made $2.5 billion in 2015, with a 25% market share. This equates to $18 billion in PBM revenue over 18 years, projecting Express Scripts out to the whole market and assuming that 10% of the profits are from the anti-cholesterol market.\footnote{Other ways to estimate this, including using a per prescription profit from PBM financial filings also leads to a similar upper-bound estimate.} This represents something of an upper-bound, as PBMs have become more...
profitable over time, and the anti-cholesterol market has stayed at or below 10% of all drug expenditures for the period. Even at the upper-bound, PBMs still reduce total expenditures on anti-cholesterol drugs by about 15%.

Table 2.3: Summary of Counterfactual Market Outcomes

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Average ($)</th>
<th>Net Revenue ($ bill)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Zocor</td>
<td>Lipitor</td>
</tr>
<tr>
<td>Baseline</td>
<td>2.83</td>
<td>2.03</td>
</tr>
<tr>
<td>Varying Model Parameters:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Merck-Medco</td>
<td>2.27</td>
<td>1.98</td>
</tr>
<tr>
<td>No inertia</td>
<td>2.17</td>
<td>1.47</td>
</tr>
<tr>
<td>No exclusion penalty</td>
<td>2.13</td>
<td>1.41</td>
</tr>
<tr>
<td>Pricing Structure Changes:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30% Coinsurance</td>
<td>4.37</td>
<td>4.53</td>
</tr>
<tr>
<td>Single Tier</td>
<td>2.96</td>
<td>1.96</td>
</tr>
<tr>
<td>Commit one preferred</td>
<td>2.71</td>
<td>1.93</td>
</tr>
<tr>
<td>Commit one exclusive</td>
<td>2.03</td>
<td>1.00</td>
</tr>
<tr>
<td>No insurance</td>
<td>1.40</td>
<td>1.49</td>
</tr>
</tbody>
</table>

Notes: Statistics summarizing the results from the counterfactual exercises. i) Baseline: Best fit model. ii) No Merck-Medco: no relationship between the two, sets $b_1 = 0$ iii) No inertia: no drug-specific inertia, with effects shifted to preferences for any drug ($\gamma' = 0, \nu' = \nu + \gamma$). iv) No exclusion penalty: PBM does not have extra aversion to exclusions $\chi = 0$. v) 30% Coinsurance: patients face coinsurance rate of 30% with no PBM involvement. vi) Single Tier: PBMs can only choose inclusion or exclusion for branded drugs, with included drugs having copay equal to the preferred copay level. vii) Commit one preferred: PBM commits to only picking formularies that have one branded drug in the preferred copay tier. viii) Commit one exclusive: PBM commits to formularies that include at most one branded drug. ix) No insurance: patients face full price, assuming same $a$ coefficient. Average prices are computed in a weighted manner across the 18 years from 1996-2013, are scaled to daily prices, and only include branded prescriptions. Total payments is computed by multiplying computed equilibrium net prices by market share and by market size, and summing over all years. Net revenue captures revenue to drug companies over the 18 year period, and excludes PBMs profits.
Two caveats are important to note here: I am not making any welfare claims about the impact of PBMs and I am also not ruling out that a more complex direct pricing model cannot generate the patterns in the net price data. In terms of welfare, a key aspect of drug pricing debates is the role of profits in encouraging investments in new therapies. As such, paying $44 billion more to drug companies instead of $18 billion to PBMs, as is the differences between the structures analyzed here, may lead to longer-term benefits. Second, the fact that my model does not allow for heterogeneity in demand opens the door for a richer model without PBMs to explain prices. For example, it could be that Zocor attracts price-insensitive consumers, and therefore exists as a high-end product relative to mass-market Lipitor. This is an implausible explanation, as characteristics such as income, age, and gender are very similar across the group of patients on each drug.\footnote{Based on MEPS data.}

Finally, I also assess additional counterfactuals related to direct pricing and formulary structure. These results are summarized in Table 2.3. First, I assess demand under a case where patients are uninsured and faced prices directly, under the assumption that their demand parameters remain the same. This unsurprisingly leads to much lower demand and correspondingly lower prices on the three drugs, as only patients with the greatest need will use the drugs. Second, I play with variants of the copay structure and formulary choice set. Having a multi-tiered formulary produces $8 billion savings, but the effects are small, given the insensitivity of patients to copays. PBMs committing to having only one preferred drug also reduces expected spending, but the savings are much smaller relative to a case where they commit to only including one branded drug.

\subsection*{2.5.2 The Effects of Vertical Integration}

Here, I briefly discuss the effects of vertical integration between PBMs and drug companies. As discussed in Section 2.2.2, Merck purchased Medco in 1993, and started winding down the relationship in 2003, mostly due to regulatory pressure and . Neither Pfizer nor AstraZenea were vertically integrated during the period.
To assess the impact of vertical integration, I run counterfactuals shutting off the Merck-Medco relationship. In Figure 2.6, I show the counterfactual prices in a setting with no relationship between Merck and Medco ($b_1 = 0$). The qualitative result that Zocor is still priced higher relative to Lipitor is maintained in the counterfactual, highlighting the impact of first-mover advantage. However, prices are significantly lower, and Table 2.3 shows that drug company profits would have been $10 billion lower absent the relationship, a far cry from the $30 million dollar settlement paid by Medco in 2003 (see footnote 24), and much closer to the $6 billion Merck acquired Medco.

Two main caveats apply here. First, my counterfactuals depend on my estimate of $b_1$, which is only identified from observational data. It is possible that other changes in the early 2000s caused the large drop in prices, but the circumstantial evidence suggests that this is unlikely. Second, I do not measure potential benefits of vertical integration, several of which were raised in 1993 and similar ones raised in recent vertical mergers involving PBMs. These include more efficient provision of drugs and reduction in transaction costs.

![Figure 2.6: No Merck-Medco relationship ($b_1 = 0$).](image)

### 2.5.3 Implications for Medicare Part D Price Negotiation

The counterfactuals surrounding PBMs inform us as to the potential pros and cons of centralized price negotiation in terms of total drug expenditures. On the one hand, the
government can replace PBMs and do the negotiation itself, saving on the significant fees currently paid to PBMs. On the other hand, the government would need to have a strong exclusion threat for popular drugs, currently captured in the model by the $\chi$ parameter, in order to counteract the inflationary pressure generated by history-dependent demand, in order to achieve prices close to the current net prices.

Exclusion is a much more drastic step than changing cost sharing, but is necessary as patients are generally quite insensitive to copays. Political pressure, especially in the US, may limit the governments ability to exclude. In fact, current Medicare Part D rules go in the opposite direction, and require the inclusion of a certain number of drugs.\footnote{Part D plans currently have to cover at least two drugs from each category, and almost all drugs in six categories, including HIV, cancer, and antipsychotics. See detailed list of rules on the CMS website: https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovContra/downloads/Chapter6.pdf}

The key assumption I am making here is that centralized negotiation would have similar bargaining structure to the one in my model. As discussed earlier, the period-by-period take-it-or-leave-it price offers from drug companies is based on the structure employed by some PBMs. In other countries, governments use other measures such as reference pricing, price increase caps, and cost-effectiveness based criteria to increase their bargaining power. For example, the government could be the ones to make a take-it-or-leave it offer to drug companies based on prices in other countries, which would give it more bargaining power. However, these options are less likely to have an impact in the US. The US typically charges more than other countries and places a much higher value on the value of a quality-adjusted life year, making reference-price and cost-effectiveness based price offers less effective. Price increase caps may be more feasible, and could reduce the need to exclude popular drugs, but it may also lead to higher initial prices.

### 2.5.4 Quantifying the Effects of History-Dependent Demand on Pricing

Finally, my framework can also help us understand pricing dynamics in other markets and the effects of generic substitution laws. Specifically, I compute counterfactuals under...
alternative values of switching cost parameters, and also assess variation in other behaviors such as generic adoption.

To begin, I evaluate the equilibrium price dynamics under a scenario with no drug-specific switching costs, and find price deflation. To implement this, I run the game under parameters \((\gamma', v')\), setting \(\gamma' = 0\) and transferring its effect over to \(v' = v + \gamma\). This helps preserve the intensive margin versus extensive margin, but removes a drug-specific demand advantage. I also recompute implied drug qualities in each period as an additional way to maintain market shares. Figure 2.7 presents a comparison of prices under the counterfactual versus those from the baseline model. Prices are generally much lower in level, but also exhibit deflation in the period before 2006. This is due to the increasing levels of competition, as each drug can no longer rely on an existing patient base. The prices in later years are only slightly reduced, as lower inertia also hurts the competitiveness of generic Zocor. All of these effects would be accentuated if I also simultaneously set \(\chi = 0\), equivalent to having the PBM customers ignore drug-specific exclusions. In the results shown, I still keep the exclusion effect, which still provides a boost to equilibrium prices. Overall, the results suggest that inertia plays a significant role in preventing deflation in drug prices when competitors or generics enter.

![Figure 2.7: No Drug-Specific Inertia](image)
Another aspect of history-dependent demand is generic adoption, but counterfactuals show this to have limited impact on pricing strategies. To do this, I modify Equation (2.2), which governs state updating, to allow for some fraction $h$ of projected share to be retained by the generic, with $h = 1$ nesting the current case and $h = 0$ nesting the case where generics enter without any incumbent advantages:

$$\tilde{x}_{j,t+1} = h[1 - \mu_{t+1}(1 - \kappa)] Pr(m_t = j), j \notin G_t, j \in G_{t+1}$$

where $G_t$ once again represents the set of drugs that have gone generic by period $t$. Therefore, this update modification only applies across periods when there is a transition from brand to generic. Figure B.3 in the Appendix presents the predicted prices in the case $h = 0$, when patients do not exhibit loyalty to molecule. The predicted prices are almost identical, which suggests that generics are inherently competitive to begin with, so making them more competitive does not affect branded drugs by much, as the companies owning them focus on preserving their current patient base. The caveat here is that, for internal consistency, one would need to also consider the role of the expiring branded drug if it does not automatically lose most of its patients.

### 2.6 Conclusion

My major contribution in this paper is to construct and estimate one of the first models of drug pricing that includes pharmacy benefit managers. The model overcomes some of the problems encountered by existing ones in the literature, while also allowing us to study the impact of the PBM industry on drug market outcomes.

The estimated model shows that PBMs reduced expected spending in the anti-cholesterol market between 1996 and 2013, but capture a significant profit while also not competing away the effects of vertical integration. Through their threats to exclude drugs, PBMs reduce drug company profits by 25% relative to a structure in which consumers are charged a fraction of the prices set by drug companies. Based on the profits stated in their financial
filings, PBMs appear to capture up to 40% of lost profits, passing along the rest to their customers. In addition, there is circumstantial evidence that PBMs are affected by vertical integration, as prices and formulary positions change significantly after Merck, seller of Zocor, breaks its relationship with the largest PBM at the time.

In terms of pricing policy, the results suggest that there is a tradeoff faced by the government if it were to take on the role of PBMs and negotiate prices. On the one hand, the government would save on the drug expenditures currently captured by PBMs by replacing them in the system. On the other hand, the government would need to be able to exclude popular drugs in order to obtain the current negotiated prices with drug companies, something that runs counter to current inclusion rules and possible political constraints.

A minor but important contribution here is the construction and usage of net price estimates for understanding the forces in prescription markets, which can be extended for a broader analysis of the industry. In the paper, I construct net prices using net revenue data reported in financial filings and earnings call transcripts. This data has become more readily available over time, and drug companies generally report net revenues by country or region for their best selling drugs in recent years. Net prices differ significantly from list prices, and likely capture the forces in the market more accurately, as list prices tend to just increase steadily over time. Future research should make use of net prices in order to better capture actual outcomes in the industry.

Finally, I aim to use the framework to formally analyze drug company investment behavior and innovation policy in future work. Various policies exist to encourage investment in drug development, and some of them, such as rare disease vouchers, exclusivity on combination drugs, and incentives to develop antibiotics, may depend on features of demand and pricing structure. In ongoing work, I look to use the framework as a basis for analyzing the investment behavior of drug companies with respect to incremental drugs.
Chapter 3

Crafting Intellectual Property Rights: Implications for Patent Assertion Entities, Litigation, and Innovation

3.1 Introduction

In this chapter, we show that the way patent rights are crafted has a large impact on several patent outcomes (including patent value, citations, and litigation) and especially on Patent Assertion Entities (PAEs), which are controversial as they drive patent licensing and litigation. We obtain variation in patent rights from the quasi-random allocation of patents to examiners: a one standard deviation change in examiner effects leads stock market capitalization to increase by 3 million dollars, citations by 24%, and litigation by 64%. PAEs overwhelmingly purchase and assert patents that were granted by “lenient” examiners; these examiners issue patents that are much more likely to be litigated but also to be deemed invalid (by the courts, the patent office, and applicants themselves). These results show the pitfalls of the influential “rational ignorance” view of the patent office, which argues that there is little at stake in the patent examination process.

1Co-authored with Xavier Jaravel
A striking feature of patent systems around the world is the enormous variation in private returns, social returns and litigation risk across patents (e.g., Pakes (1986) and Kogan et al. (2017) on firms’ returns, Toivanen and Väänänen (2012) and Bell et al. (2017) on inventors’ returns, Jaffe et al. (1993) on patent citations as a proxy for social value, and Lanjouw and Schankerman (2001) on exposure to litigation). What are the sources of this heterogeneity in patent outcomes? Scientific factors, such as the expertise of eminent scientists (e.g., Azoulay et al. (2010)) or a firm’s learning capacity (e.g., Cohen and Levinthal (1989)), are likely to be important drivers of patent outcomes. Yet, the value of a patent may not be solely determined by the quality of the idea embedded in it: a patent is not a raw idea but a carefully-worded legal document, conferring to its holder the right to sue for infringement.

Using variation in the process of writing the patent description and claims at the United States Patent Office (USPTO), this paper shows that a significant amount of heterogeneity in patent outcomes is independent of scientific determinants and results from the way patent rights are crafted. This finding is of particular relevance to understand a much-debated feature of the U.S. innovation system: the activities of Patent Assertion Entities (PAEs). PAEs, which acquire patents from third parties and generate revenue by asserting them against alleged infringers, have become controversial as they account for a large share of patent licensing and lawsuits.1 We find that PAEs disproportionately purchase and assert patents from “lenient” patent examiners, who craft patents that are more likely to be litigated and to be legally invalid.2

1For instance, RPX Corporation (2015) reports that the share of PAEs in overall patent lawsuits went from 35% in 2010 to 70% in 2015, while Federal Trade Commission (2016) documents that the share of PAE in licensing revenue was 80% in the wireless chipset sector between 2009 and 2014. Some experts conjecture that recent patent reforms (inter-partes review) and Supreme Court rulings (eBay Inc. v. MercExchange, L.L.C., Alice Corp. v. CLS Bank International) may curb PAE activity. One piece of evidence supporting this conjecture is that the number of defendants targeted by PAEs fell by 27% from 2016 to 2017. However, PAE-targeted firms still made up 56% of all defendants in IP cases, and the number of PAE lawsuits does not account for other forms of PAE activity.

2As we discuss in Section 3.4.4, surprisingly these examiners surprisingly do not, on average, increase patents’ private value, as measured by stock market responses.
indicates that PAEs are primarily the symptom of a specific friction in the patent system, the
way patent rights are crafted by lenient examiners (which affects litigation more broadly).

In the first part of the analysis, we show that the crafting of patent rights is an important
driver of a wide range patent outcomes, in particular those related to litigation. To arrive at
this result, we need variation in patent rights that is orthogonal to other determinants of
patent outcomes, such as scientific merit. Patent examiners may provide such variation as
they only affect patent rights, not the underlying idea embedded in the patent. Examiners
are heavily involved in the process of writing the patent description and claims through a
back-and-forth process with the applicant between patent filing and patent grant (known
as the “prosecution” process). By law, all examiners must ensure that the patents they
grant have clear, well-defined claims with appropriate scope. In practice, we find significant
variation in the way examiners craft patent rights (using prosecution data from Frakes and
Wasserman (2017)); we can therefore use examiner assignment as a source of variation in
patent rights, holding idea quality fixed.

A growing literature (e.g., Sampat and Williams (2015), Gaulé (2015) and Farre-Mensa
et al. (2017)) suggests that patent applications can be treated as quasi-randomly allocated to
examiners conditional on some covariates like application, year and technology class.³ Prior
research has used examiner assignment to estimate the causal effects of obtaining a patent,
as examiners differ in their grant rates. We build on this approach but differ in two ways.
First, we develop new quasi-experimental approaches to address identification concerns
about examiner specialization raised in more recent work (Righi and Simcoe (2017)); second,
we exploit variation in examiner prosecution behavior conditional on granting the patent,
rather than variation in the propensity of examiners to grant patents. We present evidence
supporting the validity of our approach after reporting a set of baseline results.

Our baseline research design estimates examiner fixed effects on the set of granted

³Conceptually, patent outcomes may vary because of heterogeneity in idea quality, heterogeneity in the
applicant’s input into patent drafting (typically via the applicant’s lawyers), and heterogeneity in the examiner’s
input into patent drafting. We use variation in patent drafting from examiners, rather than from lawyers,
because examiners are quasi-randomly assigned to patents while lawyer assignment may be correlated with
idea quality across applicants.
patents conditional on technology by year fixed effects. Our estimator uses an Empirical Bayes shrinkage correction to prevent “overfitting” of the fixed effects, which would misattribute some of the variation from the noise to causal variation across examiners. We apply this methodology to a range of patent outcomes related to private returns (stock market response from Kogan et al. (2017) and payment of maintenance fees), patent citations (total citations, self citations and external citations), patent market dynamics (patent sales, in general and specifically to PAEs) and legal disputes (patent infringement lawsuits, in general and specifically from PAEs). The estimated examiner effects are large for many outcomes, in particular for those related to PAEs and litigation. For example, a one standard deviation change in examiner effects leads stock market capitalization to increase by 3 million dollars, total citations by 24%, patent purchases by PAEs by 63%, litigation by 64%, and litigation specifically by PAEs by 46%. These estimates imply that policies affecting examiner behavior can have a substantial impact on the U.S. innovation system.4

We then validate the causal interpretation and the magnitudes of our baseline estimates in three ways. First, regarding identification, Righi and Simcoe (2017) report strong evidence that examiners working in the same technology-based group (called “art unit”)5 in fact specialize in specific sub-technologies, in ways that may be difficult to control for using observables. We develop two complementary quasi-experimental approaches to address this concern: (1) we show that there is a large subset of art units within which patent applications are assigned to examiners based on the last digit of the application’s serial number, implying that examiner assignment is orthogonal to potential confounds in these art units; and (2) we show that an examiner’s “busyness” can be used as an instrument

4As a point of comparison, the teacher value-added literature has documented sizable but much smaller effects of teachers on students’ outcomes. Chetty et al. (2014a) and Chetty et al. (2014b) estimate that a one standard deviation improvement in teacher effects in one grade raises students’ earnings by about 1% at age 28.

5Examiners at the USPTO are divided into more than 600 working groups called “art units”, each composed of about twenty examiners who handle patent applications on relatively homogeneous technologies. Following qualitative evidence on assignment of applications to examiners reported in Cockburn et al. (2003), Lemley and Sampat (2010) and Lemley and Sampat (2012), the recent literature treats assignment of patents applications to examiners within the same art unit as “as good as random” (e.g., Sampat and Williams (2015), Gaulé (2015) and Farre-Mensa et al. (2017)).
for application assignment: examiners with recently disposed applications are much more likely to be assigned the next incoming application, which provides variation in assignment even in art units with significant specialization. These two alternative sources of variation yield estimates that are similar to our baseline results. Second, we show that our results are not confounded by selection effects stemming from the decision to grant a patent. Since examiners differ in their grant rates, it could be the case that patent outcomes vary across examiners because of underlying differences across examiners’ pools of granted patents, independently of the crafting of patent rights. For instance, examiners with a low grant rate might only grant patents of high scientific merit. To establish that the bias is small empirically, we introduce flexible controls for examiners’ grant rates in our baseline specification and show that there is equally large causal variation in patent outcomes across examiners with the same grant rate. Third, we validate our baseline estimates in out-of-sample tests. We find that the Empirical Bayes shrinkage correction is important to suitably account for excess variance from noise and obtain unbiased estimates of examiner effects, in particular for rare outcomes such as PAE purchase and litigation.

In the second part of the analysis, we investigate why examiner effects are an important driver of the wave of patent purchases and lawsuits by PAEs, a major and controversial feature of the U.S. innovation system. We focus on outcomes related to PAEs because they rank among the outcomes that are most sensitive to examiner effects, and because PAEs have generated substantial academic and policy debate. There are two main hypotheses about PAEs’ behaviors: (1) PAEs may be useful intermediaries in the patent market, fostering greater incentives to innovate by lowering the cost of matching patent holders to patent buyers (e.g., Hagiu and Yoffie (2013) and Abrams et al. (2016)) and by helping enforce the patents of small inventors who lack the financial resources or legal expertise to defend...

---

6PAEs, also known as “non-practicing entities”, “patent monetization entities” or “patent trolls”, are defined as entities that generate revenue exclusively from patent licensing and litigation, without producing or selling products (Federal Trade Commission (2016)). Since there is no official list of PAEs, we follow the literature (e.g., Bessen and Meurer (2014)) and rely on a list provided by the RPX Corporation, a firm that helps companies manage risks from exposure to patent litigation. Universities, individual inventors and failed companies are excluded from the set of PAEs we consider and we show that the results are similar with alternative PAE lists from Cotropia et al. (2014).
themselves against large infringing companies (e.g., Lu (2012) and Galetovic et al. (2015)); or (2) PAEs may exploit imperfections in the legal system by acquiring patents with unclear claim boundaries and by asking innovative firms for licensing fees, whether or not the asserted patent is valid or infringed, in the hope that targeted firms will settle instead of risking a costly and uncertain trial (e.g., Miller (2013), Council of Economic Advisers (2013), Cohen et al. (2016) and Federal Trade Commission (2016)). Any plausible theory of PAEs should account for the new fact, documented in the first part of this paper, regarding the large sensitivity of PAEs to the way examiners craft patent rights. By analyzing which examiners drive patent acquisition and litigation by PAEs, we can assess which PAE theories are plausible.

We start by studying the characteristics of examiners who issue patents that are purchased and asserted by PAEs or by practicing firms. We correlate the causal examiner effects from the first part of the paper with measures of examiners’ prosecution behaviors based on the correspondence between examiners and applicants (from Frakes and Wasserman (2017)). We find that, within the same technology category, PAEs and practicing firms target patents issued by examiners with different characteristics. PAEs disproportionately purchase and assert patents that were granted by “lenient” examiners, who require applicants to make fewer changes to the text of the patent, such as clarifying a claim or withdrawing a claim deemed to be obvious or to bear on a non-patentable subject matter. Examiner leniency has a negligible impact on purchases by practicing firms, a sizable effect on litigation by practicing firms and on purchases by PAEs, and a much larger effect on litigation by PAEs. These patterns are first-order: for instance, a one standard deviation increase in a simple proxy for examiner leniency, the change in the number of words per claim between patent filing and grant, leads to an increase in litigation of 40.5% for PAEs and of 13.9% for practicing firms. These results cannot be accounted for by theories of PAEs based on a generic friction in the patent market, such as matching costs or the lack of financial resources for some inventors. They are consistent with the view that PAEs have a comparative advantage in patent litigation and therefore handle patents that are subject to a higher litigation risk,
induced by the way patent rights were crafted during the patent prosecution. The fact that examiner leniency is an important driver of litigation for both PAEs and practicing firms, although the effect is not as large for the latter, is in line with a nuanced view of PAEs (e.g., Lemley and Melamed (2013) and Schwartz and Kesan (2013)). According to this view, PAEs do not exploit imperfections of the legal system in an idiosyncratic way, but behave as litigation experts. In sum, our results show that PAEs’ activities are the symptom of the way patent rights are crafted by lenient examiners, who affect litigation more broadly.

Given the evidence that patent litigation by PAEs is strongly correlated with examiner leniency, we study whether lenient examiners tend to issue patents that are more likely to be invalid according to the standards set by current patent law. Several observers have hypothesized that PAEs assert invalid patents (e.g., Federal Trade Commission (2016)); approaching this question in terms of examiner effects has the potential to be informative about PAEs but also about patent litigation by practicing firms, who also selectively assert patents that were crafted by lenient examiners. Patent invalidity is notoriously difficult to measure because of selection effects. For instance, court rulings on patent validity are observed only for a strongly selected set of patents, as there were only a few hundred rulings over the past decade. To address this issue, we introduce a proxy for patent invalidity available in the full sample of granted patents: patent re-issuance requests, which can be filed by the applicant when a patent is deemed wholly or partly “inoperative or invalid” through an error in the document. Using this proxy as well as two common proxies for invalidity (decisions from court rulings and trials at the patent office), we document robust and quantitatively important evidence that lenient examiners issue patents that are more likely to be invalid. The evidence is therefore consistent with the view that PAEs are willing to purchase and assert patents whose validity is questionable, but PAEs are not the only entities to assert such patents: practicing firms do so as well.7

7This finding does not speak conclusively to the welfare effects of PAEs, because litigation of patents issued by lenient examiners could conceivably be socially valuable, even when these patents are deemed invalid by the courts, the USPTO, or applicants themselves. For instance, Galetovic et al. (2015) suggest that the process of litigation might be the socially-efficient dynamic process through which the patent system defines the contours of what should be patentable in highly-innovative, rapidly changing industries.
These results are surprising for several reasons. First, the very large impact of patent examiners on patent outcomes (conditional on grant) is unexpected. While previous work documented variation in examiners’ propensities to grant patents (Sampat and Williams (2015), Gaulé (2015) and Farre-Mensa et al. (2017)), we uncover the importance of the “intensive margin” of examiner effects (the crafting of patent rights, conditional on patent grant). This margin was not previously thought of as being of paramount importance for patent outcomes, as evidenced by (1) the fact that patent examiners are not very well paid;¹⁸ (2) the influential “rational ignorance” view of the patent office (Lemley (2000)), which states that there is little at stake in the patent examination process and that low-quality patent can be rationally ignored without significant consequences; (3) the focus of innovation research on the “macro-determinants” of the patent system, such as laws that establish a patent system or change the set of patentable subject matters;¹⁹ in contrast, we show the importance of the “micro-determinants” of patents by establishing that the specific way in which patent rights are crafted (by examiners who are all subject to the same patent law) has a substantial impact on a range of patent outcomes and is of first-order importance to understand certain features of the U.S. innovation system such as litigation (by PAEs in particular, but also by practicing firms). Additionally, our results on PAEs and litigation are also unexpected: (1) our finding that PAEs overwhelmingly purchase and assert patents from lenient examiners is not in line with mainstream theories of PAEs;¹⁰ (2) our findings imply that policies affecting examiner behavior could have a large impact on PAEs’ activities and litigation (in contrast with prior work that does not use quasi-experimental variation, e.g. Marco and Miller (2017));¹¹ (3) the examiners who drive PAEs’ activities do not tend to

¹⁸ In 2017, most patent examiners started with annual salaries between $54,099 and $82,094; salary can reach around $130,000 for senior patent examiners, which is much lower than the typical salary of patent lawyers.

¹⁹ For theoretical contributions, see e.g., Nordhaus (1969), Klemperer (1990) and Gilbert and Shapiro (1990); for empirical studies, see e.g., Sakakibara and Branstetter (2001), Moser (2005), Lerner (2009) and Williams (2013).

¹⁰ For recent work on PAEs, see e.g. Golden (2006), McDonough III (2006), Chien (2013), Tucker (2014b), Allison et al. (2018) and Haber and Werfel (2016).

¹¹ Our findings are also related to the pioneering study of Cockburn et al. (2003), who document relationships between some examiner characteristics and patent invalidity rulings. We show how to recover the full magnitude
create greater private value in general, as discussed in Section 3.4.\textsuperscript{12}

The remainder of the paper is organized as follows. Section II presents the data and descriptive statistics. Section III estimates examiner effects on a range of patent outcomes. Section IV studies the implications for PAEs’ activities. Section V concludes.

### 3.2 Data

In this section, we describe the data sources, define the samples and key variables we used in the analysis, and present summary statistics.

#### 3.2.1 Data Sources, Samples and Variable Definitions

*Patent Records.* We use two types of patent data to achieve two purposes. First, we rely on data on granted patents to measure a series of post-grant patent outcomes. Specifically, we build proxies for the private returns to patents, identify high-impact patents using citations, and document transactions in the patent market. Second, we use data on both granted and ungranted patent applications to identify examiners and measure their behavior during patent prosecution.

The granted patent dataset is obtained from USPTO and extends from 1975 to 2016.\textsuperscript{13} We rely on several proxies for the private returns to patents. Following the literature (e.g., Pakes (1986)), we use the payment of patent maintenance fees as a lower bound on the private valuation of the patent by the assignee. These fees are due 4 years, 8 years and 12 years after patent grant and are increasing over time.\textsuperscript{14} We also use the estimates of examiner effects on patent outcomes using a fixed effects estimator with a Bayesian shrinkage correction.

\textsuperscript{12}Therefore it is not the case that examiners with high PAE effects create more valuable patents for all agents in the patent system, as would be the case if they were just allowing greater scope.

\textsuperscript{13}This data is obtained through the Reed Tech USPTO page: \url{http://patents.reedtech.com/patent-products.php}.

\textsuperscript{14}For entities that do not benefit from reduced rates, the fees are $1,600 after 4 years, $3,600 after years and $7,400 after 12 years. The complete fee schedule is available from the USPTO at \url{https://www.uspto.gov/learning-and-resources/fees-and-payment/uspto-fee-schedule#Patent Maintenance Fee}. 

106
firm-level returns to patents from Kogan et al. (2017), who run event studies to estimate the excess stock market return realized on the grant date of patents assigned to publicly-traded firms; these estimates are available for patents granted before 2010. Moreover, we use data on patent citations to identify high-impact patents. We consider alternatively total citations, self citations (i.e. the assignee of the focal patent cites it in future patents) and citations by assignees that were not listed on the focal patent. We build these measures using the disambiguated assignee names from Balsmeier et al. (2015). To address censoring, we focus on citations that occurred in the three years following patent grant and we document in robustness checks that the results are similar when considering all citations. Finally, we measure changes in ownership of patents by merging in data on patent re-assignments from Graham et al. (2015b).\(^{15}\)

The data covering both granted and ungranted patent applications ranges from 2001 to 2015 and is obtained from the USPTO’s Patent Examination Dataset (Graham et al. (2015a)). We use this dataset to obtain unique numeric identifiers for each examiner during their tenure at the patent office, which are the critical inputs needed to estimate examiner effects. We then merge in data from Frakes and Wasserman (2017) on the correspondence between the examiner and the applicant. When asking applicants to amend patent documents, examiners need to ground their demands in specific sections of patent law, which we describe in Section 3.2.2.\(^{16}\) To characterize an examiner’s behavior during prosecution, we count the number of references made to the various sections of patent law. We also measure the examiner’s grant rate and, for granted patents, we directly measure the extent to which the text of the patent changes between application and grant by computing changes in the

---

\(^{15}\)Records of the assignments (transactions) affecting US patents are maintained by the US Patent & Trademark Office and available up from 1970 to 2014. There is no express legal requirement for parties to disclose assignments to the USPO, but patent laws provide incentives for recording. For instance, failure to record an assignment renders it void against any subsequent purchaser or mortgagee (35 USC 261). See Graham et al. (2015b) for more details.

\(^{16}\)When a patent is assigned to two examiners, a “primary” examiner with signatory authority and a “secondary” examiner who carries out most of the work, we treat the data as if the patent had been assigned to the secondary examiner only, following the example of Lemley and Sampat (2012).
number of words per claim and in the number of claims.\footnote{The USPTO’s Patent Examination Dataset only covers published patent applications. For ungranted patents, applicants are free to opt out of publications, which occurs in about 5\% of cases during the period we consider (Graham et al. (2015a)). The potential selectivity issues that could arise from the omission of “nonpublic” applications are largely orthogonal to our analysis, as we only rely on ungranted applications to measure an examiner’s allowance rate.}

Our main analysis sample is the Patent Examination Dataset merged to the patent outcomes of the granted patent dataset. We implement one important sample restriction: we exclude the so-called “continuation applications”, applications that follow an earlier-filed patent application. Those applications are assigned to the same examiner as the patent they follow and, therefore, quasi-random assignment of examiner does not hold. Our main analysis sample covers each non-continuation granted patents between 2001 and 2015, for which we observe the patent outcomes of interest as well as examiners’ identity and prosecution behaviors. For robustness, we estimate examiner effects on the full sample of (non-continuation) granted patents going back to 1975 by disambiguating examiner names (given the lack of numeric identifiers in this sample), but we lose information on examiners’ prosecution behaviors.

\textit{Patent Litigation}. We combine three data sources to obtain a comprehensive picture of patent litigation. Specifically, we combine data from LexMachina, Darts IP and RPX, which have been tracking intellectual property lawsuits since 2000 and thus offer full coverage for our main analysis sample. Although the datasets have significant overlap, it is sometimes challenging to identify all the patents involved in a given lawsuit, which creates differences in the lists.\footnote{We manually checked a few of the differences and verified that the patents were actually involved in litigation.}

\textit{Patent Assertion Entities}. Following standard practice (e.g., Bessen and Meurer (2014)), we rely on a list provided by the RPX Corporation, a firm that helps companies manage litigation risk, and exclude from the list any individual inventor, university or failed company.\footnote{Excluded entities are based on classifications from RPX and Cotropia et al. (2014).} We then build the patent portfolio of PAEs by merging the PAE list to the patent re-assignment dataset of Graham et al. (2015b) by assignee name. We only consider patents that were
purchased by PAEs (a few large PAEs, such as Intellectual Ventures, also invent their own patents). To establish that our results are robust to the choice of PAE list, we repeat the analysis using alternative PAE lists from Cotropia et al. (2014) and considering only the patent portfolio of Intellectual Ventures.\footnote{20}{Intellectual Ventures holds an estimated 25-30k US patents and released a list of around 20,000 patents on their website in November of 2013, which is available at http://patents.intven.com/data/ivpatents.csv.}

3.2.2 Summary Statistics

Table 3.1 presents the summary statistics for the variables of interest, documenting heterogeneity in patent outcomes (Panel A), the extent to which patent documents change between application and grant (Panel B) and heterogeneity in examiner behavior (Panel C).
### Panel A: Heterogeneity in Patent Outcomes

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Mean</th>
<th>Median</th>
<th>S.D.</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patent value from <em>Kogan et al. (2017)</em>, $M</td>
<td>9.0188</td>
<td>2.56</td>
<td>25.39</td>
<td>356,375</td>
</tr>
<tr>
<td>4th-year fee payment rate</td>
<td>0.8708</td>
<td>1</td>
<td>0.3354</td>
<td>1,247,958</td>
</tr>
<tr>
<td>8th-year fee payment rate</td>
<td>0.6098</td>
<td>1</td>
<td>0.4877</td>
<td>697,918</td>
</tr>
<tr>
<td>12th-year fee payment rate</td>
<td>0.2089</td>
<td>0</td>
<td>0.4065</td>
<td>373,207</td>
</tr>
<tr>
<td>Total patent citation within (3yr grant)</td>
<td>0.5256</td>
<td>0</td>
<td>1.461</td>
<td>988,585</td>
</tr>
<tr>
<td>Patent citations by same assignee within (3yr grant)</td>
<td>0.1134</td>
<td>0</td>
<td>0.7257</td>
<td>988,585</td>
</tr>
<tr>
<td>Patent citations by other assignees (3yr grant)</td>
<td>0.4122</td>
<td>0</td>
<td>1.1992</td>
<td>988,585</td>
</tr>
<tr>
<td>Rate of patent acquisition by non-PAEs</td>
<td>0.1965</td>
<td>0</td>
<td>0.3974</td>
<td>1,270,082</td>
</tr>
<tr>
<td>Rate of patent acquisition by PAEs</td>
<td>0.0102</td>
<td>0</td>
<td>0.10045</td>
<td>1,270,082</td>
</tr>
<tr>
<td>Rate of patent litigation by non-PAEs</td>
<td>0.0065</td>
<td>0</td>
<td>0.0804</td>
<td>1,270,082</td>
</tr>
<tr>
<td>Rate of patent litigation by PAEs</td>
<td>0.0004</td>
<td>0</td>
<td>0.0202</td>
<td>1,270,082</td>
</tr>
</tbody>
</table>

### Panel B: Changes to Patent Document between Application and Grant

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Mean</th>
<th>Median</th>
<th>S.D.</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in number of words per claim, %</td>
<td>57.32</td>
<td>25.24</td>
<td>84.58</td>
<td>1,110,272</td>
</tr>
<tr>
<td>Change in number of claims, %</td>
<td>-3.64</td>
<td>0</td>
<td>46.14</td>
<td>1,110,912</td>
</tr>
<tr>
<td>Use of Section 101 - Lack of utility or eligibility</td>
<td>0.0541</td>
<td>0</td>
<td>0.226</td>
<td>1,270,210</td>
</tr>
<tr>
<td>Use of Section 102(a) - Prior art exists</td>
<td>0.0174</td>
<td>0</td>
<td>0.130</td>
<td>1,270,210</td>
</tr>
<tr>
<td>Use of Section 103(a) - Obvious invention</td>
<td>0.419</td>
<td>0</td>
<td>0.493</td>
<td>1,270,210</td>
</tr>
<tr>
<td>Use of Section 112(b) - Vague claims</td>
<td>0.056</td>
<td>0</td>
<td>0.231</td>
<td>1,270,210</td>
</tr>
<tr>
<td>Patent citations added by examiner</td>
<td>0.185</td>
<td>0</td>
<td>0.388</td>
<td>1,270,210</td>
</tr>
</tbody>
</table>

### Panel C: Heterogeneity in Examiner Behavior

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of years at the U.S. Patent Office</td>
<td>6.35</td>
<td>7</td>
<td>3.19</td>
<td>10,018</td>
</tr>
<tr>
<td>Number of art units active in</td>
<td>1.80</td>
<td>2</td>
<td>0.96</td>
<td>10,018</td>
</tr>
<tr>
<td>Total patent applications processed</td>
<td>190</td>
<td>119</td>
<td>215</td>
<td>10,018</td>
</tr>
<tr>
<td>Patent grant rate</td>
<td>0.55</td>
<td>0.57</td>
<td>0.27</td>
<td>10,018</td>
</tr>
<tr>
<td>Use of Section 101 - Lack of utility or eligibility</td>
<td>0.09</td>
<td>0.02</td>
<td>0.14</td>
<td>10,018</td>
</tr>
<tr>
<td>Use of Section 102(a) - Prior art exists</td>
<td>0.02</td>
<td>0.006</td>
<td>0.03</td>
<td>10,018</td>
</tr>
<tr>
<td>Use of Section 103(a) - Obvious invention</td>
<td>0.45</td>
<td>0.48</td>
<td>0.21</td>
<td>10,018</td>
</tr>
<tr>
<td>Use of Section 112(b) - Vague Claims</td>
<td>0.19</td>
<td>0.17</td>
<td>0.15</td>
<td>10,018</td>
</tr>
<tr>
<td>Rate of patent acquisition by PAEs</td>
<td>0.011</td>
<td>0</td>
<td>0.032</td>
<td>10,018</td>
</tr>
</tbody>
</table>
Statistics on private returns, citations, patent sales and patent litigation are shown in Panel A of Table 3.1. Private returns feature high variance: the standard deviation of the firm-level patent value estimates from Kogan et al. (2017) is equal to almost three times the mean. The rates of maintenance fee payments are very high in early years but are substantially lower for the more expensive 12th-year maintenance fee payment, which also indicates heterogeneity in private valuations. Citations within three years of grant also feature high variance, indicating that patents greatly vary in their level of impact, regardless of whether we consider total citations, self-citations or citations by other assignees. The panel also shows that about 20% of all granted patents are sold to practicing (i.e. non-PAEs) firms and 1.01% to PAEs. Only 0.65% of all granted patents are litigated. Patent litigation by PAEs involves 0.04% of patents: this fraction is very small but it indicates that PAEs’ litigation rate is over six times higher than average, given that they own only about 1% of the patent stock.\textsuperscript{21} The purpose of Section 3.3 is to estimate the extent to which this heterogeneity in patent outcomes results from the way patent rights are crafted by examiners.

Panel B of Table 3.1 shows how the patent document changes between application and grant. In most cases, the examiner issues a so-called “rejection” as her first decision on the application (Williams (2017)), which is effectively an invitation for the patent applicant (or their representative, typically a patent attorney) to revise the text of the patent. Panel B shows that these changes are substantial. Through the back-and-forth with the examiner, the number of words in each claim increases by 57% on average.\textsuperscript{22} The lengthening of the claims can be interpreted as limiting the scope and clarifying the claims by making them more precise (Marco et al. (2016a)). In addition, examiner tend to ask applicants to reduce the number of claims to limit the scope of the patent: while the average change is limited (-3.64%), the standard deviation across patents is high (46.14%).\textsuperscript{23} We also observe

\textsuperscript{21}In addition, PAE patents are involved in about 7 cases per litigated patent versus about 2 cases for non-PAE litigated patents, based on a simple count of District Court cases per patent in the LexMachina data.

\textsuperscript{22}Given that the effective IP protection provided by a patent depends entirely on the content of the claims, and given that examiners affect to a great extent the words in the claims during prosecution, it is plausible that examiners may have a large impact on the legal force of the patent.

\textsuperscript{23}Following the literature, we report statistics for independent claims, leaving dependent claims aside as in
that the examiner asks the applicant to add citations to prior patents. The changes to
the patent document during the back-and-forth between the applicant and the examiner
show that the examiner is engaged in an iterative process and does not simply make a
one-time accept-or-reject decision. During this process, the examiner must substantiate her
demands by referring to specific sections of patent law corresponding to various standards of
patentability, namely that the invention is useful and its subject matter is eligible for a patent
(35 U.S.C. §101), it is novel relative to the prior art (35 U.S.C. §102(a)), it is non-obvious (35
U.S.C. §103(a)), and the claims are sufficiently clear to satisfy the disclosure requirement
(35 U.S.C. §112(b)). Panel B of Table 3.1 shows that on average non-obviousness is used
significantly more frequently than other sections.

Panel C of Table 3.1 presents statistics at the level of examiners. We observe 10,018
examiners in our main analysis sample, who work at the USPTO for 6.35 years on average.
The median number of technology areas in which an examiner works (called “art units”) is
two. The average examiner processes close to 200 patents over the course of our sample.
The panel shows that some examiners have a much higher grant rate than others, or have a
stronger tendency to invoke specific sections of patent law during the back-and-forth with
the applicant. We also observe large variation across examiners in the shares of their granted
patents that is purchased by a PAE: the standard deviation across examiners is twice the
average PAE purchase rate. This observed heterogeneity across examiners could merely
reflect noise or the fact that different examiners are working on different technologies,
or it could be driven by systematic (causal) differences in examiner behavior, which we
investigate in the remainder of the paper.

### 3.2.3 Illustration of Main Findings

Some of our main results in Sections 3.3 and 3.4 can be previewed in a simple, graphical
way. The various panels of Figure 3.1 document the relationship between patent acquisition
or litigation and a simple measure of examiners’ prosecution behavior.

Marco et al. (2016a).
Figure 3.1: The Effect of Examiners on Patent Acquisition and Litigation

Notes: In the various panels of this figure, the level of observation is a patent. The average change in the number of words per claim is measured at the level of an examiner, leaving out the focal patent. All specifications include art unit by year fixed effects and address potential extensive margin effects by controlling for the examiner’s leave-one-out patent grant rate (see text for details). The sample is the full patent grant sample described in Section 3.2.1, excluding examiners in the top 1% of the distribution of the total number of granted patents. The total number of patents granted by the examiner is used as weights in all panels. Each dot represents 5% of the data and the OLS best-fit lines are reported. Standard errors are clustered by examiner.
For each patent, we compute the average change in the number of words per claim between application and grant for all other granted patents processed by the same examiner, leaving out the focal patent. This leave-one-out examiner measure is exogenous to the focal patent. To ensure that we compare similar examiners, we include art unit by patent filing year fixed effects in all specifications. To ensure that potential extensive-margin selection effects are not confounding the results, we control for the (leave-one-out) grant rate of the examiner. Conceptually, these specifications compare patent outcomes for examiners who have the same grant rate, work in the same art unit in the same year, but differ in the way they craft property rights, as measured by the change in the number of words per claim between application and grant.

Panel (a) of Figure 3.1 shows that the probability that a patent is purchased by a PAE is a strongly negative function of the examiner’s propensity to ask applicants to add words to the patent claims (for instance to clarify them). Each dot in the binned scatter plot represents 5% of the data. The PAE purchase rate falls by about 25% of the baseline rate as we move from the left to the right along the x-axis, which shows very directly that the way examiners craft property rights is first-order for certain patent outcomes. Similarly large effects are found for litigation by PAEs and litigation by practicing firms, but not for purchases by practicing firms. The comparison of the various panels shows that PAEs and practicing firms respond in a similar way to examiners for the purpose of patent litigation (Panels (c) and (d)) but not for patent acquisition (Panels (a) and (b)).

This simple regression approach has the benefit that its robustness can immediately be assessed graphically. But the choice of the variable on the x-axis is arbitrary: this variable may capture only a small fraction of the relevant examiner behaviors and it may be correlated with examiner traits that would suggest different interpretations. To address this limitation, we turn to a research design that can recover the full impact of examiners on patent outcomes (Section 3.3), and we then correlate the examiner-level causal estimates with a range of examiner characteristics (Section 3.4).
3.3 Estimating Examiner Effects on Patent Outcomes

In this section, we estimate the impact of examiners on a range of patent outcomes. We assess the validity of the identifying assumptions in our baseline design using additional sources of variations and alternative specifications.

3.3.1 Research Design

To estimate the extent to which the heterogeneity in patent outcomes results from the way patent rights are crafted, we need variation in patent rights that is orthogonal to other determinants of patent outcomes, such as scientific merit. Through their back-and-forth with the applicant between initial filing and grant, examiners may provide such variation. By definition, examiners only affect patent rights, not the underlying idea embedded in the patent. Moreover, a growing literature suggests that patent applications can be treated as quasi-randomly allocated to examiners working in the same art unit in the same year (Sampat and Williams (2015), Gaulé (2015) and Farre-Mensa et al. (2017)).

Using quasi-random allocation of patent applications to examiners raises three empirical concerns, which were previewed in the introduction. First, since we are interested in recovering the full magnitude of examiner effects, conceptually we need to estimate fixed effects for all examiners, instead of projecting the data onto a specific examiner trait as in Figure 3.1.24 Given that we have a large number of examiners and work with rare outcomes such as litigation, it is likely that we may be “overfitting” the fixed effects: we may misattribute some of the variation from the noise to causal variation across examiners. This “excess variance” problem is well-known and we address it using a standard Bayesian

24Running a specification using examiner characteristics as regressors can only recover a lower bound for the overall effect of examiners, because the observed characteristics only capture a fraction of examiner behavior. A fixed effects estimator can recover the full effect, but it must be adequately adjusted to avoid excess variance due to overfitting of the fixed effects. In addition, the regression coefficients for the various examiner characteristics included in the specification should not be interpreted as causal, because random assignment occurs at the level of examiners and the observed examiner characteristics are likely to be correlated with other, unobserved examiner characteristics. For instance, in contemporaneous work, Kuhn (2016) and Kuhn and Thompson (2017) create an instrumental variable for patent scope based on an examiner characteristic they label “scope toughness”, but this characteristic could be correlated with other examiner traits that may affect the patent through channels other than scope.
Our baseline research design estimates examiner fixed effects on the set of granted patents with an Empirical Bayes shrinkage correction, conditional on art unit by year fixed effects. The identification assumption is that the allocation of (non-continuation) patents to examiner working in the same art unit in the same year is as good as random, i.e. it is not correlated with other determinants of patent outcomes. Given this assumption, we estimate examiner effects using the following statistical model:

\[ Y_i = \alpha_{ut(i)} + \nu_{ij}, \quad (3.1) \]

\[ \nu_{ij} = \mu_j + \epsilon_i, \]

where \( i \) indexes the patent, \( j \) the examiner, \( u \) the art unit and \( t \) the year. \( Y_i \) is the patent outcome of interest, \( \alpha_{ut(i)} \) denotes art unit by year fixed effects, \( \mu_j \) is the causal examiner effect of interest and \( \epsilon_i \) is an idiosyncratic patent-level shock. Our goal is to recover \( \sigma_\mu \equiv \sqrt{\text{Var}(\mu_j)} \).

We estimate the standard deviation of the underlying distribution of examiner effects in three simple steps. We first obtain estimates of residuals \( \{\hat{\nu}_{ij}\} \) for each patent by estimating art unit by year fixed effects in by OLS. We then compute the average estimated residual per examiner in each year:

\[ \bar{\nu}_{jt} \equiv \frac{1}{n_{jt}} \sum_{i=1}^{n_{jt}} \hat{\nu}_{ij} = \mu_j + \frac{1}{n_{jt}} \sum_{i=1}^{n_{jt}} \epsilon_i, \quad (3.2) \]

where \( n_{jt} \) is the number of patents processed by examiner \( j \) in year \( t \).
Finally, we compute the covariance between an examiner’s average residuals across consecutive years:

\[
\hat{\sigma}_\mu = \sqrt{\text{Cov}(\bar{\tilde{v}}_{jt}, \bar{\tilde{v}}_{j(t+1)})},
\]

which yields a consistent and unbiased estimate of \( \sigma_\mu \), as can be seen immediately from the second equality in (3.2). Excess variance in the average residual is handled by isolating the “systematic” component of the variation in average residuals that persists over time. If the examiner causal effects \( \{\mu_j\} \) are close to zero, we may still observe variation in the average residuals \( \{\bar{v}_{jt}\} \) across examiners in any given year because of idiosyncratic shocks, but there will be no covariance between examiners’ average residuals across years because the idiosyncratic shocks are uncorrelated. We call \( \sigma_\mu \) the “signal” standard deviation of examiner effects to contrast it with the “raw” standard deviation of residuals, which is contaminated by noise. The covariance calculation in (3.3) uses the counts of patents granted by each examiner \( \{n_{jt}\} \) as weights to increase precision.

The signal standard deviation is our primary focus because it is informative about the overall variation from examiners, but we also compute individual estimates of causal effects for each examiner. We compute an average of the residuals \( \hat{v}_{ij} \) over all years for each examiner, which we denote \( \bar{\tilde{v}}_j \). We then construct the empirical Bayes posterior estimate of each examiner effect by multiplying \( \bar{\tilde{v}}_j \) by a shrinkage factor:

\[
\hat{\mu}_j = \frac{\hat{\sigma}_\mu^2}{\text{Var}(\bar{\tilde{v}}_j)} \cdot \bar{\tilde{v}}_j.
\]

The shrinkage factor is the ratio of signal variance to total variance. We validate this

\[25\] To increase precision, \( \bar{\tilde{v}}_j \) is computed using weights that make \( \bar{\tilde{v}}_j \) a minimum variance unbiased estimate of \( \mu_j \) for each examiner. This step requires estimating the variances of other shocks in the statistical model. Specifically, we allow for an examiner-by-year shock \( \theta_{jt} \) and compute \( \hat{\theta}_j^2 = \text{Var}(v_{ij} - \bar{v}_{jt}) \) and \( \hat{\beta}_j^2 = \text{Var}(v_{ij}) - \hat{\sigma}_\mu^2 - \hat{\sigma}_e^2 \). We obtain \( \bar{\tilde{v}}_j = \sum w_{jt} \theta_{jt} \), with \( w_{jt} = \frac{h_{jt}}{\sum h_{jt}} \) and \( h_{jt} = \frac{1}{\hat{\sigma}_\mu^2 + \hat{\sigma}_e^2} \). See Online Appendix A for a complete discussion.

\[26\] Online Appendix A discusses the computation of \( \text{Var}(\bar{\tilde{v}}_j) \). Because of the precision weights in \( \bar{\tilde{v}}_j \), the shrinkage factor is lower for examiners for which more patents are observed. The estimated examiner effects \( \{\hat{\mu}_j\} \) have an empirical Bayes interpretation as the Bayesian posterior estimates of the examiner effects, starting from a normal prior distribution centered around zero with signal variance \( \sigma_\mu \). There is also a frequentist interpretation: the shrinkage factor is the OLS coefficient in a hypothetical regression of the true (unobserved)
research design by documenting in Section 3.3.3 that this approach yields unbiased estimates of examiner effects in out-of-sample tests, while ignoring excess variance delivers misleading results.

3.3.2 Baseline Estimates of Examiner Effects

Table 3.3 reports the estimates of examiner causal effects for a range of patent outcomes. We find substantial examiner effects for private value and for outcomes related to patent litigation.

Private value is strongly affected by examiner effects. The first row of Table 3.2 shows that the signal standard deviation of examiner effects corresponds to a 3.32 million dollar change in patent value, using the estimates from Kogan et al. (2017). In percentage terms, one signal standard deviation in examiner effects explains 40.8% of the average patent value for publicly-traded firms. The process of creation of patent rights therefore has a first-order impact on a patent’s private value to its assigned firm. We confirm this result in rows two to four of the table by considering other proxies. The rates of payment of patent maintenance fees at the various horizons are all responsive to examiner effects. Consistent with the notion that fee payments can only give a lower bound on private valuations, especially in earlier years when the fees are smaller, the examiner effects are smaller than with the Kogan et al. (2017) estimates; the signal standard deviations are under 10% of the average payment rate.

\[ \mu_j \text{ on the (observed) } \sigma_j. \]
<table>
<thead>
<tr>
<th></th>
<th>Signal S.D.</th>
<th>S.D. of Shrunk</th>
<th>Sample Size,</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% of Avg</td>
<td>Level</td>
<td>Effects, % of Avg</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Patent value, $M</td>
<td>40.80</td>
<td>(38.94—41.95)</td>
<td>3.32</td>
</tr>
<tr>
<td>4th-year fee payment rate</td>
<td>3.76</td>
<td>(3.64—3.91)</td>
<td>0.0328</td>
</tr>
<tr>
<td>8th-year fee payment rate</td>
<td>10.79</td>
<td>(10.40—10.82)</td>
<td>0.0658</td>
</tr>
<tr>
<td>12th-year fee payment rate</td>
<td>22.62</td>
<td>(21.44—23.37)</td>
<td>0.0472</td>
</tr>
<tr>
<td>Log total patent citation</td>
<td>23.79</td>
<td>(23.27—24.15)</td>
<td>0.0610</td>
</tr>
<tr>
<td>Log cites by same assignee</td>
<td>46.06</td>
<td>(43.62—48.63)</td>
<td>0.0278</td>
</tr>
<tr>
<td>Log cites by other assignees</td>
<td>24.47</td>
<td>(23.88—24.80)</td>
<td>0.0512</td>
</tr>
<tr>
<td>Patent acquisition, non-PAE</td>
<td>14.61</td>
<td>(13.60—15.41)</td>
<td>0.0287</td>
</tr>
<tr>
<td>Patent acquisition, PAE</td>
<td>62.96</td>
<td>(52.95—70.93)</td>
<td>0.0064</td>
</tr>
<tr>
<td>Patent litigation, non-PAE</td>
<td>64.25</td>
<td>(52.79—72.73)</td>
<td>0.0042</td>
</tr>
<tr>
<td>Patent litigation, PAE</td>
<td>46.04</td>
<td>(0—147.76)</td>
<td>0.0002</td>
</tr>
</tbody>
</table>

Notes: This table reports the signal standard deviations of examiner effects as a percentage of the mean (Column 1) and in level (Column 2), as well as the standard deviations of shrunk examiner effects (Column 3). The Bayesian shrinkage methodology used to obtain these estimates is presented in Section 3.3.1. In Column 2, 95% confidence intervals are obtained by bootstrapping. The log patent citation variables refer to the log of one plus the number of citations within three years of grant. The patent value variable is taken from Kogan et al. (2017), which is based on stock market response for patents granted to public firms. Patent value here is right-winsorized at the 99th percentile. See Section 3.2.1 for details on the sample and variable definitions.
Citations also respond to examiner effects. Considering in turn the signal standard deviations for total patent citations, self citations and citations by other assignees, we consistently find significant effects. The impact is strongest for self-citations, with a signal standard deviation of 46.06%, while the signal standard deviation for citations by other assignees is only 24.47%. This finding points to the role of cumulative innovation by the assignee.\textsuperscript{27}

We find particularly strong examiner effects for litigation and PAEs’ activities. The signal standard deviation of examiner effects accounts for over 60% of the baseline rate of patent purchases by PAEs. In contrast, the impact of examiners on the probability that a patent is sold to a practicing firm is much smaller: the signal standard deviation is 14.6% of the baseline rate. The impact of examiners on the probability that a patent is litigated is very large: the signal standard deviation is about 65% of the baseline rate. Considering the raw standard deviation of examiner effects would be very misleading: for rare outcomes like patent litigation or PAE purchase, the raw standard deviation is implausibly high, over four times larger than the signal standard deviation (Online Appendix Table C.1).

We use a bootstrapping procedure for inference. We re-draw samples with replacement at the application level and repeat the estimation of the signal standard deviations.\textsuperscript{28} The 95% confidence intervals are reported in Column (1) of Table 3.2. The signal standard deviations are all precisely estimated, except for one extremely rare outcome, patent litigation by PAEs.

The standard deviation of shrunk examiner effects obtained from equation (3.4) are also substantial. Column (3) of Table 3.2 reports these results. For instance, the standard deviation of shrunk examiners effects accounts for 29.48% of the average patent value from Kogan \textit{et al.} (2017), 31.11% of the baseline rate of PAE patent purchases, and 27.43% of the average rate of patent litigation.

The large signal standard deviations indicate that examiners have a first-order impact

\textsuperscript{27}Although this finding may also reflect strategic self citations, the literature on strategic self citations has emphasized the importance of strategic continuation filings, while we focus on non-continuation patents.

\textsuperscript{28}We also bootstrapped by re-sampling within examiner or within examiner by filing year and obtained similar results (not reported).
on patent outcomes. Consequently, policies affecting examiners have the potential to greatly affect the U.S. innovation system, for instance regarding litigation rates or the activities of PAEs. The large standard deviations of shrunk examiner effects indicate that, based on historical data, one can identify examiners who have a particularly large or low impact on specific outcomes.\textsuperscript{29} Our analysis so far is silent on the characteristics of these examiners, which we turn to in Section 3.4. Before doing so, we establish the validity of our identification assumptions with a series of tests and robustness checks.

3.3.3 Validation of Baseline Design: Addressing Non-Random Assignment and Selection

In this subsection, we use alternative research designs and specifications to investigate potential limitations of the baseline research design.

Alternative source of variation \#1: allocation of applications to examiners using the last digit of the application’s serial number. A potential concern with our baseline research design is that there is specialization even across examiners working in the same art unit at the same time (Righi and Simcoe (2017)). If specialization patterns are correlated with other factors that affect patent outcomes, then the examiner effects document in Table 3.2 may reflect omitted variable bias.

To address this potential concern, we identify art units where application assignment to examiners is determined by the last digit of the serial number of the patent application. The last digit of an application’s serial number, ranging from 0 to 9, is determined by the order of submission of applications and is therefore orthogonal to potential confounding variables such as scientific factors.\textsuperscript{30} Anecdotal evidence suggests that some art units assign applications to examiners based on the last digit of the serial number (Lemley and Sampat

\textsuperscript{29}We found that examiner effects do not tend to “average out” across outcomes; for instance, there is a large share of examiners who produce patents with systematically lower value, fewer citations and higher probabilities of litigation or of PAE purchase (not reported).

\textsuperscript{30}When a patent application is filed, the Office of Patent Application Processing assigns it a serial number. The first part of the serial number indicates the technology category while the last digits reflect the order of arrival of applications.
To determine which art units do so at different points in time, we compute an index of “concentration” of last digits across examiners working in the same art unit in the same year. If some examiners systematically get specific last digits, we will find a high degree of concentration. We use the concentration index initially developed by Mori et al. (2005) to study industry agglomeration, which was recently applied by Righi and Simcoe (2017) to the context of patents to study examiner specialization. Applied to our purposes, the test delivers a Chi-square statistic asking whether applications’ last digits are less dispersed across examiners than one would expect if last digits were not used for application assignment. We carry out the test in each year and in each art unit.

Figure 3.2 presents the results. Panel A shows the distribution of the p-values of the Chi-square tests across art units. There is a large number of art units with a p-value below 1%, indicating that these art units use application last digit to assign patents. The test only rejects the null that last digits are not used and it can of course not guarantee that in art units with a p-value below 1% all applications are assigned to examiners solely based on last digits. To address this limitation, we use a split-sample procedure to quantify the extent to which examiners get consistently assigned the same last digits. We split our main sample into two 50% samples at random. For each of the two subsamples, we compute the share of each last digit in an examiner’s pool of assigned applications. We then test whether the shares computed in the first subsample are predictive of those in the second subsample (comparing assigned shares for the same examiner in the same year in the two samples). Panel B of Figure 3.2 presents the results. For the art units that use last digits to allocate applications according to the Chi-square test (p-value < 0.01), we find a strong correlation between the last digit shares that were independently estimated in the two subsamples, with

\footnote{Righi and Simcoe (2017) use this test to document specialization of examiners in the same art unit and year, specifically testing for failure of random assignment with respect to technological features of the patent. We use the same test, but for the opposite purpose: we use the test to identify art units that allocate applications based on their last digits, which implies quasi-random allocation with respect to technological features of the patent.}

\footnote{Formally, we are testing the null that applications assignment is independent of their last digit; this test can be viewed as a multivariate generalization of a t-statistic comparing observed frequencies to the distribution under random assignment. For details, see Online Appendix A as well as Mori et al. (2005) and Righi and Simcoe (2017).}
a slope close to one. This result indicates that the use of last digits for allocation of patents is quantitatively important (i.e. the Chi-square tests are not identifying statistically significant but quantitatively small rejections of the null that last digits are not used for application assignment). In contrast, in the art units for which we cannot reject that last digits are not used for application assignment (p-value > 0.01), there is no relationship between the last digit shares across the two samples. The two panels of Figure 3.2 thus establish that there is a large number of art units that use last digits for application assignment and that they do so in a quantitatively important way.
Panel A: Distribution of p-values of Chi-square tests

Panel B: Graphical Evidence on Allocation by Application’s Last Digit

Figure 3.2: The Allocation of Patent Applications to Examiners by Application’s Last Digit

Notes: In Panel A, the level of observation is an art unit. This panel reports the distribution of the p-values of the Chi-square tests described in the main text; a p-value below 0.01 indicates excess concentration of patent applications across examiners by application’s last digit. In Panel B, the level of observation is an examiner-by-application’s last digit cell. Two binned scatter plots are reported with the corresponding best fit lines; each cell is weighed by the total number of applications processed by the examiner.
Panel A of Table 3.3 shows that the signal standard deviations estimated for art units that allocate patents using last digits are quantitatively similar to those from the baseline design. Column (1) shows the signal standard deviations for various outcomes in the sub-sample of art units with a p-value below 0.01 in the Chi-square test. Moreover, Column (2) repeats the estimation of the signal standard deviation in the subsample of art units belonging to Information Technologies.\textsuperscript{33} The results are similar in this subsample as well, which is comforting because Righi and Simcoe (2017) report that they find no evidence of examiner specialization in Information Technologies.\textsuperscript{34}

\textsuperscript{33}This subsample includes the following technology centers: Computer Architecture and Software (21); Computer Networks, Multiplex, Cable and Cryptography/Security (24); Communications (26); and Business Method art units (3620s, 3680s, 3690s). We exclude technology center 2800 (Semiconductors), which Righi and Simcoe (2017) identify as having significant examiner specialization.

\textsuperscript{34}The signal standard deviation for patent value from Kogan \textit{et al.} (2017) is smaller in the IT subsample (3.3) than in the full sample (Table 3.2). But this is due to heterogeneity in the signal standard deviation of examiner effects across technology categories, rather than to endogeneity concerns: Online Appendix Table C.2 reports smaller signal SDs for patent value in IT-related technology categories.
Table 3.3: Validation of Baseline Estimates of Examiner Effects

Panel A: Accounting for Violations of Random Assignment

<table>
<thead>
<tr>
<th></th>
<th>Signal S.D., % of Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Patent value</td>
<td>44.32</td>
</tr>
<tr>
<td>Log cites</td>
<td>21.10</td>
</tr>
<tr>
<td>Patent acquisition by non-PAEs</td>
<td>15.38</td>
</tr>
<tr>
<td>Patent acquisition by PAEs</td>
<td>40.01</td>
</tr>
<tr>
<td>Patent litigation by non-PAEs</td>
<td>55.65</td>
</tr>
</tbody>
</table>

Sample

<table>
<thead>
<tr>
<th>Art units allocating by last digit, according to $\chi^2$ test</th>
<th>Art units in Information Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of art units</td>
<td>243</td>
</tr>
<tr>
<td></td>
<td>254</td>
</tr>
</tbody>
</table>

Panel B: Accounting for Extensive Margin Selection Effects

<table>
<thead>
<tr>
<th></th>
<th>Signal S.D., % of Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Patent value from <em>Kogan et al. (2017)</em></td>
<td>40.46</td>
</tr>
<tr>
<td>Log total patent citation</td>
<td>18.66</td>
</tr>
<tr>
<td>Rate of patent acquisition by non-PAEs</td>
<td>14.31</td>
</tr>
<tr>
<td>Rate of patent acquisition by PAEs</td>
<td>62.64</td>
</tr>
<tr>
<td>Rate of patent litigation by non-PAEs</td>
<td>63.06</td>
</tr>
</tbody>
</table>

Controls

<table>
<thead>
<tr>
<th>Examiner grant rate</th>
<th>Examiner grant rate and application characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel C: Accounting for Excess Variance with Empirical Bayes Beta-Binomial Count Model

<table>
<thead>
<tr>
<th></th>
<th>S.D. of Shrunk Examiner Effects, % of Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate of patent acquisition by PAEs</td>
<td>46.72</td>
</tr>
<tr>
<td>Rate of patent acquisition by non-PAEs</td>
<td>7.99</td>
</tr>
<tr>
<td>Rate of patent litigation by non-PAEs</td>
<td>48.95</td>
</tr>
</tbody>
</table>

Notes: Panel A reports the signal standard deviations of several examiner effects using the Bayesian shrinkage methodology in two subsamples in which there is no examiner specialization within art unit. Panel B repeats the calculation of the signal standard deviations of examiner effects in the same sample as Table 3.2, but adding controls to address potential selection effects. Panel C reports the standard deviation of average shrunk examiner effects using the Empirical Bayes Beta-Binomial Count model. See Section 3.3.3 for a description of the methodologies underlying each panel.
Alternative source of variation #2: a busyness instrument. A limitation of using art units that allocate applications using last digits is that these art units account for only about a third of all art units. There is anecdotal evidence that some art units allocate applications to examiners based on the timing of arrival of applications (Lemley and Sampat (2012)). When a new application arrives at the patent office, an examiner who recently finished processing another application may be particularly likely to be assigned the new application, because they happen to have more time on their hands.

To proxy for how busy an examiner is when a given new application arrives, we measure the number of cases closed by the examiner in the two preceding weeks. For each incoming application, we compute assignment probabilities across all examiners working in the relevant art unit and time period based on the number of cases closed in the previous two weeks, art unit by year fixed effects and examiner fixed effects. Within an art unit and a year, assignment probabilities vary only because of changes in (relative) busyness across examiners. We estimate assignment probabilities using a simple linear probability model, presented in Online Appendix A.

Using the estimated assignment probabilities across examiners, we instrument for the characteristics of the examiner who actually processed the application. For instance, if an application arrives in the art unit at a time when only “lenient” patent examiners (who tend to ask the applicant to make only a few changes to the patent) happen to be free, then the application should be more likely to receive a more lenient treatment. Using this source of variation, we can document the relationship between any given examiner characteristic and any patent outcomes. Specifically, we can use the estimated assignment probabilities to compute the expected examiner characteristic, which we can relate to the actual characteristic of the examiner who handled the application (the “first stage”) and to any patent outcome of interest (the “reduced form”).

Figure 3.3 presents the results of the busyness approach. The panels are based on the
following specifications:

\[
E_{j(i)} = \beta_1 \left( \sum_{j \in at(i)} p_{ij} E_j \right) + a_{ut(i)} + v_{ir}
\]  
\[
Y_i = \beta_2 \left( \sum_{j \in at(i)} p_{ij} E_j \right) + a_{ut(i)} + \kappa_{ir}
\]

(3.5)  
(3.6)

where \(i\) indexes the patent, \(j\) the examiner, \(u\) the art unit and \(t\) time. \(p_{ij}\) denotes the application-specific examiner assignment probability; \(E_j\) denotes the examiner characteristic, measured using a leave-one-out procedure that does not use information on patent \(i\); \(E_{j(i)}\) is the (leave-one-out) characteristic of the examiner who actually processed application \(i\); and \(Y_i\) is the patent outcome of interest. Figure 3.3 estimates these specifications, considering the (leave-one-out) change in the number of words per claim as the examiner characteristic and the (actual) purchase by a PAE as the outcome of interest. This choice of variables allows for a comparison with Figure 3.1, which did not use the busyness instrument and was using raw variation in the examiner’s propensity to change the number of words per claim between application and grant.
Panel A: First Stage

Panel B: Reduced Form

**Figure 3.3: A Busyness Instrument for the Effect of Examiners on Patent Acquisition by PAEs**

*Notes:* Panel A shows the relationship between the busyness instrument (described in the main text) for an examiner’s propensity to change the number of words per claim during application and grant and the propensity of the examiner whom the application was actually assigned to. Panel B depicts the relationship between the busyness instrument and the purchase rate by PAEs. On both panels, each dot represents 5% of the data and OLS best-fit lines are reported.
Panel A of Figure 3.3 reports the relationship between the actual and expected examiner characteristics, as in (3.5). The slope is strong and positive and the binned scatter plot is close to linear, indicating that the busyness instrument has power. Panel B of 3.3 shows the relationship between PAE purchase and the expected examiner propensity to increase the number of words per claims: there is a strong downward relationship. These patterns are similar to Figure 3.1, which used the raw variation in examiner characteristic instead of the busyness instrument. These results provide additional evidence that departures from random assignment of examiners to applications do not bias our estimates.

Accounting for potential selection effects on the extensive margin. Another potential concern with our baseline research design is that our estimates may be confounded by selection effects stemming from the decision to grant a patent. Examiners differ in their grant rates, therefore it could be the case that patent outcomes vary across examiners because of underlying differences across examiners’ pools of granted patents, independently of the crafting of patent rights. For instance, examiners with a low grant rate might only grant patents of high scientific merit. To investigate this possibility, we introduce controls for the examiner’s leave-one-out grant rate in equation (3.1) and then repeat the estimation of the signal standard deviation using equation (3.3). With this specification, we are now estimating the amount of systematic variation in patent outcomes across examiners who work in the same art unit, in the same year, and have the same grant rate.

Panel B of Table 3.3 reports the results and shows that our baseline estimates remain virtually unaffected. Column (1) controls for the grant rate in (3.3).\(^{35}\) The estimated signal standard deviations are very similar to our baseline estimates from Table 3.2. In principle, it may be possible for extensive margin effects to operate even across examiners with the same grant rate. For instance, an examiner may systematically grant patents with underlying technological characteristics that appeal to PAEs, while another examiner (with a similar overall grant rate) may tend to systematically reject those patents and grant others. To assess

\(^{35}\)To flexibly control for the grant rate, we introduce a quartic polynomial in the grant rate. The results are similar when controlling linearly for the grant rate or introducing higher-order polynomials (not reported).
how strong this effect might be empirically, Column (2) introduces controls for a host of initial characteristics of the patent application, namely: the application’s initial number of independent claims and number of words per claim; the assignee’s number of applications, grants and citations prior to the filing date; and the first inventor’s number of applications, grants and citations prior to the filing date. The estimates of signal standard deviations are not sensitive to these controls, indicating that extensive margin effects are unlikely to bias our estimates in any meaningful way.

**Accurately accounting for excess variance.** The preceding discussion indicates that our results are robust to failures of random assignment and extensive margin selection effects. A remaining potential concern is that the Empirical Bayes shrinkage correction used in our baseline research design may fail to account for noise perfectly. To address this point, we first discuss some plausible limitations of our baseline design, in particular for rare binary outcomes such as litigation; we then present an alternative approach which addresses these limitations and produces similar results. Finally, we use out-of-sample tests to directly show that our baseline design accurately accounts for excess variance.

Our baseline research design yields very large signal standard deviation estimates for rare binary outcomes, such as litigation or purchase by a PAE, but the Bayesian shrinkage correction may not be appropriate in such cases. Indeed, for binary outcomes our statistical model in equation (3.1) may be misspecified as it does not impose the constraint that the predicted value should lie between zero and one. Given that rare binary outcomes have a particularly high estimated signal standard deviation in Table 3.2, it appears important to assess whether these results are sensitive to a change in the underlying statistical model.

We repeat the analysis using an Empirical Bayes Beta-Binomial count model, a common statistical model that can fit count data in a flexible way (Ellison and Swanson (2010)). To see how this framework operates, consider the example of patent purchases by PAEs. For each examiner $j$, we observe data of the form $(n_j, r_j)$, where $n_j$ is the examiner’s total number of granted patents and $r_j$ is the number of patents granted by the examiner that were purchased by PAEs. We assume that the probability $p$ of granting a patent purchased
by a PAE follows a Beta distribution across examiners working in the same art unit in the same year:  
\[ p \sim \text{Beta}(\alpha, \beta). \]

Given that we are examining the count of PAE purchases across examiners, the likelihood function for the data is a binomial distribution. Using the fact that the beta distribution is the conjugate prior of the binomial distribution, we show in Online Appendix A that the integrated likelihood is:

\[
L(r_j|n_j, \alpha, \beta) = \binom{n_j}{r_j} \frac{\Gamma(\alpha + \beta) \Gamma(r_j + \alpha) \Gamma(n_j - r_j + \beta)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(n_j + \alpha + \beta)},
\]

which we estimate via maximum likelihood in each art unit by year. Having recovered estimates of the hyperparameters, \( \hat{\alpha} \) and \( \hat{\beta} \), we compute the posterior mean for each examiner:

\[
\hat{\mu}_j^{\text{BetaBinomial}} = \frac{\hat{\alpha} + r_j}{\hat{\alpha} + \hat{\beta} + n_j}. \tag{3.7}
\]

Panel C of Table 3.3 reports the standard deviation of the estimates: we continue to find large examiner effects. This finding indicates that our large estimates for the impact of examiner on patent litigation and purchase by PAEs is not an artifact of the statistical model used in our baseline design.

To conclude this section, we conduct out-of-sample tests of the examiner effects estimated in our baseline research design to check that we have recovered estimates of the correct magnitude. After splitting the main analysis sample into two 50% samples at random, in each subsample we compute the raw examiner effects using equation (3.2) and the shrunk examiner effects using equation (3.4). To test predictive accuracy, we regress the raw examiner effect from the first subsample on the shrunk examiner effects from the second subsample.\(^37\) We also regress the raw examiner effect from the first subsample on the raw examiner effect from the second subsample to assess whether a standard regression approach would suffer from excess variance. We do so in the full sample but also in a

\(^{36}\)Intuitively, this procedure shrinks an examiner’s PAE share towards the mean PAE share in the art unit. The amount of shrinkage is larger for examiners who have granted fewer patents.

\(^{37}\)We regress raw effects on shrunk effects because the shrinkage factor in the shrunk effects addresses measurement error, which poses an issue for the independent variable but not for the dependent variable.
reduced sample of examiners who granted more than fifty patents, as measurement error may no longer be a problem if sufficiently many patents are observed per examiner.

Figure 3.4 reports the results and shows that the Empirical Bayes shrinkage approach yields unbiased estimates of examiner effects, in contrast with standard regression analysis. A regression coefficient of one indicates unbiased prediction, while a coefficient below one indicates attenuation bias and implies that the estimates suffer from excess variance due to noise. Figure 3.4 shows that our baseline design delivers unbiased estimates of examiner effects even for rare outcomes such as patent purchase by PAEs or patent litigation. The point estimates are very close to one and are precisely estimated. In contrast, the specifications without shrinkage always deliver a coefficient well below one, indicating that the raw variation in examiner effects contains a lot of noise. This problem is less acute for outcomes that are more common, such as the patent value measure of Kogan et al. (2017) (with a regression coefficient close to 0.5 full sample), than for rare outcomes like patent litigation (with a regression coefficient close to 0.1 in the full sample). Restricting the analysis to examiners who handle a lot of patents does not solve the problem, which offers another vindication of our baseline research design.
Figure 3.4: Out-of-Sample Tests of Baseline Estimates of Examiner Effects

Notes: This figure reports the OLS coefficients in examiner-level out-of-sample regressions. After splitting the main analysis sample into two halves at random, we compute the raw and shrunk examiner effects on each half following the methodology described in Section 3.3.3. To test predictive accuracy, we regress the raw examiner effect from the first half on examiner effects estimated in the second half, using in turn as regressors the shrunk examiner effects (“shrinkage”), the raw examiner effects (“no shrinkage, all examiners”) and the raw examiner effects for the subset of examiners who have granted more than fifty patents (“no shrinkage, examiners with > 50 patents only”). A regression coefficient of one indicates unbiased prediction. The heteroskedasticity-robust 95% confidence interval is reported.

3.3.4 Robustness Checks

Table 3.4 shows the robustness of the signal standard deviations when using alternative samples and specifications. The first row repeats the analysis including continuation applications; the second row includes all granted patents from 1976 to 2015; the third row controls for the length of time between filing and grant to assess whether the results may be driven by delays rather than by the way patent rights are crafted; the fourth row includes fixed effects for examiner experience as in Frakes and Wasserman (2017).\textsuperscript{38} The results are

\textsuperscript{38}An alternative to the inclusion of examiner experience fixed effects in our baseline specification is to look for discontinuities in patent outcomes around examiners’ promotions; we find no discontinuity (Online
very similar across samples and specifications. Finally, the Online Appendix shows that the signal standard deviations are of comparable magnitudes across technology categories (Online Appendix Table C.2) and reports the distributions of the shrunk examiner effects (Online Appendix Figure C.2).

Table 3.4: Robustness Checks on Examiner Causal Effects on Patent Outcomes

<table>
<thead>
<tr>
<th></th>
<th>Signal S.D., % of Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patent value from Kogan et al. (2017)</td>
</tr>
<tr>
<td>(A) Incl. continuations</td>
<td>41.8%</td>
</tr>
<tr>
<td>(B) Granted patents 1976-</td>
<td>36.7%</td>
</tr>
<tr>
<td>(C) Review-Time Ctrl</td>
<td>40.8%</td>
</tr>
<tr>
<td>(D) Examiner Career Ctrl</td>
<td>41.5%</td>
</tr>
</tbody>
</table>

Notes: Row (A) adds continuation applications which were filed and granted between 2001 and 2012 to our baseline analysis sample, which covers the same period. Row (B) uses the sample of all non-continuation granted patents from 1976 to 2015. Row (C) controls for “time under review” in equation (3.1) with a quadratic polynomial in the number of years between filing and grant. Row (D) controls for examiner career effects in equation (3.1) with experience fixed effects, as defined in Frakes and Wasserman (2017) (namely, the examiner’s GS level by bins corresponding to 0-1, 2-3, 4-5, 6-7, 8+ years experience at that level). All reported values are normalized by the average in the relevant sample.

3.4 Implications for Patent Assertion Entities

Our analysis so far has established that the crafting of patent rights is an important driver of a wide range of patent outcomes, in particular those related to PAEs and litigation. In this section, motivated by the large sensitivity of PAEs to the way examiners craft patent rights, we investigate the features of examiner behavior that drive PAEs’ responses. We find that “lenient” examiners, who issue patents with higher litigation and invalidity risks, produce a much higher share of patents purchased and asserted by PAEs. We discuss how this evidence helps discipline theories of PAE behavior.

Appendix Figure C.1), which confirms that examiner experience effects play a second-order role compared with the examiner fixed effects we focus on.
3.4.1 Research Design

There are two standard views of the role played by PAEs in the patent market. According to the first view, PAEs could be useful intermediaries who address standard frictions in the patent market by lowering transaction costs and solving liquidity problems (Hagiu and Yoffie (2013), Abrams et al. (2016), Lu (2012) and Galetovic et al. (2015)). The second view suggests that PAEs do not help address any particular friction but, rather, exploit limitations of the legal system by asserting patents of questionable validity in the hope that targeted firms will pay them settlement fees instead of risking a costly and uncertain trial (Miller (2013), Council of Economic Advisers (2013), Cohen et al. (2016) and Federal Trade Commission (2016)).

We investigate the extent to which the two standard views can account for the (quantitatively large) patterns related to examiners in the data. The way examiners craft patent rights has a first-order impact on PAEs: a one standard deviation change in examiner effects shifts the probability of patent acquisition by a PAE by over 60% of the baseline rate (Table 3.2). This fact may not be incompatible with the two standard views of PAE behavior. For instance, the process of creation of patent rights may create frictions affecting both PAEs and practicing firms (in line with the first view) or may lead to the issuance of questionable patents that only PAEs are willing to purchase and exploit via frivolous litigation (in line with the second view).

We examine this question using detailed data on the prosecution behaviors of examiners, drawing a contrast between the responses of PAEs and practicing firms. We start by characterizing the prosecution behaviors that are predictive of future purchase or litigation by a PAE or practicing firms (Section 3.4.2); we then investigate whether these prosecution behaviors are predictive of patent invalidity (Section 3.4.3). Specifically, we run regressions of the following form:

\[ Y_i = \beta E_{j(i)} + a_{ut(i)} + \epsilon_i, \]  

(3.8)

where \( i \) indexes the patent, \( j \) the examiner and \( ut \) the art unit-by-year; \( Y_i \) is the patent outcome of interest; and \( E_{j(i)} \) is a (vector of) examiner behavior(s), estimated using a leave-
one-out procedure that does not use information on patent $i$. We scale the examiner behavior measures $E_{j(i)}$ by their signal standard deviations, which are estimated using (3.3). This standardization gives us the proper scaling to compare the quantitative importance of various examiner traits.\(^{39}\)

We rely on a variety of proxies reflecting different aspects of examiner behavior to isolate robust correlations with the potential to inform theories of PAE behavior. The estimates from specification (3.8) cannot be interpreted as causal because quasi-random assignment occurs at the level of examiners working in the same art unit at the same time, and not at the level of examiners’ traits. Given that quasi-random assignment is at the level of examiners, the only causal effect that can be recovered is the effect of the examiner “as a whole” on patent outcomes (as in Section 3.3).\(^{40}\) In contrast, the relationships between specific examiner traits and patent outcomes may be biased by potential omitted variables (i.e. other traits of the examiner that are unobserved). To address this limitation, we use several proxies to control for various aspects of examiner behavior and we focus on establishing correlations which (1) are quantitatively large and robust to the inclusion of additional controls; and (2) can be interpreted as reflecting a more general trait of the examiner, such as the propensity to let the applicant keep the text of the claims relatively unchanged between application and grant (“leniency”).

### 3.4.2 PAEs and Examiner Behavior

In this subsection, we document which examiner traits correlate with patent acquisition or litigation by PAEs and practicing entities. We use specification (3.8) and consider seven measures that capture different aspects of examiner behavior.

We use three general proxies for the degree of “leniency” of the examiner. By examiner leniency, we refer to the extent to which the examiner makes demands on the applicant

---

\(^{39}\)Specification (3.8) is analogous to the regression underlying Figure 3.1, except that we are now using properly scaled regressors.

\(^{40}\)One would need a quasi-experiment that directly affects specific behaviors (e.g., a training program) in order to recover more granular causal impacts.
during prosecution. First, the percentage change in the number of words per claim (averaged across claims) indicates the extent to which the examiner asks the applicant to refine the claims. Second, the percentage change in the number of claims reflects the extent to which the examiners affects the overall structure and scope of the patent document. Third, the examiner’s grant rate can be interpreted as another proxy for leniency, given that examiners who are more demanding on applicants also have lower grant rates.

To characterize in greater detail the examiner behaviors that drive PAEs’ activities, we measure examiners’ propensities to cite specific sections of patent law when asking the applicant to revise the patent. As mentioned previously, the examiner must substantiate any demand by referring to specific sections of patent law corresponding to various standards of patentability. An examiner who is less lenient should tend to refer more often to any of the sections compared with other examiners working in the same art unit at the same time. The relative frequency of usage of the various sections may differ across examiners depending on their examination styles. Examiners who place more emphasis on the invention being useful and eligible for a patent should use section 101 more often; those who particularly care about prior inventions should refer section 102 frequently; section 103 should be invoked more often by examiners who are particularly sensitive to the requirement that the invention should be non-obvious to someone who knows the field; and section 112(b) should be used by examiners who focus on the requirement of claim clarity.41

Table 3.5 presents the results with patent acquisition as the outcome.42 In both panels, the first seven columns run univariate regressions, while columns (8) and (9) consider multivariate regressions. Panel A shows that all proxies of examiner leniency deliver a similar message: more lenient examiner grant substantially more patents that are eventually purchased by PAEs. The regression coefficients are standardized by the signal standard

41 Although all examiners are supposed to apply the same standards for patent grant, which are determined by patent law, we find large causal variation across examiners in terms of their propensity to refer to the various sections (Online Appendix Table C.3)

42 The sample is restricted to art units that are part of Information Technologies since PAEs are primarily active in these art units (Online Appendix Table C.4). All results reported in this section are similar in the full sample (Online Appendix Tables C.5, C.6 and C.7).
deviations of the regressors and expressed as a percentage of the outcome. Column (1) shows that a one standard deviation increase in the distribution of examiner effects for the change in number of words per claim implies a 13.9% decrease in the probability of purchase by a PAE. This fraction is relatively large, given that a one standard deviation change in the overall examiner effect accounts for about 60% of the baseline rate (Table 3.2). Columns (2) and (3) show that the effect goes in the same direction, with a similar magnitude, for the other broad proxies for examiner leniency: a one standard deviation increase in the change in number of claims implies a 7.3% increase in the probability of PAE purchase;\textsuperscript{43} the corresponding number for grant rates is 11.4%. Columns (4) to (7) show that the same finding holds when considering the use of various sections of patent law: examiners who use sections more often tend to have a lower rate of purchase by PAEs (although some specifications are noisy). Column (8) presents the results of a specification that simultaneously includes all types of references to patent law. In this specification, the section relating to the obviousness of the invention is the most important. Finally, specification (9) includes all regressors simultaneously. The results become more noisy because of collinearity, but the coefficient on the change in the number of words per claim remains large, significant and similar in magnitude to the univariate regression in Column (1). These findings show that PAEs have a preference for purchasing patents that were issued by lenient examiners.

\textsuperscript{43}More lenient examiners tend to reduce the number of claims by less, which means that a higher change in the number of claims (in absolute value) reflects higher leniency. In contrast, a more lenient examiner increases the number of words per claim by less, i.e. a higher change in the number of words per claim reflects lower leniency.
### Table 3.5: Patent Acquisition and Examiner Behavior

#### Panel A: Patent Acquisition by PAEs

<table>
<thead>
<tr>
<th>Leave-one-out Examiner Effects</th>
<th>Patent Purchase by PAE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>% Change in Words/Claim</td>
<td>-0.139***</td>
</tr>
<tr>
<td>% Change in # Claims</td>
<td>0.073**</td>
</tr>
<tr>
<td>Grant Rate</td>
<td>0.114***</td>
</tr>
<tr>
<td>Use of Section 101</td>
<td>-0.061*</td>
</tr>
<tr>
<td>- Not patentable</td>
<td>0.007</td>
</tr>
<tr>
<td>Use of Section 102(a) -</td>
<td>-0.0602***</td>
</tr>
<tr>
<td>- Prior art exists</td>
<td>(0.024)</td>
</tr>
<tr>
<td>Use of Section 103(a) -</td>
<td>-0.037</td>
</tr>
<tr>
<td>- Obvious invention</td>
<td>(0.027)</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>274,464</td>
</tr>
</tbody>
</table>

#### Panel B: Patent Acquisition by Practicing Firms

<table>
<thead>
<tr>
<th>Leave-one-out Examiner Effects</th>
<th>Patent Purchase by Practicing Firm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>% Change in Words/Claim</td>
<td>0.0071</td>
</tr>
<tr>
<td>% Change in # Claims</td>
<td>-0.003</td>
</tr>
<tr>
<td>Grant Rate</td>
<td>0.022***</td>
</tr>
<tr>
<td>Use of Section 101</td>
<td>0.0147**</td>
</tr>
<tr>
<td>- Not patentable</td>
<td>(0.0065)</td>
</tr>
<tr>
<td>Use of Section 102(a) -</td>
<td>-0.0037</td>
</tr>
<tr>
<td>- Prior art exists</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Use of Section 103(a) -</td>
<td>0.0065</td>
</tr>
<tr>
<td>- Obvious invention</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Use of Section 112(b) -</td>
<td>0.002</td>
</tr>
<tr>
<td>- Vague claims</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>274,464</td>
</tr>
</tbody>
</table>

**Notes:** The sample is restricted to IT patents. Regressors are standardized by their standard deviations and coefficients are expressed as a fraction of the mean of the outcome. Standard errors are clustered by examiners. *p < 0.1, **p < 0.05, ***p < 0.01.
Panel B of Table 3.5 shows the results for patent purchase by practicing firms, which stand in sharp contrast with the patterns for PAEs. First, the effects are all much smaller in magnitude than in Panel A. In the first seven columns of the table, the effects are almost all insignificant and are never larger than 2%. Second, the relationship with examiner leniency does not appear to be robust: it switches signs across proxies or specifications. For instance, in the univariate regression in Column (1) we obtain a precisely estimated zero for the correlation with the change in the number of words per claim. But the regression coefficient becomes positive in specification (9), suggesting that practicing firms may have a preference for less lenient examiners, although the coefficient is relatively small (3.49%). Overall, there appears to be no quantitatively large or statistically robust relationship between purchases by practicing firms and examiner leniency.

The fact that only PAEs selectively purchase patents issued by lenient examiners is not consistent with the view that PAEs solve a generic friction in the patent market. If PAEs were primarily lowering transaction costs or solving liquidity problems, there would be no reason for them to selectively purchase patents from lenient examiners, which in contrast do not affect patent acquisitions by practicing firms. To examine whether PAEs may rather be addressing a patent-specific friction related to the patent examination process itself, we now investigate the correlates of patent litigation.

Table 3.6 presents the results with patent litigation as the outcome. Panel A reports the results for patent litigation by PAEs. The patterns are similar to those found in Table 3.5 for PAEs, except that the magnitudes are much larger. Column (1) shows that a one standard deviation increase in the examiner effect for the change in the number of words per claim implies a 40.5% increase in the rate of litigation by PAEs. This effect is very large in itself but also relative of the overall examiner effects documented in Table 3.2, according to which the signal standard deviation of examiner effects for PAE litigation is 46% (although it is imprecisely estimated). This result suggests that a simple proxy for examiner leniency can account for most of the relationship between examiner effects and PAE litigation. Moreover, the other columns of Table 3.5 indicate that this pattern is very robust. The other general
proxies for examiner leniency, the change in the number of claims and the grant rate, go in the same direction and are larger in magnitude than when considering patent purchases. Considering the use of the various sections of patent law, as for patent purchase by PAEs the section relating to the obviousness of the invention is the most important, but the magnitude of the effect is now substantially larger. In the multivariate regression including all examiner effects simultaneously in Column (9), the patterns still point to the role of leniency as the predictive power loads on the grant rate, with a coefficient indicating that a one standard deviation increase in the grant rate implies an increase in the rate of PAE litigation close to 50%.
### Table 3.6: Patent Litigation and Examiner Behavior

#### Panel A: Patent Litigation by PAEs

<table>
<thead>
<tr>
<th>Leave-one-out Examiner Effects</th>
<th>Patent Litigation by PAE</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Change in Words per Claim</td>
<td>-0.405*** (0.083)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Change in Number of Claims</td>
<td>0.127*** (0.067)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grant Rate</td>
<td>0.567*** (0.099)</td>
<td>0.48*** (0.14)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of Section 101</td>
<td>-0.105 (0.077)</td>
<td>-0.09 (0.08)</td>
<td>0.05 (0.08)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Lack of utility or eligibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of Section 102(a)</td>
<td>0.0178 (0.089)</td>
<td>0.019 (0.08)</td>
<td>0.023 (0.082)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Prior art exists</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of Section 103(a)</td>
<td>-0.156*** (0.075)</td>
<td>-0.176** (0.083)</td>
<td>-0.039 (0.08)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Obvious invention</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of Section 112(b)</td>
<td>-0.0003 (0.079)</td>
<td>0.102 (0.08)</td>
<td>0.085 (0.086)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Vague claims</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** The sample is restricted to IT patents. Regressors are standardized by their standard deviations and coefficients are expressed as a fraction of the mean of the outcome. Standard errors are clustered by examiners. *p < 0.1, **p < 0.05, ***p < 0.01.

#### Panel B: Patent Litigation by Practicing Firms

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>% Change in Number of Word per Claim</td>
<td>-0.138*** (0.043)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Change in Number of Claims</td>
<td>0.022 (0.031)</td>
<td>0.015 (0.034)</td>
<td>0.026 (0.034)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grant Rate</td>
<td>0.24*** (0.045)</td>
<td>0.23*** (0.067)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of Section 101</td>
<td>-0.068* (0.037)</td>
<td>-0.0205 (0.0397)</td>
<td>0.005 (0.04)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Lack of utility or eligibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of Section 102(a)</td>
<td>-0.008 (0.04)</td>
<td>0.0150 (0.0406)</td>
<td>0.026 (0.04)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Prior art exists</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of Section 103(a)</td>
<td>-0.073** (0.034)</td>
<td>-0.0387 (0.0370)</td>
<td>0.008 (0.04)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Obvious invention</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of Section 112(b)</td>
<td>-0.118*** (0.032)</td>
<td>-0.0978** (0.0366)</td>
<td>-0.065 (0.04)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Vague claims</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Fixed Effects**

| N | 274,464 274,537 311,615 311,470 311,470 311,470 311,470 311,470 274,464 |

**Year by Art Unit**
Panel B of Table 3.6 reports the results for patent litigation by practicing firms, which are qualitatively similar to the patterns for PAEs but are smaller in magnitude. Across all proxies and specifications in this panel, we consistently find that lenient patent examiners — who increase the number of words per claim by less, have a higher grant rate and reference patent law less often — issue patents with a higher litigation risk. The magnitude of the effects is less strong than for litigation by PAEs but is comparable to the magnitude of the effects for purchases by PAEs (Panel A of Table 3.5). For instance, a one standard deviation fall in the examiner effect for the change in the number of words per claim implies a 13.8% increase in the rate of litigation and a 13.9% increase in the rate of PAE purchase.

The finding that patent litigation by both practicing firms and PAEs is driven by examiner leniency challenges the view that PAEs engage in idiosyncratic frivolous lawsuits. The merit of the lawsuits involving patents issued by lenient patent examiners may be questionable, but PAEs are not the only entities to selectively assert patents from lenient examiners: practicing firms do so as well. PAEs purchase patents that are different from those handled by practicing firms in the market for patents (Table 3.5) but their propensity to assert patents issued by lenient examiners is merely a more extreme version of the litigation behavior of practicing firms (Table 3.6).

The patterns in the data are therefore difficult to reconcile with the mainstream views of PAEs, either as intermediaries solving a generic friction in the patent market or as perpetrators of frivolous lawsuits. Rather, it appears that much of the activities of PAEs is driven by a specific friction in the patent market, which is caused by the way examiner craft patent rights and which strongly correlates with examiner leniency. Our findings are therefore in line with a nuanced view of PAEs, suggesting that PAEs’ activities are the symptom of features of the patent system that affect litigation more generally (e.g., Lemley and Melamed (2013) and Schwartz and Kesan (2013)). PAEs behave as litigation experts and much of their activities stem from the way patent rights are crafted by lenient examiners, who affect litigation more generally. Although we can only document correlations with examiner traits, we emphasize that the underlying causal examiner effects are quantitatively
large and should therefore be accounted for by any convincing theory of PAEs’ activities.\textsuperscript{44}

### 3.4.3 PAEs and Patent Invalidity

In this subsection we study whether lenient examiners, who play an important role for litigation in general and for litigation by PAEs in particular, tend to issue patents that are more likely to be invalid. Various observers (e.g., Federal Trade Commission (2016)) have hypothesized that PAEs may be asserting patents that are “invalid”, in the sense that these patents should not have been issued in the first place because they do not comply with the standards set by U.S. patent law. Given the evidence that patent litigation by PAEs is very strongly correlated with examiner leniency, we can re-cast this question in terms of examiner effects: do lenient examiners tend to issue patent that are more likely to be invalid? Approaching this question in terms of examiner effects has the potential to be informative about PAEs but also about patent litigation by practicing firms, since they also selectively assert patents that were crafted by lenient examiners.

**Proxies for Patent Invalidity.** Patent invalidity is notoriously difficult to measure because of selection effects (e.g., Miller (2013)). To assess whether a robust relationship exists between examiner leniency and patent invalidity, we rely on three complementary proxies for patent invalidity. We consider two restricted samples to study two common proxies for patent invalidity, which are subject to substantial sample selection but are standard in the literature. We also introduce a third proxy available in the full sample of granted patents.

First, for a small number of cases, patent litigation does not result in a settlement and a court trial closes the case (see Allison \textit{et al.} (2013) for a review). We obtain this data from Lex Machina. The sample of cases for which trial outcomes are available is very selected: in our main analysis sample, there are only 516 cases with information on whether the court deemed the patent invalid or found an infringement.

\textsuperscript{44}Of course, even though the causal examiner effects from Table 3.2 are large, they do not account for the entirety of PAEs’ patent acquisition and assertion behaviors. We only speak to the (substantial) part of PAEs’ activities which is caused by examiner effects and point out that the two standard views of PAEs cannot account for these patterns.
The second common proxy for patent invalidity is a procedure for challenging the validity of a patent at the USPTO, known as “inter partes review” (IPR). IPRs were introduced in 2012 as a defensive tool for those seeking to defeat meritless infringement claims (see Chien and Helmers (2015) for a review). The procedure can be initiated by any party other than the patent owner and requires the patent office to review the validity of the patent based on specific sections of patent law. This sample is also very selected: there are 989 IPR cases in our main analysis sample.

Third, we use patent re-issuance requests as another proxy for patent invalidity. A re-issue application can be filed by the applicant “whenever any patent is, through error, deemed wholly or partly inoperative or invalid”. We obtain this information from the continuation data in the Patent Examination Dataset. Re-issue applications are a useful metric for our purposes as they are available for all granted patents and provide a direct measure of examiner mistakes from the perspective of the patent applicant.

Table 3.7 reports summary statistics on our proxies for patent invalidity. Court rulings are observed for only 516 patents, or about 0.0004% of our sample. Conditional on observing a court ruling, the rate of invalidity is close to 19%. In 31.9% of cases, the court declares that the patent is infringed, which indirectly attests to its validity. The panel also indicates that an IPR procedure is filed for 0.0003% of patents. Conditional on filing, 78.5% of IPRs are “instituted”, meaning that the patent office deems it likely that the patent is at least in part invalid. Because the “institution” rate of IPRs is very high, close to 80%, either the occurrence of an IPR or the institution of an IPR can be used as proxies for patent invalidity.

For both court rulings and IPRs, the invalidity rates appear to be high, but they are observed

---

45 Patent law states that “Whenever any patent is, through error, deemed wholly or partly inoperative or invalid, by reason of a defective specification or drawing, or by reason of the patentee claiming more or less than he had a right to claim in the patent, the Director shall, on the surrender of such patent and the payment of the fee required by law, reissue the patent for the invention disclosed in the original patent, and in accordance with a new and amended application, for the unexpired part of the term of the original patent.” (35 USC 251(a)). Re-issue applications can petition for an increase in the scope of claims only if they are filed within two years from grant of the original patent (35 USC 251(d)). We repeat our analysis considering only re-issues applications beyond this threshold to establish that attempts to increase claim scope are not driving the patterns.

46 According to patent law, “An inter partes review may be instituted upon a showing that there is a reasonable likelihood that the petitioner would prevail with respect to at least one claim challenged” (35 USC Ch. 31, §311 - §319).
conditional on a stringent form of sample selection.

Table 3.7: Summary Statistics on Proxies for Patent Invalidity

<table>
<thead>
<tr>
<th>Proxy</th>
<th>Mean</th>
<th>Median</th>
<th>S.D.</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invalidity decision by court, conditional on court ruling</td>
<td>0.1880</td>
<td>0</td>
<td>0.3911</td>
<td>516</td>
</tr>
<tr>
<td>Infringement decision by court, conditional on court ruling</td>
<td>0.3198</td>
<td>0</td>
<td>0.4668</td>
<td>516</td>
</tr>
<tr>
<td>IPR filing</td>
<td>0.0003</td>
<td>0</td>
<td>0.0164</td>
<td>1,833,464</td>
</tr>
<tr>
<td>IPR institution, conditional on IPR filing</td>
<td>0.7858</td>
<td>1</td>
<td>0.4105</td>
<td>719</td>
</tr>
<tr>
<td>Re-issuance</td>
<td>0.0020</td>
<td>0</td>
<td>0.0458</td>
<td>1,833,464</td>
</tr>
<tr>
<td>Re-issuance &gt;2 years after grant</td>
<td>0.0004</td>
<td>0</td>
<td>0.0206</td>
<td>1,833,464</td>
</tr>
</tbody>
</table>

Notes: This table reports summary statistics on several proxies for patent invalidity. See Section 3.4.3 for variable definitions and Section 3.2.1 for information on the sample.

Finally, Table 3.7 shows that re-issue applications are submitted for about 0.002% of patents. According to patent law, a re-issue application indicates that the applicant believes that the patent is wholly or in part invalid because of a mistake in the document. To address the potential concern that some applicants may violate patent law and strategically exploit re-issue applications to obtain greater scope, instead of correcting a mistake, we consider re-issue applications that are submitted more than two years after grant. After the two-year delay, re-issue applications cannot petition for an increase in scope; they account for about 0.0004% of all granted patents. This fraction is very small but it is comparable in magnitude to the number of observations for court rulings and IPRs and has the advantage of being available for the full sample of granted patents.

Results. We run specification (3.8) with our patent invalidity proxies as outcomes. The regressors are examiner effects for the change in the number of words per claim and the grant rate, which were the most powerful univariate predictors of patent acquisition and assertion by PAEs in Tables 3.5 and 3.6. We also consider the best linear predictor for patent purchase by PAEs using the specification in Column (9) of Table 3.5. The results are reported in Table 3.8.
Table 3.8: Examiner Behavior and Likelihood of Patent Invalidity

Panel A: Reissuance of Granted Patents

<table>
<thead>
<tr>
<th>Examiner Effects</th>
<th>Reissuance Rate</th>
<th>Reissuance Rate Two Years or More after Grant</th>
</tr>
</thead>
<tbody>
<tr>
<td>(separate regressions)</td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td></td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td></td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td><strong>(A)</strong> % Change in Words per Claim</td>
<td>-0.26***</td>
<td>-0.24***</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.06)</td>
</tr>
<tr>
<td></td>
<td>-0.25***</td>
<td>-0.55***</td>
</tr>
<tr>
<td></td>
<td>(0.068)</td>
<td>(0.15)</td>
</tr>
<tr>
<td><strong>(B)</strong> Grant Rate</td>
<td>0.29***</td>
<td>0.27***</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.061)</td>
</tr>
<tr>
<td></td>
<td>0.28***</td>
<td>0.54***</td>
</tr>
<tr>
<td></td>
<td>(0.061)</td>
<td>(0.13)</td>
</tr>
<tr>
<td><strong>(C)</strong> Linear Predictor for PAE Acquisition</td>
<td>0.139***</td>
<td>0.136***</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td>(0.035)</td>
</tr>
<tr>
<td></td>
<td>0.142***</td>
<td>0.24***</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.13)</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Year</td>
<td>Year by Year by Year by Year by Year by Year by Art</td>
</tr>
<tr>
<td></td>
<td>Art Unit</td>
<td>Art Unit by Class Art Unit by Class Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Class</td>
</tr>
<tr>
<td>N</td>
<td>274,464</td>
<td>273,839</td>
</tr>
</tbody>
</table>

Panel B: Trials at the Patent Office (“Inter Partes Reviews”)

<table>
<thead>
<tr>
<th>Examiner Effects</th>
<th>IPR Rate</th>
<th>Institution Rate of IPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>(separate regressions)</td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td></td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td></td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td><strong>(A)</strong> % Change in Words per Claim</td>
<td>-0.38***</td>
<td>-0.43***</td>
</tr>
<tr>
<td></td>
<td>(0.098)</td>
<td>(0.094)</td>
</tr>
<tr>
<td></td>
<td>-0.42***</td>
<td>(0.087)</td>
</tr>
<tr>
<td></td>
<td>(0.057)</td>
<td>(0.27)</td>
</tr>
<tr>
<td></td>
<td>-0.23</td>
<td>(0.29)</td>
</tr>
<tr>
<td><strong>(B)</strong> Grant Rate</td>
<td>0.41***</td>
<td>0.44***</td>
</tr>
<tr>
<td></td>
<td>(0.085)</td>
<td>(0.081)</td>
</tr>
<tr>
<td></td>
<td>0.44***</td>
<td>(0.082)</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.11)</td>
</tr>
<tr>
<td></td>
<td>0.034</td>
<td>(0.13)</td>
</tr>
<tr>
<td><strong>(C)</strong> Linear Predictor for PAE Acquisition</td>
<td>0.28***</td>
<td>0.34***</td>
</tr>
<tr>
<td></td>
<td>(0.088)</td>
<td>(0.086)</td>
</tr>
<tr>
<td></td>
<td>0.33***</td>
<td>(0.077)</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.21)</td>
</tr>
<tr>
<td></td>
<td>0.21</td>
<td>(0.26)</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td>Year</td>
<td>Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Year by Class</td>
</tr>
<tr>
<td>N</td>
<td>274,537</td>
<td>180</td>
</tr>
</tbody>
</table>

Notes: The sample is restricted to IT patents. Regressors are standardized by their standard deviations and coefficients are expressed as a fraction of the mean of the outcome. The linear predictor for PAE acquisition is given by specification (9) in Table 3.5. Examiner effects are computed leaving out the focal patent. Standard errors are clustered by examiners. *p < 0.1, ** p < 0.05, *** p < 0.01.
We find a very strong and robust relationship between examiner leniency and our preferred proxy for patent invalidity, the reissuance of granted patents. Panel A of Table 3.8 reports this finding. The various rows of this panel correspond to separate univariate regressions. The first row of Column (1) indicates that, conditional on year fixed effects, a one standard deviation increase in the examiner effect for the change in the number of words per claim (i.e. less leniency) leads to a 26% decline in the probability of reissuance. Columns (2) and (3) show that the coefficient is very stable as art unit by year fixed effects and art unit by year by technology class fixed effects are introduced. Similarly strong and robust patterns are documented in the other rows of the tables for the grant rate and the linear predictor for PAE acquisition. Column (4) to (6) show that the patterns are even stronger when we consider the reissuance rate two years or more after grant, the delay beyond which a reissuance request cannot petition for an increase in the scope of the claims. For instance, the coefficient for the change in the number of words per claim hovers between 55% and 61% across specifications. Since PAEs selectively assert patent granted by lenient examiner (more so than practicing firms), they are more likely to assert patents that are likely to contain mistakes, as reflected by the reissuance rates.

Panel B of Table 3.8 uses IPR occurrence and IPR institution as proxies for patent invalidity from the perspective of the Patent Office. Columns (1) to (3) of Panel C of Table 3.8 document that examiner leniency is a very strong predictor of the occurrence of an IPR. For instance, the first row of Column (2) indicates that a one standard deviation increase in examiner effects for the change in the number of words per claims (lower leniency) implies a 41% fall in the probability of an IPR. The regression coefficients are all large and very stable across specifications that include different sets of fixed effects. In contrast, Columns (4) to (6) do not deliver conclusive results regarding IPR institution, because the selected sample of patents that go through an IPR is too small to provide adequate power.

In sum, Table 3.8 indicates that, when using suitable proxies for patent invalidity that do not suffer from small sample issues, there is strong and robust evidence that lenient examiners issue patents that are more likely to be invalid. These examiners account for a
disproportionate share of patent litigation, in particular by PAEs. This finding indicates that examiner behavior during patent prosecution is a quantitatively important determinant of patent invalidity, suggesting that PAEs specialize in purchasing and asserting patent that should not have been issued as such in light of the standards set by current patent law.\textsuperscript{47}

3.4.4 Robustness Checks and Additional Results

In the final part of this section, we discuss the robustness of our PAE results across samples, specifications, and PAE types. In addition, we use data on patent value, auction prices, and European Patent Office decisions to shed further light on PAE behavior.

Robustness across samples and specifications. Table 3.9 documents the robustness of the signal standard deviations of examiner effects for PAE purchases across alternative specifications and subsamples. Row (A) reports the baseline estimate in our main analysis sample, as in Table 3.2. Row (B) shows that the signal standard deviation remains similar when introducing assignee fixed effects in equation (3.1): PAEs selectively purchase patents coming from specific examiners even within the portfolio of a given assignee. Row (C) to (E) show that the signal standard deviation is very similar across PAE lists. Row (C) reports similar estimates when excluding from the sample the patents purchased by the largest PAE, Intellectual Ventures. Conversely, Row (D) shows that the results are comparable when considering only patents purchased by Intellectual Ventures.\textsuperscript{48} The estimates also remain stable when using the list of PAEs defined by Cotropia et al. (2014), as shown in Row (E).

\textsuperscript{47}This finding does not speak conclusively to the welfare effects of PAEs, because litigation of patents issued by lenient examiners could conceivably be socially valuable, even when these patents are deemed invalid by current patent law. The standards set by current patent law may not be socially optimal and are dynamically evolving. For instance, Galetovic et al. (2015) point out that the process of litigation helps defines the contours of patent law in highly-innovative, rapidly changing industries.

\textsuperscript{48}The estimates reported in rows (C) and (D) do not average out to the estimate in (A), implying that there is not as much covariance between the two outcomes (purchase by Intellectual Ventures and purchase by a PAE other than Intellectual Ventures) as there is within outcomes. This result indicates that there is some segmentation of the market between PAEs, and that examiners effects are strong everywhere.
Table 3.9: Heterogeneity in Examiner Causal Effects on Patent Acquisition by PAEs

<table>
<thead>
<tr>
<th></th>
<th>Signal SD, % of Average</th>
<th>Average Purchase Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Baseline</td>
<td>63.0%</td>
<td>1.02%</td>
</tr>
<tr>
<td>(B) Including Assignee Fixed Effects</td>
<td>44.5%</td>
<td>1.02%</td>
</tr>
<tr>
<td>(C) Excluding Intellectual Ventures</td>
<td>82.7%</td>
<td>0.55%</td>
</tr>
<tr>
<td>(D) Intellectual Ventures Only</td>
<td>82.0%</td>
<td>0.47%</td>
</tr>
<tr>
<td>(E) PAE list from Cotropia et al. (2014)</td>
<td>67.8%</td>
<td>0.60%</td>
</tr>
<tr>
<td>(F) Small PAEs</td>
<td>40.7%</td>
<td>0.07%</td>
</tr>
<tr>
<td>(G) PAEs purchasing from small entities/unassigned patents</td>
<td>70.8%</td>
<td>0.11%</td>
</tr>
</tbody>
</table>

Notes: This table reports the signal standard deviation of examiner effects using different specifications and different PAE outcomes, using the full sample. The Bayesian shrinkage methodology used to obtain these estimates is presented in Section 3.3.1. Row (A) reports the baseline estimate from Table 3.2. In row (B), the specification includes assignee fixed effects. Row (C) uses purchase by a PAE other than Intellectual Ventures as the outcome. Row (D) considers purchases by Intellectual Ventures as the outcome. Row (E) uses the PAEs list from Cotropia et al. (2014). Row (F) examines purchases by PAEs with a small patent portfolio, as defined by Cotropia et al. (2014). Row (G) considers purchases by PAEs whose portfolios have more than 50% of patents that either were unassigned (i.e., the inventor is the owner) or that were assigned to a firm that the USPTO classifies as a small entity.

Results by PAE type. Existing research has hypothesized that large and small PAEs may behave differently (Cotropia et al. (2014)). Row (F) of Table 3.9 shows that PAEs with a small portfolio of patents, as defined by Cotropia et al. (2014), are as responsive to examiner effects as the average PAE. Another plausible hypothesis is that PAEs that primarily work with small firms or individual inventors may have a different behavior with respect to examiners, for instance because they may be focused on addressing frictions that specifically affect these firms and inventors. Row (G) considers a subset PAEs which bought over 50% of their patents from small entities.\footnote{We define patents from small entities as patents that either were unassigned (i.e., the inventor is the owner) or that were assigned to a firm that the USPTO classifies as a “small entity” (if there is an assignee, each patent reports whether it was initially assigned to a small entity, i.e. a small firm). On average, PAEs purchase only 15.7% of their patents from small firms (19% when excluding continuation applications). Likewise, the share of unassigned patents in PAEs’ purchases is low, ranging from 6.2% when including continuations to 10.7% without continuations. These low shares are difficult to reconcile with the view that the typical PAEs is addressing frictions that specifically affect small firms or individual inventors.} In this subsample as well, the signal standard deviation is very similar to the baseline.
Furthermore, Figure 3.5 investigates whether different types of PAEs react differently to examiner leniency. Using specification (3.8), this figure reports the correlation between PAE purchase rates and the main proxy for examiner leniency from Section 3.4.2, the change in the number of words per claim between application and grant. We consider in turn Intellectual Ventures, all PAEs but Intellectual Ventures, small PAEs, and PAEs which purchased over 50% of their patents from small entities. We find that they all selectively purchase patents from more lenient examiners, with relationships of very similar magnitudes across PAE groups. The leniency-bias of PAEs is therefore a very stable feature.

![Figure 3.5: Heterogeneity in Patent Acquisition Behavior across Groups of PAEs](image)

**Notes:** The sample is restricted to IT patents. The regression coefficients indicate the percentage change in the probability of PAE acquisition (relative to the baseline rate) for a one standard deviation increase in the examiner effect for the change in the number of words per claim during prosecution. The methodology is described in Section 3.4.2 (see specification (3.8)). Regression coefficients are reported separately for four samples of PAEs: Intellectual Ventures, PAEs other than Intellectual Ventures, PAEs with a small patent portfolio according to the classification of Cotropia et al. (2014), and PAEs which primarily buy patents from small entities (specifically, as described in the main text they purchase more than half of their patents from small firms or individual inventors). The 95% confidence intervals are based on standard errors clustered by examiners.

50 The results are similar with other proxies for examiner leniency, such as the grant rate, as well as when considering the full sample of patents instead of IT patents only (not reported).
Additional results. Although we have documented that examiners with high PAE effects are lenient, we also find that these examiners do not create greater private value for patent holders, suggesting a distinction between breadth and vagueness of claims. Online Appendix Table C.8 shows a small and statistically insignificant relationship between examiner PAE effect and examiner private value effect, as measured by stock market response. This result suggests that these examiners are not simply granting patents with greater scope, which should create higher private value. An alternative explanation is that the patents they grant contain less well-defined or vaguer language, which is consistent with the negative relationship between Section 112(b) blocking action usage and non-PAE litigation shown in Panel B of Table C.6.51

Finally, the Online Appendix reports additional results shedding light on the mechanisms leading PAEs to selectively purchase patents from more lenient examiners. First, in a subsample of patents for which auction prices are available, we find that patents issued by more lenient examiners tend to sell at a lower price (Online Appendix Figure C.3). Second, considering all patents that were jointly filed at the USPTO and at the European Patent Office (EPO), we find that PAEs are much more likely to purchase patents that were rejected by the EPO (panel A of Online Appendix Table C.9). Interpreting EPO grant decisions as a measure of a patent’s inventive step size (Picard and Van Pottelsberghe de la Potterie (2011)), this result suggests that PAEs target patents that bear on more incremental, less innovative technology. Furthermore, we find that PAEs selectively purchase patents that were rejected by the EPO only when these patents were issued by specific examiners, with a large causal impact on PAE purchases (panel B of Online Appendix Table C.9). It is therefore plausible that these patents are particularly productive for litigation, as they are closer to existing intellectual property than average (given the small step size revealed by EPO rejections) and their claims may be less well-defined and harder to interpret than average (given the examiners who granted them).

51Section 112(b) is typically used to clarify indefinite claims language. Under a simple model, there would only be litigation in equilibrium if there is disagreement between parties, which would not happen if claims were broad but clear.
3.5 Conclusion

In this paper, we have shown that significant heterogeneity in patent outcomes results from the process of creation of patent rights and is independent of scientific determinants. We established this result by using the allocation of patent applications to examiners as a source of quasi-random variation in patent rights. To address identification concerns, we accounted for potential examiner specialization within detailed technology categories by developing new sources of quasi-experimental variation, based on assignment mechanisms at the patent office related to patent application serial numbers and examiner busyness. These techniques could be used to investigate issues related to the crafting of patent rights in future research.

We have also shown that the process of creation of patent rights is of first-order importance to understand a central and much-debated feature of the U.S. innovation system, the activities of PAEs. We found that PAEs selectively purchase and litigate patents issued by “lenient” examiners; these examiners tend to issue patents that are more likely to be litigated, but not purchased, by practicing firms. These patterns are quantitatively important and cannot be accounted for by standard PAE theories, which describe PAEs either as intermediaries solving a generic friction in the patent market (such as transaction costs and illiquidity) or as perpetrators of frivolous lawsuits. Instead, we found that the activities of PAEs are best characterized as a response to a specific friction in the patent system, which is caused by the way lenient examiners craft patent rights and which affects litigation more broadly. These findings imply that policies affecting the behaviors of patent examiners, and specifically of lenient examiners, have the potential to greatly affect PAEs and litigation. In contrast, the current policy debate has focused on a possible reform of patent law to reduce PAEs’ activities and litigation, which observers have noted may be difficult to implement (Schwartz and Kesan (2013)).

More broadly, our results call for a greater focus on understanding the impact of the crafting of patent rights on innovation dynamics. This paper provided a set of tools to conduct such investigation and showed the explanatory power and potential policy relevance of this line of inquiry in the context of the debate over PAEs.
References


Appendix A

Appendix to Chapter 1

A.1 Suggestive Evidence of History-Dependent Demand Based on Firm Behavior

As alluded to in the background section, several types of firm behavior are also suggestive of history-dependent demand. These include entry decisions surrounding me-too and incremental drugs and the amount firms pay to shorten review time by a few months. The observations also suggest that history-dependent demand may have important effects beyond pricing impacts.

Previous studies have shown that the entry of me-too and incremental drugs are time-dependent. Me-too drugs are branded drugs that have similar properties to existing drugs on the market, and tend to be introduced by a non-incumbent firm, whereas incremental drugs are tweaks to an existing incumbent drug introduced by the incumbent firm. For me-too drug entry, Gilchrist (2016) finds potentially causal evidence that longer patent terms on first-in-class drugs equate to more me-too entry.\(^1\) He argues this is due to the threat of having to compete with generics, an effect that could be amplified by any frictions in switching from a generic to a newly introduced branded drug.

Timing also appears to matter for incremental drug entry. Huckfeldt and Knittel (2011)

\(^1\)He uses delays between initial patenting and first clinical trial as a source of variation.
document that they tend to be introduced and heavily advertised in the two years before patent expiration. These occur at around the same time as the price increases documented by Caves et al. (1991), consistent with the practice of “soft switches” discussed in industry circles. A soft switch is when a company introduces an incremental version of a basic drug with little patent term left and raise prices on the basic drug in hopes of establishing users on the incremental drug, which usually has longer protection remaining. A more extreme strategy is the hard switch, where a company pulls the basic drug altogether, as was the case with Namenda, an Alzheimer’s treatment. Both strategies are valuable, according to CEOs and analysts, because very few people move from the incremental drug onto the generic version of the basic drug, consistent with the findings in this paper. Recently, the FTC criticized these tactics as anti-competitive in an amicus brief on the case Mylan vs. Warner Chilcott.

A final piece of observational evidence is the size of first-mover advantage in the industry, evident in the sale value of FDA priority review vouchers. Ridley and Régnier (2016) discuss the sizable value of FDA priority review vouchers, which were introduced by Congress in 2007 as a reward for firms that commercialize treatments for rare diseases. Each voucher gives firms the right to convert a standard FDA review, a ten-month process mostly for me-too drugs, into a priority review, a six-month process usually reserved for drugs deemed to be major improvements. These vouchers are transferable, and five have been sold at prices ranging from $67.5 million to $350 million. For some context, one of the best-selling drugs of all time, Lipitor, made an estimated $583 million in its entire first year, probably around $65 million in its first four month, given its sales trajectory. Therefore, the amounts paid for these vouchers suggest that there must be other factors beyond just the mechanical four months worth of sales and the discounted value of extra sales from the extended patent term. One key factor could be demand-driven first-mover advantage.  

---


3Ridley and Régnier (2016) offer a simple estimate suggesting that a third to a half of the voucher value comes from competitive considerations.
A.2 Additional Data Background

MarketScan User Panels - Unbalanced and Balanced

Given the significant entry and attrition of users in the MarketScan data, I work with both an unbalanced and a balanced panel of users in my analysis. Each has its advantages, which I factor in to my later analyses.

The unbalanced panel encompasses all users in the Truven dataset, which yields great statistical power but raises issues surrounding user history and weighting. Using the raw data is the simplest approach and provides the greatest statistical power for my analysis, especially in the 2000s, when the dataset covers a sizable fraction of the US population. However, the rapid growth in people covered creates issues in terms of weighting and composition. For example, if we compute average copays or prices for all users in the dataset, changes could just be driven by the nature of entering users rather than changes in price facing existing users. Similarly, when we pool all entry events to study history-dependence, events later on in the period will have a higher weight.

To alleviate some of these problems, I also construct a smaller balanced panel from 2003-2012. I choose the period based on the 10-year period that gives us the most number of users, which turns out to be around 2.5 million. Starting any earlier is problematic, because most users from the 1990s have exited the data by the end of the coverage period. Of course, a balanced panel selects on specific types of employers that consistently use Truven services and on employees who tend to stay at the same firm over time.

A.3 Robustness - Methodology, Other Events, and Other Chronic Drug Markets

Next, I turn to the issue of the robustness of the results. My estimates are not particularly sensitive to the window I use and the simplification to quarterly data. In addition, the same inertial patterns are present around negative clinical trial announcements and in other
chronic drug markets, suggesting that inertia is an important force in many drug-related settings.

One concern is the robustness of the results to the window around the launch events, which I address by using a narrower window to select treatment and control users. As mentioned earlier, the main motivation for selecting users starting at a similar time is to avoid picking up on any changes over time in diagnosis guidelines that may shift the composition of new users. In Table A.1, I report the same coefficients when selecting users based on 60-day windows around the launch date. The coefficients are again noisier, as our sample is cut by a third, but generally similar to the ones using 180-day windows. The increase in coefficients after period 9 is influenced by the composition of drug entry events and their associated user attrition rates.\textsuperscript{4}

I also find little evidence that existing patients try new drugs once and switch back to their old drug, a behavior that I might be attributing to inertia if quarterly aggregates obscure fast switches. Under a learning model, it’s possible that the control group may try the new drug at the same rate as the treatment group, but then switch back to what they started on. I verify that this is not the case by re-running the analysis with the outcome “prescribed new drug in current or previous period,” which can also be thought of as an “ever tried” variable. To make sure that my quarterly simplification does not miss users who try the new drug once, I generate the variable using raw claims data. This ends up not being a large issue, as it only affects the data in 0.3% of all user-quarters, and, unsurprisingly, the estimates are almost identical to those in Table 1.3.

Next, I turn to other types of variation in starting conditions beyond drug availability, focusing on starting cohorts around the announcement of influential clinical trials. In the anti-cholesterol market, drug companies routinely run post-approval clinical trials to ascertain long-term health outcomes, particularly in chronic disease areas. Some clinical trials are influential in changing beliefs about drug quality, and if inertia is present, one

\textsuperscript{4}For example, there was very high attrition in 2005, so Crestor and Vytorin would dominate the early coefficients, but many users for that event drop out, leaving other events to influence the coefficient.
might expect some differences in cohorts who start before or after a big announcement. I test this by looking at the announcement of the results of the ENHANCE trial in January 2008, which found that the combination drug Vytorin did not reduce heart attacks more than statins like Zocor, which had gone generic in 2006. Prior to the study, Vytorin was thought to be more effective, but the results suggested that they were a much less cost-effective option. Figure A.7 shows the market share of Vytorin in cohorts starting treatment around the announcement of the trial. Consistent with the core results, the “before” cohort has a steadily higher Vytorin market share relative to the “after.”

Finally, Figure A.6 shows that these patterns hold to different degrees in other major chronic disease areas. The graphs report results from applying my methodology to other areas, and suggest that history-dependent demand may also play a significant role in the multiple sclerosis, diabetes, and chronic obstructive pulmonary disease (COPD) markets. The main caveat here is that these markets are more medically complex, which brings into question my identification assumption. Some drugs may target patients for whom the existing therapies are inappropriate, so many patients may wait for the new drug to launch before starting, violating my identification assumption.

A.3.1 Heterogeneity by User Characteristics

Next, I turn to documenting heterogeneity in history-dependence across user characteristics such as age, gender, and observable health characteristics, which will be important for demand modeling and gaining additional insight into microfoundations, given previous findings about demographic predictors of risk aversion and adoption.

My approach is exactly the same as the instrumental variables one above, except now I add in interactions. What I’m doing in effect is to separate out the same analysis for different groups of patients. For example, we would be looking at difference in female market share after 4 years, scaled by the initial difference.

The first set of characteristics I look at are purely descriptive, based on user demographics

---

5https://www.wsj.com/articles/SB120000578748383895929
including age and gender. I find small negative effects for both age and gender in terms of history-dependence at all points in time. In other words, women and older individuals are more likely to switch than their counterparts. This may be surprising given other results about the correlation of age and gender with risk aversion and adoption in the literature. However, it is possible that these characteristics are correlated with unobserved health state or experience with generics.

The second set of characteristics are based on health at the time of first prescription, which can lend insight into the micro-foundations. I find coefficients suggesting that healthier individuals display much lower levels of history-dependence. For each additional drug a patient is on at the time they start anti-cholesterol medication, the causal effect up by 10 percentage points. Similarly, patients starting treatment during an inpatient episode exhibit about a 30 percentage point increase in causal effect.

Finally, I put all of these factors together in a horse race regression in Table A.2, finding that the health characteristics appear to be the main drivers, and that the female and age coefficients are now very weakly positive, more in line with expectations given a risk aversion-based story of history-dependence.

A.4 Using the Framework to Assess Long-Term Health and Spending Impacts of New Drugs

An interesting byproduct of my empirical strategy is a way to estimate the long-term treatment effects of a given drug, relative to existing treatments. I discuss the approach, including the assumptions and conditions needed for it to work, and offer a proof of concept.

The approach is to use the same treatment and control groups as in my main methodology, but to compare average health outcomes instead of usage across the two groups. The intuition behind the approach is that if users are randomly assigned to a drug, and exhibit a high degree of inertia, any health outcome differences should be attributable to long-term usage of different drugs. One can use the same instrumental variables framework as in
Equation (1.1), but alter it to be

\[ H_{it} = \alpha + \gamma S_{it} + \epsilon_{it} \]  

(A.1)

where \( H \) represents health outcomes including hospitalization, health care spending, and common side-effects. \( S \) can be some indicator of usage of the new drug between period 0 and period \( t \), such as continual usage or quarters of use.\(^6\) Once again, we would instrument for \( S_{it} \) using the before or after treatment variable \( Z_i \).

The interpretation of the coefficient would be the effect of taking the new medication over the existing set of medications. For example, if one were analyzing the entry of Crestor, \( \gamma \) would measure the health differences generated by Crestor relative to the existing standards of treatment. This is again a treatment-on-the-treated estimate, normalizing the average difference by the rate of adoption. Another issue here is that there could be local average treatment effects at play, as the complier group may be especially well-matched to the entering medication. In this case, the coefficient should be interpreted as the value of the new treatment for the complier population.

As a proof of concept, I analyze the impact of new drugs on hospitalizations using this framework, and find very little difference in outcomes. Hospitalizations are a key health outcome in many settings, as reducing occurrences can lead to large savings.\(^7\) In Figure A.8, I plot hospitalization rates for the treatment and control groups associated with Crestor entry, and find very little long-term differences between the two groups.\(^8\) Similar analysis of other drug entries in this period yield similar results.\(^9\) This results is not too surprising, as statins are generally very similar in efficacy and side-effect profiles.

\(^6\)One might use “uninterrupted usage” as the predictor if there may be non-linear effects, whereas one would use “quarters of use” if benefits are linear in usage.

\(^7\)I also look at differences in general health spending as an outcome.

\(^8\)The initial difference comes from the fact that many new users start after a hospitalization, so the control group has a spike about two quarters before the treatment group. The results are similar when I restrict to heart-related hospitalizations.

\(^9\)The one surprise is that the results also hold for Baycol, a drug that was later pulled from the market due to safety concerns.
The framework can also be used to assess spending, the other half of the cost-effectiveness equation, and I find more interesting dynamic spending differences that point to possible long-run cost-reducing strategies on the part of PBM or payers. For my basic analysis, I just plot mean annual spending on cholesterol medication by group.\(^1\) I find that newer branded entries generally lead to persistently lower annual costs for the treated group, which is unsurprising, given that later drugs tend to have lower prices. One also finds interesting patterns in spending for cohorts around generic entry events. Figure A.9 presents two entry events with significant and time-varying differences across cohorts in spending. For the entry of Livalo, a less popular statin launched in 2010, we see an initially lower spending level for the treatment group, which then jumps above the control group when generic Lipitor enters and most Lipitor users move to the cheaper generic. Therefore, it may be long-run cost-effective to incentivize users to stay on Lipitor. A similar story emerges by analyzing groups of users around the entry of generic Zocor, which I show in the second graph in Figure A.9. Here, I find that the patients in the control group who picked a different branded drug (mostly Lipitor and Crestor) end up spending more over time. Control patients who pick branded Zocor start off spending more, but converge to a much lower price in the long-run.

### A.5 Micro-Foundations - A Detailed Discussion

Here, I discuss the possible micro-foundations driving the history-dependence uncovered above. There are two general classes of explanations for the results, rational and behavioral, both of which have support in the data. In addition, there is the issue of the role played by doctors, which is important in thinking about how much market position carries over when the market size is growing and how large inertia may be in markets for acute medication.

There is some support in the data that the observed inertia is generated by experience good effects. Experience good models typically involve a risk-averse user learning about the

\(^1\)For now, I use pharmacy prices in the MarketScan dataset.
qualities of products through consuming them. The higher the risk aversion, the lower the likelihood of switching. In the results discussed in Section A.3.1, I find that heterogeneity in inertia mainly stems from differences in health state, with unhealthy patients and those on many other drugs much less likely to switch. These patients are probably more risk averse in their decisionmaking, bolstering the case for experience good effects.

As in many dynamic individual behavior settings, habit formation can also explain inertia. Once a user develops a routine of which drug to take, they may find that repeated usage brings them higher utility. This explanation is harder to test, although the additional evidence I provide on the relationship of switching to duration of usage does suggest this may be driving part of the effects.

Another explanation of the results is the presence of physical switching costs driven by health considerations. Physical costs are less obvious in the anti-cholesterol market setting, where treatments are fairly uniform. However, there is some discussion and research in the medical literature on poor compliance in patients after switching, which is a form of switching cost. In addition, patients need to go to their doctors to obtain a change in prescription, another form of switching cost. In other areas, such as multiple sclerosis, there are clearer medical costs to switching. Patients need to go through a tapering of the old treatment and a waiting period before starting on the new one.

Behavioral explanations for inertia are led by inattention. One possibility is that patients do not pay attention to prices and new options, and stick with their existing option unless put into an active choice environment. In refinements to my basic demand estimation, I find that patients are more likely to switch in periods where they are hospitalized, providing some evidence that inattention may be at play. However, this is also consistent with the rule of thumb discussed previously. More definitive evidence on this, through exogenous shocks to attention such as exposure to advertising, would open up the possibility of informational interventions in changing observed inertia.

Other behavioral phenomena such as the endowment effect or loss aversion may also be factors in this setting. The main evidence pointing to this is the product hopping phenomena
discussed earlier. People appear to switch from incremental products to generics of the base product and vice versa, suggesting that they dislike either having to pay more or to lose out on quality.

A.6 Standard Claims Datasets and Multi-Agent Models

One major limitation of standard claims datasets is that it is hard to identify a multi-agent model of demand. As discussed in Section 1.2.1, prescription drug demand is driven by the joint decision of patient, doctor, and pharmacist. However, standard datasets such as MarketScan and MEPS only record patient-level filled prescriptions. Therefore, researchers will have a difficult time identifying the contributions of each agent. For example, the data does not allow researchers to easily distinguish between patients who were not prescribed a medication from patients who were but did not fill it.

The ideal dataset for understanding drug demand would record all decisions made by doctors, pharmacists, and patients. This would include the original prescription by the doctor, any subsequent changes made at the pharmacy, and the existing data on filled prescriptions. This would allow us to study how copayments and history affect each of the three agents, and opens up the ability to study the impacts of more fine-grained policy interventions.

---

11To my knowledge, Medicare Part D data contains identifiers for doctors and pharmacists, but is still at the filled prescription level.
A.7 Additional Tables and Figures

Table A.1: Estimates of the Impact of Inertia - 60-Day Window

<table>
<thead>
<tr>
<th></th>
<th>t+3</th>
<th>t+6</th>
<th>t+9</th>
<th>t+12</th>
<th>t+15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initially Chose Drug</td>
<td>0.815***</td>
<td>0.786***</td>
<td>0.628***</td>
<td>0.769***</td>
<td>1.057***</td>
</tr>
<tr>
<td></td>
<td>(0.0376)</td>
<td>(0.0557)</td>
<td>(0.0672)</td>
<td>(0.0751)</td>
<td>(0.0784)</td>
</tr>
<tr>
<td>Observations</td>
<td>540903</td>
<td>414062</td>
<td>320228</td>
<td>256722</td>
<td>186507</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.180</td>
<td>0.083</td>
<td>0.055</td>
<td>0.037</td>
<td>.</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

* p < 0.05, ** p < 0.01, *** p < 0.001

Notes: Estimates of Equation (1.1), but using a 60-day window each side of an event to classify treatment and control users rather than the baseline 180-day window. Comparable to Table 1.3.

Figure A.1: MarketScan Sample Size Over Time

Notes: A plot of the number of people covered by MarketScan data over the sample period. The number increases significantly over the period, reaching a peak of over 50 million individuals in the 2008-2012 period.
<table>
<thead>
<tr>
<th></th>
<th>t+3</th>
<th>t+6</th>
<th>t+9</th>
<th>t+12</th>
<th>t+15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initially Chose New Drug</td>
<td>0.758*** (0.0381)</td>
<td>0.870*** (0.0648)</td>
<td>0.702*** (0.0798)</td>
<td>0.816*** (0.0963)</td>
<td>0.929*** (0.100)</td>
</tr>
<tr>
<td>Chose x Start Age</td>
<td>-0.00331*** (0.000723)</td>
<td>-0.00697*** (0.00123)</td>
<td>-0.00586*** (0.00153)</td>
<td>-0.00704*** (0.00186)</td>
<td>-0.00999*** (0.00195)</td>
</tr>
<tr>
<td>Chose x Female</td>
<td>-0.0528*** (0.0125)</td>
<td>-0.0995*** (0.0200)</td>
<td>-0.0777*** (0.0239)</td>
<td>-0.0767** (0.0281)</td>
<td>-0.0934** (0.0285)</td>
</tr>
<tr>
<td>Chose x Other Drugs</td>
<td>0.0870*** (0.00888)</td>
<td>0.110*** (0.0142)</td>
<td>0.116*** (0.0172)</td>
<td>0.107*** (0.0199)</td>
<td>0.111*** (0.0204)</td>
</tr>
<tr>
<td>Chose x Inpatient</td>
<td>0.0827** (0.0261)</td>
<td>0.278*** (0.0404)</td>
<td>0.420*** (0.0540)</td>
<td>0.296*** (0.0644)</td>
<td>0.372*** (0.0616)</td>
</tr>
<tr>
<td>Observations</td>
<td>1354497</td>
<td>1005048</td>
<td>774485</td>
<td>611098</td>
<td>419943</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.421</td>
<td>0.222</td>
<td>0.150</td>
<td>0.121</td>
<td>0.087</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.2: Heterogeneity in Inertia

Notes: Estimates of Equation (1.1), but with additional covariates to test for heterogeneity in inertia. “Start Age” refers to the age at which the patient started on anti-cholesterol drugs. “Other Drugs” refers to the number of other drugs the patient is taking at the time he or she starts on anti-cholesterol drugs. “Inpatient” refers to whether or not the patient started on anti-cholesterol medication in the same quarter as a hospitalization.
Figure A.2: Impact of Inertia By Period

Notes: A plot of the $\beta_1$ coefficients from Equation (1.1), with and without instrumenting with treatment group indicator. The data sample is the unbalanced panel of MarketScan individuals. The results suggest significant inertia effects that decrease with time, with pooling generating some non-monotonicities that are not present when analyzing individual entry events.
Figure A.3: Smoothness of Flow of New Users (Unbalanced Panel)

Notes: A plot of new patients starting on cholesterol treatment each month in the complete sample. The vertical lines highlight the entry of branded Lipitor, Crestor, Zetia, generic Zocor, and generic Lipitor. The trend generally follows the overall increase in users covered in the MarketScan data, but does not exhibit noticeable spikes around entry events.
Figure A.4: Event study results broken down for major anti-cholesterol entry events

Notes: Graphs depicting differences between treatment and control groups for major entry events from 1996-2013 in the anti-cholesterol market. Treatment group contains patients who start on treatment in the 180 days after a drug launches and the control group contains patients who start in the 180 days before launch.
Figure A.5: Market Share - All Cholesterol Drugs

Notes: A raw breakdown of total pills sold for the three major statins, using estimates from MEPS. Unlike Figure 2.1, I aggregate over all treatment intensity levels and use a quantity-based measure. Striking patterns include that fact that Lipitor begins to dominate the market right away after its launch in 1997. Another key pattern is that generic Zocor and generic Lipitor begin to take over the market later in the period. Another aspect to note is the brief popularity of Vytorin, a combination of Zocor and Zetia, from 2006 to 2008, which is mostly used for high intensity treatment.
Figure A.6: History-Dependence in Other Chronic Drug Markets

Notes: Graphs depicting OLS and IV estimates of $\beta_i$ in Equation (1.1), for branded drug entry events in other chronic drug markets. For multiple sclerosis, I include all new product entry events, including interferons that are introduced that have the same generic name as existing drugs on the market. The results are noisier for most of these, as there are fewer people in the dataset, but the general pattern of inertia appears to hold to varying degrees across areas.
Figure A.7: Event Study Using Announcement of ENHANCE Trial

Notes: A plot of Vytorin market share for cohorts that start before and after the announcement of the ENHANCE trial results, which showed that Vytorin was not superior to other statins, as was thought to be the case based on LDL reduction properties. The after group (blue) has a persistently lower share of patients on Vytorin, again suggestive of history-dependence in demand. The graph shows that other types of variation in initial conditions, in this case medical information, can also have lasting impacts on choice patterns.
Figure A.8: Incremental Effect of Crestor on Hospitalization

*Notes:* A plot of hospitalization rates by treatment and control group for the entry of Crestor in August 2003. The aim is to study the long-run health impacts of Crestor, which we can uncover given random initial conditions and strong inertia, as discussed in Section A.4. The two groups generally have similar long-term rates of hospitalization, suggesting minimal difference between Crestor and the existing standard of care before.
Figure A.9: Impact of Drug Entry on Spending

Notes: Plots of average cholesterol drug spending for treatment and control groups for (a) the entry of Livalo in 2010 and (b) generic Zocor in 2006. The Livalo example illustrates initial savings from some fraction of users taking a cheaper branded drug (Livalo) versus the general baseline, but this difference reverses once generic Lipitor enters and Lipitor users, who comprise a larger share of the control group, switch to generics. Therefore, keeping users on Lipitor in the short-term may save money in the long-run. The generic Zocor breakdown illustrates a similar idea, by breaking down spending by group. Control patients who pick a different branded drug (mostly Lipitor and Crestor) start off spending less than those who start on branded Zocor, but spend much more over time relative to all other groups. Control group patients who start on branded Zocor quickly converge towards the average spending levels of the group on generic Mevacor, which was already on the market.
Appendix B

Appendix to Chapter 2

B.1 Constructing Net Price Estimates

In this section, I provide a more detailed description of the issues in constructing net price estimates, the data I collect, and my methodology for dealing with problematic aspects of the data.

A major challenge here is that net sales estimates are at the drug level, but different dosages of the same drug are sometimes priced differently. For example, a 40MG Lipitor tablet has a gross price that is $1 more expensive than the more commonly used 10MG tablet.\footnote{Crestor actually sets the same price for all dosage levels.} Since I am focusing on the aggregate patterns in the medium intensity dosages, I want to avoid interpreting shifts in other dosage levels or in form of purchase as changes in rebate levels, so I construct overall weighted net and gross prices, and assume that rebates are roughly constant at the drug level.

To construct my estimates, I collect revenue data from SEC filings and IMS Top-Line data, US aggregate demand estimates from the Medical Expenditure Panel Survey (MEPS), and additional net price estimates from SSR Health. I then construct net prices based on the data available in each year.

The financial filing data comes from annual company SEC filings (10-K) and conference
call transcripts, between 1996-2013. At the start of the period, companies generally report

global sales by disease area, but progressively offer more granular data, with the current

standard being a breakdown by drug and geographic area (US, Europe, global). Roughly,

the financial data coverage is as follows:

1. 1996-2000: Merck reports global aggregate sales for its cholesterol drugs (Mevacor and

 Zocor) in its 10-K filings; Pfizer reports global Lipitor sales in its 10-K filings

2. 2001-2003: Merck begins to report Zocor global sales separately in 10-K filings


4. 2003-present: Merck (at the time Schering-Plough) reports global sales of Zetia and

 Vytorin.


In addition, I collect additional net revenue data from IMS Top-Line Market Data\(^2\) and

from media reports on the cholesterol market. The IMS Top-Line data is currently available

from 2007

In terms of demand data, I construct estimates of average demand from MEPS. I break

this down by different dosage levels and seller (pharmacy vs. mail order) for each of the

cholesterol drugs. MEPS covers the entire period from 1996-2013.

Next, I focus on gross prices. To do this, I collect average wholesale price (AWP) and

median price paid (labeled as “pay” in the MarketScan data) for each drug by dosage by

seller from the MarketScan data. Each entry in the data contains these two quantities. The

AWP is the same across all entries with the same NDC number, while the “pay” variable

partially reflects pharmacy markups.

Finally, I obtained estimates of net sales and quantity sold from SSR Health, a pharma-

ceutical consulting company. They estimate net prices for a large number of drugs since

2007 at the quarterly level, and shared with me their estimates for the cholesterol market. I

---

use their numbers for the post-2007 period, but my estimates using raw financial filing data also lines up with their estimates.

After collecting this data and aggregating it to an annual level, I take the following procedure to calculate net prices:

- In years with reported net US sales, I just divide net sales by estimated US quantity from MEPS.

- In years where only net global sales are reported, I arrive at a net US sales number by making the assumption that all of the global rebates are from rebates paid in the US. I verify this assumption in years with both net global and net US sales numbers.

\[
\text{Net US Sales} = \text{Net Global Sales} - (\text{Gross Global Sales, IMS} - \text{Gross US Sales, IMS})
\]

I also make the following corrections to account for issues created by aggregating to an annual level:

- Zocor generics enter in June 2006. Using aggregate quantity for the whole year creates a misleadingly low price for pre-generic net price. This is because after generics enter, branded drugs typically price at a duopoly level to compete with the exclusive generic.\(^3\) Instead, I estimate the fraction of units sold in the first two quarters, and use quarterly revenue data to arrive at a pre-generic net price.

To arrive at a discount per pill, I compare the overall net price estimates to overall gross sales estimates, both by looking at IMS gross sales numbers when available and by constructing my own estimate of gross sales using pharmacy values. The pharmacy payment should roughly be an upperbound on the true net sales number, as illustrated in Figure.

\(^3\)The first generic entrant typically has an 180-day exclusivity period based on provisions in the Hatch-Waxman Act.
\[
\text{Avg. Gross Price} = \frac{\sum_{ij} P_{ij} X_{ij}}{\sum_{ij} X_{ij}}
\]

where \( i \) is the drug and \( j \) is the seller (pharmacy or mail order), \( X \) is the sales estimate from MEPS and \( P \) is either the AWP or "pay" variable from MarketScan.

B.2 Illustrative Model - Inertia and Copay Tiers

In this section, I provide a simple model to capture the qualitative effects from the interaction between patient inertia, PBMs, and the copay tier structure. The key implication is that entrants will either price close to existing competitors, or choose to undercut the incumbent by a significant amount to push them into the higher copay tier.

B.2.1 Model Outline

The simplified model contains the following features:

1. There are two firms. Firm 1 has an established drug with price \( p_1 \) (exogenous for now) that is being taken by \( \gamma \) of the current patient pool (normalized to have unit mass).

2. Firm 2 has an entering drug that is identical to Firm 1’s drug. To begin with, I assume that Firm 2 is the only active price setter.

3. The intermediary (PBM) takes \( p_1, p_2 \), and assigns each to a copay tier (no possibility of exclusion).

4. The formulary arrangements are based on two copay tiers: \( \zeta, \bar{\zeta} \) with \( \zeta < \bar{\zeta} \)

5. The intermediary’s objective function is summarized as a weighted average between consumer surplus (simplification of demand for PBM services) and expenditures:

\[
\max_{f} W(f) = \theta \sum_{i=1}^{2} p_i q_i(f)
\]

where \( W(f) \) is the value to the consumer (normalize to 1) minus the copay.
6. Demand is assumed to be very simple: patients already on the drug owned by Firm 1 will keep taking it, and new patients will pick the drug with lower copay, unless they are equal, in which case they split evenly across the two drugs.

B.2.2 PBM Decision

We start with the PBM problem. Assume for now they pick between three formulary arrangements:

\[ \mathcal{F} = \{(\bar{c}, c), (c, \bar{c}), (\bar{c}, \bar{c})\} \]

Given \( p_1, p_2 \), the \((c, \bar{c})\) formulary yields:

\[ v_{\text{same}} = (1 - \bar{c}) - \theta \left[ \gamma p_1 + \frac{1}{2}(1 - \gamma)p_1 + \frac{1}{2}(1 - \gamma)p_2 \right] \]

while the \((\bar{c}, c)\) arrangement (“2 preferred”) yields:

\[ v_{2p} = \gamma(1 - \bar{c}) + (1 - \gamma)(1 - \bar{c}) - \theta \left[ \gamma p_1 + (1 - \gamma)p_2 \right] \]

and the \((\bar{c}, \bar{c})\) arrangement (“1 preferred”) yields (everyone picks Firm 1):

\[ v_{1p} = (1 - \bar{c}) - \theta p_1 \]

B.2.3 Entering Firm Pricing

Given the PBM’s behavior, we then evaluate the incentives facing the entering firm. Firm 2 then sets its price \( p_2 \) relative to a fixed \( p_1 \) to maximize profits, taking into account the PBMs incentives.

In the simplest case, setting \( p_2 > p_1 \) will mean that the PBM will pick the “1 preferred arrangement,” as \( v_{1p} \) is larger than both of the other options (cheaper vs. “same”; cheaper and higher welfare vs. “2 preferred”). This means that they will get no profits from this strategy.
At \( p_2 = p_1 \), the PBM is indifferent between \( v_{same} \) and \( v_{1p} \). Firm 2 can price \( \epsilon \) below to push the PBM to pick the “same” arrangement. One can also modify the model to have a small fraction of users who really like drug 2, thus boosting the welfare under the “same” arrangement.

So the decision then boils down to the price at which Firm 2 can persuade the PBM to choose the preferred tier.

\[
v_{same} - v_{2p} = \gamma(1 - \bar{c}) - \gamma(1 - \hat{\epsilon}) + \theta \left[ \frac{1}{2}(1 - \gamma)p_1 - \frac{1}{2}(1 - \gamma)p_2 \right]
\]

\[
= \gamma(\hat{\epsilon} - \bar{c}) + \frac{1}{2}\theta(1 - \gamma)(p_2 - p_1)
\]

Define \( d^* \) is the discount \( p_1 - p_2 \) at which the difference becomes zero. Setting the expression to zero and re-arranging, we get

\[
d^* = \frac{2\gamma(\hat{\epsilon} - \bar{c})}{\theta(1 - \gamma)}
\]

This means that the discount Firm 2 has to offer is:

1. Increasing in \( \hat{\epsilon} - \bar{c} \): intuitively how unhappy incumbent users would be if they had to pay a higher copay
2. Increasing in \( \gamma \): intuitively the degree of incumbent advantage
3. Decreasing in \( \theta \): how much PBMs are focused on savings vs. keeping consumers happy.

Finally, we can think about Firm 2 profits. At \( p_2 = p_1 \), they obtain:

\[
\pi_{p_2=p_1} = \frac{1}{2}p_1(1 - \gamma)
\]

while at \( p_2 = p_1 - d^* \), they obtain:
\[ \pi_{p_2 = p_1 - d} = (p_1 - d^*)(1 - \gamma) \]

Pricing a little bit below either of those two points is unprofitable, as they obtain the same demand at lower profits per unit.

Therefore, Firm 2 picks the \( p_2 = p_1 \) strategy if:

\[ p_1 < 2d^* = \frac{4\gamma(\bar{c} - \underline{c})}{(1 - \gamma)\theta} \]

The results suggest that there’s more room to undercut existing prices if there is a lower copay differential or lower levels of patient inertia.

\section*{B.2.4 Summary}

The main takeaway from the toy model is that it is difficult to gain market share through undercutting a competitor with an existing consumer base. PBMs are less likely to want to move a drug with a consumer base to a higher copay tier, or exclude it entirely, as is possible in my full model. Therefore, a company that wants to gain dominant market share will have to undercut competition by a significant amount.

The model also speaks to the entry strategy of firms, who will either price the same or undercut competitors by a large amount. The key mechanism driving this result is the interaction of PBMs with discrete copay tiers and patient inertia. Specifically, the two key features of copay setting are that the formulary tiers are discrete (\( \bar{c} - \underline{c} \neq 0 \)) and that the PBM has the option to put both drugs in the same formulary tier. In conjunction with stickiness, this pushes the Firm 2’s pricing to either be equal or to be dramatically different, because Firm 1 has an advantage by having a pool of captured consumers. The PBM is likely to put Firm 1 in the low copay tier, because it has trouble moving people to capture the discounts, and also has to take into account the copay differential they are imposing on the captured consumers. Therefore Firm 2 can choose either to overcome this by setting a much lower price, or choose to split the share of new patients by setting an equal price.
If the equal tier setup was not in the set of possible formularies\(^4\), then Firm 2 would have more of an incentive to price lower, as pricing the same as Firm 1 will push them into the higher copay tier.

The prescription drug market is different to other markets with intermediaries, because in other contexts such as supermarket shelf space and sponsored search auctions, positions are naturally rival. Here, it’s an option to place drugs on the same tier, and in the data, we see that a large majority of formularies exhibit this parity.

### B.3 Additional Graphs

![Implied Quality Estimates](image)

**Figure B.1: Implied Quality Estimates**

*Notes: A plot of the \(\delta_{jkt} \) coefficients from Equation (2.1), with generic Zocor and generic Lipitor qualities replacing their branded counterparts in the series once they enter. The data shows that Zocor has lower implied quality than Lipitor until its generic version enters. Both Lipitor and Crestor experience declines in implied quality once generic Zocor enters. This is because the fraction of diagnosed patients taking a drug stays steady, but generic Zocor begins to take market share from its competitors. Other quality fluctuations are driven by advertising campaigns and changes in medical evidence.*

\(^4\) \(F\) in the notation of my empirical model.
Notes: Graphs capturing the equilibrium formulary distribution predicted by the model. The first plots the share of formularies in which each drug is in the non-preferred tier and the second plots the rate of exclusion for each drug. The predictions generally reflect trends found in the MarketScan data and industry reports: i) no exclusion in the early periods ii) Zocor is in the non-preferred tier at higher rates relative to Lipitor iii) after generic Zocor enters, higher copays and formulary restrictions are put on the remaining drugs.
Notes: Counterfactual prices under a scenario with no carryover in previous market share to entering generics \((h = 0)\). The two price series are almost identical, as effects from generics not benefiting from existing market share of the branded drug are minimal. This is reflective of the inherent competitiveness of generic medication and the size of inertia protecting branded competitors.
Appendix C

Appendix to Chapter 3

C.1 Examiner Effect Estimation Methods

In this appendix, we report additional steps involved in the research designs described in Section 3.3 to recover the causal effect of examiner on patent outcomes, namely (1) the baseline research design with the Bayesian shrinkage correction, (2) the design using applications’ last digits as a source of variation, (3) the design using examiners’ busyness as a source of variation, and (4) the baseline research design with the Beta-Binomial count model for binary outcomes.

1. Bayesian shrinkage. In what follows, we describe the two steps we take to estimate the shrunk examiner effects introduced in Section 3.3.1. These two steps help increase the precision of the Empirical Bayes posterior estimate of each examiner effect in equation (3.4).

In the first step, we amend the statistical model to allow for an examiner-by-year shock $\theta_{jt}$, i.e.

$$Y_i = a_{ut(i)} + v_{ij},$$
$$v_{ij} = \mu_j + \theta_{jt(i)} + \epsilon_i,$$

where $i$ indexes the patent, $j$ the examiner, $u$ the art unit and $t$ the year. We compute $\bar{v}_{jt}$ using (3.2), $\hat{\sigma}_\mu$ using (3.3), as well as $\hat{\sigma}_\epsilon^2 = Var(v_{ij} - \bar{v}_{jt})$ and $\hat{\sigma}_\theta^2 = Var(v_{ij}) - \hat{\sigma}_\mu^2 - \hat{\sigma}_\epsilon^2$. 

193
In the second step, for each examiner we compute a weighted average of the yearly average residuals \( \{ v_{jt} \} \) that has the property of being a minimum variance unbiased estimate of \( \mu_j \). This average uses weights such that years in which the examiner granted more patents are given a higher weight:

\[
\bar{v}_j = \sum_t w_{jt} \bar{v}_{jt},
\]

where

\[
w_{jt} = \frac{h_{jt}}{\sum h_{jt}}, \quad h_{jt} = \frac{1}{\tilde{\sigma}_\theta^2 + \frac{\tilde{\sigma}_e^2}{n_{jt}}},
\]

We then compute the Empirical Bayes posterior estimate of each examiner as in (3.4):

\[
\hat{\mu}_j = \frac{\tilde{\sigma}_\mu^2}{\text{Var}(\bar{v}_j)} \cdot \bar{v}_j,
\]

with

\[
\text{Var}(\bar{v}_j) = \tilde{\sigma}_\mu^2 + \left( \sum h_{jt} \right)^{-1}.
\]

The shrinkage factor is the ratio of the signal variance to the total variance and varies across examiners depending on the total number of patents they granted.

The shrunk examiner effects \( \{ \hat{\mu}_j \} \) have two noteworthy properties. First, under the assumption that \( \mu_j \sim N(0, \sigma_\mu^2) \), \( \theta_{jt} \sim N(0, \sigma_\theta^2) \) and \( e_{it} \sim N(0, \sigma_e^2) \), the shrunk examiner effect is the optimal Bayesian posterior expectation of an examiner’s effect given the history of patent outcomes up to the current period. The derivation is a standard application of Bayes’ rule. Intuitively, since there is no drift in examiner effects, we can use the average patent outcome in all years prior to the current year as a sufficient statistic to form the posterior distribution of examiner effects. Second, the shrunk examiner effects also have a frequentist interpretation. The shrinkage factor is the regression coefficient in the hypothetical regression of the true (unobserved) \( \mu_j \) on \( \bar{v}_j \). The regression coefficient is naturally the ratio of the covariance of \( \mu_j \) and \( v_j \) (given by \( \tilde{\sigma}_\mu^2 \) because the other components of \( \bar{v}_j \) are noise) to the variance of \( \bar{v}_j \).
2. Allocation of applications to examiners using the last digit of applications’ serial numbers. To identify art units assigning applications based on the last digit of their serial numbers, we use the USPTO patent examination database and follow three steps.

First, we prepare the data. We exclude continuation applications, since these applications are almost always assigned to the examiner who processed the parent application. For each patent application we record the “docket date”, which is the date on which the application was assigned to the relevant art unit. After an application is filed, the USPTO assigns the application to a specific art unit according to its technological features, which takes some time; therefore the docket date is typically different from the filing date and is the relevant point in time when examiner assignment occurs within the art unit.

Second, we compute the key statistics of interest. We count the number of applications falling in each last-digit-by-examiner-by-art-unit-by-docket-year cell; we denote the application count in each cell by $n_{djut}$, where $d$ indexes the last digit of the application’s serial number (ranging from 0 to 9), $j$ the examiner, $u$ the art unit and $t$ the docket year. In addition, for each examiner we record the total number of applications they were assigned in each art unit in each docket year, denoted by $n_{jut}$.

Third, for each art unit in each year, we implement a statistical test of the null hypothesis that examiner assignment does not depend on last digits. If last digits are not used, the expected number of application with last digit $d$ assigned to examiner $j$ is simply one tenth of the total number of application assigned to this examiner in this art unit and docket year, which we denote by $n_{djut}^E = \frac{n_{jut}}{10}$. To assess whether the data rejects the null in a given art unit and docket year, we compute the following Pearson’s Chi-squared statistic:

$$\chi^2_{ut} = \sum_{j \in art} \sum_{d=0}^{9} \frac{(n_{djut} - n_{djut}^E)^2}{n_{djut}^E}.$$ 

Intuitively, we compare the actual number of applications with last digit $d$ assigned to examiner $j$ in art unit $u$ in docket year $t$ ($n_{djut}$) to the expected number ($n_{djut}^E$). If the actual assignment patterns are “too concentrated” relative to what may happen simply by chance, we reject the null. Formally, under the null that art unit $u$ did not use last digits for examiner
assignment in docket year $t$, $\chi^2_{ut}$ has a Chi-squared distribution with $9 \cdot (J_{ut} - 1)$ degrees of freedom, where $J_{ut}$ is the number of examiners in art unit $u$ in docket year $t$. The degrees of freedom follow from the fact that there are ten possible last digits per examiner, minus the constraints for the total number of applications within each examiner and for the total number of applications by last digit cells. Accordingly, we compute the p-value for the null by comparing the value we obtain in the data for $\chi^2_{ut}$ with a Chi-squared distribution with $9 \cdot (J_{ut} - 1)$ degrees of freedom.\(^1\)

Fourth, we draw the list of art units for which we reject the null that last digits are not used for assignment at the 1% level, i.e. with a p-value of the Chi-squared test below 0.01. We draw this list by docket years, i.e. based on statistics $\{\chi^2_{ut}\}$ that are specific to both art units and docket years, so that each art unit is allowed to change its assignment mechanism over time. To make it simple for other researchers to use this source of variation in future work, we make publicly available on our websites the list of art units by docket years for which we rejected the null at the 1% level (click here for the list).

*Busyness instrument.* We describe below our methodology to recover the application-specific examiner assignment probabilities $p_{ij}$ used in equations (3.5) and (3.6) in Section 3.3.3. Our approach delivers variation in examiner assignment solely from changes in examiner busyness over time, which is useful to validate the baseline estimates of examiner effects.

We start by preparing the data. We aggregate total disposals (grants plus abandonments) for examiners in each two-week period in a given year. As before, we exclude continuation applications because they tend to be assigned to the examiner who handled the parent application. For each incoming application, we create the list of all examiners that it could have been assigned to, which is given by the set of examiners who processed at

\(^1\)Our Chi-squared test is similar in spirit to the divergence statistics used by Righi and Simcoe (2017) to provide evidence of examiner specialization based on the dispersion of patent technology classes across examiners working in the same art unit. The exact formulas for their divergence tests differ from ours because they allow for technology-specific patterns of specialization within an art unit (i.e., within a given art unit, it could be that only a subset of all technology classes feature examiner specialization); we use a similar but technically different test for assignment by last digits, because it seems implausible that only a subset of last digits would be used for examiner assignment.
least one application in that art unit and in that year. As a proxy for how an examiner’s busyness changes over time, we compute the number of patent application cases closed by the examiner in each two-week period. Intuitively, an examiner may be assigned more applications as they become less busy, i.e. in periods when they just finished working on other applications.

Next, we estimate the following linear probability model by OLS:

\[ Y_{ij} = \beta D_{jt} + \delta_i + \gamma_j + \epsilon_{ijt}, \] (C.1)

where \( i \) indexes the application, \( j \) the examiner and \( t \) the two-week period. \( Y_{ij} \) is an indicator variable for the assignment of application \( i \) to examiner \( j \); \( D_{jt} \) is the number of patent application cases closed by the examiner during the relevant two-week period; \( \delta_i \) is an application fixed effect which captures the fact that a larger or smaller number of examiners may be available when a given application arrives; and \( \gamma_j \) is an examiner fixed effect which accounts for the possibility that some examiners might be systematically assigned a large of smaller number of applications (e.g., due to seniority). The coefficient \( \beta \) estimates the extent to which an examiner is more likely to be assigned an application (relative to the baseline captured by the fixed effects) in a period when they just finished working on other applications.

Finally, we use the estimates from (C.1) to compute the predicted assignment probabilities \( p_{ij} \equiv \widehat{Y}_{ij} \), which are used in equations (3.5) and (3.6) in Section 3.3.3. If the estimate of \( \beta \) were zero, there would be no variation in the application-specific examiner assignment probabilities \( \{ p_{ij} \} \) across applications received in the same art unit in the same year and the research design would have no power. In fact, we estimate \( \beta > 0 \) and obtain sufficient variation in examiner assignment probabilities to implement the busyness research design presented in Section 3.3.3.

**Beta-Binomial count model.** In what follows, we derive the integrated likelihood for the Beta-Binomial count model used in Section 3.3.3. As a reminder, we aggregate data for each examiner \( j \) in year \( t \) and art unit \( u \) into the form \( (n_{jau}, r_{jut}) \), where \( n \) denotes the
total number of granted patents for a given examiner and \( r \) the total number of patents purchased by PAEs (or some other binary outcome) for this examiner. We then model the data generating process with a binomial likelihood on each examiner for each art unit and year: each examiner has some true probability \( p_{jut} \) of patent purchase by a PAE (conditional on grant). We drop the \( ut \) subscripts below for brevity.

For each art unit in each year, we specify a flexible Beta prior distribution on examiner effects: \( p \sim \text{Beta}(\alpha, \beta) \). We then compute the integrated likelihood:

\[
L(r|n, \alpha, \beta) = \int_{p=0}^{1} \binom{n}{r} p^r (1-p)^{n-r} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1} dp \\
= \binom{n}{r} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} \int_{p=0}^{1} p^r (1-p)^{n-r} p^{\alpha-1} (1-p)^{\beta-1} dp \\
= \binom{n}{r} \frac{\Gamma(\alpha + \beta) \Gamma(r + \alpha) \Gamma(n - r + \beta)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(n + \alpha + \beta)},
\]

where the second step conjugates the inside to integrate to one based on the probability density function of the Beta distribution. Using this expression, we estimate the hyperparameters \( \alpha \) and \( \beta \) by maximum likelihood.

Finally, we compute the shrunk examiner effect for each examiner. A shrunk examiner effect is simply a posterior mean: we start from the prior, which is governed by the estimates for \( \alpha \) and \( \beta \) for each art unit and year, and apply Bayesian updating with the examiner’s data in that art unit and year according to equation (3.7). We then aggregate these shrunk effects across years within each examiner by taking a weighted average (weighting by number of cases) to obtain the overall shrunk effect for each examiner.

In this appendix, we describe the procedure to we use to build the patent portfolio of Patent Assertion Entities. We proceed in four steps:

1. We start with the list of PAE names from RPX for our main sample and from Cotropia et al. (2014) for robustness checks. We exclude universities (e.g. Wisconsin Alumni Research Foundation) and academic hospitals (Children’s Medical Center Corporation). For the Cotropia et al. (2014) list, we only include entities in categories 3 (Large aggregator) and 5 (Patent holding company). This excludes failed companies and technology development companies.

2. We normalize entity names from both the PAE list and the USPTO Assignment Database from Marco et al. (2015). We do so by capitalizing all names, removing punctuation, and removing the following standard entity terms: INC, CO, COMPANY, COMPANIES, CORP, CORPORATIONS, DIV, GMBH, LLC, LC, INCORPORATED, KG, LIMITED, LIMITED PARTNERSHIP, LP, LTD, NV, PLC, SA, SARL, SNC, SPA, SRL, TRUST USA, CENTER, BV, AG, AB, GROUP, FOUNDATION, INSTITUTE, and TECHNOLOGIES.

3. We collect the identifiers of patent transactions in the USPTO Assignment Database (“Reel/Frame IDs” in the USPTO assignment data, which corresponds to one transaction) that have a normalized entity name matching the normalized name of a PAE in Step 2.

4. Using the patent transaction identifier from Step 3, we know from the USPTO Assignment Database whether the patent was assigned to the employer of the inventor(s). We only keep transactions that are non-employer assignments, to mitigate any PAE classification errors that might cause us to include patents filed by failed companies.
and technology development companies. We exclude transactions such as securitiza-
tion, mergers, and name changes. We end up with a list of patents that were sold to
PAEs on the patent market.
### C.3 Additional Tables and Graphs

#### Table C.1: Raw Standard Deviations of Patent Outcomes across Examiners

<table>
<thead>
<tr>
<th>Raw S.D.</th>
<th>% of Average (1)</th>
<th>Level (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patent value from Kogan et al. (2017), $M</td>
<td>106.14</td>
<td>8.65</td>
</tr>
<tr>
<td>4th-year fee payment rate</td>
<td>11.44</td>
<td>0.0996</td>
</tr>
<tr>
<td>8th-year fee payment rate</td>
<td>18.05</td>
<td>0.1101</td>
</tr>
<tr>
<td>12th-year fee payment rate</td>
<td>34.57</td>
<td>0.0722</td>
</tr>
<tr>
<td>Log total patent citation</td>
<td>42.71</td>
<td>0.20</td>
</tr>
<tr>
<td>Log patent citations by same assignee</td>
<td>92.37</td>
<td>0.19</td>
</tr>
<tr>
<td>Log patent citations by other assignees</td>
<td>46.43</td>
<td>0.09</td>
</tr>
<tr>
<td>Rate of patent acquisition by non-PAEs</td>
<td>68.41</td>
<td>0.1344</td>
</tr>
<tr>
<td>Rate of patent acquisition by PAEs</td>
<td>286.88</td>
<td>0.0292</td>
</tr>
<tr>
<td>Rate of patent litigation by non-PAEs</td>
<td>439.17</td>
<td>0.0285</td>
</tr>
<tr>
<td>Rate of patent litigation by PAEs</td>
<td>1359.17</td>
<td>0.0055</td>
</tr>
</tbody>
</table>

**Notes:** This table reports the raw standard deviations of examiner effects as a percentage of the mean (Column 1) and in level (Column 2). The raw standard deviations refer to the standard deviations of the average residuals (defined by equation (3.2)) across examiners. The raw standard deviations account for art unit by year fixed effects but not for excess variance from noise. The results in this table are directly comparable to those of Table 3.2, which account for excess variance. See Section 3.2.1 for details on the sample and variable definitions.
### Table C.2: Signal Standard Deviations of Examiner Causal Effects by Technology Categories

<table>
<thead>
<tr>
<th>Technology Category</th>
<th>Patent value</th>
<th>Log total patent citation</th>
<th>Non-PAE purchase</th>
<th>Change in words per claim</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Biotechnology and organic chemistry</td>
<td>25.33</td>
<td>30.58</td>
<td>15.39</td>
<td>18.55</td>
</tr>
<tr>
<td>(B) Chemical and materials engineering</td>
<td>74.89</td>
<td>24.85</td>
<td>23.11</td>
<td>19.30</td>
</tr>
<tr>
<td>(C) Computer architecture, software and information security</td>
<td>9.00</td>
<td>24.69</td>
<td>6.43</td>
<td>25.60</td>
</tr>
<tr>
<td>(D) Computer networks, multiplex communication, etc.</td>
<td>11.80</td>
<td>31.80</td>
<td>6.10</td>
<td>18.06</td>
</tr>
<tr>
<td>(E) Communications</td>
<td>24.39</td>
<td>20.78</td>
<td>11.68</td>
<td>30.97</td>
</tr>
<tr>
<td>(F) Semiconductors, electrical and optical systems and components</td>
<td>30.83</td>
<td>17.59</td>
<td>12.08</td>
<td>26.47</td>
</tr>
<tr>
<td>(G) Transportation, construction, electronic commerce, agriculture, national security, and license and review</td>
<td>21.95</td>
<td>21.42</td>
<td>12.77</td>
<td>19.15</td>
</tr>
<tr>
<td>(H) Mechanical engineering, manufacturing</td>
<td>29.78</td>
<td>23.08</td>
<td>19.41</td>
<td>19.73</td>
</tr>
</tbody>
</table>

**Notes:** This table reports the signal standard deviations of examiner effects (as a percentage of the mean) for four patent outcomes across the eight technology centers of the USPTO. The means are re-computed within each technology center. The table shows that examiner effects are substantial in all technology centers. We have studied heterogeneity in signal standard deviations for the other outcomes reported in Table 3.2 and did not find large differences across technology centers, except for patent acquisitions by PAEs, which occur primarily in computers, software and communications (not reported). The Bayesian shrinkage methodology used to obtain these estimates is presented in Section 3.3.
Table C.3: Signal Standard Deviations of Examiner Prosecution Behaviors

<table>
<thead>
<tr>
<th></th>
<th>Signal S.D.</th>
<th>S.D. of Shrunken Effects, % of Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% of Average</td>
<td>Level</td>
</tr>
<tr>
<td>Change in number of words per claim, %</td>
<td>23.37</td>
<td>13.39</td>
</tr>
<tr>
<td>Change in number of claims, %</td>
<td>136.83</td>
<td>4.99</td>
</tr>
<tr>
<td>Use of Section 101 - Lack of utility or eligibility</td>
<td>60.43</td>
<td>0.032</td>
</tr>
<tr>
<td>Use of Section 102(a) - Prior art exists</td>
<td>108.69</td>
<td>0.018</td>
</tr>
<tr>
<td>Use of Section 103(a) - Obvious invention</td>
<td>25.27</td>
<td>0.105</td>
</tr>
<tr>
<td>Use of Section 112(b) - Vague claims</td>
<td>47.72</td>
<td>0.088</td>
</tr>
<tr>
<td>Patent citations added by examiner, %</td>
<td>14.53</td>
<td>7.95</td>
</tr>
<tr>
<td>Citations to non-patent literature added by examiner, %</td>
<td>39.70</td>
<td>5.73</td>
</tr>
</tbody>
</table>

Notes: This table reports the signal standard deviations of examiner effects as a percentage of the mean (Column 1) and in level (Column 2), as well as the standard deviations of shrunk examiner effects (Column 3). The Bayesian shrinkage methodology used to obtain these estimates is presented in Section 3.3. See Section 3.2.1 for details on the sample and variable definitions.
Table C.4: PAE Patent Portfolios and Litigated Patents across Technology Categories

Panel A: Patents Owned by PAEs

<table>
<thead>
<tr>
<th>Technology Category</th>
<th>Number of Patents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical</td>
<td>1,669</td>
</tr>
<tr>
<td>Computers &amp; Communications</td>
<td>27,156</td>
</tr>
<tr>
<td>Drugs &amp; Medical</td>
<td>1,312</td>
</tr>
<tr>
<td>Electrical &amp; Electronic</td>
<td>10,660</td>
</tr>
<tr>
<td>Mechanical</td>
<td>2,709</td>
</tr>
<tr>
<td>Others</td>
<td>1,453</td>
</tr>
</tbody>
</table>

Panel B: Patents Litigated by Non-PAEs

<table>
<thead>
<tr>
<th>Technology Category</th>
<th>Number of Patents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical</td>
<td>1,626</td>
</tr>
<tr>
<td>Computers &amp; Communications</td>
<td>4,175</td>
</tr>
<tr>
<td>Drugs &amp; Medical</td>
<td>2,609</td>
</tr>
<tr>
<td>Electrical &amp; Electronic</td>
<td>2,497</td>
</tr>
<tr>
<td>Mechanical</td>
<td>2,859</td>
</tr>
<tr>
<td>Others</td>
<td>4,611</td>
</tr>
</tbody>
</table>

Notes: This table reports the number of patents owned by PAEs (Panel A) and the number of litigated patents by non-PAEs (Panel B) across technology categories. The technology categories are based on the primary USPTO technology class for each patent, following Hall et al. (2001). These panels show that PAEs tend to be most active in technology areas related to computers, communications and electronics, where patent litigation by non-PAEs is also frequent.
Table C.5: Patent Acquisition and Examiner Behavior, Full Sample

Panel A: Patent Acquisition by PAEs

<table>
<thead>
<tr>
<th>Leave-one-out Examiner Effects</th>
<th>Patent Purchase by PAE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>% Change in Words per Claim</td>
<td>-0.076***</td>
</tr>
<tr>
<td>Number of Claims</td>
<td>(0.015)</td>
</tr>
<tr>
<td>Grant Rate</td>
<td>0.0022</td>
</tr>
<tr>
<td>- Lack of utility or eligibility</td>
<td>(0.005)</td>
</tr>
<tr>
<td>- Prior art exists</td>
<td>0.006</td>
</tr>
<tr>
<td>- Obvious invention</td>
<td>(0.004)</td>
</tr>
<tr>
<td>- Vague claims</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td>Year by Art Unit</td>
</tr>
<tr>
<td>N</td>
<td>1,109,882</td>
</tr>
</tbody>
</table>

Panel B: Patent Acquisition by Practicing Firms

<table>
<thead>
<tr>
<th>Leave-one-out Examiner Effects</th>
<th>Patent Purchase by Practicing Firm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>% Change in Words per Claim</td>
<td>0.0022</td>
</tr>
<tr>
<td>Number of Claims</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Grant Rate</td>
<td>0.001</td>
</tr>
<tr>
<td>- Lack of utility or eligibility</td>
<td>(0.005)</td>
</tr>
<tr>
<td>- Prior art exists</td>
<td>0.0045</td>
</tr>
<tr>
<td>- Obvious invention</td>
<td>(0.004)</td>
</tr>
<tr>
<td>- Vague claims</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td>Year by Art Unit</td>
</tr>
<tr>
<td>N</td>
<td>1,109,882</td>
</tr>
</tbody>
</table>

Notes: Regressors are standardized by their standard deviations and coefficients are expressed as a fraction of the mean of the outcome. Results are similar with patent-level controls and assignee fixed effects (not reported). Standard errors are clustered by examiners. *p < 0.1, **p < 0.05, ***p < 0.01.
Table C.6: Patent Litigation and Examiner Behavior, Full Sample

Panel A: Patent Litigation by PAEs

<table>
<thead>
<tr>
<th>Leave-one-out Examiner Effects</th>
<th>Patent Litigation by PAE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>% Change in Words per Claim</td>
<td>-0.27***</td>
</tr>
<tr>
<td>% Change in Number of Claims</td>
<td>0.16***</td>
</tr>
<tr>
<td>Grant Rate</td>
<td>0.28***</td>
</tr>
<tr>
<td>Use of Section 101</td>
<td>-0.17</td>
</tr>
<tr>
<td>- Lack of utility or eligibility</td>
<td>-0.014</td>
</tr>
<tr>
<td>Use of Section 102(a)</td>
<td>-0.077</td>
</tr>
<tr>
<td>- Prior art exists</td>
<td>-0.077</td>
</tr>
<tr>
<td>- Obvious invention</td>
<td>-0.077</td>
</tr>
<tr>
<td>Use of Section 103(a)</td>
<td>-0.077</td>
</tr>
<tr>
<td>- Obvious invention</td>
<td>-0.077</td>
</tr>
<tr>
<td>N</td>
<td>1,109,882</td>
</tr>
</tbody>
</table>

Panel B: Patent Litigation by Practicing Firms

<table>
<thead>
<tr>
<th>Leave-one-out Examiner Effects</th>
<th>Patent Litigation by Practicing Firm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>% Change in Words per Claim</td>
<td>-0.066***</td>
</tr>
<tr>
<td>% Change in Number of Claims</td>
<td>0.029*</td>
</tr>
<tr>
<td>Grant Rate</td>
<td>0.073**</td>
</tr>
<tr>
<td>Use of Section 101</td>
<td>-0.069***</td>
</tr>
<tr>
<td>- Lack of utility or eligibility</td>
<td>-0.0016</td>
</tr>
<tr>
<td>Use of Section 102(a)</td>
<td>-0.021</td>
</tr>
<tr>
<td>- Prior art exists</td>
<td>-0.021</td>
</tr>
<tr>
<td>- Obvious invention</td>
<td>-0.021</td>
</tr>
<tr>
<td>Use of Section 112(b)</td>
<td>-0.012</td>
</tr>
<tr>
<td>- Vague claims</td>
<td>-0.012</td>
</tr>
<tr>
<td>N</td>
<td>1,109,882</td>
</tr>
</tbody>
</table>

Notes: Regressors are standardized by their standard deviations and coefficients are expressed as a fraction of the mean of the outcome. Results are similar with patent-level controls and assignee fixed effects (not reported). Standard errors are clustered by examiners. *p < 0.1, ** p < 0.05, *** p < 0.01.
<table>
<thead>
<tr>
<th>Panel A: Reissuance of Granted Patents</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Examiner Effects</strong></td>
</tr>
<tr>
<td>(separate regressions)</td>
</tr>
<tr>
<td><strong>Reissuance Rate</strong></td>
</tr>
<tr>
<td>(1)</td>
</tr>
<tr>
<td>(2)</td>
</tr>
<tr>
<td>(3)</td>
</tr>
<tr>
<td><strong>Reissuance Rate Two</strong></td>
</tr>
<tr>
<td>Years or More after Grant</td>
</tr>
<tr>
<td>(4)</td>
</tr>
<tr>
<td>(5)</td>
</tr>
<tr>
<td>(6)</td>
</tr>
<tr>
<td><strong>(A) % Change in Words per Claim</strong></td>
</tr>
<tr>
<td>-0.12*** (-0.031)</td>
</tr>
<tr>
<td>-0.10*** (0.029)</td>
</tr>
<tr>
<td>-0.12*** (0.03)</td>
</tr>
<tr>
<td><strong>(B) Grant Rate</strong></td>
</tr>
<tr>
<td>0.13*** (0.03)</td>
</tr>
<tr>
<td>0.11*** (0.02)</td>
</tr>
<tr>
<td>0.16*** (0.031)</td>
</tr>
<tr>
<td><strong>Fixed Effects</strong></td>
</tr>
<tr>
<td>Year Year by Year by Year by Art Unit Unit by Class</td>
</tr>
<tr>
<td><strong>N</strong></td>
</tr>
<tr>
<td>1,109,882</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Court Rulings</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Examiner Effects</strong></td>
</tr>
<tr>
<td>(separate regressions)</td>
</tr>
<tr>
<td><strong>Invalidity Rate</strong></td>
</tr>
<tr>
<td>(1)</td>
</tr>
<tr>
<td>(2)</td>
</tr>
<tr>
<td>(3)</td>
</tr>
<tr>
<td><strong>Infringement Rate</strong></td>
</tr>
<tr>
<td>(4)</td>
</tr>
<tr>
<td>(5)</td>
</tr>
<tr>
<td>(6)</td>
</tr>
<tr>
<td><strong>(A) % Change in Words per Claim</strong></td>
</tr>
<tr>
<td>0.019 (0.028)</td>
</tr>
<tr>
<td>0.11 (0.11)</td>
</tr>
<tr>
<td>0.17 (0.179)</td>
</tr>
<tr>
<td><strong>(B) Grant Rate</strong></td>
</tr>
<tr>
<td>0.005 (0.02)</td>
</tr>
<tr>
<td>-0.025 (0.09)</td>
</tr>
<tr>
<td>-0.097 (0.16)</td>
</tr>
<tr>
<td><strong>Fixed Effects</strong></td>
</tr>
<tr>
<td>Year Year by Year by Year by Art Unit Unit by Class</td>
</tr>
<tr>
<td><strong>N</strong></td>
</tr>
<tr>
<td>479</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: Trials at the Patent Office (“Inter Partes Reviews”)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Examiner Effects</strong></td>
</tr>
<tr>
<td>(separate regressions)</td>
</tr>
<tr>
<td><strong>IPR Rate</strong></td>
</tr>
<tr>
<td>(1)</td>
</tr>
<tr>
<td>(2)</td>
</tr>
<tr>
<td>(3)</td>
</tr>
<tr>
<td><strong>Institution Rate of IPR</strong></td>
</tr>
<tr>
<td>(4)</td>
</tr>
<tr>
<td>(5)</td>
</tr>
<tr>
<td>(6)</td>
</tr>
<tr>
<td><strong>(A) % Change in Words per Claim</strong></td>
</tr>
<tr>
<td>-0.286*** (0.067)</td>
</tr>
<tr>
<td>-0.283*** (0.063)</td>
</tr>
<tr>
<td>-0.27*** (0.06)</td>
</tr>
<tr>
<td><strong>(B) Grant Rate</strong></td>
</tr>
<tr>
<td>0.286*** (0.061)</td>
</tr>
<tr>
<td>0.2613*** (0.058)</td>
</tr>
<tr>
<td>0.28*** (0.06)</td>
</tr>
<tr>
<td><strong>Fixed Effects</strong></td>
</tr>
<tr>
<td>Year Year by Year by Year by Art Unit Unit by Class</td>
</tr>
<tr>
<td><strong>N</strong></td>
</tr>
<tr>
<td>1,109,882</td>
</tr>
</tbody>
</table>

Notes: Regressors are standardized by their standard deviations and coefficients are expressed as a fraction of the mean of the outcome. The linear predictor for PAE acquisition is given by specification (9) in Table C.6. Results are similar with patent-level controls and assignee fixed effects (not reported). Examiner effects can computed leaving out the focal patent. Standard errors are clustered by examiners. *p < 0.1, **p < 0.05, ***p < 0.01.
Table C.8: Examiner Behavior and Other Patent Outcomes

Panel A: Patent Value from Kogan et al. (2017)

<table>
<thead>
<tr>
<th>Leave-one-out Examiner Effects</th>
<th>Patent Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) (2) (3) (4) (5) (6) (7) (8) (9)</td>
</tr>
<tr>
<td>Purchase by PAE</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
</tr>
<tr>
<td>% Change in</td>
<td>-0.0033</td>
</tr>
<tr>
<td>Words per Claim</td>
<td>(0.008)</td>
</tr>
<tr>
<td>% Change in</td>
<td>-0.0026</td>
</tr>
<tr>
<td>Number of Claims</td>
<td>(0.0226)</td>
</tr>
<tr>
<td>Grant Rate</td>
<td>0.0398***</td>
</tr>
<tr>
<td></td>
<td>(0.0149)</td>
</tr>
<tr>
<td>Use of Section 101</td>
<td>0.0308**</td>
</tr>
<tr>
<td>- Lack of utility or eligibility</td>
<td>(0.0150)</td>
</tr>
<tr>
<td>Use of Section 102(a) -</td>
<td>-0.0071</td>
</tr>
<tr>
<td>- Prior art exists</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Use of Section 103(a)</td>
<td>-0.011</td>
</tr>
<tr>
<td>- Obvious invention</td>
<td>(0.007)</td>
</tr>
<tr>
<td>Use of Section 112(b)</td>
<td>-0.008</td>
</tr>
<tr>
<td>- Vague claims</td>
<td>(0.011)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Year by Art Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>356,250 310,264 310,332 356,318 356,250 356,250 356,250 356,250 310,332</td>
</tr>
</tbody>
</table>

Panel B: Patent Citations

<table>
<thead>
<tr>
<th>Leave-one-out Examiner Effects</th>
<th>Log Total Patent Citations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) (2) (3) (4) (5) (6) (7) (8) (9)</td>
</tr>
<tr>
<td>Purchase by PAE</td>
<td>0.015**</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
</tr>
<tr>
<td>% Change in</td>
<td>-0.065***</td>
</tr>
<tr>
<td>Words per Claim</td>
<td>0.006</td>
</tr>
<tr>
<td>% Change in</td>
<td>0.010</td>
</tr>
<tr>
<td>Number of Claims</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Grant Rate</td>
<td>0.111***</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
</tr>
<tr>
<td>Use of Section 101</td>
<td>-0.03***</td>
</tr>
<tr>
<td>- Lack of utility or eligibility</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Use of Section 102(a) -</td>
<td>0.001</td>
</tr>
<tr>
<td>- Prior art exists</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Use of Section 103(a)</td>
<td>-0.054***</td>
</tr>
<tr>
<td>- Obvious invention</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Use of Section 112(b)</td>
<td>-0.027***</td>
</tr>
<tr>
<td>- Vague claims</td>
<td>(0.006)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Year by Art Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>988,249 848,162 848,527 988,432 988,249 988,249 988,249 988,249 848,162</td>
</tr>
</tbody>
</table>

Notes: Regressors are standardized by their standard deviations and regression coefficients are expressed as a fraction of the mean of the outcome. The sample includes all technology categories. Standard errors are clustered by examiners. *p < 0.1, **p < 0.05, ***p < 0.01.  

208
Table C.9: Patent Acquisition by PAEs and Grant Decision at the European Patent Office

Panel A: Full Sample of Patents with Joint Filing at USPTO and EPO

<table>
<thead>
<tr>
<th></th>
<th>Patent Purchase by PAE</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>EPO Grant</td>
<td>-0.211***</td>
<td>-0.199**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.057)</td>
<td>(0.059)</td>
<td></td>
</tr>
<tr>
<td>Art unit by year</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Fixed Effects</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Examiner Fixed Effects</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>218,867</td>
<td>217,491</td>
<td></td>
</tr>
</tbody>
</table>

Panel B: Analysis in Subsamples

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>EPO Grant</td>
<td>-0.2144**</td>
<td>0.0037</td>
<td>-0.0831</td>
</tr>
<tr>
<td></td>
<td>(0.1001)</td>
<td>(0.0133)</td>
<td>(0.1074)</td>
</tr>
<tr>
<td>Subsample of examiners with PAE effect below median</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Art unit by year Fixed Effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Examiner Fixed Effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Assignee Fixed Effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>109,383</td>
<td>109,484</td>
<td>109,383</td>
</tr>
</tbody>
</table>

Notes: The two panels of this table examine whether PAEs selectively purchase higher-quality patents, using patent grant decisions at the European Patent Office (EPO) as a proxy for patent quality (as in Lei and Wright (2017) and Picard and Van Pottelsberge de la Potterie (2011)). Regressors are standardized by their standard deviations and regression coefficients are expressed as a fraction of the mean of the outcome. In Panel A, the sample includes all patents that were jointly filed at the USPTO and EPO. Column (1) of Panel A shows that PAEs are much more likely to purchase patents that were rejected by the EPO, controlling for art unit by year fixed effects. This result suggests that PAEs target lower-quality patents, which bear on more incremental, less innovative technology. As shown in Column (2), this pattern continues to hold, with a similar magnitude, even within the portfolio of each examiner (i.e. controlling for examiner fixed effects). Panel B documents the within-examiner patterns in more detail. Columns (1) and (2) of Panel B show that PAEs selectively purchase patents that were rejected by the EPO only in the patent portfolios of examiners who have a relatively large causal impact on PAE purchases (specifically, their PAE purchase effect is above median; the examiner-specific PAE purchase effects were estimated using equation (3.4)). These patterns suggest that PAEs selectively purchase patents with two features: (i) these patents are close to existing intellectual property because they bear on incremental technologies (hence their higher likelihood of rejection at EPO); (ii) these patents were issued by specific (lenient) examiners at the USPTO, and their claims may be less well-defined are harder to interpret than average. As shown in Columns (3) to (6), there is no such effect for patent acquisition by practicing firms (for which we obtain precisely estimated zeros) or for non-PAE litigation. Standard errors are clustered by examiners. *p < 0.1, **p < 0.05, ***p < 0.01.
Panel A: Prosecution Behavior

Panel B: Patent Acquisition by PAEs

Figure C.1: Examiner Career Effects

Notes: Following Frakes and Wasserman (2017), this figure examines the behavior of individual examiners over the course of various promotions (indicated by red bars on the figure) that carry with them reductions in examination time allocations. In each panel, we regress the outcome on a series of dummy variables reflecting examiners’ experience within a grade level. For each grade level — GS-level 12, GS-level 13 without signatory authority, GS-level 13 with signatory authority, and GS-level 13 —, we track examiners for 1-2, 3-4, 5-6, 7-8, and over 9 years of experience. Specifications include examiner and year fixed effects, and standard errors are clustered by examiners. Panel A shows that after being promoted (and having less time for examination), examiners tend to make fewer demands on the applicant during the prosecution process, as evidenced by the reduction in the issuance of 103(a) blocking action (which is consistent with the findings on grant rates in Frakes and Wasserman (2017)). In contrast, Panel B reports that the rate of purchase by PAEs does not vary significantly around promotion events. This result indicates that examiner career effects have a second-order impact on PAE purchase, relative to the examiner fixed effects estimated in Section 3.3.
Figure C.2: Distributions of Shrunk Examiners Effects

Notes: This figure reports histograms of the shrunk examiner effects for four patent outcomes. The shrunk examiner effects are computed according to equation (3.4). In each panel, the shrunk examiner effects are expressed as a percentage of the mean of the outcome and the histogram is reported for shrunk effects that are within 2.5 signal standard deviations of the mean. This figure shows that there is substantial variation in shrunk examiner effects, i.e. the data delivers very different Bayesian posterior expectations across examiners. The distribution is more concentrated towards zero for rare outcomes like purchase by a PAE or litigation, because the shrinkage factors are higher for such outcomes.
Notes: This figure reports the relationship between the (leave-one-out) examiner grant effect and the patent price in an auction. The examiner grant effects (on the x-axis) are computed as in Section 3.3 and are standardized by their signal standard deviation. The auction price (on the y-axis) is provided by Ocean Tomo. Each dot on the figure represents 5% of the underlying data and the OLS best-fit line is reported. Since patents are sometimes auctioned as a group rather than individually, we include fixed effects for lot size. The specification also includes art unit and year fixed effects. The negative slope shows that more lenient examiners (with a higher grant rate) issue patents that sell for less in the patent market. A patent issued by an examiner with a grant rate one signal standard deviation above the mean is sold for $48,000 less on average.