Essays on Health Economics

by

Yufei Wu

B.S. Economics and Finance, Tsinghua University (2011)

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

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2016

Author ..... Signature redacted .....
Department of Economics
April 15, 2016

Certified by. Signature redacted .....
Amy Finkelstein
Ford Professor of Economics
Thesis Supervisor

Certified by. Signature redacted .....
Heidi Williams
Class of 1957 Career Development Assistant Professor
Thesis Supervisor

Accepted by ..... Signature redacted .....
Ricardo Caballero
Ford International Professor of Economics
Chairman, Departmental Committee on Graduate Studies

Essays on Health Economics

by

Yufei Wu

Submitted to the Department of Economics
on April 15, 2016, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

The first chapter explores strategic insurer pricing in response to consumer inertia. A growing literature has documented evidence that consumers in health insurance markets are inertial, or behave as though they face substantial switching costs in choosing a health insurance plan. I investigate whether the private firms that provide prescription drug insurance through Medicare Part D exploit this inertia when setting prices. I first document descriptive evidence consistent with insurers initially setting low prices in order to "invest" in future demand before later raising prices to "harvest" inertial consumers. I then apply a two-step estimation approach following Bajari, Benkard and Levin (2007) to explore the implications of these invest and harvest incentives for equilibrium pricing, finding that on net, demand inertia reduces equilibrium prices (i.e. the invest incentive dominates the harvest incentive). Finally, I evaluate welfare consequences of policies that could be used to constrain insurers' ability to conduct such "invest-then-harvest" pricing patterns. I find, for example, that a policy change to cap premium increases would improve consumer welfare by both lowering average premiums and smoothing prices over time.

Motivated by prior work on market size spurring innovation, the second chapter (co-authored with Manuel Hermosilla) explores the role of increased downstream demand in facilitating inter-firm cooperation in the pharmaceutical industry, where licensing is a common form of collaboration between upstream innovators and downstream commercializers. We propose a simple model of licensing with heterogeneous match quality which predicts that positive demand shocks will increase the likelihood of licensing and improve match quality by reducing the relative importance of transaction costs. We then use the differential impacts of the introduction of Medicare Part D across drug categories targeting different ages of consumers as a source of variation in demand, and document empirical evidence consistent with the model.

Using US county-level data on physician stock from the Area Resource File, the third chapter is devoted to uncovering and understanding the differential effects of medical schools on the supply of physician across regions. I use a difference-in-difference framework to compare changes in physician supply in areas closer to new medical school entries with regions further away. I find that a new medical school increased the physician supply by one to three times the county average level in the county where the medical school was located, relative to other counties. The broader regional effect was smaller but still substantial – a new medical school increased physician supply by one fourth to two thirds of the sample average in counties within 50 miles, relative to counties farther away. When tracking the effect over time, I find that a new medical school had the same impact in the year of entry and in the following 20 years, which indicates that most of the impacts could be attributed to the immediate responses. I find no effect on the physician supply in most of the pre-entry years, which supports the identifying assumption that locations of new medical schools were not correlated with other underlying determinants of physician supply.

Thesis Supervisor: Amy Finkelstein
Title: Ford Professor of Economics

Thesis Supervisor: Heidi Williams
Title: Class of 1957 Career Development Assistant Professor

Contents

[1]**This Chapter is co-authored with Manuel Hermosilla.** We gratefully acknowledge the support of Kauffman Dissertation Fellowship and the Rustgi Family Fund in Entrepreneurship.

Acknowledgments

First and foremost, I am deeply grateful to my dedicated advisors Amy Finkelstein, Heidi Williams, and Paulo Somaini for their continuous help and support throughout the past few years. Words cannot express the gratitude I feel for the time they spent discussing my research and providing detailed comments, supporting my exploration of different fields and ideas, filling me with faith as well as suggestions when I ran into challenges and obstacles, and providing life lessons when I most needed them.

Amy Finkelstein and Heidi Williams both helped shape my interest in health economics from the very beginning of my PhD experience, and Heidi Williams also inspired my interest in the economics of innovations. There was always something new that I could learn from them, including the importance of staying focused on the big picture research question while also keeping track of every piece of institutional detail, how to keep presentations concise and to the point, and much, much more. Amy not only sets the golden standard for research, but she is also a role model - a caring teacher and a sharp-witted advisor who never fails to ask great questions. When I first met Heidi, I mistook her for a fellow student, and although she wasn't a student anymore, I was amazed at how she could empathize with all my concerns from a student's perspective and provide the most thoughtful advice. Paulo Somaini introduced me to the structural world of industrial organization and continued to offer guidance throughout my exploration of structural methods. I cannot recall how many times he spent precious hours discussing my work and offering invaluable advice, and even after he moved, I never felt his absence: he was always available whenever I needed feedback. I am incredibly fortunate to have had Amy, Heidi, and Paulo as mentors.

I am also thankful for the generosity of a long list of other faculty members. Jonathan Gruber was more than kind in helping me get access to the data I needed, provided invaluable feedback on my research throughout my program, always helped me to think further about the big picture and the policy implications, and offered life-saving tips as well as comforting words during my job search. Pierre Azoulay, Mark Shepard, and Tavneet Suri each offered helpful suggestions about how to talk about my research on the job market and provided encouragement during that stressful time. I also thank Nikhil Agarwal, David Autor, Joseph Doyle, Esther Duflo, Glenn Ellison, Sara Fisher Ellison, Ariel Pakes, Jim Poterba, Scott Stern, and Michael Whinston, whose helpful comments helped shape and improve this thesis.

If I had to name one person who was responsible for inspiring my original interest in economics, that person would certainly be my undergraduate advisor, Yingyi Qian. I decided to become an economics major after sitting in an intriguing and enlightening introductory lecture by Yingyi Qian, and I was fortunate to have him as my advisor. While fully devoted to reforming undergraduate education as the dean, he also frequently checked in with students, offering advice on both academics and life. His continued mentorship throughout my time in college and graduate school has been invaluable, and I will always be grateful to him for being both a mentor and a role model. I am also indebted to other faculty members at Tsinghua, especially Chong-En Bai and Xiaohan Zhong, for their support when I embarked on my first attempts to conduct empirical research, when I decided to apply to graduate school, and when I went back to discuss my current research with them. It was also my good fortune to work and discuss my research interests with Botond Koszegi, Ulrike Malmendier, Emmanuel Saez, and Richard Zeckhauser in college. Their advice and assistance were of great help to me when applying to PhD programs.

Moreover, I am deeply thankful for Kate Easterbrook, Maria Polyakova, and Pian Shu, who might not have realized that they were being incredible peer mentors to me, not only when they were at MIT, but also after they graduated. I could always talk to them when I had concerns or

doubts and needed someone who had more experience for guidance, and all of them helped me keep my sanity in check. I am also indebted to Joshua Krieger, Manuel Hermosilla, and Wei Huang, great people outside my department who were always happy to discuss my research with me. Their generous help and support has been a boon to both me and my research.

My thesis would not have been possible without the creative and friendly atmosphere that my peers at MIT cultivated. I thank Jie Bai, Enrico Cantoni, Ludovica Gazze, Manisha Padi, Brendan Price, and Mahesh Shrestha, all wonderful empiricists to whom I've grown close ever since our first year, when we began sharing homework, laughter, and anxiety. I also feel grateful for friendships with German Gieczewski, Mark Hou, Gaston Illanes, Yan Ji, Yusuke Narita, and Hongkai Zhang, all wonderful people who are like a superb "think tank" that I can go to whenever I have questions on structural methods and/or theory. I thank Tianjiao Dai, Fei Song, and Mengxi Wu for our women's hang-outs and chats. I thank Manasi Deshpande for detailed guidance on econometrics in my first year and for the many helpful conversations we shared on the school shuttle. I am grateful for all the wonderful people that I have had the pleasure of sharing an office with. I thank Brendan Price for kindly letting me "free ride" on the desk lotteries, for constantly providing helpful feedback on my empirical research, and most importantly, for being a great friend with whom I can always share deep conversations and feelings. I thank Alex Bartik and Richard McDowell for sharing useful tips on research and productivity. I thank Adrien Auclert, Matt Rognlie, and Daan Struyven, who saved me a lot of time by generously sharing their presentation templates and teaching me life-saving tricks in Lyx and Latex. I thank Michael Stepner for having a picture by the desk portraying an ice cream cone saying "it will be fine", which brought me joy during the most stressful times. I am also indebted to all the staff at the department, especially Carl Anderson, Andrew Dorner, Gary King, and Beata Shuster, for kindly addressing my never-ending questions and requests. I am grateful to all these people for having made my years at MIT special.

Outside research work, I am grateful to all my friends for always being there for me. I have been lucky to have Xinyu Cao, Yu Shi, and Ying Zhao as my roommates, all of whom strictly dominate me in terms of cooking skills (among many other things) and have filled my life outside work with wonderful tastes, bright colors, and joyful cheers. I thank Danyan Zha, my close friend since college, for staying close all these years, for always being available for deep conversations, and for going out of her way to cheer me up during my moments of self-doubt and misery. I am thankful to Xiaoqian Yu, who always blew me away with her superb cooking skills, for all the delicious homemade dinners she invited me to, for the sometimes philosophical conversations, and most importantly, for inspiring me with her unmatched passion and curiosity to explore her interests in research and in life.

Last but by no means least, words cannot possibly express my gratitude to my family for their love and support throughout the years. I am thankful for my parents and grandparents for caring for me and raising me to be who I am, for encouraging me to do what I like even when it is not what they hoped for me to do, and for feeling proud of what I have become but never giving me any stress about what to accomplish. I thank my father, who is kind, wise, and down-to-earth, for never treating me like a child even when I was young, for teaching me the right values using his own words and deeds, and for supporting my pursuit of a PhD and providing encouragement during my moments of self-doubt. I thank my mother for taking good care of me, for letting her only child find her own place far away even though it is difficult for her, and for her passion in learning English and fitting better into my current life. I am more than grateful to my grandparents for sincerely adoring me ever since I was born (despite the prevalence of gender bias back then), for teaching me to be kindhearted and thoughtful, and for their never-ending unconditional love.

List of Tables

List of Figures

Chapter 1

Supply Response to Consumer Inertia: Strategic Pricing in Medicare Part D

1.1 Introduction

A growing literature has documented evidence that consumers in health insurance markets behave as if they face substantial switching costs when choosing health insurance plans. In this paper, I investigate whether private firms exploit this type of consumer inertia when setting prices for health insurance products and analyze the resulting welfare and policy implications. My empirical setting is Medicare Part D, a public program in which private insurers under contract with the government provide outpatient prescription drug insurance to more than 30 million Medicare beneficiaries (Hoadley et al., 2014).

Consumer inertia is a well-recognized feature of Medicare Part D, where standard enrollees only need to actively choose plans when they first join the program and are subsequently defaulted into previous choices unless they choose to switch. Hoadley et al. (2012) document the low frequency of switching despite large changes in plan premiums. Miller and Yeo (2015a) and Polyakova (2015) both identify substantial switching costs among Part D enrollees and estimate significant welfare loss because switching costs tend to prevent consumers from re-optimizing and to lock them into sub-optimal plans.

[0]I am greatly indebted to my advisors Amy Finkelstein, Heidi Williams and Paulo Somaini for their invaluable guidance and support throughout this project. I thank Nikhil Agarwal, Jie Bai, Joseph Doyle, Esther Duflo, Glenn Ellison, Sara Fisher Ellison, Jonathan Gruber, Gaston Illanes, Yusuke Narita, Manisha Padi, Ariel Pakes, Maria Polyakova, Jim Poterba, Brendan Price, Michael Whinston, Hongkai Zhang and participants at the MIT Public Finance and Industrial Organization field lunches and workshops for very helpful comments, suggestions and discussions. I also thank Mohan Ramanujan and Jean Roth at NBER, and Sarah Brunsberg and Kelly McMaken at ResDAC for their help with navigating the Medicare data. I gratefully acknowledge the generous support from the MIT George and Obie Shultz Fund.

Building on prior evidence of inertia in consumer demand in the context of Medicare Part D, my paper proceeds in three steps. First, I use administrative micro-data on Medicare beneficiaries and their plan choices to document descriptive facts that are consistent with the theoretical framework outlined by Klemperer (1987). The key idea of Klemperer (1987)'s framework is that, in the presence of demand inertia, insurers initially set low prices in order to "invest" in future demand before raising prices to "harvest" inertial consumers. I start by testing for this "invest-then-harvest" pricing pattern (also known as "bait-and-switch" or "bargains-then-ripoffs" pricing) by using a measure of markup or variable profit margin to eliminate potential confounding variations in cost and subsidy that affect insurers' pricing decisions. I document descriptive evidence consistent with insurers initially setting low prices in order to invest in future demand before later raising prices to harvest inertial consumers. Indeed, insurers charge lower markups when plans first enter and increase markups afterward; even the same insurer charges 148 dollars lower in annual markups for entrant plans than incumbent plans offering similar coverage. This difference in markup is over 30 percent of the average annual premium during the sample period. Such a striking invest-then-harvest pricing pattern rejects the null of no strategic response and provides suggestive evidence that insurers account for inertia when setting prices.

My finding confirms the public suspicion of Part D sponsors' "bait-and-switch" tactic. According to a Boston Globe article by Krasner (2006), the start-up of Medicare Part D was seen as "a once-in-a-lifetime opportunity" to attract new customers for Humana, one of the biggest insurers operating in Medicare Part D. The article notes that Humana tends to introduce plans at low prices, which are subsequently increased by a large margin.[1] One health-care analyst's response to Humana's pricing is very telling: "That's not an acceptable inflationary increase in prices. That's sucker them in and you just start raising the prices." A Humana spokesman blamed the price increase on the government's subsidy formula, but that contention was disputed by an actuary from the Centers for Medicare and Medicaid Services (Krasner, 2006).

Insurers' invest-then-harvest pricing has important welfare implications. On one hand, dynamic choice inefficiency arises as consumers' plan choices tend not to remain optimal after price changes, but switching frictions prevent many from taking advantage of re-optimization. On the other hand, it is an empirical question whether the invest or harvest incentive dominates and whether prices are higher or lower compared with the benchmark with no inertia. To explore the implications of these invest and harvest incentives for equilibrium pricing, I propose and estimate a dynamic model of insurers' pricing decisions that incorporates consumer inertia and adverse selection. Following Bajari et al. (2007)'s two-step estimation approach, I uncover insurers' discount factor, which tells us how much firms value future profits relative to current profits and quantifies their incentive to invest in future demand. As a result, the identification comes from the observed price or markup levels. Intuitively, the more insurers care about the future, the stronger invest incentive they face and the lower they set their premiums. The structural estimation reveals a strong invest incentive

[1]For example, premiums of Humana Standard, with over 2 million enrollees, increased by 60 percent on average between 2006 and 2007 and by 466 percent in seven regions (Krasner, 2006).

for insurers, which is consistent with low markups observed early on.

I apply this dynamic model to answer two important economic questions. First, what is the net effect of strategic pricing in response to inertia on equilibrium prices? In other words, do switching costs toughen or soften competition? It is an empirical question and depends on which of the following incentive dominates – the incentive to price low to invest in future demand, or the incentive to harvest inertial incumbent consumers. To quantify insurers' trade-off between these counteracting incentives, I decompose observed pricing patterns by comparing this dynamic model with a counter-factual benchmark without inertia and with a counter-factual in which insurers are myopic and do not invest in future demand. Comparisons show that on net, demand inertia toughens competition and reduces equilibrium prices in this setting, i.e. the invest incentive dominates the harvest incentive.

Finally, I apply the model to understand the potential role of government regulations. What are the price and welfare consequences of policies that could be used to constrain insurers' ability to exploit inertia using the "invest-then-harvest" pricing tactic? Even if there are policies that can effectively reduce the scope of investing and harvesting, the effects of government intervention are not directly intuitive. In fact, the effects are ambiguous ex-ante because pricing response to inertia creates two offsetting effects on consumer welfare. On one hand, price increases create dynamic choice inefficiency in consumer choice in the presence of switching frictions. On the other hand, the structural estimation suggests that inertia reduces prices as insurers face very strong incentives to invest in future profits. The desirability of government intervention depends on how effectively each policy can smooth prices over time without increasing average price levels. In order to assess desirability of government intervention, I first consider the most natural and straightforward policy, which is to cap the rate of annual premium increases at ten percent.[2] A second policy I consider is to offer a public option at a low price to compete with private insurers. An inexpensive public option would not only restrain room for increasing prices later on, but would also reduce the incentive to invest in future demand early on. I also consider the effects of removing risk sharing and fully exposing insurers to excessive losses and gains from their pricing decisions. A caveat with the last two policies is that public options and risk corridors are important policy instruments with many potential effects other than influencing insurer response to inertia, and my analysis here only speaks to one of many aspects of their effects. Among these policies, I find that a policy change to cap premium increases would be the most effective in improving consumer welfare by both lowering average premiums and smoothing prices over time. Offering a low-price public option lowers average prices and increases consumer welfare, but such welfare gains are dominated by the extra social cost of offering the public plan. Removing risk sharing has little impact on welfare but transfers money from the government to insurers because with risk sharing, taxes on excessive gains outweigh subsidies on excessive losses, both in the model and empirically.

[2]This policy experiment is similar in nature to the "Effective Rate Review" policy under the Affordable Care Act, which ensures that "in any State, any proposed rate increase by individual or small group market insurers at or above 10 percent will be scrutinized by independent experts to make sure it is justified". See CMS report: `http://www.cms.gov/CCIIO/Resources/Fact-Sheets-and-FAQs/rate_review_fact_sheet.html`.

My work builds on multiple literatures and contributes to the general understanding of supply in privatized health insurance markets, often with switching frictions. In recent years, we have seen a growing role for non-group insurance over typical employer-based and traditional government-provided insurance. For example, the Affordable Care Act establishes state-based health insurance exchanges where individuals and small business can choose from plans provided by private insurers. Therefore, it is increasingly important to understand how the private supply side operates in health insurance markets with switching frictions. First, my paper builds on the growing literature on consumer inertia and choice frictions in general[3] in insurance markets, including Medicare Part D and other health insurance settings. Polyakova (2015) models inertial consumers facing a switching cost and estimates switching costs to be two to four times as high as annual premiums among Medicare Part D enrollees. Ho et al. (2015) study inattention as a crucial driver of observed inertia and analyze its implications for prices, consumer out-of-pocket costs and government subsidy. Switching friction is a general feature of a variety of insurance markets with defaults, not limited to prescription drugs for the elderly. For example, Nosal (2012) estimates switching cost in Medicare Advantage, while Handel (2014) provides evidence of consumer inertia among a large firm's employees in choosing from employer-provided insurance plans. The contribution of my paper is to build on these studies and develop a structural model of dynamic pricing that allows me to simulate supply-side policy counter-factuals.

Furthermore, my work contributes to recent studies on insurance supply in privatized health insurance markets and its interactions with government regulations. Abaluck and Gruber (2013) conclude that the increased welfare loss from choice inconsistency in Medicare Part D is largely driven by supply-side changes, indicating the importance of understanding insurers' behavior. Ericson (2014) is the first to examine strategic pricing in response to inertia in Medicare Part D, documenting evidence of increasing premiums that is consistent with insurers exploiting inertia in pricing. Miller (2014) studies the role of inertia as well as government subsidy in insurers' plan offering and welfare in Medicare Advantage. Shepard (2015) studies insurers' competition over hospital networks in response to adverse selection. Starc (2014) analyzes the impact of imperfect competition on consumer welfare in Medigap. Decarolis (2015) identifies insurers' strategic response to the low-income subsidy system in their plan offering and pricing. Decarolis et al. (2015) and Miller (2015) study the welfare impacts of the current subsidy policy in Medicare Part D. Ericson and Starc (2015) examine the impacts of pricing regulations in Massachusetts's health insurance exchange. Miller and Yeo (2015b) analyze the effect of introducing a public option alongside private insurers in Medicare Part D. Building on these papers, my study investigates insurers' pricing response to inertia and analyzes policy counter-factuals where such strategic pricing interacts with pricing regulation, additional competition from a public option, etc.

Finally, this study is related to both theory and empirical literatures on firm strategy in the

[3]This growing body of literature examines inefficiency or sub-optimality of enrollees' plan choices in Medicare Part D (Heiss et al., 2008, 2013; Abaluck and Gruber, 2011, 2013; Kesternich et al., 2013; Kling et al., 2012; Ketcham et al., 2012, 2015).

presence of switching costs (Farrell and Klemperer, 2007, provide a review). Klemperer (1987) uses a two-period model to discuss the general intuition for firms' pricing incentives when consumers face switching cost – the incentive to invest in future demand by charging low prices, and the incentive to harvest inertial incumbent enrollees by charging high prices. Beggs and Klemperer (1992) illustrate how these two incentives interact in an infinite-period model with horizontal differentiation and infinite switching costs, and show analytically that the harvest incentive always dominates and switching costs soften competition. My study builds on Dubé et al. (2009) and Arie and Grieco (2014), both of which relax the crucial assumption of infinite switching cost and show that switching costs do not necessarily soften competition and can actually reduce equilibrium prices. Empirical evidence of strategic pricing in response to inertia is established in the bank deposit market (Sharpe, 1997), in the credit card market (Stango, 2002), in electricity markets (Waterson, 2003), in phone services (Knittel, 1997; Shi et al., 2006; Viard, 2007; Park, 2010), in the software market (Larkin, 2008), and recently in insurance markets (Ericson, 2014; Miller, 2014). My paper adds to the recent extension of this literature to the health insurance sector, an important market featuring consumer switching cost.

The rest of this paper is organized as follows. Section 1.2 describes the empirical setting and data. Section 1.3 discusses important intuitions from relevant theory papers. Section 1.4 presents descriptive evidence and discusses alternative explanations. Section 1.5 lays out the structural model. Section 1.6 describes the empirical strategy and presents estimation results. Section 1.7 conducts counter-factual analysis of policy experiments. Section 1.8 concludes.

1.2 Empirical Setting and Data

1.2.1 Institutional Features of Medicare Part D

Medicare is a public health insurance program for the elderly and the disabled in the US. Medicare Parts A and B have covered hospital and physician services since the program's inception in 1965, but prescription drug coverage was not provided until the introduction of Medicare Part D in 2006. Providing outpatient prescription drug insurance to the elderly and the disabled, Medicare Part D is a large program in terms of both enrollment and spending. The Congressional Budget Office reports that in 2014, there were 37 million Medicare beneficiaries enrolled in Part D (Hoadley et al., 2014), and the Congressional Budget Office (CBO) estimates the program cost around 65 billion dollars[4].

Unlike Medicare Parts A and B and other traditional government insurance programs, Part D is not delivered directly by the government, but rather by private insurers under contract with the government. These companies offer Medicare beneficiaries a choice between two types of prescription drug plans: bundled medical insurance and prescription drug benefits through the Medicare Advantage Prescription Drug plans (MA-PDs) that were in place prior to the deployment of Part D

[4]See Congressional Budget Office's Medicare Baseline Projection Reports in March 2015: `http://www.cbo.gov/publication/44205`.

under Medicare Advantage[5] and the stand-alone prescription drug coverage-only plans introduced in 2006. These stand-alone prescription drug plans (PDPs) are the focus of the present study.

Of all Medicare beneficiaries who have private prescription drug coverage, about 62 percent were enrolled in stand-alone prescription drug plans in 2012 (Hoadley et al., 2012). Stand-alone plans are offered in 34 geographically-defined markets within the continental United States. Plans in each market are offered by private insurers that are regulated by the government through the Centers for Medicare and Medicaid Services (CMS). In a typical market, approximately 20 firms offer more than 30 plans that are differentiated in terms of coverage. There are two broad types of prescription drug plans: basic plans that provide coverage actuarially equivalent to the required minimum coverage as per the defined standard benefit set by the CMS and so-called "enhanced-benefit" plans that offer offer supplemental coverage on top of the minimum required coverage. Supplemental coverage relative to the defined standard benefit includes reduced deductible, partial or full coverage in the donut hole, reduced cost sharing, etc.

There are two types of Medicare beneficiaries, and I conceptualize the demand systems for both types based on the institutional setting in the structural model. Standard beneficiaries become eligible for Medicare at age 65. Enrollment takes place annually during an open enrollment period. After standard enrollees become eligible and first join Part D, they have to actively choose their prescription drug plans. In years after this initial enrollment, standard beneficiaries are defaulted into their previous plans unless they actively switch. Low-income enrollees are eligible through the low-income subsidy (LIS) system. Unlike standard beneficiaries, low-income enrollees do not need to choose their own plans or pay their own premiums and out-of-pocket costs. Instead, the government pays all or part of their premiums and out-of-pocket costs and randomly assigns them to basic plans priced below market average. Within both groups of beneficiaries, a small fraction of beneficiaries leave and a slightly higher fraction of new beneficiaries arrive annually: the annual attrition rate is around eight percent for standard enrollees and around ten percent for low-income enrollees; the annual arrival rate is around ten percent for standard enrollees and around thirteen percent for low-income enrollees.

Insurers can enter any market and offer one or more plans in each market. Within each market, price discrimination is not allowed and the same plan must be offered at the same price to both incumbent enrollees and newcomers. Premiums are set annually in two components – a basic premium for basic coverage, which applies to all plans, and a supplemental premium for supplemental coverage, which applies only to enhanced-benefit plans. Basic and supplemental premiums are set simultaneously, but in different manners. Supplemental premiums are set directly by insurers, while basic premiums are set through a centralized bidding process. Each year before the new enrollment cycle starts, insurers submit bids to the CMS for basic premiums. The CMS then computes the basic premiums for each plan as the insurer's bid minus the national average bid plus some base

[5]Medicare Advantage (MA) is a health insurance program of managed health care (preferred provider organization (PPO) or health maintenance organization (HMO)) that serves as a substitute for Medicare Parts A and B Medicare benefits.

premium adjustments.[6] This is referred to as a bidding process because basic plans that bid below market average win a share of low-income enrollees.

Insurer revenue is generated by enrollee premiums and three types of government subsidies. The government provides these subsidies to mitigate adverse selection and to partially insure insurers against excessive losses. First, plans are paid risk-adjusted subsidies based on each enrollee's health status or risk in terms of drug spending. Second, individual reinsurance covers 80 percent of catastrophic spending. Finally, risk corridors provide risk sharing between the government and insurers – excessive losses are partially compensated and excessive profits are taxed. Despite the complexity of the subsidy regime, variable profit or markup is still an increasing function of premiums, given any enrollee. In other words, the standard trade-off between a higher markup versus a higher market share still holds in this setting.

1.2.2 Data

I use administrative data provided by the Centers for Medicare and Medicaid Services (CMS) on Medicare beneficiaries (henceforth "beneficiary files") and insurance plans (henceforth "plan files"). The beneficiary files cover a 20 percent random sample of Medicare beneficiaries from 2006 to 2011. For each year, this sample includes on average about 2.2 million standard enrollees and about 2 million low-income enrollees who are enrolled in stand-alone prescription drug plans. These beneficiary files include variables on enrollee demographics such as age, gender and race; on prescription drug plan choices in each year; and details on drug expenditures. The plans files include information on plan premiums and financial characteristics, such as a plan's deductible, gap coverage and tiered cost sharing. The recently released plan bridge files provide a crosswalk to unencrypted insurer and plan names, which allows me to identify the same insurer and plan across markets.

In descriptive evidence, I focus on a measure of markup or variable profit margin among standard enrollees to get rid of potential confounding variations in cost and subsidy that affect insurers' pricing decisions. I construct markups for each plan, averaged across its standard (low-income) enrollees, as the plan's premium plus the average risk-adjusted subsidy minus the expected cost, where the expected cost is defined as expected claims cost adjusted for pharmaceutical rebates, variable administrative cost and individual reinsurance. Details on the construction of expected claims cost will be discussed in Section 1.5.2. This markup measure incorporates the above-mentioned individual reinsurance and risk-adjusted subsidy. Individual reinsurance for catastrophic drug expenditure is computed using information on drug claims. To compute risk-adjusted subsidies in each year, I use the corresponding risk adjustment software from the CMS to compute the "risk score" for each enrollee. The CMS computes this risk score as a comprehensive summary of enrollee risk in terms of predicted drug spending and uses it to determine the amount of direct subsidy to pay insurers for each enrollee.[7]

[6] Base premium is about one third of the enrollment-weighted national average bid.

[7] The CMS computes risk scores in each year using the corresponding software to predict each beneficiary's

Table 1.1 reports market-year level summary statistics. Panel A of Table 1.1 reports summary statistics on market structure, including numbers of insurers and plans, the Herfindahl index, and enrollment-weighted average premiums. The average number of number of insurers offering stand-alone plans in a market is 21, and there is meaningful variation across markets, ranging from 11 to 29. Most markets have more than 30 stand-alone plans, about half of which are basic plans and the other half of which are enhanced-benefit plans. There is also substantial cross-market variation in these numbers of plans. Part D markets are on average moderately concentrated, with a Herfindahl index of 0.22. There is some variation in premium levels across markets and a general increasing trend over time: enrollment-weighted average premium increased from 329 dollars in 2006 to 507 dollars in 2011.

Panels B and C of Table 1.1 report summary statistics on standard and low-income Medicare beneficiaries at the market level, including population size, annual attrition and arrival rates, and the share choosing stand-alone plans. Numbers of beneficiaries correspond to the 20% random sample, and should be multiplied by 5 to get actual Medicare population sizes. Arrival and attrition rates are relative to one-year lagged population sizes. Shares in stand-alone plans are calculated as out of the entire population of standard or low-income Medicare beneficiaries, including those with stand-alone plans, those with bundled coverage under Medicare Advantage, those with coverage provided by employers or third parties, and those without any prescription drug coverage. For both types of enrollees, the average attrition and arrival rates, as well as the share choosing stand-alone plans, do not vary much over time within each market. Therefore, I take these rates as constant for each market for the structural estimation.

Table 1.2 reports summary statistics on stand-alone plans, first pooled across years and then by year. The first row summarizes pooled data from 2006 to 2011, and each other row corresponds to a single year. The first column is the total number of plans. There were 1429 plans in 2006, and the decline in the number of plans over time is mostly driven by consolidations rather than exits. The second column reports numbers of plan entries, which were concentrated in 2006 and 2007. There were relatively few entrants overall after 2009. The third column reports numbers of plan exits, which are relatively low compared with the number of plans. Total premium is the annual total premium, which is the combined value of basic premium and supplemental premium.

As a simplification of my analysis, I focus on strategic pricing in this paper and abstract away from a second strategic response to inertia: since price discrimination across new and old enrollees is banned, firms face an incentive to continuously introduce new plans that can be priced low to "invest" in future demand while charging higher premiums to incumbent consumers. Many plans

---

prescription drug spending in year t as a function of their inpatient and outpatient diagnoses from year t-1 and demographic information and uses these risk scores to determine risk-adjusted subsidy to insurers for each enrollee. To compute risk-adjusted subsidy in each year between 2006-2011, I use the corresponding RxHCC risk adjustment model from `http://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Risk-Adjustors-Items/Risk2006-2011.html` (retrieved June, 2014; last accessed October, 2015). Einav et al. (forthcoming) also use CMS software to compute risk scores as a proxy for individual predicted drug spending, and they use the 2012 model to compute risk scores for enrollees in 2006-2011 to consistently compare health status across years.

similar in coverage were forced to consolidate to comply with the "meaningful-difference" regulation, which was introduced in 2010 by the CMS to limit strategic entry behavior by requiring new plans to be sufficiently differentiated in coverage from existing plans by the same insurer. Although strategic plan entry is another important margins of firms' strategic behavior[8], I abstract away from it and focus on pricing here – i.e., conditional on the set of plans being offered, withdrawn and consolidated each year, how do firms price their plans? I include all prescription drug plans, including both non-consolidated and consolidated plans for the analysis. In order to link consolidated plans over time and to control for plan fixed effects for regression analysis later, I use each plan's most recent plan ID as its unique identifier.

1.3 Conceptual Framework

Klemperer (1987) discusses the general intuition for two counteracting incentives that firms face in the presence of consumer inertia, or when consumers behave as if they face switching costs. In the benchmark case without switching costs, demand in different periods is independent and so are firms' optimal strategies. However, when consumers face switching costs, demand is sticky over time, which creates two opposing incentives for firms: on one hand, firms want to charge low prices to "invest" in future demand, but on the other hand, firms want to charge higher prices to "harvest" inertial incumbent consumers. In a simple two-period model, firms only face the "invest" incentive in the first period and only face the "harvest" incentive in the second period. As a result, equilibrium price follows an "invest-then-harvest" pattern – firms charge low prices initially and increase prices afterward.

While a two-period model highlights the key trade-off firms face, a more realistic approximation of real markets is an infinite-period model, in which the invest and harvest incentives coexist. Beggs and Klemperer (1992) show how these two incentives interact in an infinite-period model with horizontal-differentiated products and consumers who are subject to switching costs. They solve for the unique Markov Perfect Nash Equilibrium under a critical assumption of perfect lock-in – i.e., consumers never switch because they are subject to infinitely large switching costs. In this equilibrium, prices are higher than the benchmark case without switching costs. This is not likely the case in a real-world context such as Medicare Part D, in which switching costs are not infinite as evidenced by the fact that some consumers do switch plans.

Dubé et al. (2009) relax this crucial assumption of infinite switching costs and examine an infinite-period model with switching costs and vertical differentiation. The authors establish the existence of a Markov Perfect Equilibrium and numerically solve for equilibrium prices as functions of switching costs. Applying this model to the markets of orange juice and margarine and using the empirically estimated level of switching costs in model simulations, they find equilibrium prices to be 18% lower than the case without switching costs, which reflects that the invest incentive dominates

[8]In a separate project, I document descriptive evidence consistent with such strategic entry and product proliferation.

because of "the strategic effects of firms lowering their prices to defend themselves against other firms' attempts to steal customers". Moreover, the authors show that depending on the magnitude of switching costs, equilibrium prices can be higher or lower than the case without switching costs. For example, when switching costs are sufficiently large or even infinite, the harvest incentive dominates and prices are higher than the case without switching costs. In other words, with finite switching costs, it is an empirical question whether the invest or harvest incentive dominates. My supply model in Section 1.5 similarly features finite switching costs, and I will investigate this question empirically in the setting of Medicare Part D.

Applying these intuitions to the setting of Medicare Part D, plans only face the invest incentive when they first enter and face both the invest and harvest incentives in subsequent periods. Therefore, prices or markups should be lower when plans initially enter than in subsequent periods, and they should also be lower compared to incumbent plans. Because the invest and harvest incentives coexist every year except for the first, whether inertia leads to higher or lower equilibrium prices is an empirical question.

1.4 Descriptive evidence

1.4.1 Switching Costs

To lay the framework for my analysis of strategic pricing, I first summarize the existing evidence of consumer inertia and present some corroborative evidence. Polyakova (2015) documents evidence of consumer inertia in Medicare Part D, and estimates the magnitude of switching costs to be two to four times as high as average annual premiums. As corroborative evidence of this type of consumer inertia, Table 1.3 shows, separately for different cohorts of standard beneficiaries, the enrollment shares as of 2011 by plans introduced in each year. Note that most plans were introduced in 2006 and fewer plans were introduced in subsequent years, partly contributing to higher shares in plans introduced in 2006 among all cohorts of enrollees. Interestingly, the percentage enrolled in the oldest plans (introduced in 2006) declines for younger cohorts of beneficiaries (84% among the 2006 cohort of enrollees and 72% among the 2011 cohort). Moreover, new cohorts of consumers are more likely to choose newly introduced plans. For example, the 2007 cohort is more likely to choose plans introduced in 2007 than the 2006 cohort. Similarly, the 2008 cohort is more likely than the 2006 and 2007 cohorts to choose plans introduced in 2008, and so forth. These statistics provide corroborative evidence that consumer inertia matters from the insurers' perspective.

1.4.2 Invest-then-Harvest Pricing

Figure 1-1, which displays enrollment-weighted markups over time for plans introduced in different years, shows two notable patterns. First, there is a general increasing trend: markups tend to increase as plans age. As I will show below, this increasing trend is robust across a variety of specifications. Second, within most years, entrants are priced lower than incumbent plans by a

substantial margin. These patterns are consistent with the invest-then-harvest predictions discussed in Section 1.3.

To formalize this invest-then-harvest pricing pattern, I estimate regressions that compare markups between plans that have just entered the market and incumbent plans.

$$Markup_{kjmt} = \alpha + \beta \mathbf{1}\{Entry\}_{kjmt} + Coverage_{kjmt}\gamma + \delta_m + \lambda_t + \xi_k + \epsilon_{kjmt} \tag{1.1}$$

$Markup_{kjmt}$ is markup averaged across standard enrollees for plan j offered by firm k in market m in year t. $\mathbf{1}\{Entry\}_{kjmt}$ is a dummy indicating whether plan j first entered market m in year t. $Coverage_{kjmt}$ includes plan features, such as deductible level, whether the plan offers gap coverage and tiered cost sharing.

Table 1.4 reports ordinary least squares estimates of β from four specifications. Column (1) shows the raw correlation between markup and the entry dummy. Column (2) adds market and year fixed effects to account for potential differences across markets and over time that affect both entry and plan pricing. Column (3) adds insurer fixed effects to account for unobserved time-invariant heterogeneity at the insurer level. Column (4), which adds controls for plan coverage, compares plans within insurer and controlling for coverage. Consistent with the prediction in Section 1.3, these regression estimates suggest that the same insurer charges significantly lower markups on new plans than on incumbent plans with the same characteristics. Within the same insurer and year controlling for plan characteristics, markup is 148 dollars lower on entrant plans than incumbent plans, which is high relative to the average annual premium of 372 dollars.

1.4.3 Addressing Alternative Explanations

While the empirical invest-then-harvest pricing patterns documented in Section 1.4.2 are consistent with firms exploiting consumer inertia to maximize profits dynamically, such pricing patterns might also be rationalized by alternative explanations. First, since Medicare Part D is a new market, insurers might not be well-informed about cost, which could cause them to under-price initially and adjust prices upward as they learn about cost over time. Relatedly, in a learning-by-doing story, insurers might set low prices and invest in market shares in order to learn more quickly about cost. Finally, as Decarolis (2015) shows, the low-income-subsidy system also contributes to premium increases over time.

I have no intention of running a horse race to rule out these potential alternative explanations. It seems likely that insurer responses to inertia as well as these alternative stories are all empirically relevant to some extent. However, I argue that these alternative stories seem unlikely to explain the pricing patterns documented in Section 1.4.2. First, in the story of learning about cost, it is not clear why firms would systematically underestimate cost. Moreover, I find that even within the same firm, significantly lower markups are charged for entrant plans than for incumbent plans, contradicting the notion of learning about cost. Such within-firm comparisons also help contradict the learning-by-doing story.

In order to formally assess the robustness of my results to these alternative explanations, I test for evidence of the invest-then-harvest pricing pattern on subsamples of plans for which these alternative explanations are arguably not relevant. Table 1.5 summarizes these estimates for Equation 1.1 using subsamples. To address the first alternative explanation that firms learn about cost, the first three columns of Table 1.5 focus on subsamples of plans offered by insurers who are arguably well-informed about the cost of supplying prescription drug insurance to Medicare beneficiaries. Column (1) restricts the sample to plans offered by insurers that were major sponsors[9] of Medicare Advantage prior to 2006 and that provided prescription drug insurance bundled with medical insurance to Medicare beneficiaries. Column (2) restricts the sample to plans offered by insurers with prior experience in Medicare Advantage. Column (3) restricts the sample to plans offered by insurers with prior experience in providing insurance coverage to Medicare beneficiaries. These three subsamples are not subject to the concern that insurers are not informed about cost. To address the second alternative explanation or the learning-by-doing story, Column (4) reports estimates using a subsample of plans offered by insurers that are already experienced in Part D. Specifically, I assume that the benefit from such learning diminishes after the insurer serves many enrollees, which motivates restricting the sample of plans to those offered by insurers that have served at least 5000 enrollees before in the same market. Finally, to address potential confounding effects from the low-income-subsidy system, I use the subsample of enhanced-benefit plans, which are not eligible to receive low-income enrollees.

As shown in Table 1.5, the estimated coefficient–or the difference in annual markup between entrants and incumbent plans, holding the insurer and plan coverage as fixed, ranges from -$134 to -$187, which is not much different from the estimate of -$148 on the full sample of plans. In other words, the result that markups are much lower on entrant plans than incumbent plans is robust to focusing on subsamples of plans where these alternative explanations are less relevant. This suggests that the empirical pricing pattern we observe is largely driven by strategic responses to consumer inertia rather than by these alternative explanations.

1.5 Model

My descriptive evidence in Section 1.4 rejects the null of no strategic response to inertia and is consistent with firms exploiting consumer inertia to maximize profits dynamically. To explore insurers' trade-offs between the invest and harvest incentives, I develop a dynamic model of insurers' pricing decisions that incorporates demand inertia and adverse selection. Structural estimation of this model in Section 1.6 uncovers insurers' discount factor, which quantifies the strength of the invest incentive. In Section 1.7 I further decompose observed pricing patterns to quantify insurers' trade-offs between invest and harvest incentives by comparing this model with a counter-factual benchmark with no inertia and with a counter-factual case where insurers are myopic and face no invest incentive. Finally, in Section 1.7 I simulate the price and welfare consequences of several

[9]Seven biggest sponsors in terms of market shares as of 2005 according to Gold (2006)

policy experiments that could be used to constrain insurers' ability to exploit inertia.

1.5.1 Demand

As described in Section 1.2, there are two types of Medicare beneficiaries – standard enrollees and low-income enrollees – and I conceptualize the demand system for each type based on the institutional setting. I start with demand for standard beneficiaries, who make their own plan choices and are defaulted into their previous choices unless they actively switch. Since the focus of this study is on understanding firm pricing, I use the demand model and estimates from Polyakova (2015) for standard enrollees. In her model, standard enrollees are myopic[10] and choose a plan to maximize current utility, subject to switching costs. Let i denote individual, j plan, k insurer, m market (region) and t year. Individual i's utility[11] from choosing plan j in year t is as follows, where p_{kjt} is annual premium, ϕ_{kjt} is characteristics of the plan, and $\mathbf{1}\{\text{Default}\}_{ikjt}$ is an indicator of whether consumer i is defaulted into plan j at time t. This default dummy is omitted for new enrollees, who are not defaulted into any plans.

$$u_{ikjt} = -\alpha p_{kjt} + \beta_{it}\phi_{kjt} + \gamma_{it}\mathbf{1}\{\text{Default}\}_{ikjt} + \lambda_{it}\mathbf{1}\{\text{Insurer}\}_{k} + \epsilon_{ikjt} \tag{1.2}$$

In this logit model, ϵ_{ikjt} is independent and identically distributed with a Type 1 Extreme Value distribution function.[12] ϕ_{kjt} includes the following characteristics that are feasibly observed by beneficiaries when they are making their choices: the deductible, the initial coverage limit, whether the plan offers coverage in the gap, whether the plan uses fixed dollar co-payments or coinsurance percentages, and whether the plan is eligible for getting low-income subsidy enrollees. Preferences over plan coverage β_{it} depend on the individual's demographics and health risk, D_{it} =

[10] One concern is that consumers can be forward-looking about changes in plan prices and their own health risk in the future. The latter is allowed by controlling for age, while consumers forward-looking about future price changes will be less likely to start with a cheap plan, which should dampen insurers' incentives to invest in market shares. However, as Handel (2014) argues, consumers make very poor decisions if we consider forward-looking demand. Moreover, dynamic demand adds additional complexity while dynamic supply is already computationally demanding. In fact, in dynamic games literature on durable goods, experience goods and network goods, it is fairly standard to assume myopic demand.

[11] Polyakova (2015) points out "this formulation assumes that individuals choose the option with the highest "perceived" utility, which may not necessarily correspond to the highest "objective" valuation of plans as financial contracts (indeed, Abaluck and Gruber (2011, 2013) suggest that beneficiaries are choosing their plans inconsistently with the objective efficiency frontier)".

[12] Polyakova (2015) models choice among stand-alone prescription plans and does not include the outside option. I use a similar linear regression to predict, separately for incumbent and new beneficiaries, the share choosing to enroll in stand-alone plans instead of bundled coverage or no coverage, based on prices, number of plans, market fixed effects, etc. Estimates show that market fixed effect explains 66 percent for newcomers and 95 percent for incumbent enrollees. Details will be discussed in the Appendix. Alternatively, I can re-estimate the demand model using a nested logit model, in which beneficiaries first choose between not enrolling in prescription coverage, enrolling in bundled coverage and enrolling in stand-alone coverage, and then choose a plan if they choose any coverage in the first step. Since I focus on supply of stand-alone plans, I choose to take the simplistic approach instead to abstract away from the complexity of modeling demand for both bundled and stand-alone coverage.

$\{\text{age}_{it}, \text{gender}_i, \text{race}_i, \text{risk score}_{it}, \text{esrd indicator}_{it}\}$, where risk score is a measure of each beneficiary's health risk in terms of drug spending and esrd indicator is a dummy for end-stage renal disease. There are random coefficients in preferences over deductible, initial coverage limit and gap coverage: $\beta_{it} = \pi^{\beta} D_{it} + \psi_i^{\beta}$, where $\psi_i^{\beta} \sim N(\psi^{\beta}, \sigma^2)$. Switching costs γ_{it} and preference over insurers λ_{it} also depend on individual demographics and health risk: $\gamma_{it} = \pi^{\gamma} D_{it} + \psi^{\gamma}$, and $\lambda_{it} = \pi^{\lambda} D_{it} + \psi^{\lambda}$.[13]

Standard consumer i's probability of choosing plan j depends on plan features as well as the default plan l, which I denote as $P_j(p, l)$ and which follows the logit form. Active consumers without default plans face an unconditional choice probability $P_j(p)$. Aggregating individual choice probabilities, the share of standard enrollees choosing plan j in year t $S_{kjt}()$ is derived as follows.

$$S_{kjt}(p, \mathbf{S}_{t-1}) = \frac{1-\lambda}{1-\lambda+\mu} \sum_{l \in J(m)} S_{lt-1} P_j(p, l) + \frac{\mu}{1-\lambda+\mu} P_j(p) \tag{1.3}$$

$S_{kjt}()$ is a function of prices p, lagged shares S_{t-1}, the attrition rate of standard enrollees λ and the arrival rate μ. Because attrition and arrival rates do not vary much empirically within each market, I take λ and μ as exogenous and suppress their notations for this share function $S_{kjt}(p, \mathbf{S}_{t-1})$. Let $J(m)$ denote the set of plans in market m. The first component of Equation 1.3 is the summation of shares across incumbent enrollees defaulted into different plans ($\sum_{l \in J(m)} S_{lt-1} P_j(p, l)$) weighted by the fraction of incumbent beneficiaries ($\frac{1-\lambda}{1-\lambda+\mu}$), while the second component represents the share among new beneficiaries without defaults (or the unconditional choice probability $P_j(p)$) weighted by the fraction of new beneficiaries ($\frac{\mu}{1-\lambda+\mu}$). In other words, other than prices, lagged market shares are important in determining current shares of standard enrollees because incumbent consumers' choice probability $P_j(p, l)$ is biased toward the lagged choice or default plan l. The importance of lagged shares is slightly depreciated by attrition of incumbent enrollees (λ) and arrival of new enrollees (μ): empirically $\frac{1-\lambda}{1-\lambda+\mu}$ is approximately 0.90 in my data.

Although only around 20 percent of Medicare beneficiaries are low-income enrollees, they account for over 40 percent of enrollment in stand-alone prescription plans. Therefore, it is important to include profits from the population of low-income enrollees when modeling insurers' profit maximization problem. Unlike standard beneficiaries, low-income enrollees do not need to choose their own plans. Instead, the Centers for Medicare & Medicaid Services randomly assigns them to eligible plans when they first qualify for the low-income subsidy or when their previous plans are no longer eligible for receiving low-income enrollment. Low-income enrollees are evenly divided into eligible plans – basic plans priced below market average – except that an insurer eligible both in the last period and the current period keeps its incumbent low-income enrollees on top of this random assignment.[14] Based on how low-income enrollees are automatically allocated across plans in reality,

[13] More specifically, preference over the two biggest insurers depend on D_{it} while preferences over other insurers follow the form of standard fixed effects.

[14] In 2006, low income enrollees were randomly assigned to basic plans pricing below market average. In subsequent years, insurers keep previously assigned low-income enrollees conditional on having a basic plan pricing below the benchmark, where the benchmark is weighted by lagged low-income enrollment. Except for these enrollees who stay with a below-benchmark basic plan, low-income enrollees are again randomly assigned to basic plans pricing below

I model their discrete and mechanical demand, which depends on lagged low-income shares other than current plan bids for basic premiums.

Let λ^{LIS} denote the attrition rate of low-income enrollees and μ^{LIS} the arrival rate. The share of low-income enrollees assigned to basic plan j in year t $S^{LIS}_{kjt}()$ is computed as follows, where $\omega = \frac{1-\lambda^{LIS}}{1-\lambda^{LIS}+\mu^{LIS}}$ is the share of incumbent enrollees and $1-\omega = \frac{\mu^{LIS}}{1-\lambda^{LIS}+\mu^{LIS}}$ is the share of newcomers. The benchmark $\bar{b}_{mt}$ is the average bid among basic plans weighted by lagged low-income enrollment. $J_B(m)$ is the set of basic plans in market m. $N_{mt} = \sum_{l \in J_B(m)} \mathbf{1}\{b_{lt} \leq \bar{b}_{mt}\}$ is the number of basic plans pricing below benchmark. $S^{Reassign}_{mt}$ is the share of low-income enrollees who need re-assignment because their former insurers lost below-benchmark status.[15]

$$S^{LIS}_{kjt}(p, \mathbf{S}_{t-1}) = \begin{cases} 0 & b_{kjt} > \bar{b}_{mt} \\ (\omega S^{Reassign}_{mt} + 1 - \omega)/N_{mt} & b_{kjt} \leq \bar{b}_{mt}, b_{kj't-1} > \bar{b}_{mt-1} \forall j' \in J_B(k) \\ \omega S^{LIS}_{kj,t-1} + (\omega S^{Reassign}_{mt} + 1 - \omega)/N_{mt} & b_{kjt} \leq \bar{b}_{mt}, b_{kjt-1} \leq \bar{b}_{mt-1} \\ \omega S^{LIS}_{kj',t-1} + (\omega S^{Reassign}_{mt} + 1 - \omega)/N_{mt} & b_{kjt} \leq \bar{b}_{mt}, b_{kjt-1} > \bar{b}_{mt-1}, b_{kj't-1} \leq \bar{b}_{mt-1} \end{cases} \tag{1.4}$$

In the first case in Equation 1.4, basic plan j prices above the benchmark and receives no low-income enrollees. In the second case, basic plan j prices below benchmark, and the insurer k had no plans pricing below benchmark in the previous year. In this case, plan j receives an even share of incumbent low-income enrollees who need to be re-assigned ($\omega S^{Reassign}_{mt}$) plus new low-income enrollees ($1-\omega$). In the third case, basic plan j prices below benchmark, and it also priced below benchmark in the previous year. In this case, plan j receives an even share of randomly assigned enrollees as in the second case, while keeping its incumbent low-income enrollees ($\omega S^{LIS}_{kj,t-1}$). In the final case, basic plan j prices below benchmark, and another plan j' by the same insurer prices above benchmark but priced below benchmark in the previous year. In this case, plan j receives an even share of randomly assigned enrollees as in the second case, plus it keeps incumbent low-income enrollees within the same insurer ($\omega S^{LIS}_{kj',t-1}$).

1.5.2 Cost

Medicare Part D is a health insurance market with the potential for adverse selection. In my setting, health risk correlates with consumer preference as well as switching costs, as suggested by the demand estimates. Moreover, Handel (2014) and Polyakova (2015) both conclude that the

benchmark. Low-income enrollees can choose to opt out of their assigned plans and choose a different plan and pay the difference in premiums. Among low-income enrollees choosing stand-alone prescription drug plans, the empirical fraction of "choosers" who have ever opted out increases over time from around 6% in 2006 to around 20% in 2010 (Summer et al., 2010). Such opting out behavior is not flagged in the administrative data, and cannot be identified except for those who choose plans not eligible for low-income enrollees. Decarolis et al. (2015) model the demand for such "choosers" based on the subsample for which opting out is observed in the data. I choose to model only the random assignment and not such opting out behavior because it is not essential to my focus of strategic pricing in response to inertia among standard enrollees.

[15] This can be computed as $S^{Reassign}_{mt} = \sum_{l \in J_B(m)} \mathbf{1}\{b_{kj't} > \bar{b}_{mt}, \forall j' \in J_B(k)\} S^{LIS}_{klt-1}$.

interaction between adverse selection and switching costs has important welfare implications, which depend on the specific market setting. In order to account for this well-recognized issue, I follow Starc (2014) to model adverse selection and allow claims cost to be endogenous to price. This complication is a nuance rather than the focus of my model.

Conceptually, in the presence of adverse selection, consumers with different risks in terms of drug expenditure sort into different plans based on coverage and prices. As a result, insurers' claim costs depend on the types of consumers each plan gets, and therefore they are endogenous to price, which affects consumers' sorting behavior. I start with constructing the claims cost measure at the level of individual-plan pairs, before formulating endogenous claims cost at the plan level. For each enrollee's drug expenditure, an insurer is responsible for covering the remaining after subtracting the part paid out-of-pocket by the enrollee, the part covered by the government and the part rebated by the pharmacy and pharmaceutical manufacturers. As described in Section 1.2, the government provides three types of subsidies: risk-adjusted subsidies based on each enrollee's health risk, individual reinsurance for catastrophic drug spending, and risk corridors that partially compensate excessive losses and tax excessive profits. Individual reinsurance lowers insurer claims cost, while risk-adjusted subsidy and risk corridors do not enter claims cost directly but enter the profit function in other ways in Section 1.5.3. As for rebates, the claims data already incorporates rebates from pharmacies but not rebates form pharmaceutical manufacturers, which I will adjust for later using summary statistics from government reports.

Each individual's claims cost is constructed as total drug expenditure net of pharmacy rebates, individual reinsurance from the government and the enrollee's out-of-pocket spending. Intuitively, individual i's claims cost to plan j offered by insurer k is a function of both plan coverage (X_{kjt}) and consumer demographics and health risk (W_{it}). Insurer fixed effects δ_k are included to account for unobserved time-invariant heterogeneity in coverage or quality, such as broadness of pharmacy network, generosity of formularies and quality of customer service, which vary across insurers.

$$c_{ikjt}(X_{kjt}, W_{it}) = \alpha + X_{kjt}\beta + W_{it}\gamma + \delta_k + \xi_{ijt} \tag{1.5}$$

There are two important identifying assumptions embedded in this individual claims cost function. First, selection only works through observables. This is not a terrible assumption in this setting, as W_{it} includes enrollee risk score, which is a comprehensive risk measure in terms of expected drug spending. Second, this function assumes that there is no plan-individual specific moral hazard: while cost may depend on the plan's characteristics, the unexplained part of an individual's cost does not depend on the plan chosen. To the extent that the variation in plan coverage is well captured by both the detailed plan characteristics X_{kjt} and the insurer fixed effects included in the cost function, this assumption is justified because the notion of moral hazard in insurance markets typically refers to the fact that enrollees utilize more services with more generous coverage as they face a lower marginal price.

Based on this individual cost function and the demand system outlined in Section 1.5.1, I aggregate individual costs to get plan-level expected claims costs as follows. The cost (per enrollee)

of plan j offered by insurer k depends on its coverage as well as the average characteristics of its enrollees, which is endogenous to price and the resulting selection.

$$C_{kjt} = \alpha + X_{kjt}\beta + E[W_{it}|\text{Choose Plan j}]\gamma + \delta_k \tag{1.6}$$

For tractability of the supply model, in which the state space includes lagged market shares by consumer type, I discretize types of standard enrollees based on risk score and gender.[16] In other words, instead of controlling for enrollee characteristics W_{it} continuously in estimating the individual cost function 1.5, I drop W_{it} and estimate this function separately for each type of consumers. Within each type, consumers are assumed to be homogeneous (up to random coefficients in preferences) with cost realizations drawn from a common distribution. Low-income enrollees are taken as homogeneous with a common cost distribution because of the automatic random allocation.

In order to get expected variable cost at the plan level, I adjust for two sources of variable cost other than expected claims cost. First, I take variable administrative cost to be 16% of claims cost based on estimates from other studies on similar markets: Starc (2014) estimates administrative cost to be 16% of premiums on average in Medigap; Ho et al. (2015) use data from the National Health Expenditure Survey to compute administrative cost to be 14-16% of total cost, and 16-19% of non-administrative cost, averaged across Medicare Advantage plans and plans in Medicare Part D. Second, I take rebates from pharmaceutical manufacturers to be 10% of total drug spending based on summary statistics from government reports: Boards of Trustees (2012) reports that the average manufacturer rebate rate, as a percentage of total prescription drug costs, ranged between 8.6 percent and 11.3 percent between 2006 and 2010.[17]

Other than expected variable cost, cost realizations also matter for insurers' dynamic profit maximization because of risk corridors. Risk corridors provide risk sharing between the government and insurers by partially compensating excessive losses and taxing excessive profits. In order to account for this when estimating the supply model, I randomly draw realized cost from normal distributions centered around expected cost and average across these random draws to get expected insurer profit. The standard deviation of this distribution of plan-level cost is calculated using standard deviation of individual cost and plan enrollment.

1.5.3 Supply

My model of insurers' strategic response to inertia in pricing builds on the work of Beggs and Klemperer (1992) and Dubé et al. (2009) and incorporates new features based on my empirical setting. As in Dubé et al. (2009), I consider an overlapping generations model with imperfect lock-in. In this model, both single-product and multi-product insurance firms offer differentiated plans

[16] Although the risk score is computed using demographics including gender, an OLS regression shows a small difference in cost by gender even conditional on risk score. To fully capture the cost difference across genders, I group standard consumers by gender in addition to risk score. I do not divide consumers by other demographic variables because they do not appear significant in predicting cost after controlling for risk score.

[17] There is no need to adjust for pharmacy rebates, which are already net out in the claims data.

and compete for consumers subject to switching cost. In each period, a fraction of old consumers leave the market and new consumers arrive.

In order to focus on insurers' dynamic pricing decisions in the presence of inertia, I make the following simplifying assumptions. First, I take plan characteristics as given, which is not a bad approximation in my setting, as empirically insurers tend to adjust premiums instead of plan characteristics. Second, I take market structure as given and abstract away from strategic entry. This assumption is less innocuous because entry does happen empirically. Since price discrimination is not allowed, firms face an incentive to continuously introduce new plans that can be priced low to invest in future demand while also charging higher premiums to incumbent consumers. However, the "meaningful-difference" regulation essentially put an end to such strategic entry, and the number of plans remains quite stable afterward. Although the timing of this regulation is close to the end of the sample period, it suffices in confirming that there will not be unobserved entry after the sample period, as the supply model involves forward simulation for many more periods. In other words, entry is less common in recent years and will continue to be less common looking-ahead. Relatedly, I only model variable profits of insurers and not fixed costs, which are sunk costs and therefore not relevant for pricing decisions. Third, I assume that insurers take the regulation environment as given, without foresight of future policy changes. Finally, I assume pricing decisions are separately made for stand-alone prescription drug plans and Medicare Advantage plans that bundle medical insurance and prescription benefits. Although cannibalization between these two segments is a concern, it is not essential to the invest-then-harvest pricing story, and I focus on the pricing of stand-alone prescription drug plans and abstract away from modeling the demand and supply for MA-PD plans.

1.5.3.1 Value Function

Insurers account for consumer inertia and choose bids and supplemental premiums to maximize discounted profits. As calculated in Equation 1.7, $V(\sigma_k, \sigma_{-k}, \delta, \mathbf{SV}_{mt_0})$ is the expected present value of profit for firm k in market m in year t_0, where σ denotes pricing strategies, δ denotes insurers' annual discount factor, SV includes the state variables and Π denotes annual variable profit.

$$V(\sigma_k, \sigma_{-k}, \delta, \mathbf{SV}_{mt_0}) = \mathrm{E}[\sum_{t=t_0}^{\infty} \delta^t \Pi_{kt}(\sigma_k, \sigma_{-k}, \mathbf{SV}_{mt})] \tag{1.7}$$

Since firms account for demand inertia, profits and pricing strategies are state-dependent. Besides exogenous state variables, including plan characteristics and enrollee characteristics, because of inertia (Beggs and Klemperer, 1992), SV_{mt} also includes lagged market shares by consumer type, which evolve deterministically based on the demand system in Section 1.5.1. Insurer profits and pricing strategy depend on lagged market shares among both standard and low-income enrollees. First, lagged standard enrollee shares matter because, intuitively, the harvest incentive depends on how many standard enrollees an insurer has locked in. Second, lagged low-income enrollee shares

also matter for insurer profits and pricing strategy due to the way low-income enrollees are assigned as described in Section 1.5.1. In the presence of adverse selection, different types of enrollees differ in cost and demand. Therefore, lagged shares of different types of enrollees arguably affect insurer pricing differently, and I include lagged shares by consumer type to account for this. As a side note, lagged market shares by consumer type pin down expected cost, and therefore there is no explicit cost term in the value function.

Insurer k's pricing strategy σ_k for all its plans ($j \in J(k)$) is a mapping from states $\mathbf{SV}_{mt}$ to bids for basic premiums (b_{jt}) and supplemental premiums (PS_{jt}). More specifically, σ_k includes bids for basic premiums for each plan $b_j(\mathbf{SV}_{mt}, \epsilon_{jt}) = f(\mathbf{SV}_{mt}) + \epsilon_{jt}$ and supplemental premiums for each enhanced-benefit plan $PS_j(\mathbf{SV}_{mt}, \epsilon_{jt}) = h(\mathbf{SV}_{mt}) + \epsilon_{jt}$.[18]

1.5.3.2 Annual Profit Function

Firm k's annual variable profit consists of profits from all its plans in market m, $j \in J_m(k)$. In other words, multi-product firms jointly maximize profit across all plans. Plan j's total profits include profits from different groups of enrollees Π^{θ}_{kjt}, where θ represents discrete types of standard enrollees and the group of low-income enrollees. $\Gamma(.)$ is a function representing adjustments from the risk corridors, which partially compensate for excessive losses and tax excessive gains.

$$\Pi_{kt}(b, PS, \mathbf{SV}_{mt}) = \sum_{j \in J(k)} \Gamma(\sum_{\theta} \Pi^{\theta}_{kjt}(b, PS, \mathbf{SV}_{mt})) \tag{1.8}$$

Plan j's (pre-risk-corridor) profit from each enrollee type can be calculated as enrollment times markup.

$$\Pi^{\theta}_{kjt}(b, PS, \mathbf{SV}_{mt}) = M^{\theta}_{mt} S^{\theta}_{kjt}(b, PS, \mathbf{S}_{t-1}) Markup^{\theta}_{kjt}(b, PS, \mathbf{SV}_{mt}) \tag{1.9}$$

M^{θ}_{mt} denotes the population of each type of enrollee within market m in year t. S^{θ}_{jt} denotes shares of each type of enrollee choosing plan j in year t, which is calculated based on the demand system as shown in Equations 1.3 and 1.4. Markup on each enrollee type is equal to total premium minus expected cost plus risk-adjusted subsidy. Expected cost is constructed in Section 1.5.2, while premiums and subsidies are computed following the actual process of setting prices and government subsidy. Each year before enrollment takes place, for each plan $j \in J_m(k)$, insurer k submits a bid b_{jt} for its basic premium and sets directly the supplemental premium PS_{jt} if it is a enhanced-benefit plan. The CMS computes basic premium as $PB_{jt} = b_{jt}-$ (national average bid $-$ base premium), where base premium is a fixed fraction of national average bid. Enrollees face a total premium $p_{jt} = PB_{jt} + PS_{jt}$, where $PS_{jt} = 0$ for basic plans. In order to mitigate adverse selection in this market, the government computes a risk score for each enrollee r_{it}, based on demographics and medical history, and pays risk-adjusted subsidy $r_{it}b_{jt} - PB_{jt}$ to the insurer. For an average enrollee

[18]The interpretation for ϵ_{jt} is managerial mistake or specification error, and is assumed to be drawn independently across plans and years from a normal distribution centered around zero.

with a risk score of one, the sum of basic premium and risk-adjusted subsidy is equal to the bid for basic premium. In other words, although risk-adjusted subsidy is endogenous to plan bids for basic premiums, insurers still face the standard trade-off between a higher markup (as a result of both a higher enrollee premium and a higher government subsidy) and a higher market share when setting prices.

I restrict insurers' strategies to be Markovian because the full set of dynamic Nash equilibria is unbounded and complicated. The Markov-Perfect Nash Equilibrium requires $V(\sigma_k, \sigma_{-k}, \delta, \mathbf{SV}_{mt}) \geq V(\sigma'_k, \sigma_{-k}, \delta, \mathbf{SV}_{mt})$ given competitors' strategies σ_{-k} for all states and alternative strategies σ'_k, i.e. each insurer's strategy has to be optimal given competitor's strategies.

1.6 Structural Estimation

1.6.1 Demand Estimation

Table 1.6 reports Polyakova (2015)'s simulated maximum-likelihood estimates on a few important demand parameters. Estimates for the price coefficient and the switching cost dummy are relatively robust across specifications. Besides including a rich set of plan features in all specifications, Columns (3) and (4) include more insurer fixed effects than the first two columns[19] to address the concern with unobserved insurer quality affecting both premiums and demand. Moreover, Columns (2) and (4) use lagged cost as an instrument for plan premium to address the concern with unobserved plan quality affecting both premiums and demand. Both instrumenting and controlling for more insurer fixed effects only increase the magnitude of the premium coefficient slightly, which confirms that including rich plan characteristics leaves little room for unobserved insurer and plan quality to affect both pricing and demand.

Controlling for more insurer fixed effects reduces the magnitude of the intercept for the switching cost term from 5.6 to 5.1, or reduces the implied switching cost for a 75 year-old female enrollee with average risk from \$1330 to \$1164. This difference suggests that there is unobserved quality at the insurer level that enrollees persistently value over time, and it is important to account for those unobservables with insurer dummies. Therefore, I choose the last specification with instruments for premium and ten insurer fixed effects as input for my supply estimation.

1.6.2 Cost Estimation

Figure 1-2(a) visually summarizes the individual cost estimation results. This figure reports, for each type of enrollee, expected cost to a basic plan offering minimum coverage and to an enhanced-benefit plan with more generous coverage (zero deductible and gap coverage). Standard enrollees are divided into groups with low, medium and high risk scores. In addition to these expected cost measures, Figure 1-2(b) adds switching cost in dollars and willingness-to-pay for more generous

[19]Columns (1) and (2) include three insurer fixed effects by including dummies for the two biggest insurers (the omitted category consists of all other insurers), while Columns (3) and (4) include three insurer fixed effects by including dummies for the nine biggest insurers (the omitted category consists of all other insurers).

coverage. There is a significant correlation between cost and willingness-to-pay for extra coverage across different types of plans. There is also a small positive correlation between cost and switching cost, but this is less strong than the correlation between cost and willingness-to-pay.

Figure 1-2 pools female and male enrollees for simplicity, and Appendix Figure 1-3 also breaks down by gender in addition to risk score. The patterns look similar – there is a lot of cost heterogeneity across enrollee types, and cost correlates strongly with preference and weakly with switching cost. More details on the estimation results are reported in Appendix Table 1.13.

1.6.3 Supply Estimation

Estimating parameters of dynamic games and computing equilibria are computationally demanding (Benkard, 2004; Bajari et al., 2010). The large number of insurers in Part D markets makes it even more difficult computationally. Instead of solving for the equilibrium of the supply model, I follow Bajari et al. (2007)'s two-step approach to uncover insurers' valuation of future profits. Essentially, this approach minimizes the violation of insurer rationality by finding the parameter value or insurers' discount factor such that the observed pricing strategies are closest to equilibrium strategies. This approach is implemented in two steps. In the first step, I empirically estimate how insurers price their plans by regressing premiums on relevant state variables. Such reduced-form estimates empirically correlate insurers' actions to states and characterize insurers' strategies $\sigma(SV)$, which are also referred to as the empirical policy functions.

In the second step, I take competitors' strategies as given by these empirical strategies characterized in the first step and forward simulate to construct insurers' discounted profits $V(\sigma_k, \hat{\sigma}_{-k}, \delta, \mathbf{SV})$ as in Equation 1.7 given a discount factor. This simulated value function can be constructed using both each insurer's empirical strategy and alternative strategies. Imposing rationality or optimality on insurers' decisions based on the definition of MPNE in Section 1.5.3, I estimate the discount factor δ such that profitable deviations from empirical policies are minimized, i.e. the empirical strategies reflect minimum violation of rationality.

In other words, I assume the insurers solve the dynamic pricing game in Section 1.5.3 and set their pricing strategies accordingly, and I look for parameters of the supply model such that insurers' pricing behavior is optimal. Besides model assumptions in Section 1.5.3 and the following functional form assumption in Section 1.6.3.1, this estimation approach requires that insurers in all markets play the same equilibrium strategies so that data from all markets can be used to jointly characterize empirical pricing strategies in the first step.

1.6.3.1 Step One: Empirical Pricing Policy Function

I let the data reveal insurers' empirical pricing strategies by estimating prices or premiums as functions of shares as well as other determinants of pricing decisions as in equation 1.10. Premiums p_{jt} include bids for basic premium b_{jt} for basic plans and supplemental premiums PS_{jt} for enhanced-benefit plans. The controls include own lagged shares by enrollee type S^{θ}_{kjt} and shares of other plans offered by the same insurer S_{k-jt}, plan characteristics X_{jt}, and insurer fixed effects to account for

unobserved heterogeneity across insurers that affect both shares and pricing decisions. The residual is assumed to be normally distributed, and I use the estimated standard deviation to get random draws for competitors' prices for forward simulations in the second step.

$$p_{kjt} = \alpha + \sum_{\theta} \beta^{\theta} S^{\theta}_{kjt} + \sum_{\theta} \gamma^{\theta} S^{\theta}_{k-jt} + X_{jt}\lambda + \xi_k + \epsilon_{jt} \tag{1.10}$$

These empirical policy functions condition on a coarser set of state variables than what is required to compute a Markovian strategy and are similar in nature to the notion of oblivious strategy as formalized by Weintraub et al. (2008). As an approximation for Markov perfect equilibria, Weintraub et al. (2008) define oblivious equilibrium as an equilibrium in which each firm is assumed to make decisions based on its own state and knowledge of the long-run average industry state. The rationale for using a coarser set of state variables in my setting is the same as that for computing oblivious equilibrium: realistically it is computationally infeasible to compute Markov perfect equilibria when market sizes are large and the state space explodes even with 20 firms. Such simplifications can actually provide good approximation to firms' equilibrium behavior. In fact, Weintraub et al. (2008) show that the oblivious equilibrium approximates a Markov perfect equilibrium as the number of firms grows.

I estimate empirical pricing strategies separately for three clusters of plans: basic plans offered by single-product firms, basic plans offered by multiple-product firms, and enhanced-benefit plans offered by multiple-product firms. Different factors are relevant for pricing across these clusters of plans – for example, controls are different for single- versus multiple-product firms (shares of other plans within firm are not relevant for the former). Therefore, I estimate the empirical pricing functions separately for these three clusters of plans.

Table 1.7 summarizes my key coefficient estimates. Not surprisingly, plans with higher coverage are more expensive: premiums decrease with deductible amount and increase with gap coverage. Premiums also depend on lagged shares, but the coefficient varies across clusters of plans and types of enrollees. Finally, the key takeaway is that the adjusted R^2 is reasonably high, meaning that this first step is doing a good job at predicting what firms do based on these observable factors, which is a prerequisite for feeding these empirical policy functions into the second step to estimate firms' discount factor.

1.6.3.2 Step Two: Uncover Insurers' Discount Factor

Given the discount factor and pricing strategies, I can forward simulate to get the empirical value function for insurer k, the empirical counterpart to the value function in Equation 1.7.

$$\hat{V}(\sigma_k, \sigma_{-k}, \delta, SV_{mt_0}) = \mathrm{E}_n[\sum_{t=t_0}^{\infty} \delta^t \Pi_{kt}(\sigma_k, \sigma_{-k}, \mathbf{SV}_{mt}) | \mathbf{SV}_{mt}] \tag{1.11}$$

I take competitors' strategies σ_{-k} as given by empirical pricing strategies estimated from the first step and consider each insurer's optimization problem separately. In order to compute this

simulated value function for each possible σ_k, including the empirical strategy and alternative strategies, I forward simulate 500 times and take the average across simulations to get $\hat{V}$. The discount factor can be estimated using the simulated minimum distance estimator as follows, where N is the number of states times the number of alternative-strategies considered.

$$\hat{\delta} = \text{argmin} \frac{1}{N} \sum_{\tilde{\sigma}_k, \mathbf{SV}_{mt}} (\min\{\hat{V}(\tilde{\sigma}_k, \hat{\sigma}_{-k}, \delta, \mathbf{SV}_{mt}) - \hat{V}(\hat{\sigma}_k, \hat{\sigma}_{-k}, \delta, \mathbf{SV}_{mt}), 0\})$$

Intuitively, the discount factor reflects minimum violation of insurer rationality by minimizing room for profitable deviations. The objective function is the average forgone profit by choosing empirical strategies $\hat{\sigma}_k$, compared with alternative strategies $\tilde{\sigma}_k$. Since MPNE requirement applies to all possible alternative strategies, alternative strategies can be any perturbations of empirical strategies. Therefore, I consider single-period deviations from the empirical policy functions for simplicity and consider 100 alternative strategies for each insurer.

Conceptually, the discount factor tells us how much insurers care about future profits and therefore how strong the invest incentive is. The identification comes from the observed price or markup levels – intuitively, the more insurers care about the future, the stronger invest incentive they face and the lower they set the premiums. Table 1.8 reports the estimated $\hat{\delta}$ of 0.946^{20}, which suggests that insurers value future profits strongly and therefore face a strong invest incentive. The standard error is bootstrapped.

1.7 Counter-factual Analysis

Section 1.4 shows a striking invest-then-harvest pricing pattern that is consistent with insurers exploiting consumer inertia. Structural estimation in Section 1.6 uncovers a high discount factor, indicating that insurers have very strong incentives to invest in future demand. Should we worry about such invest-then-harvest pricing among Part D sponsors? On one hand, price increases over time create dynamic choice inefficiency in consumer choice in the presence of inertia. On the other hand, the net effect on consumer welfare also depends on whether switching costs toughen or soften competition. In this section, I apply the dynamic supply model above to answer two important economic questions. First, what is the net effect of strategic pricing in response to inertia on equilibrium prices? This is an empirical question and depends on whether the invest or the harvest incentive dominates. To quantify insurers' trade-off between these counteracting incentives, I decompose the observed pricing patterns into components attributed to invest and

[20] One potential concern is that such a high annual discount rate cannot be reconciled with the fact that many Part D sponsors are publicly traded and have high rates of returns on investment. However, it should be noted that I estimate a common discount rate for all insurers in this market for computational feasibility. While the discount factor or rate of return might vary across insurers empirically, this estimate represents the average discount factor across insurers. Furthermore, even for big insurers such as Humana, the annual rate of return is not much higher than that implied than the estimated discount factor. For example, Humana's recent annual return on investment ranges from 6.37% to 7.91% based on `http://csimarket.com/stocks/HUM-Return-on-Investment-ROI.html`.

harvest incentives in Section 1.7.1. Second, what are consequences of policies that could be used to constrain insurers' ability to exploit inertia using the "invest-then-harvest" pricing tactic? To evaluate the desirability of government intervention, I simulate the effects of three policies on prices and welfare, including two policies implemented or proposed under the Affordable Care Act.

1.7.1 Do Switching Costs Lead to Higher or Lower Prices?

The competitive effect of switching costs is ambiguous and depends on whether the invest incentive or the harvest incentive dominates. While Beggs and Klemperer (1992) show that the harvest incentives always dominates when consumers are perfectly inertia and switching costs are infinite, this is not necessarily true when consumers are subject to finite switching costs. In fact, Dubé et al. (2009) show that depending on the magnitude of switching costs, equilibrium prices can be higher or lower than the case without switching costs. Contrary to conventional wisdom that switching costs soften competition, the authors show examples where inertia reduces equilibrium prices. When switching costs are finite, firms face incentives to price low not only to attract new consumers but also to attract consumers currently attached to competitors. Arie and Grieco (2014) highlight the "compensating" effect, or the incentive to induce competitors' consumers to switch products, as the key contributing factor to lower price levels.

In order to decompose the effects of the invest and harvest incentives on driving prices, I compare prices in the model with inertia with two counter-factual benchmarks, one without inertia and one in which insurers are myopic. In the dynamic model with inertia, insurers are subject to both the invest and harvest incentives when setting prices. In the counter-factual benchmark with no inertia, insurers are subject to neither the invest incentive nor the harvest incentive. In the counter-factual with myopic insurers, insurers face no invest incentive and only the harvest incentive. The comparison between the dynamic model with consumer inertia and these two counter-factual benchmarks helps decompose insurers' trade-off between the invest and harvest incentives.

In the counter-factual benchmark with no inertia, standard enrollees' demand is different from Section 1.5.1 because their utility, which is described below by Equation 1.12, no longer includes switching costs as in Equation 1.2.

$$u_{ikjt} = -\alpha p_{kjt} + \beta_{it}\phi_{kjt} + \lambda_{it}\mathbf{1}\{\text{Insurer}\}_k + \epsilon_{ikjt} \tag{1.12}$$

In the counter-factual with myopic insurers, demand is the same as in Section 1.5.1, but now the discount factor $\delta = 0$ in the supply model, and insurers set prices only to maximize annual variable profits without any consideration for future profits.

Table 1.9 reports enrollment-weighted equilibrium markups among standard enrollees in a simplified two-period model with inertia corresponding to the actual setting, in the counter-factual benchmark with no switching costs and in the counter-factual with myopic insurers. Consistent with the invest-then-harvest intuition, in the model with inertia we see low prices (small and negative average markup) in the first year but high prices (high average markup) in the second period.

Interestingly, average markup is lower than the benchmark with no inertia, which indicates that the invest incentive dominates the harvest incentive and that inertia toughens competition. A comparison between the model with inertia and the counter-factual benchmark without inertia in the first year shows that the invest incentive accounts for a drop of around $300 in markup. Another comparison between the model with inertia and the counter-factual with myopic insurers in the second year shows that the harvest incentive accounts for an increase of around $100 in markup. These comparisons show that the invest incentive dominates the harvest incentive and that switching costs make the market more competitive. These findings contradict the conventional wisdom that switching costs soften competition, and confirm the conclusions of Dubé et al. (2009) and Arie and Grieco (2014).

1.7.2 Policy Experiments

The effects of government intervention are not directly intuitive and are in fact ambiguous ex-ante, because pricing response to inertia creates two offsetting effects on consumer welfare. On one hand, price increases create dynamic choice inefficiency in consumer choice in the presence of switching frictions. On the other hand, the structural estimation suggests that inertia reduces prices as insurers face very strong incentives to invest in future profits. The desirability of government intervention depends on how effectively each policy can smooth prices over time without increasing average price levels. In this section, I apply my model to understand the potential role of government intervention by simulating the price and welfare consequences of policy experiments where the government restricts insurers' ability to exploit consumer inertia with the invest-then-harvest pricing tactic.

First, the most straight-forward way to constrain insurers' invest-then-harvest pricing strategy is to cap the annual increase in plan bids and supplemental premiums by a certain percentage. This cap directly curbs insurers' ability to harvest and raise prices later, and therefore also dampens the invest incentive upon entry. In fact, the Affordable Care Act implements an "Effective Rate Review" policy closely resembling a cap on annual premium increase: "any proposed rate increase by individual or small group market insurers at or above 10 percent will be scrutinized by independent experts to make sure it is justified" (U.S. Department of Health and Human Services, 2014). Motivated by this policy, I consider a policy experiment in which firms can only increase bids and supplemental premiums by up to 10 percent each year. The model is set up similarly to that in Section 1.5, except that now there is a constraint that insurers' bids and supplemental premiums cannot exceed 110 percent of those in the previous year.

A second policy I consider is to offer a public option at a low price to compete with private insurers. Widely discussed in privatized insurance markets, public options were proposed as part of the Affordable Care Act but were removed in the final reconciled bill. Intuitively, offering an inexpensive public option increases competition, which restrains room for insurers to harvest consumer inertia or charge high prices later on and, as a result, also weakens the invest incentive early on. I consider a policy experiment in which the government adds a public option to the

market, offering the minimum required coverage and priced at $300 in all years. The model is similar to that in Section 1.5, except that now there is additional competition from this public option.

Finally, I also consider the effects of removing risk sharing and fully exposing insurers to excessive losses and gains from their pricing decisions. The risk corridors might have exacerbated insurers' invest-then-harvest pricing tactic by making it less costly for insurers to price low initially to attract consumers. However, risk corridors might also weaken the invest-then-harvest incentives because insurer profits exceeding a threshold are taxed. The net effect is ambiguous and is an empirical question. Given the importance of the risk corridors in this setting, I analyze this counter-factual to understand the effect of this regulation (or its removal) on insurer pricing and consumer welfare. The supply model remains the same as in Section 1.5, except that risk corridor adjustment Γ is removed from the profit function in Equation 1.8. A caveat with the last two policy experiments is that public options and risk corridors are big policy instruments with many potential effects other than influencing insurer response to inertia, and my analysis here only speaks to one of many aspects of their effects.

1.7.3 Implementation

The empirical policy functions estimated in Section 1.6.3.1 only characterize equilibrium strategies in the empirical setting and can no longer be used as competitors' strategies in the counter-factuals. Instead, I need to solve for the Markov Perfect Nash Equilibrium in each counter-factual. It is computationally difficult to solve for the equilibrium in each game, given the large number of insurers and the large number of parameters to solve for in the equilibrium strategy. For computational feasibility, I restrict the set of strategies to follow the functional form of the empirical policy function. I assume that in each counter-factual, equilibrium pricing strategies (bids for basic premiums and supplemental premiums) take the functions form of the following, where I constrain the coefficients on shares to change by the same proportion relative to the empirical coefficients and constrain the coefficients on plan characteristics to change by the same proportion relative to the empirical coefficients.

$$\tilde{p}_{kjt} = \tilde{\alpha} + \tilde{\xi}_k + \sum_{\theta} \tilde{\beta}^{\theta} S^{\theta}_{kjt} + \sum_{\theta} \tilde{\gamma}^{\theta} S^{\theta}_{k-jt} + X_{jt}\tilde{\lambda} + \epsilon_{jt} \tag{1.13}$$

With this simplification, I forward simulate to get the expected value functions given the price vector in the first year P_0 and the coefficients that guide pricing strategies in subsequent periods. I then iterate over insurers' optimal choices of initial prices and these parameters using the simulated value functions until a fixed point is reached, which provides the equilibrium pricing strategies. Given the solved equilibrium strategies, I move on to Section 1.7.4 and compute welfare.

For computational feasibility, I conduct counter-factual analysis on one representative market with around a quarter of a million enrollees annually choosing stand-alone plans.[21] The population

[21] Although I only perform the simulations on one representative market, I assume the same policy experiment is

sizes of both standard and low-income enrollees are close to cross-market averages as reported in Table 1.1. In addition, for computational feasibility given the large number of insurers, I report results from simplified two-period models. Although the price levels would be more comparable to the data in the model with a longer time horizon, two-period models already highlight intuitions for the key economic forces, and the following qualitative interpretations are not an artifact of the two-period set up.[22] As a benchmark for comparison, I also solve for the equilibrium in a two-period game with the current setting, i.e. with consumer inertia, no cap on price increase, no public option and with the current risk corridor set up.

1.7.4 Welfare Metrics

I evaluate effects of three policy experiments on prices as measured by enrollment-weighted premiums and markup levels, as well as on social welfare, including consumer welfare, insurer profit and government subsidy. I compute consumer welfare, insurer profit and government subsidy as relevant to standard enrollees because the automatic allocation of low-income enrollees makes it difficult to infer their preferences and to compute their welfare. In order to consistently compare static and dynamic counter-factuals, I define these welfare measures on a per-period basis. Total surplus W on a per-period basis can be calculated as follows, where CS denotes consumer surplus in money metrics, Π denotes total insurer profits, $Subsidy$ denotes government subsidy, and λ is the social cost of raising public funds. I take $\lambda = 1.3$, based on estimates in Hausman and Poterba (1987).

$$W = CS + \Pi - \lambda Subsidy \tag{1.14}$$

Whether consumer surplus should include switching costs depends on the underlying causes for observed inertia (Handel, 2014, provides a detailed discussion). Such distinction is not crucial for welfare analysis in my policy experiments, which all directly influence the supply side rather than the demand side. Here I take switching costs as welfare-neutral, but when evaluating the robustness of my results, I plan to consider switching costs as partly or fully welfare-relevant. When treated as welfare neutral, switching costs do not count towards actual utility but do affect choice probabilities. I simulate individual utilities and choices to compute consumer welfare in monetary terms after the normalization over the absolute value of the price coefficient.

Insurer profits are calculated based on individual choice simulations above and using the relevant profit functions. I compute pre-risk-corridor profits because risk corridor payments will be reported separately as part of government subsidy. There are three types of subsidies relevant to standard enrollees – risk-adjusted direct subsidy for each enrollee to insurers; risk corridor payments to

implemented throughout all markets and the national-average bid (used to transform bids to premiums and subsidies) changes by the same proportion as in this market.

[22] While the two-period models suffice in illustrating the key insights, I am working on models with a longer time horizon in order to verify the robustness of the conclusions and to provide a more realistic comparison with the data.

insurers in cases of excessive losses (but the payment can go the other way if the insurer earns excessive profits); and individual reinsurance to insurers, which covers 80 percent of catastrophic drug expenditures. In the counter-factuals, I focus on direct subsidy and risk corridor payments, which are endogenous to firms' pricing strategies. These subsidies can be computed similarly to the profit term above on an per-year basis. The other subsidy, individual reinsurance, is not likely to change much across counter-factuals and therefore is less interesting for the counter-factual exercise. Following Decarolis et al. (2015), I compute government subsidy as relative to what would have been spent subsidizing the same individuals in MA-PD instead, assuming that in the absence of stand-alone plans, an enrollee would get prescription drug coverage through MA-PD plans instead.

1.7.5 Policy Experiment Results

Table 1.10 reports enrollment-weighted equilibrium markups among standard enrollees, first predicted by the model with inertia corresponding to the actual setting as the benchmark for comparison, then from three policy experiments: setting a cap on the percentage of annual premium increase, providing a public option at a low price, and removing risk sharing between the government and insurers. All three policies dampen the invest incentive in the first year and lead to higher markups on average, to different degrees. All three policies also dampen the harvest incentive in the second period and lead to lower markups on average, to different degrees.

In the counter-factual with a 10 percent cap on annual bid and supplemental premium increase, average markup is higher relative to the benchmark case in the first year, but is lower in the second year, which is consistent with the intuition that such a policy dampens the invest-then-harvest incentives. The average price level is slightly lower than the benchmark without any policy experiment. In the counter-factual where a public option is offered at a fixed price of $300, markups in both periods are lower compared to the benchmark, but price in the second year is still much higher than that in the first year. In the last counter-factual, where the risk corridor is removed, prices are only slightly different from the benchmark, if noticeable. To sum up, capping annual premium rise is the most effective in terms of both smoothing prices over time and constraining average price levels. Offering a low-price public option constrains markup rise in the second year but not by much. Removing the risk corridors has little impact on markups.

The prediction that the cap on annual premium increase rate can lower average prices is not as intuitive as the other prediction on smoothing pricing dynamics. On one hand, capping the annual premium increase constrains insurers' ability to raise premiums and harvest inertial incumbent enrollees, which therefore tends to decrease price levels. On the other hand, given the reduced room for harvesting, insurers now face weaker incentives to set low prices to invest in future demand. The net impact on average price levels depends on the interactions of these two channels. Because there is strong competition among Part D sponsors, the invest incentive turns out to be less sensitive to this policy change than the harvest incentive, leading to a lower price level on average. One caveat with this prediction is that it might be specific to the market structure of Part D, and therefore it needs to be re-evaluated in other market settings.

Table 1.11 reports simulated per-period consumer surplus for the benchmark model as well as the three policy counter-factuals. For consumer surplus, I also show the difference between each counter-factual and the benchmark and decompose this difference into the component driven by changes in the share of standard enrollees choosing stand-alone prescription drug plans, the component driven by changes in prices and the component driven by changes in choice efficiency. Capping annual price increase results in the highest consumer welfare, which is largely due to the direct effect of lower average prices and the resulting increase in enrollment, but there is also a noticeable reduction in dynamic choice inefficiency. Offering a low-price public option increases consumer surplus, which works mostly through enrollment in the public option, but there is also a small reduction in dynamic choice inefficiency. Removing the risk corridors has little impact on all margins.

Table 1.12 reports results on social welfare, including consumer surplus, insurer profit, subsidy and social surplus. Capping premium increase is the most desirable in terms of both consumer and social welfare. Offering a low-price public option improves consumer welfare, but such welfare gains are dominated by the extra social cost of providing the public option. Removing the risk corridors has little impact on consumer and social welfare but transfers money from the government to insurers because with risk sharing, taxes on excessive gains outweigh subsidies on excessive losses both in the model and empirically.

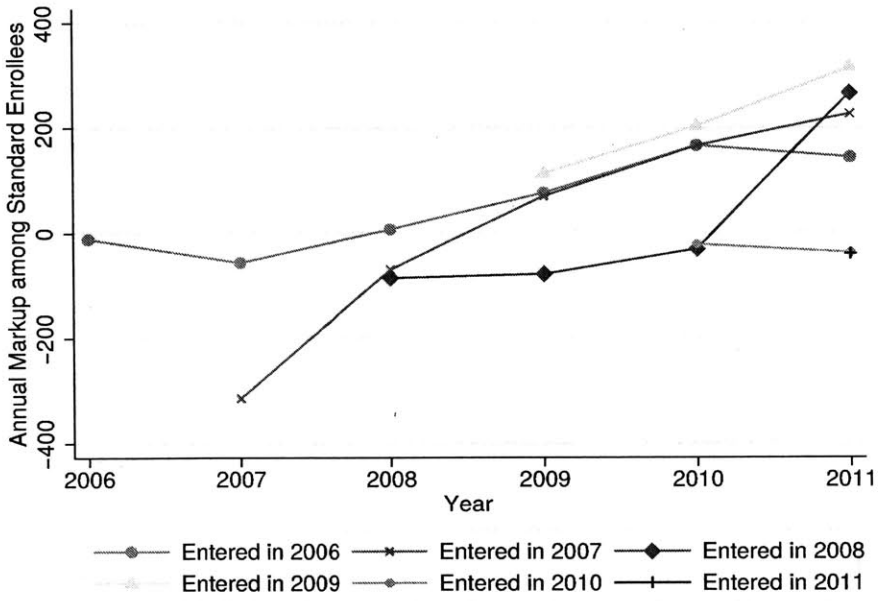
These policy experiments are informative about the desirability of each policy in terms of restricting invest-then-harvest pricing and in terms of improving consumer welfare. Among the policies I consider, a policy change to cap premium increases would be the most effective in improving consumer welfare. This welfare increase comes from both smoother price dynamics and lower average premiums. There are two important next steps to check the robustness of the policy implications. First, I am working on models with a longer time horizon to evaluate the robustness of my conclusions. While key insights from these policy experiments are intuitive and the qualitative interpretations are not an artifact of the current two-period set up, some numbers might be different when we consider a longer time horizon. For example, prices in the second year in the model with inertia are higher than the counter-factual without inertia, which will change after allowing for a longer and more realistic time horizon. Second, the timing of policy intervention matters and I plan to evaluate effects of introducing the cap on annual premium increases at different points of time.

1.8 Conclusion

A growing literature has documented evidence that consumers in health insurance markets behave as if they face substantial switching costs when choosing health insurance plans. In this paper, I investigate whether private insurers in Medicare Part D exploit this type of consumer inertia when setting prices for insurance plans. I first document descriptive evidence consistent with insurers initially setting low prices in order to invest in future demand before later raising prices to harvest

inertial consumers. To explore the implications of these invest and harvest incentives for equilibrium pricing, I develop and estimate a dynamic model of insurers' pricing decisions that incorporates demand inertia and adverse selection. I estimate a high discount factor among insurers, which is indicative of a strong incentive to invest in future demand and is consistent with low prices observed early on. I also find that on net, demand inertia reduces equilibrium prices, i.e. the invest incentive dominates the harvest incentive. Finally, I evaluate welfare consequences of policies that could be used to constrain insurers' ability to conduct such invest-then-harvest pricing patterns. Among the policies that I analyze, I find that a policy change to cap premium increases would be the most effective in improving consumer welfare by both lowering average premiums and smoothing prices over time.

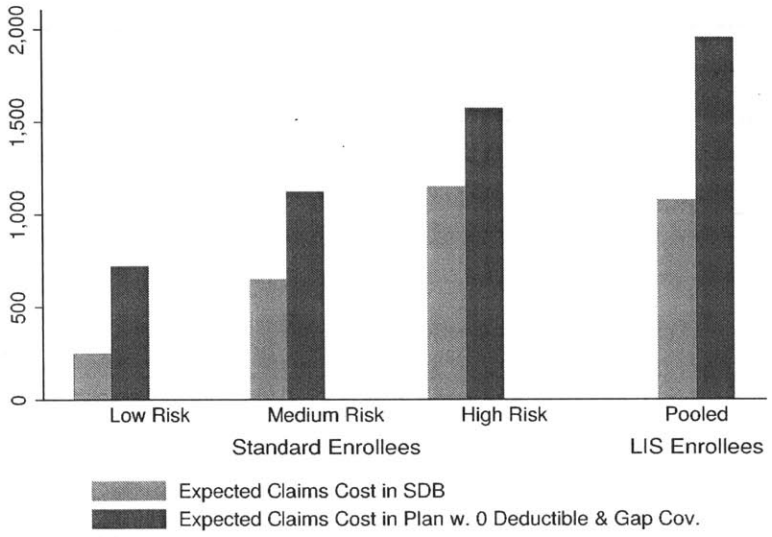
Figure 1-1: Markup By Year of Entry



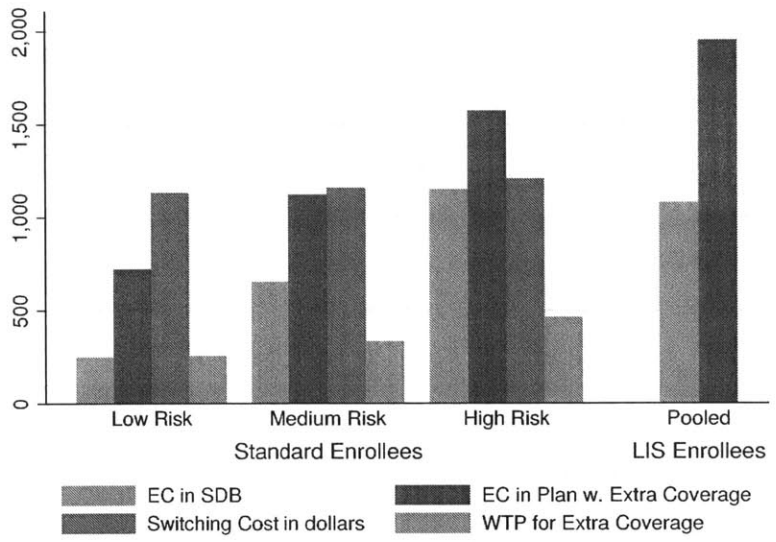
Notes: This figure shows trends in average annual markup for plans introduced in each year. The horizontal axis is year and the vertical axis is annual markup among standard enrollees. Each data point represents the enrollment-weighted average markup for a given cohort of plans in a given year. Each line shows the trend in average annual markup for plans introduced in a specific year.

Figure 1-2: Expected Claims Cost, Switching Cost and Preference by Consumer Type

(a) Expected claims cost by consumer type: low, medium, and high risk scores



(b) Expected claims cost, switching cost and preference by consumer type



Notes: Panel (a) reports, for each type of enrollee, their expected claims cost to a basic plan providing the required minimum coverage or standard defined benefit (SDB) and to an enhanced-benefit plan offering zero deductible and gap coverage on top of basic coverage. Standard enrollees are divided into low-, medium- and high-risk groups based on their risk scores. Panel (b) reports, for each type of enrollee, the expected claims cost as in Panel (a), switching cost in dollars, and willingness-to-pay for extra coverage (zero deductible and gap coverage) relative to basic coverage.

Table 1.1: Summary Statistics

	Mean	SD	Min	Max
Panel A: Stand-alone Prescription Drug Plans				
# Insurers	21	2.5	11	29
# Plans	47	8.2	27	66
# Basic Plans	24	4.0	15	36
# Enhanced-Benefit Plans	23	4.9	12	31
Herfindahl index	0.22	0.06	0.13	0.48
Average Annual Premium, 2006	329	36.5	270	413
Average Annual Premium, 2011	507	46.6	393	583
Panel B: Standard Medicare Beneficiaries				
# Beneficiaries	217,794	160,446	8,635	762,538
Annual Rival Rate	0.10	0.015	0.078	0.150
Annual Attrition Rate	0.08	0.007	0.061	0.098
Share in Stand-alone Plans, 2006	0.24	0.081	0.042	0.43
Share in Stand-alone Plans, 2011	0.26	0.063	0.18	0.42
Panel C: Low-income Medicare Beneficiaries				
# Beneficiaries	59,738	51,148	2,822	267,848
Annual Rival Rate	0.13	0.022	0.092	0.25
Annual Attrition Rate	0.10	0.012	0.073	0.15
Share in Stand-alone Plans, 2006	0.88	0.098	0.52	1
Share in Stand-alone Plans, 2011	0.80	0.11	0.46	0.99

Notes: Panel A reports summary statistics for stand-alone plans across markets. The Herfindahl Index is computed using enrollment of standard enrollees. Average premiums are weighted by standard enrollment. Panels B and C report summary statistics on standard and low-income Medicare beneficiaries at the market level. Numbers of beneficiaries correspond to the 20% random sample and should be multiplied by 5 to get actual Medicare population size. Arrival and attrition rates are relative to lagged population size. Shares in stand-alone plans are calculated as out of the entire population of standard or low-income Medicare beneficiaries, including those with stand-alone plans, those with bundled coverage under Medicare Advantage, those with coverage provided by employers or third parties, and those without any prescription drug coverage.

Table 1.2: Summary Statistics on Plans by Year

	# Plans	# Entries	# Exits	Total Premium	Markup
Pooled	9490			371.9	47.5
2006	1429	1429	NA	329.0	-5.5
2007	1865	594	3	362.7	-55.4
2008	1824	201	86	408.1	4.2
2009	1687	53	83	476.6	71.9
2010	1576	107	24	503.3	148.8
2011	1109	22	95	507.1	121.3

Notes: The table reports summary statistics for all stand-alone prescription drug plans, first pooled across years and then by year. Total premium is the total of annual basic premium and annual supplemental premium. Premiums and markups are reported in enrollment-weighted averages. Markup is defined in Section 1.2.2.

Table 1.3: Enrollment Shares as of 2011 by Beneficiary Cohort

Beneficiary Cohort	2006	2007	2008	2009	2010	2011
# Standard Beneficiaries in PDPs	1,412,073	126,234	121,569	111,783	109,327	118,387
Share in plans introduced in 06 (%)	83.68	79.14	76.79	71.90	71.67	71.67
Share in plans introduced in 07 (%)	7.15	8.10	9.22	10.42	8.42	7.06
Share in plans introduced in 08 (%)	2.60	2.79	3.09	3.97	3.32	2.03
Share in plans introduced in 09 (%)	1.22	1.78	2.16	3.23	2.34	0.97
Share in plans introduced in 10 (%)	3.13	4.76	5.15	6.30	8.85	9.44
Share in plans introduced in 11 (%)	2.23	3.43	3.58	4.18	5.41	8.83

Notes: This table reports shares of standard enrollees choosing stand-alone plans by cohort of beneficiaries (the columns) and by year of plan entry (the rows). Beneficiary cohort is the year that the beneficiary first enrolls in a stand-alone prescription drug plan. Each cell reports the share of the corresponding enrollee cohort choosing plans introduced in a certain year. The enrollment shares in this table are computed as of 2011. For example, the first column shows that there were about 1.4 million standard Medicare beneficiaries who first enrolled in stand-alone prescription drug plans as standard enrollees in 2006. In 2011, among these standard beneficiaries, 83 percent were enrolled in plans introduced in 2006, 7 percent were enrolled in plans introduced in 2007, and so forth.

Table 1.4: Comparing Markups Between Entrant Plans and Incumbent Plans

Markup	(1)	(2)	(3)	(4)
1{Entry}, =1 for entrants	-82.7	-182.8	-158.1	-147.7
	(64.9)	(42.8)	(30.8)	(27.9)
Market FE		x	x	x
Year FE		x	x	x
Insurer FE			x	x
Plan Features				x
N	9312	9312	9312	9312
Adjusted R^2	0.024	0.173	0.268	0.688

Notes: The table reports the regression results for equation 1.1, using plan-year level observations for all plans in the sample period 2006-2011. The regressor of interest is a dummy variable equal to one if the plan enters in that year. All standard errors are clustered at the plan level. Column (1) reports estimates without any controls. Column (2) controls for market and year fixed effects. Column (3) also controls for insurer-fixed effects in addition to market and year fixed effects. Column (4) adds controls for plan coverage, including deductible amount, whether the plan offers gap coverage and tiered cost sharing. Standard errors are clustered at the insurer level.

Table 1.5: Comparing Markups Between Entrant Plans and Incumbent Plans on Subsamples

	Experienced Prior to Part D			(4) Experienced in Part D	(5) Enhanced benefit plans
	(1) Major MA sponsors	(2) MA sponsors	(3) Medicare Sponsors		
1{Entry}, =1 for entrants	-186.6	-168.9	-168.6	-192.0	-134.3
	(29.6)	(25.1)	(24.9)	(30.5)	(62.1)
N	2881	5339	5657	2918	4573
Adjusted R^2	0.732	0.707	0.706	0.567	0.766

Notes: The table reports the regression results for equation 1.1, using plan-year level observations for subsamples of plans in 2006-2011. The regressor of interest is a dummy variable equal to one if the plan enters in that year. Controls include plan coverage, and market, insurer and year fixed effects. Column (1) reports estimates using the subsample of plans offered by insurers that were major sponsors in Medicare Advantage prior to 2006 and offered prescription drug coverage bundled with medical coverage to the Medicare population. Column (2) reports estimates using the subsample of plans offered by insurers with some experience in Medicare Advantage prior to 2006. Column (3) reports estimates using the subsample of plans offered by insurers that provided insurance to the Medicare population prior to 2006. Column (4) reports estimates using the subsample of plans offered by insurers that have served at least 5000 Part D enrollees in the same market before. Column (5) reports estimates using the subsample of enhanced-benefit plans only, which are not eligible for random assignment of low-income beneficiaries. Standard errors are clustered at the insurer level.

Table 1.6: Polyakova (2015)'s Demand Estimates for Standard Enrollees

	(1)	(2)	(3)	(4)
Annual Premium, $100	-0.39	-0.45	-0.41	-0.50
	(0.01)	(0.01)	(0.01)	(0.01)
Default plan, 1/0	5.45	5.61	5.07	5.09
	(0.25)	(0.26)	(0.26)	(0.26)
x Health Risk Score	0.23	0.22	0.36	0.37
	(0.06)	(0.07)	(0.07)	(0.07)
Number of insurer FE	3	3	10	10
Use lagged cost as IV for Premium	No	Yes	No	Yes
Implied SC for 75yo female, av. risk	$1506	$1330	$1392	$1164

Notes: The table reports estimation results for a few key coefficients from Polyakova (2015)'s simulated maximum likelihood estimation – coefficients on plan premium, on the default dummy, and on the interaction of the default dummy and enrollee risk score. Columns 1 and 3 do not instrument for annual premium, while columns 2 and 4 use lagged cost as an instrument for premium. Columns 1 and 2 control for insurer dummies for the 2 biggest insurers in each market (the omitted category consists of all other insurers), while columns 3 and 4 control for insurer dummies for the 9 biggest insurers in each market (the omitted category consists of all other insurers).

Table 1.7: Empirical Pricing Strategies

	Single-plan Insurers	Multi-plan Insurers		
	Basic Plans	Basic Plans	Enhanced-Benefit Plans	
	Bid for P_{basic}	Bid for P_{basic}	Bid for P_{basic}	$P_{supplemental}$
Intercept	231.4	1065.5	1244.7	66.7
	(133.9)	(679.2)	(238.91)	(268.0)
Plan Coverage (%)				
Deductible (in $100)			-56.1	-18.9
			(17.3)	(16.5)
1{Gap coverage}			146.5	95.5
			(29.3)	(18.0)
Lagged Shares By Type of Standard Enrollees (%)				
Low risk share	4.1	-4.2	-16.6	3.2
	(3.3)	(3.9)	(6.7)	(3.2)
Female low risk	-10.6	7.2	3.1	-1.1
	(3.9)	(6.1)	(4.0)	(2.5)
Male medium risk	3.2	0.8	-5.1	-0.5
	(5.9)	(2.7)	(10.9)	(5.7)
Female medium risk	-10.2	-11.8	-9.8	-16.9
	(5.3)	(4.7)	(7.3)	(6.9)
Male high risk	-1.2	-1.7	11.5	20.5
	(8.7)	(4.2)	(8.8)	(6.7)
Female high risk	15.3	10.7	1.9	-5.5
	(6.4)	(4.1)	(9.5)	(8.9)
Number of obs.	677	2804	3265	3278
Adjusted R^2	0.78	0.56	0.72	0.69

Notes: The table reports key coefficient estimates from empirical pricing policy function estimation for three clusters of plans, controlling for plan coverage and insurer fixed effects. All standard errors are clustered at the insurer level.

Table 1.8: Structural Parameter Estimate

	Coefficient	Standard Error
Discount Factor ($\hat{\beta}$)	0.946	0.073

Notes: The table reports the minimum-distance estimate for the discount factor and the bootstrapped standard error.

Table 1.9: Decomposition Results: Equilibrium Markup Levels

Average Markup	Model w. Inertia	Restrict Insurers' Ability to Exploit Inertia		
		Cap Annual Price Rise	Add Public Option	Remove Risk Corridors
2006	-8	157	45	7
2007	403	164	346	398

Notes: The table reports enrollment-weighted average markups (among standard enrollees) in the two-period model with consumer inertia, in the counter-factual benchmark without inertia and in the counter-factual with myopic insurers.

Table 1.10: Counter-factual Policy Experiments: Equilibrium Markup Levels

Average Markup	Model w. Inertia	Restrict Insurers' Ability to Exploit Inertia		
		Cap Annual Price Rise	Add Public Option	Remove Risk Corridors
2006	-8	157	45	7
2007	403	164	346	398

Notes: The table reports enrollment-weighted average markups (among standard enrollees) in the two-period model with consumer inertia and in the counter-factual policy simulations.

Table 1.11: Counter-factual Policy Experiments: Consumer Welfare

Consumer Surplus ($Millions)	Model w. Inertia	Restrict Insurers' Ability to Exploit Inertia		
		Cap Annual Price Rise	Add Public Option	Remove Risk Corridors
Consumer Surplus	78.68	109.23	93.24	78.88
ΔCS		30.54	14.56	0.20
ΔCS due to ΔP		11.30	0.48	-1.42
ΔCS due to Δchoice efficiency		6.87	3.60	1.52
ΔCS due to ΔPDP share		12.37	10.48	0.10

Notes: The table reports consumer welfare estimates for the actual policy setting (as the benchmark for comparison) and for policy counter-factuals (relative to the benchmark), all with a simplified two-period setting. Consumer Surplus is as defined in Section 5.3 and is converted to a per-year value so that it can be consistently compared across counter-factuals. The change in consumer surplus resulting from each counter-factual policy is decomposed into three components – the difference due to enrollees opting in and out of stand-alone plans, the difference due to the resulting equilibrium price changes, and the difference due to changes in consumers' plan choices.

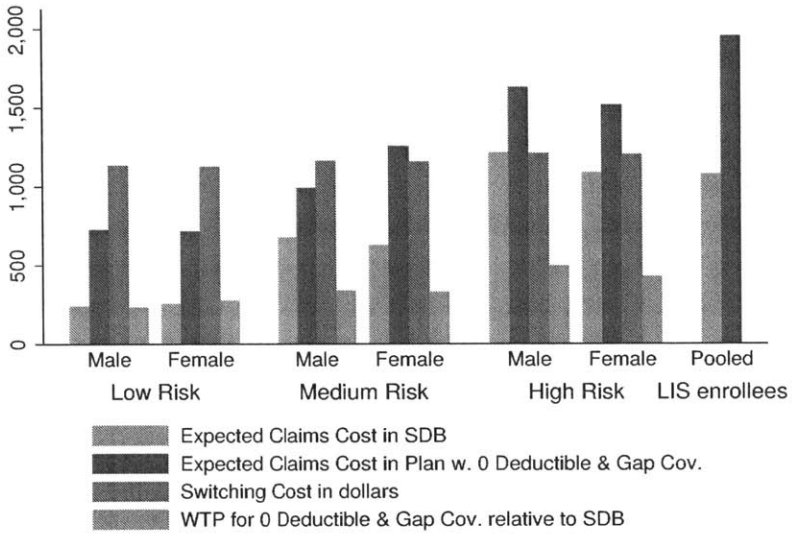
Table 1.12: Counter-factual Policy Experiments: Social Welfare

Social Welfare ($Millions)	Model w. Inertia	Restrict Insurers' Ability to Exploit Inertia		
		Cap Annual Price Rise	Add Public Option	Remove Risk Corridors
Consumer Surplus	78.68	109.23	93.24	78.88
Insurer Profit	14.43	16.01	9.98	46.60
Direct Subsidy	182.81	165.64	157.06	183.60
Subsidy for Public Option	NA	NA	13.75	NA
Total Direct Subsidy, rt MA-PD	2.88	-19.01	5.82	2.88
Risk Corridor Payments to insurers	-30.34	-27.17	-31.30	0.00
Total Surplus	119.71	177.12	109.09	121.74
Public Option Share	NA	NA	0.08	NA

Notes: The table reports welfare estimates for the actual policy setting (as the benchmark for comparison) and for policy counter-factuals, all with a simplified two-period setting. Consumer Surplus, Insurer Profit, Direct Subsidy and Risk Corridor Payments are as defined in Section 5.3 and are all converted to a per-year value so that they can be consistently compared across counter-factuals. Positive risk corridor payments mean that the government pays insurers in aggregate while negative risk corridor payments mean that insurers pay the government in aggregate.

1.9 Appendix

Figure 1-3: Expected claims cost, switching cost and preference by consumer type



Notes: This figure reports, for each type of enrollees, their expected cost, switching cost in dollars, and willingness-to-pay for extra coverage: zero deductible and gap coverage, relative to basic coverage. In addition to the division into low-, medium- and high-risk groups based on their risk scores as in Figure 1-2, standard enrollees are also grouped by gender, which makes a small difference in terms of expected cost.

Table 1.13: Individual Cost Estimation

	Low Risk-score		Medium Risk-score		High Risk-score		Low-income
	Male	Female	Male	Female	Male	Female	Enrollees
Intercept	383.4	383.2	750.2	706.9	1145.2	1078.6	1018.4
	(5.11)	(4.44)	(4.91)	(3.50)	(9.04)	(4.85)	(3.35)
Deductible	-0.585	-0.451	-0.297	-0.234	0	0.056	0.248
	(0.015)	(0.013)	(0.014)	(-0.010)	(0.02)	(0.013)	(0.009)
Partial Gap Cov.	206.2	204.5	192.2	226.8	266.4	344.1	594.2
	(9.3)	(7.8)	(6.5)	(5.0)	(10.0)	(6.0)	(14.2)
Full Gap Cov.	637.3	653.9	716.8	694.3	1051.8	1032.8	1490.0
	(19.4)	(15.2)	(12.1)	(8.8)	(15.6)	(8.9)	(26.4)
Number of obs.	568,117	670,788	1,195,847	1,856,589	1,160,410	2,451,539	9,928,135
Adjusted R^2	0.08	0.09	0.07	0.08	0.05	0.07	0.02

Notes: The table reports some key coefficients from the individual cost estimation for Equation 1.5. Insurer fixed effects are included. Standard errors are clustered at the insurer level.

Chapter 2

Market Size and Innovation: The Intermediary Role of Technology Licensing[1]

2.1 Introduction

Technological innovation is regarded as a primary source of improvements to economic welfare and growth. This notion is particularly palpable in healthcare, where the availability of new treatments can be directly linked to higher longevity, better clinical outcomes, and overall health improvements (Murphy and Topel, 2003, 2006; Lichtenberg, 1996, 2010). Early insights (Schumpeter, 1942; Nordhaus, 1969) and subsequent research converge on the idea that firms' aspiration to rip the monetary rewards derived from commercialization constitutes a leading factor propelling the innovation of new technologies. This suggests that larger market sizes exert a stronger *pulling* force, increasing R&D investment and consequently, enabling higher innovation rates.

The causal impact of market size on innovation has been widely studied, primarily by an empirical literature focusing on the identification of a key statistic: the elasticity of innovation to market size. Due to its data richness and paramount importance, the pharmaceutical industry has become a preferred arena for this type of research. Despite adopting a wide variety of empirical approaches, studies in this realm can be partitioned in two broad groups according to the type of metric employed to track innovative activity. The first group (Grabowski and Vernon, 2000; Giaccotto et al., 2005; Lichtenberg and Waldfogel, 2008; Yin, 2008, 2009; Civan and Maloney, 2009; Kyle and McGahan, 2012; Blume-Kohout and Sood, 2013; Dranove et al., 2014) uses metrics based on the amount R&D investment, or the quantity of inputs used in the innovation process. A second group (Acemoglu and Linn, 2004; Finkelstein, 2004; Acemoglu et al., 2006; Cerda, 2007; De Mouzon et al., 2015), concerns itself instead with the number of new therapies that reach the market. That

[1]**This Chapter is co-authored with Manuel Hermosilla.** We gratefully acknowledge the support of Kauffman Dissertation Fellowship and the Rustgi Family Fund in Entrepreneurship.

is, the output of the process. Despite the fact that disparities in the assortment of empirical approaches and used metrics makes it hard to assess the consistency of the elasticity estimates in this literature (De Mouzon et al., 2015), the role of market size has been invariably constrained to the determination of R&D investment decisions, excluding the possibility that it could also operate by altering the rate at which inputs are converted into outputs.

We argue that market size may not only impact innovative outcomes through the determination of R&D investment decisions, but also through an intermediation effect: the facilitation of inter-firm cooperation oriented at the development and commercialization of developing new technologies. It has long been argued this type of cooperation can increase R&D productivity by pooling firms' complementary capabilities (e.g., Teece, 1986; Gans et al., 2002; Spulber, 2014) implying that, by fostering the exploitation of inter-firm complementarities, larger market sizes could be associated to an improved rate of conversion of inputs into outputs. That is, market size may not only exert *pulling* force on innovation, but also a *catalyzing* one. A direct corollary of this argument is that a larger market sizes could be associated to higher rates of innovative output (i.e., new technologies available to consumers) even if the amount of inputs (R&D investments) were to remain constant.

Productivity gains from inter-firm cooperation appear particularly relevant in the pharmaceutical industry, where clinical development requires the application of a wide range of skills and there are important returns to experience and diversification (Cockburn and Henderson, 2001; Dranove and Meltzer, 1994). Under the current industry configuration, highly specialized entrepreneurial biotech firms focus on the early stages of the innovation process, translating novel scientific insights into embryonic technologies. These firms, however, typically lack a set of important capabilities needed to develop compounds through late stages (Powell and Brantley, 1992; Powell, 1996). By licensing their developing technologies, biotech firms are able to access these capabilities from experienced pharmaceutical commercializer partners, increasing the probability that a new treatment will become available to consumers (Danzon et al., 2005) and/or do so in a shorter time frame. This rationale is widely recognized and often made explicit when licensing agreements are announced. For example, after a recent new alliance for the development of cancer targets between the young biotech firm iTeos and the large pharmaceutical firm Pfizer was made public, an officer of the former firm stated that "the oncologic expertise of Pfizer will help enable the acceleration and expansion of the scope of iTeos' IDO1 and TDO2 programs."[2]

Set in the context of the pharmaceutical industry, we uncover a causal impact of market size on the extent of inter-firm cooperation oriented at the innovation of new technologies. Our empirical strategy exploits the impacts of the 2003 passage of Medicare Part D (henceforth "Part D") on the patterns of drug candidate licensing. This program constituted a significant expansion of prescription drug expenditure coverage for Medicare enrollees, increasing the expected US market size for treatments targeting conditions that are more prevalent among the enrolled population. Since Medicare primarily serves people 65 years and older, treatments for conditions with higher preva-

[2]See http://www.iteostherapeutics.com/iTeos-Therapeutics-Announces-License-and-Collaboration-with-Pfizer-Inc.

lence among elderly populations had a higher degree of exposure to the Part D shock. Following the approach of previous research, we use health insurance and drug expenditure data to produce a measure of shock exposure for each of the developing treatments in our sample. We then compare rates of licensing activity before and after the program's enactment, across treatments with varying degrees of exposure to the shock. Due to the international character of the drug candidate licensing market and the irrelevance of Medicare insurance outside the US, we introduce a third dimension of comparison: whether or not licensing agreements included the US among licensed territories.

Econometric results obtained with this triple-difference strategy suggest that the number of licensing deals encompassing treatments with higher exposure increased by about 60% in the years that followed the program's enactment. The effect was short-lived (5 years) and can be traced back to as early as the couple years following the program's announcement. We use these results to derive what is for us a clean estimate of a novel statistic –the elasticity of licensing-based cooperation to market size– which results in value of 0.71. Even attaching modest productivity gains to inter-firm cooperation, this estimate suggests that the intermediary role of market size could be a meaningful determinant of overall innovative productivity.

The immediacy of the cooperation surge is an important component of the analysis: it implies that the effect unfolded over a period of time in which, due to the relatively long times required for the completion of each stage of the drug development process, the supply of developing treatments available for licensing was fixed. Blume-Kohout and Sood (2013) and Dranove et al. (2014) analyze the impact of Part D on R&D investment decisions, identifying a significant increase in new clinical trials for treatments with higher exposure to the shock –a robust *pull* effect. This effect, however, manifested itself with a lag: it was not present before 2006, and mostly noticeable after 2008, once the bulk of the Part D-fueled surge in cooperation had already taken place. Consequently, our results cannot be explained by an increased supply of developing compounds available for licensing and therefore point to an impact operating over the intensive margin –an increase in probability that each developing technology will be the subject of cooperation. Consistent with this finding, we present a simple theoretical framework which establishes the intermediary role of technology licensing holding constant the amount of inputs (R&D investment) used in the process.

We draw on the literature on Markets for Technology (MFT) to rationalize these results. A central message from this stream from research is that inter-firm cooperation may be hindered by the presence of important contracting frictions –broadly labeled as *transaction costs*– rooted on problems such as costly search and negotiation, asymmetric information and bargaining power, among others (Arora et al., 2001). The presence of transaction costs reduce the return to cooperation and can often times preclude it (Spulber, 2014; Agrawal et al., 2014). As shown by the model, transaction costs imply the existence of a pool of developing technologies for which cooperation is not valuable enough. For these technologies, cooperation is precluded at the baseline market size level. A larger market size facilitates cooperation by reducing the importance of transaction costs relative to gains of cooperation. The model shows that, in absence of transaction costs, the identified Part D-fueled licensing surge could not be rationalized because all technologies would have

been the subject of cooperation regardless of market size. A direct corollary is that the identified elasticity of cooperation to market size is a function of both transaction costs and productivity gains derived from cooperation.

Our model and empirical results further suggest that a larger market size help to materialize a second layer of productivity gains. These are based on an improved matching between the characteristics of each developing technology and the capabilities contributed by the cooperating pharmaceutical commercializer. Drug candidates being developed at early stages (which in the model we refer to as "technological cores") are developed into a set of different treatments ("sub-technologies") –fine-tuned versions of the compounds, optimized and tested in late stage clinical trials for the treatment of specific conditions. First-best matching would pair each of these treatments with the commercializer who possesses the best capabilities to develop each of them. But since the treatments associated to a compound's may span diverse therapeutical areas, first-best matching may require different complementary capabilities, and thus a different cooperating commercializer for each of them. This *unbundling* of treatments into individual cooperation agreements, however, would require incurring additional transaction costs. In analogous fashion to the mechanics underlying the main effect, a larger market size may reduce the relative importance of these additional transaction costs, prompting the unbundling of treatments into individual licensing agreements. Our empirical results show that, following the program's passage, treatments licensed for territories including the US and targeting a population with greater participation of Medicare enrollees were significantly more likely to be packaged into narrower scope (often single-indication) licensing deals than their non Medicare-oriented counterparts.

Gains derived from cooperative development and commercialization are not restricted to higher productivity in the technical sense used here. The literature on MFT (Arora et al., 2001) and its precursors (e.g. Teece, 1986) make this point emphatically: cooperation gains can also be derived from the avoidance of duplicative investment in co-specialized assets (Teece, 1986) or the preservation of downstream market power (Gans et al., 2002). Both of these types of gains are likely to be relevant in the pharmaceutical industry, as prescription drug market tend to be concentrated within each therapeutic area (Malerba and Orsenigo, 2002) and costly co-specialized assets (manufacturing facilities, branded reputation, specialized sales-forces) are required for commercialization (Levine, 2009). Our analysis of section 2.7, however, suggests that these types of cooperation gains were unlikely to have prompted the surge in cooperation following Part D's enactment. [better closing for this paragraph: it must have been productivity gains, but say that we cannot prove that undisputably.]

Our study offers three main contributions. First, it adds to the literature studying the functioning of Markets for Technology (MFT) by shedding some light on the role of downstream demand. This literature has maintained a strong supply side emphasis or, in the words of Arora and Gambardella (2010), "the factors that lead companies to license or sell technology, the implications thereof [...] and the conditions that facilitate the rise of technology specialists." In the recent years there have been several contributions (Cassiman and Veugelers, 2006; Forman et al., 2008; Cecca-

gnoli et al., 2010, 2014; Ceccagnoli and Jiang, 2013) that address this gap by focusing on questions related to how firms' characteristics (or the environment they operate on) may affect their demand for external technology. We offer a more panoramic view, one in which firms on the demand side of the MFT act as agents for commercialization, whose main role is to source embryonic technologies from the MFT, contribute their capabilities to help and speed development, and allocate products to consumer demand. Although we do not directly analyze these firms' behavior, we interpret the vigorous cooperation response to Part D as a sign of them performing this role effectively. Carefully interpreted, the strong response is also a good sign in that it suggests that MFT could be effectively performing a screening role: failed cooperation could in part signal the "weeding out" of technologies *pushed* by innovators, for which market potential is not large enough to justify the commercialization effort.

A second contribution is made to the literatures of endogenous growth (e.g., Romer, 1990; Grossman and Helpman, 1991; Aghion and Howitt, 1992) and directed technical change (e.g., Samuelson, 1965; Drandakis and Phelps, 1966; Acemoglu, 2002), which view the pace of innovation as function of expected market profits. Our results show that the formation of alliances can mediate the relationship between R&D investments fueled by the expectations of higher returns and their outcomes in terms of growth and technical change. With the exception of recent research by Akcigit et al. (2013), this aspect has been largely missing from these literatures.

Our final contribution is to the analysis of the impacts of public policies on innovation. Existing research has identified various channels by which public policies could impact the rate and direction of innovation. Finkelstein (2004) differentiates between static effects (utilization of currently available technologies) and dynamic effects (increased availability of new technologies in the future) and shows that the latter can constitute an important fraction of the overall welfare effects induced by the enactment of a public policy affecting health care utilization. The above-mentioned research of BKS and DHG identify the dynamic effects triggered by Part D, which operate through public insurance-based market size expansions. Lichtenberg and Waldfogel (2008) and Yin (2008) quantify the dynamic impacts of the Orphan Drug Act, a public policy that increased market size through a mixture of tax incentives, expedited development and longer market exclusivity for therapies targeting rare diseases. Kyle and McGahan (2012) find that effective market size can also be affected by the adherence of developing countries to protocols of intellectual property right protection, and this can in turn fuel R&D spending. In all these cases, a common conclusion is that the economic analysis of public policies should not only be based on their potential static impacts, but also consider potentially large dynamic effects. Our results suggest that these dynamic impacts may not only operate through the rate of future R&D investment, but also through the extent to which inter-firm collaboration can exploit its potential returns.

The rest of the paper is organized as follows. Section 2.2 provides industry background, describing the sources and nature of gains derived from inter-firm cooperation. Section 2.3 presents an analytical framework to guide the interpretation of results and formalize the intermediary role of market size. Section 2.4 describes the Part D program and section 2.5 the available data samples.

Our main results are presented in section 2.6. In section section 2.7 we investigate the hypothesis that the licensing surge was driven by cooperation gains other than those related to innovative productivity. In section 2.8 we perform a series of robustness checks, and in section 2.9 we conclude.

2.2 Industry background

During the late seventies and early eighties, the field of biotechnology produced huge leaps. They opened the door to an alternative route to drug discovery –one that did not rely on access to the proprietary chemical libraries that acted as potent entry barriers (Pisano, 2006). These events prompted a reformatting of the pharmaceutical industry, which transitioned from a fully integrated scheme (in which a few large pharmaceutical firms performed all stages of the R&D process and commercialized approved compounds) into a vastly vertically-disintegrated one.[3] Under the resulting industry configuration, large pharmaceutical firms focus primarily on late stage R&D and commercialization, while a fringe of highly-specialized entrepreneurial biotech firms on the early stages of the process.

By several accounts, this industry ranks today amongst those with higher reliance on inter-firm cooperation. Licensing of drug candidates constitutes a leading vehicle for it. In 2010 only, there were transactions associated with potential value in excess of $40B (Giovanetti and Jaggi, 2013). In the remainder of this section we describe the source and nature of cooperation gains in the pharmaceutical industry, as well as the nature of transaction costs that may hinder their realization.

Innovative Productivity Gains

At a broad level, inter-firm cooperation create gains by enabling the exploitation of complementary capabilities (Teece, 1986; Spulber, 2014). Biotech innovators possess narrow, compound-specific technical expertise that is crucial to inform the design of clinical trials and modify the compound when needed.[4] However, as noted by Cockburn and Henderson (2001), clinical development requires the application of wide range of skills (e.g., clinical pharmacology, biostatistics, management and logistics of large clinical trials, etc.), many of which biotech innovators often lack (Powell and Brantley, 1992; Powell, 1996). This implies that development may either be less likely or take longer to come to fruition when conducted in absence of cooperation. Licensing-based cooperation grant biotech firms access to these complementary skills from large pharmaceutical partners.

Despite its wide recognition by academics and analysts, empirical identification of the causal impact of pharmaceutical alliances on innovative productivity is challenging and supporting evidence is mostly indirect. In a couple of highly cited studies, Rebecca Henderson and Iain Cockburn

[3] A full account of these events is beyond the scope of this paper. See Pisano (2006) for an excellent review.

[4] Pisano (2006) notes that at the time a biotech compound is being developed, there is usually no other scientific team in the world other than that of the originating biotech firm that has good enough working knowledge to navigate the development process.

show that large pharmaceutical firms benefit from both scale (i.e., total R&D spending) and scope (i.e., diversity of research programs across therapeutical areas) economies. These may apply to the number of compounds identified at discovery stage (Henderson and Cockburn, 1996), but also to the probability of obtaining regulatory approval conditional on discovery (Cockburn and Henderson, 2001). In this case, the primary sources of gains are scope economies and overall experience filling of regulatory approval requests.

The importance of scope economies can be rationalized because, given the interconnectedness of biological systems, medical knowledge gained in one therapeutic area may also be useful in a different one, or because the expertise devising and implementing complex clinical trials can also transfer across targeted diseases and mechanisms of action. While these studies do not differentiate between compounds developed under alliances and those discovered within developers' own labs, they show that the experience gained by a long and diversified history of drug development confers large pharmaceutical firms an innovative productivity advantage. Dranove and Meltzer (1994) further show that these types of firms may also have the ability to speed up development of the candidates they deem as more important.[5] These findings resonate with the common assessment of industry analysts, in that alliances with large pharmaceutical firms may "broaden and accelerate their clinical development programs through the application of dedicated capital and clinical expertise" (Kessel and Frank, 2007).

Danzon et al. (2005) provide what is, to the best of our knowledge, the most direct evidence of the impact of alliances on productivity. They employ a research design along the lines of Cockburn and Henderson (2001), but analyze a more recent sample of drug candidates, in a time period when alliances are much more prevalent. Their basic results are largely consistent with those of Cockburn and Henderson (2001). This is not only because overall and therapeutic-area specific experience enable higher success rates, but also because these effects are shown to be stronger when, according to the skill complementarity hypothesis, they are more needed (i.e., at more complex phase II and III stages). More importantly, they observe that, in these late-stage trials, indications developed by alliances have higher probabilities of success. This study, however, does little to address endogeneity issues and thus results may represent biased estimates of the causal impact of alliances.

The primary source of bias to these estimates lies on a potential adverse selection problem (Pisano, 1997; Arora et al., 2009; Hermosilla, 2015). The novelty and complexity of underlying science may provide biotech innovators with an informational advantage over compounds' inherent probabilities of success. These firms may exploit this informational asymmetry by licensing those candidates with relatively poorer (privately observed) development success prospects, while developing those with better prospects on their own. In this way, biotech innovators would be able to share the higher risks of candidates with poorer prospects, while retaining full ownership of those

[5] As in the Henderson and Cockburn studies, Dranove and Meltzer (1994) do not differentiate between compounds originated within the sponsoring firm and those in-licensed through an alliance. However, their sample covers a time period in which the industry was vertically integrated and most of compounds were developed in-house. The productivity advantages these studies identify accrue to the firms which now primarily act as commercializers (Pisano, 2006).

with higher chances to reach the market. While available evidence for the relevance of this adverse selection problem is mixed, it is clear that, however strong the problem really is, it does not negate the existence of productivity gains derived from alliance-based cooperation. Much to the contrary, the stronger a potential adverse selection problem may be, the larger productivity gains of alliances would have to be in order to rationalize the results of Danzon et al. (2005). In the context of our study, the relevant question is whether the Medicare-oriented licensing surge may have aggravated a potential adverse selection problem and thus cancelled out innovative productivity gains. We address this concern by looking at differences in contract termination rates but find no evidence to support it.

Other Gains

While improved productivity is generally pointed out as a major source of cooperation gains in this industry, it is by no means the only one. A second relevant source resides on the preservation of downstream market power (Teece, 1986; Gans et al., 2002; Gans and Stern, 2003). In absence of an alliance between an innovator and an incumbent commercializer, a new technology's market launch creates competition and reduces market-wide profitability. Our results therefore imply that while a larger market size may further the availability of new technologies to consumers; this may not translate into overall lower prices.

The avoidance of duplicative investments constitutes another important source of gains Teece (1986). In the drug development industry, large investments may be required to build production facilities. Important investments may also be needed to assemble and train sales forces of medical representatives who promote drugs among prescribing physicians. Large pharmaceutical firms invariably possess these assets, implying that cooperation avoids duplication of investment.

The presence of this type of gains introduces a challenge to our promoted implication, namely, that a larger market size can prompt higher productivity and thus increase the number of new technologies available to consumers even if R&D spending remains constant. This is so because it suggests that the surge in licensing may not necessarily be driven by the gains associated to the exploitation of innovative productivity gains, but instead, by the avoidance of duplicative investment. To address this concern we develop a proxy for the importance of investments in distribution channels and examine the strength of the licensing surge across ranges of this distribution. If gains based on avoiding duplicative investments were the primary driver we should observe that the surge is concentrated (or at least, more pronounced) among compounds for which distribution requirements are relatively larger. Our results, however, point to the opposite scenario. We therefore conclude that the exploitation of productivity gains plays the primary role.

Transaction costs

Despite the large potential gains, licensing-based cooperation may be precluded by the existence of important contracting frictions, generically referred to as transaction costs (Arora et al., 2001; Spulber, 2014). A primary friction resides on imperfect intellectual property rights. If this type

of protection is not available, innovators may not be willing to engage in negotiations fearing the risk of idea expropriation (Arrow, 1962). Even when intellectual property rights can be obtained, uncertainty about its extent and scope may delay timely contracting (Gans et al., 2008).

Costly and lengthy processes of search for partners (Hellmann, 2007; Bessen and Meurer, 2008; Agrawal et al., 2014) and arduous contract negotiations imply monetary and alternative costs (Lerner and Merges, 1998; Agrawal et al., 2014). For biotech firms, search may imply attending industry conferences, preparing briefings, hiring legal counsel and intermediaries. These costs may be compounded by the bargaining power asymmetry between biotech innovators and large pharmaceutical commercializers (Lerner and Merges, 1998; Malerba and Orsenigo, 2002; Bosse and Alvarez, 2010), which prompts some simultaneous negotiations with multiple potential partners in order to improve their bargaining position against any one of them.

Negotiations entail comprehensive due diligence and agreement over a wide range of control and residual rights (Lerner and Merges, 1998; Lerner and Malmendier, 2010), which aim to moderate the issues of contract incompleteness and imperfections in technology transfer.[6] The negotiation process is complex, primarily because the type of novel scientific knowledge underlying biotech candidates is often hard to codify, communicate and absorb (Von Hippel, 1994; Spulber, 2014). Even when negotiations come to fruition, it may be hard to provide detailed specifications to break the set of required activities into independent tasks, prompting continuous engagement between parties and further increasing the cost to collaborate (Arora et al., 2001). Anticipating these complications, many firms may therefore choose to abstain from collaboration (Powell, 1996; Spulber, 2014).

In the next section we present an analytical framework that highlights the roles of productivity gains and transaction costs as determinants of technology licensing. Consistent with the facts described here, and in the spirit of previous research(Arora and Fosfuri, 2003; Spulber, 2014), we model contracting frictions as fixed cost to be paid whenever a new licensing agreement takes place. Given the nature of our empirical exercise, however, we depart from previous literature by broadening its definition. Previous literature has been primarily concerned with the innovator firm's decision to license, and therefore defined the transaction cost as the cost incurred by this firm solely. Instead, we are interested in characterizing the market fundamentals that make a licensing agreement a jointly profitable enterprise. Consequently, we define the transaction cost as the sum of costs incurred by both partnering firms.

2.3 Analytical framework

Basics

We outline a simple framework that illustrates the highlighted impacts of market size on licensing incidence, unbundling of sub-technologies, and innovative productivity. We formulate these out-

[6] Previously we asserted that the novelty and complexity of technologies may informational asymmetries and a consequent adverse selection problem. These informational problems could also create moral hazard if biotech firms become disengaged with the development process after licensing takes place.

comes as a function of market size, transaction costs, and the existence of productivity gains from cooperation. Given the nature of our empirical exercise, we find it useful to abstract from strategic interactions and the mechanics by which rents are divided between contracting parties. We instead assume that licensing-based cooperation will take place if productivity gains from cooperation are large enough to offset transaction costs and that, in that case, firms will find a way to split rents.

We consider a development process comprised two stages, early and late. A unit mass continuum of upstream innovator firms conducts early stage R&D. The input of this stage is a fixed amount of R&D investment (equal for all firms); the output, a technological core X. Each innovator firm develops a single core. Cores are understood as the basic implementation of a novel set of technological ideas, which have the potential to provide value for consumers but do not do so at the time the early stage concludes.

Immediately after the early stage is concluded, a set of I sub-technologies is identified from each core at no cost. These are developed independently in the late stage in order to achieve its potential of providing consumer value and so be commercialized. We normalize the cost of developing each sub-technology through the late stage to zero. As we explain below, the outcome of this stage is uncertain in that, after the stage is concluded, some sub-technologies are revealed as unsuitable for commercialization. As mentioned above, technological cores can be assimilated to drug compounds before they are optimized and tested for the treatment of a given condition, whereas sub-technologies represent compounds' indications –fine-tuned versions of compounds that are independently tested in clinical trials to assess their safety and efficacy in the treatment of specific conditions.

There are as many downstream markets as sub-technologies across cores, and a one-to-one mapping between sub-technologies and their specific targeted market. The identity of the market targeted by each sub-technology is a priori uncertain, and revealed immediately prior to the start of the late stage. For each sub-technology, the probability of targeting each market is the same. The size of each market (denoted by d) is known throughout the process and distributed uniform standard in the population. For analytical tractability, we set $I = 2$ and write $X = (x_1, x_2)$, where x_i represents X's sub-technology i $(i = 1, 2)$. The sizes of the targeted markets by each of these are denoted d_1 and d_2, respectively.

The late stage can be carried out individually by each upstream innovator or cooperatively under a licensing agreement between the innovator and a downstream commercializer. If licensing takes place, a transaction cost $\kappa \geq 0$ must be incurred. This transaction cost represents the sum of costs incurred by both firms. There is a finite set of identical commercializers, but each has infinite capacity to cooperate with innovators.

Late stage uncertainty plays in through the probability of observing a successful outcome in the late stage. Each sub-technology is commercialized only if such outcome is observed. If this stage is

carried out in absence of cooperation, this probability equals p.[7,8] If there is a licensing agreement in place, cooperation increases the probability of success to $p + \delta \in (0, 1)$. We further assume that there are no production or commercialization costs, so total rewards to commercialization equal market size. We refer to each sub-technology that reaches the market as a new technology product. Panel A of table 2.1 summarizes the key elements from this environment. The expressions for the expected number of new technology products (column 4) make it immediately clear that overall innovative productivity depends on the extent of licensing.

To analyze the role of downstream demand in a way that fits our empirical context, we assume that development and cooperation decisions are cleared in two consecutive periods. The economic environments are as described above and identical in both periods, except for that in the second period the downstream demand for a subset of markets is affected by a positive shock. In particular, we assume that a randomly chosen half of all downstream markets is exposed to the shock. When the shock is in place, the sizes of these markets increase permanently by a factor $\theta > 1$. This structure generates a non-degenerate distribution of total core exposure to the shock in the second period: half the cores has one of their sub-technologies exposed to the shock, a quarter has both sub-technologies exposed, while the remaining quarter has none. Since exposed markets are randomly determined, cores' total baseline market size (d_1+d_2) is orthogonal to their degree of shock exposure.

Bundled licensing only

We first analyze the basic scenario in which licensing encompasses of all a core's sub-technologies. To simplify the analysis, we focus on the case where $\delta > \kappa \geq 0$, although qualitative results follow through in other cases. Licensing takes place if the expected total gains of cooperation exceed transaction costs. For cores with only one sub-technology exposed to the shock, this condition is either $(\theta d_1 + d_2)\delta > \kappa$ or $(d_1 + \theta d_2)\delta > \kappa$ (depending on which sub-technology is exposed). For those cores with both sub-technologies exposed to the shock, the condition is $(d_1 + d_2)\theta\delta > \kappa$, whereas for cores with none of their sub-technologies exposed, it is $(d_1 + d_2)\delta > \kappa$. Integrating across the distributions of market potential and shock exposure, the expected number of licensed cores when the shock is in place is

$$L = 1 - \frac{1}{2}\left(\frac{\kappa\Gamma(\theta)}{\delta}\right)^2. \tag{2.1}$$

$\Gamma(\theta) = \frac{\theta+1}{2\theta}$ represents the shock's impact on the incidence of licensing. Before the shock is in place ($\theta = 1$), $\Gamma = 1$ and $L = 1 - \frac{1}{2}\left(\frac{\kappa}{\delta}\right)^2$, showing that the extent of licensing increases the smaller transaction costs are relative to productivity gains derived from cooperation.

Through Γ, the arrival of the demand shock ($\theta > 1$) reduces the importance of transaction costs relative to productivity gains ($\Gamma' < 0$ in the relevant range), increasing the number of licensing

[7] That is, we implicitly assume that the likelihood of commercialization is independent of each sub-technology's targeted market's size. To insure the upstream innovators are willing to engage in the innovation process, we assume that the R&D investment required to undertake the early stage does not exceed p.

[8] The main conclusions can also be obtained assuming heterogeneous success probabilities, or by assuming that uncertainty operates as a determinant of realized market size through realized quality.

deals. Note, however, that in the extreme case where transaction costs equal to zero, such effect is not observed because all cores are the subject of cooperation at the baseline demand level. It follows that the observed surge in licensing described by figure 2-4 (panel A) can only be rationalized if transaction costs are large enough with respect to productivity gains derived from cooperation.

The intermediary role of technology licensing is transparentized by the expression for the expected number of new technology products, $T = 2(p + \delta L)$ –a linear combination of the lower and upper productivity bounds, $2p$ and $2(p+\delta)$, respectively. A higher rate of licensing brings T closer to the upper bound. Using (2.1) we can write T as a function of the model's fundamentals, as

$$T = 2(p + \delta) - \frac{\kappa^2}{\delta}\Gamma(\theta)^2. \tag{2.2}$$

The negative term reflects a deviation from the upper productivity bound caused by the presence of transaction costs, which preclude some cooperation and therefore thwart the full realization of productivity gains. Through the filter of Γ, a larger market size reduces the relative importance of transaction costs, fosters inter-firm cooperation, and thus increases the amount of output obtained from a each unit of input entered into the innovation process. Expression (2.2) also suggests that, at least in the short term when early stage R&D expenditure remains constant, changes in the number of observed licensing deals can be used as a proxy for changes in the expected number of new technology products.

Bundled and unbundled licensing

We now relax the assumption of bundled-only licensing to illustrate the second layer of productivity gains –the improvement of the quality of matching between sub-technologies and cooperating downstream commercializers. We do so by introducing the possibility of unbundled licensing: the situation in which a core's two sub-technologies are licensed through separate deals, to two independent cooperating commercializers.[9]

Our analysis is based on the following trade-off. Matching each of a core's sub-technologies to its *own specific best-suited* cooperating commercializer translates into additional productivity gains but requires incurring additional transaction costs. For the analysis, we set up an environment in which the best-suited cooperating commercializer differs between each core's two sub-technologies, and each licensing deal implies a separate transaction cost κ. Thus, first-best matching always implies unbundling but its viability requires productivity gains to be valued in excess of κ compared to bundled licensing and in excess of 2κ to no licensing.

We introduce heterogeneity in matching quality by assuming that there are two types of commercializers, A and B. These differ in their capabilities: those of A are best suited for the development of x_1, while those of B are best suited for x_2. While arbitrary and unrealistic, this specification allows us to formulate the condition determining the incidence of unbundling as one solely depending

[9]This phenomenon is known as "indication splitting" in the drug development industry (Longman, 2006).

on the elements already contained by the model.[10,11]

The suitability of each firm's capabilities with respect to the development of each sub-technology operates through productivity profiles, which we characterize with the pairs $(\delta+\phi, \delta-\phi)$ for A and $(\delta-\phi, \delta+\phi)$ for B, with $\phi \in (0, \min\{1-p-\delta, \delta-\kappa\}$. This means that, when cooperating with A, the probability that x_1 will reach the market is $p+\delta+\phi$, while that of x_2 is $p+\delta-\phi$ (and vice versa for cooperation with B).

Panel B of table 2.1 summarizes the key elements in this environment. The expected number of new technology products (column 4) shows that in this case the upper bound on innovative productivity is given by $2(p+\delta+\phi)$. By imposing $\delta > \phi$ through the upper bound on the domain of ϕ, we rule out the situation in which all licensing is unbundled and insure that unbundled licensing always imposes a more stringent requirement over $d_1 + d_2$ than bundled licensing. Thus, to determine the set of unbundled cores we only need to check that unbundled licensing yields higher expected total gains than bundled licensing.[12]

This structure implies a simple assortment, which we describe graphically in figure 2-1 for cores whose first sub-technology (x_1) is exposed to the shock. Cores licensed as a bundle imply cooperation with the best-suited commercializer for that sub-technology with the higher market size, whereas unbundled licensing pairs each sub-technology with its best-suited commercializer. When the shock is in place, this means that cores with $d_1 > \theta d_2$ are licensed to A, whereas those for which $d_2 > \theta d_1$ are licensed to B. As shown by figure, the arrival of the demand shock increases overall and unbundled licensing, moving all lines closer to the origin. This illustrates our claim that a larger market size translates into a two-layered impact on innovative productivity.

As in the bundled-only licensing case, integrating across the distributions of market potential and the probability of exposure to demand shocks we can compute the expected number of new technology products as a function of the modelâs fundamentals, as

$$T = 2(p+\delta+\phi) - \Sigma(\kappa, \theta, \delta, \phi),$$

with $\Sigma(\kappa, \theta, \delta, \phi) = 2\kappa\Gamma(\theta) + \left(\kappa\Gamma(\theta)\right)^2 \frac{\delta-\phi}{2\phi(\delta+\phi)}$. This formulation retains the key properties from the bundled licensing only case, in that (i) the number of new technology products as a deviation from the upper productivity bound, and (ii) a larger market size has no productivity impact if transaction costs are null.

A direct implication of the demand shock's arrival is a shift in the relative unbundling incidence of sub-technologies with and without exposure. This occurs because the unbundling rate of cores with both their sub-technologies exposed to the shock increases, while that of cores with no exposed

[10]This characterization implies that commercializer A will always be better than B at developing x_1 (and vice versa for x_2) despite the fact that the targeted market by each sub-technology is randomly determined.

[11]A more realistic framework would, for example, sort sub-technologies according to a separate technical dimension (e.g., drug delivery method), and then anchor commercializers' comparative advantages to it. We see no analytical gains to introducing such additional structure.

[12]The upper bound on the domain of ϕ imposes $\phi < \delta - \kappa$ to simplify the algebra. Replacing this with $\phi < \delta$ produces no additional insight and does not change our conclusions.

sub-technologies remains unaffected. This observation yields the prediction that when the demand shock is in place, sub-technologies with shock exposure will be more likely to be observed in stand-alone licensing deals than those without exposure.

To formalize this effect we construct Δ as the difference between the number of sub-technologies with and without exposure to the shock that are licensed through stand-alone (unbundled) deals. This results in

$$\Delta = \left(\frac{\kappa}{2\phi}\right)\left(1 - \frac{1}{\theta}\right) + \left(\frac{\kappa}{2\phi}\right)^2\left(1 - \frac{1}{\theta^2}\right) \tag{2.3}$$

Before the shock arrives ($\theta = 1$), this difference is null, as exposure is random and affects 50% of sub-technologies. After the shock arrives ($\theta > 1$), in the presence of positive transaction costs, this difference becomes positive. Again in this case, it can easily be seen that such effect can be rationalized if transaction costs operate in the market. Our analysis of section 2.6 provides evidence consistent with $\Delta > 0$.

2.4 Medicare Part D

Medicare is an important social insurance program in the United States, which provides medical insurance primarily for the elderly (65 years and older) and disabled. Since its creation in 1965, Medicare has covered beneficiaries' inpatient and outpatient expenditure through Medicare Part A and Part B, but offered little prescription drug coverage until recently. In December 2003, Medicare Part D was enacted as part of the Medicare Modernization Act (MMA) to provide outpatient prescription drug insurance to Medicare beneficiaries. The Act went through political debates and almost got killed in the voting process. Until it was finally passed in December 2003, it was not clear whether and how prescription drug coverage would be added. The program went into effect in January 1, 2006.

Part D is a large-scale program, both in terms of the number of enrollees and its cost. In 2006, there were 26 million Medicare beneficiaries enrolled in Part D. The annual program cost was about $50 million in 2008 and about $63 billion in 2012, implying that the average expenditure per-patient was close to $2,000 in 2008. The Office (2014) predicts that the total program costs will grow to $76 billion by 2015.

By various accounts, Part D was a significant shock to the industry. Blume-Kohout and Sood (2013) and Dranove and Meltzer (1994) find the program incentivized the innovation of more prescription drugs targeting the conditions that are more prevalent among enrollees. The program was also found to have increased prescription drug usage among enrollees by between 4.7% and 5.9% (Ketcham and Simon, 2008; Yin et al., 2008). Because of the large percentage of the population enrolled in Medicare and the fact that the government (as represented by Center for Medicare and Medicaid Services) is prohibited by law from directly bargaining with pharmaceutical firms, the anticipated increase in drug utilization could have been reasonably predicted to translate into rising higher drug expenditure and commercialization profits for the diseases that are more prevalent

among Medicare enrollees.

2.5 Data

Licensing deals

The main data source is Thomson Reuters Cortellis Life Sciences, a comprehensive repository of licensing contracts. This data subscription service is widely used by industry practitioners to inform strategic development decisions and prepare for negotiations. Our main set of results is drawn from the "Recap â DEAL Builder" tool offered by the Cortellis subscription, which tracks strategic alliance activity in the sector.[13] Recap is known as the gold standard for actionable data on biopharmaceutical deal making, as it contains information of over 40,000 alliances struck since the early 1970s.[14]

Our analysis focuses licensing-based alliances aimed to the development and commercialization drug candidates. We focus on the deals that Cortellis reports with the labels of "development and commercialization" and "commercialization" delas, as these refer to candidates for which the pursued therapeutical applications have already been defined. These can be thought of as vertical agreements, in which an innovator firm grants exclusive commercialization rights (within certain territory) to another firm, which will further contribute to the development process and commercialize the compound (if regulatory marketing approval is obtained). Other types of licensing based alliances (joint ventures, contracted research, research tools) are better described as horizontal alliances and are typically observed at early stages, when a specific targeted disease has not been defined. Our empirical strategy crucially relies on the ability to link the downstream market exposure to the Part D shock (computed at the targeted disease level) to each compound's licensing propensity, which is why we focus on the type of alliances that encompass a set of well-defined targets.

Cortellis contains records for such deals 12,846. Each of these specifies a set of indications to be developed and commercialized by the alliance. We drop the deals for which the list of licensed indications is or contains missing data, as well as those for which we were not able to establish a link to the variable measuring the exposure to the Part D shock. Since our focus is on the impact on developing technologies, we also drop those that contained one or more indications with regulatory marketing approval at the time the alliance was struck. Finally, we restrict the period of analysis to 1998-2014, which leaves us with a total of 4,533 contracts, trading the commercialization rights to 6,942 indications. On the resulting sample, each alliance encompasses 1.5 indications on average.

Because Part D impacted the demand of US consumers only, we differentiate between alliances that included the US among licensed territories and those that did not. An alliance is coded as

[13]The Recap service was originally provided by a San Francisco based independent company (Recombinant Capital) and later acquired by Deloitte, which later sold it to Thomson Reuters in 2013.

[14]The company obtains alliance information through Freedom of Information Act requests to the Securities and Exchange Commission (SEC). Publicly traded firms are required by law to submit this information, while privately held firms also have to in some states if they provide employees with stock option plans.

including US territories if the "Included Territories" variable in the data contains "US," âNafta," "North America," or "World." About 32% of the alliances include the US territories and among these, 85% provide worldwide commercialization rights. For ease, in the remainder we refer to alliances including the US simply as "US alliances."

Table 2.2 presents the main descriptive statistics of the sample. Columns 1 and 2 present the average number of new alliances registered each year. The number of US alliances is consistently lower than that of non US alliances, reflecting the fact that the required capabilities to commercialized compounds in different regions of the world besides the US may be reside in different firms (Kyle, 2006). The number of contracts and indications (presented by columns 3 and 4) decreases the more advanced the stage is. This reflects the joint effect of many indications "exiting the market" early by striking licensing deals at the preclinical or phase I stages, and that due to failing to deliver promising testing results and consequently seeing their development terminated.[15] Because of this feature, all employed empirical specifications will include development stage fixed effects. Finally, we note that the proportion of US alliances remains roughly constant throughout time periods and across stages and do not significantly change when we instead look at the number of licensed indications.

Exposure to the Part D shock

In order to size the magnitude of the demand shock posed by Part D across targeted conditions, we adopt the approach used by previous research exploring the impacts of Part D, Duggan and Scott-Morton (2010), Blume-Kohout and Sood (2013), and Dranove et al. (2014). In particular, we create a variable that measures the participation of Medicare enrollees (as of 2003) within each targeted condition. As in these previous papers, we label this variable "Medicare Market Share" (MMS).

To construct this variable we utilize data from the Medical Expenditure Panel Survey (MEPS),[16] a large and representative sample of US individuals' medical services utilization, including information about the types prescription drugs used and the availability and type of insurance. Using MEPS insurance and conditions files, we compute MMS as the percentage of prescriptions in each therapeutical class that were issued to enrolled in Medicare in 2003.[17] Thus, the domain of MMS is the unitary interval and a value of 0.5 for a specific condition indicates that 50% of the prescriptions used to treat that condition in 2003 were issued to Medicare enrollees. This variable therefore reflects the exposure of each of the indications licensed in our data had to the Part D shock: higher MMS indications were more exposed to the shock than lower MMS ones. In section 2.8 we develop two alternative measures of shock exposure and show that our qualitative results remain invariant. Most of our analysis will rely on a dichotomic indicator that is activated if an indication's MMS

[15] Development attrition is a well-known and pervasive feature of the drug development industry. See, for example, DiMasi et al. (2003).

[16] MEPS data are available for download at http://meps.ahrq.gov/mepsweb/.

[17] For all calculations involving the MEPS data we weight individual quantities using reported individual representativeness weights.

score is above its distributionâs median. We refer to the indications that satisfy this condition as "Medicare-oriented."

A limitation of the MEPS data is that specific conditions suffered by each individual are not reported by name or coded with high granularly. Instead, they are reported by their corresponding ICD-9 therapeutic category.[18] At its most granular level, the ICD9 categorization achieves a great deal of precision. However, MEPS reports conditions are reported at their least granular, 3-digit level, which has about 800 broad therapeutical categories. We thus bridge indications in our data to MEPS insurance variability at this level. The 648 different conditions targeted by the indications in the data cover 254 of such categories.[19]

Figure 2-2 presents the kernel distribution of MMS scores across indications associated to the compounds in the licensing data. Overall, there is substantial heterogeneity. Patterns of variation result fairly intuitive, as the epidemiological characteristics of conditions with high MMS scores are those one would tend associate with Medicare enrollees (i.e., 65 years and above). For example, at the bottom of the distribution, with MMS scores below 0.1, we observe conditions like growth hormone deficiency, attention deficit disorder and acne. With MMS scores between 0.2 and 0.3 there are, for example, conditions like myopia and psoriasis. More Medicare-oriented diseases include hyperlipidemia, chronic bronchitis and hypertension (MMS between 0.4 and 0.5). At the top of the distribution there are conditions like cardiac failure, cataracts, Alzheimer's and Parkinson's disease (all with MMS scores above 0.8), which are typically suffered by relatively older people. The median of this distribution is 0.39.

One question regarding the MMS measure lies on its relationship with baseline demand. If MMS exhibited a degree of correlation with baseline demand levels for targeted diseases our inference with respect to the causal effect of Part D on licensing activity could be biased. To explore this possibility, we again use the 2003 MEPS dataset to generate three proxies for baseline demand at the 3-digit ICD9 level. For each of these categories we compute the (i) total number of patients consuming medications, (ii) total number of prescriptions, and (iii) total prescription drug expenditure (i.e., sum of all payments issued to cover consumption). Figure 2-3 presents the scatter plot of each of these variables (ranked) and MMS. There are no obvious correlations. We therefore continue our analysis under the assumption that indications' exposure to the Part D demand shock (as measured by MMS) is uncorrelated to baseline demand levels.

[18]The International Statistical Classification of Diseases and Related Health Problems (ICD) is a widely used therapeutical classification system maintained by the World Health Organization. According to the Organization's website, about 70% of worldwide medical expenditure is codified using this system.

[19]The average number of conditions per category is 2.6.

2.6 Empirical results

Licensing incidence

The empirical analysis of the impacts Part D on the drug candidate licensing market faces two main challenges. The first corresponds to an aggregation problem: licensed indications through a single deal may vary in their degree of Medicare orientation or shock exposure. As implied by our theoretical framework, a second problem stems from the possibility that the set of licensed indications may be endogenously determined. We first adopt the simple approach of assuming that each indication's exposure to the shock is dichotomic and that the list of licensed indications is exogenously determined. With this, we measure a deal's overall shock exposure through the dichotomic variable DMMS, which equals one if at least one Medicare-oriented indication is included in the deal. We let DMMS=1 if such type of indication is included, and DMMS=0 otherwise. Below in this section we address the exogeneity assumption, and in section 2.8 relax the dichotomic exposure measurement.

Figure 2-4 describes the total number of new licensing-based cooperation agreements signed each year. Panel A focuses on deals including the US among licensed territories, whereas panel B, on those deals that did not include the US. On each case, solid lines represent the number of DMMS=1 deals and dashed lines that of DMMS=0 deals.

Among deals including the US (panel A), the number of deals with higher exposure to the shock (DMMS=1) jumped immediately following the program's enactment, whereas those with lesser exposure retained its gently increasing trend. Furthermore, since Medicare insurance is irrelevant outside the US, the impact is only observed among deals that include the US among licensed territories, as can be seen from comparing the patterns of panel A with those of panel B.

To investigate these trends econometrically, we generate the dependent variable $N_{DMMS,US,t,s}$ by aggregating the total new number of licensing deals observed within cells defined by the sub-indexes. Differencing across stages (s) is important because, as seen from table 2.2, there is lower availability of drug candidates at more advanced stages. This is also important because, due to the large development attrition rates, the relevance of a shock to consumer demand is higher for compounds at more advanced stages. Variation across years (t) is relevant because a firm's propensity to license their technologies may be determined by the fluctuations of public investment markets, as restricted access to public funds during macroeconomic downturns may render licensing as a more viable source of funding (Lerner et al., 2003). Finally, we need to differentiate among licenses that include the US (US=1) and those that do not (US=0).

This approach produces a total of 2 (DMMS) $\times$ 2 (US) $\times$ 17 (t) $\times$ 5 (s) = 340 observations. We consider the time periods listed in table 1: 1998-2003 ($j = 1$), 2004-2008 ($j = 2$), 2009-2014 ($j = 3$), defining the indicators PER_j for each. The definition of the second period is given by the date of the program's enactment (December 2003) and by the fact that it encompasses a short enough time frame in which the increased availability of Medicare-oriented indications or *pull* effects were had not been meaningfully manifested (Blume-Kohout and Sood, 2013; Dranove et al., 2014).

We estimate a negative binomial model with the following specification[20]

$$\begin{aligned} N_{DMMS,US,t,s} &= \alpha_1 + \alpha_2 \cdot DMMS + \alpha_2 \cdot US \cdot DMMS + \\ &\sum_{j=2,3} \beta_j \cdot DMMS \cdot PER_j + \sum_{j=2,3} \sigma_j \cdot DMMS \cdot PER_j \cdot US + \\ &\lambda_t + \mu_s + \varepsilon_{DMMS,US,t,s}. \end{aligned} \tag{2.4}$$

The inclusion of the DMMS level controls for differing proportions of DMMS=1 and DMMS=0 new deals signed each year. Baseline effects associated to the variable US account for the differences shown in table 2.2 –a systematically larger number of deals not including the US. The interaction between US and DMMS controls for potential differences rooted on epidemiological profiles in the US versus the rest of the world, but additionally, for those differences arising between the capabilities of firms commercializing indications with high or low Medicare orientation in or outside the US (Kyle, 2006).

The set of interactions between DMMS and the period indicators may capture differences grounded on supply side factors (e.g., patterns of innovation imply a relative increase in the availability indications at different ends of the MMS distribution). Year fixed effects λ_t control for the effects of macroeconomic conditions described above and licensing stage μ_s fixed effects for the varying availability of compounds at different stages. In all models, the main parameters of interest are those associated with the coefficients for the triple interactions (σ_j) as they reflect, for each time period, the increased number of Medicare-oriented licensing deals that include the US

This model falls short from a triple-diff specification as it does not contain interactions between US and time period (or year) fixed effects, which may capture differences in the impacts of macroeconomic cycles for firms for US and ex-US firms. The reason behind this omission is that the number of degrees of freedom per parameter significantly drops if including these interactions, causing the maximum likelihood routine to enter non-convex zones. We nevertheless cluster standard errors within DMMS/US/t cells to capture some of these effects. We also estimate linear models with the same specification, both including and not including US/t interactions. In all cases and relevant dimensions, results are qualitatively similar to the ones presented here.[21]

Results are presented in table 2.3. Coefficients of column 1 correspond to the full sample. Parameters associated to the levels of DMMS and US are strongly significant and have the expected signs. The coefficient for the interaction between DMMS and US is small and estimated imprecisely. The negative and statistically significant coefficient associated to DMMS×PER$_2$ suggests that in 2004-2008 there was a relative decline in the number of Medicare-oriented deals not including the US However, during this period, immediately following the passage of Part D, there was a relative increase of such deals but which included the US, as reflected by the triple interaction σ for that same period. Indeed, this coefficient implies that during these years the average number of Medicare oriented deals that included the US increased by an extra 68% relative to those that not include

[20]The Poisson specification is not supported by the over-dispersion test. Only a small fraction of observations (about 3%) are zeroes.

[21]These results are omitted here but available upon request.

this territory.

As stated in the introduction, one important facet of this result is that the Part D-fueled surge of US occurred before endogenous supply effects could have been manifested. Blume-Kohout and Sood (2013) and Dranove et al. (2014) find that Part D spurred an increase in the number of indications entered to clinical trial development. Both these studies find that these endogenous supply effects manifested after 2006 and were most mostly visible after 2008. Due to the relatively long development time frames, and according to the estimates presented by these studies, results of column 1 suggest that the increase in licensing activity preceded the increased availability of candidates being tested in clinical trials. Further evidence to this point is provided by the estimates of column 2, which correspond to the model estimated on a shorter sample, 1998-2005. While smaller in magnitude, the key parameter remains positive and strongly significant. When we re-estimate the model on an even shorter sample (1998-2004), the effect remains robust and with a similar size.[22]

Parameter estimates presented in columns 3 and 4 further characterize the effects by narrowing down the estimation samples according to the stage of development at licensing. The first contains alliances for compounds in earlier stages of development (Discovery and Phase I clinical trials); the second, those for compounds in later stages (Phase II and III). In both cases the coefficient for the triple interaction for 2004-2008 is positive and estimated with high precision. Their comparison suggests that the effect was stronger for late stage compounds, which is both consistent with the fact that these face a higher conditional probability of reaching the market (hence their effective exposure to the shock was higher) and that the more complex Phase II and Phase III trials are those which may benefit the most from inter-firm collaboration (Danzon et al., 2005).

These estimates provide some insight regarding the nature of contracting frictions. If the most prominent source of frictions preventing the timely occurrence of cooperation resided on factors related to the process of search for a partner, such immediate reaction would be difficult to rationalize. In this case, we would expect US Medicare-oriented alliance formation to unfold gradually as the results of more intense search efforts translate into new alliances. Similarly, if frictions in this industry were primarily rooted on the lack of well-defined intellectual property protection due to pending patents (Gans et al., 2008), we would not expect a reaction among compounds in early stages, as these are still pending for many compounds in these stages (Mossinghoff, 1999; Patrick, 2013). Yet, we observe an immediate effect that also includes early stage candidates. These results seem more consistent with frictions related to the negotiation of agreements. In particular, we conjecture that the increase in expected demand for Medicare-oriented indications may have widened of the bargaining core, making it easier for parties to reach "fair terms" and so consummate alliances.[23] A related mechanism âreducing Medicare-oriented indicationâs market size uncertainty and thus increasing its certainty equivalentâ may have also helped firms reach successful negotiations.

[22]These results are available upon request.

[23]Agrawal et al. (2014) present evidence that "agreeing to fair terms" is an important factor behind unsuccessful alliance negotiations.

Elasticity of Licensing to Market Size

Our approach to measuring a deal's overall shock exposure does not allow us to directly compute a licensing-to-market size elasticity. However, a back-of-the-envelope estimate is available by considering the individual MMS scores of all indications traded by deals including the US in the period between 2004 and 2008. To do this, we compute the average market size change (AMSC) of each set of deals ($\mathcal{D}_{DMMS}$, for DMMS=0,1) licensed in the US as

$$AMSC_{DMMS} = \frac{1}{|\mathcal{D}_{DMMS}|} \sum_{d \in \mathcal{D}_{DMMS}} \left\{ \left[\prod_{i \in \mathcal{I}_d} 1 + MMS_i \right] - 1 \right\}.$$

Where d indexes deals and i indications, $\mathcal{I}_d$ represents the set of indications licensed through deal d, and $|\mathcal{D}_{DMMS}|$ denotes the cardinality of each set. This gives us $\text{AMSC}_0 = 0.21$ and $\text{AMSC}_1 = 1.17$, suggesting that the market size of deals associated to DMMS=1 deals increased 96% more than that of DMMS=0 deals as a result of the programâs enactment. Given our estimate above, this implies an elasticity of licensing to market size of 0.71.

Unbundling

Our theoretical framework shows that one potential consequence of a larger expected market size is the unbundling of sub-technologies with higher exposure to the shock. In this section we provide additional context and evidence for this effect. In particular, based on expression (2.3), we provide evidence for the idea that the passage of Part D may have prompted contracting firms to single out Medicare-oriented indications into narrower-scope licensing agreements. As suggested by the model, we interpret this evidence as a sign of improved matching between firms, and thus, as the source of a second layer of productivity gains enabled by larger market sizes.

Likely due to data constraints, the issue of optimal bundling of indications into licensing deals remains notoriously understudied. As shown by our model, a primary factor favoring bundling is transaction costs. Unbundling implies that a compound's indications will be licensed through a larger number of deals, which require firms to incur in larger transaction costs in the form of effort and time spent searching for partners and negotiating contract terms.

Anecdotal evidence suggests additional frictions may arise at the post-licensing stage, as independent commercializers of the compound's different indications can interfere which each otherâs development[24] or cannibalize sales. For example, a widely publicized dispute between Amgen and Johnson & Johnson[25] arose because both companies were simultaneously commercializing different indications (under different packages and names) of recombinant human erythropoietin (EPO). Amgen, who had the rights for kidney dialysis, alleged that Johnson & Johnson (with rights for anemia, cancer, AIDS and surgery) marketed its product to induce off-label prescription to canni-

[24]These situations may arise because one company's clinical trials for the development of one of the compoundâs indications may produce adverse toxicity data, forcing companies developing other indications to respond with countervailing evidence or by changing the product label.

[25]See "Amgen Wins One Against Johnson & Johnson," *The New York Times*, December 19th, 1998.

balize dialysis patients. On its part, Johnson & Johnson alleged that Amgen was slowing down the approval process for the indications to which it had rights.

On the other hand, unbundling indications into independent agreements may be beneficial because it enables a better matching between each indication and the capabilities of the partnering company. Longman (2006) makes this rationale explicit writing that "[...] no company is equally good at developing or selling these molecules in all the different indications." In our data, the potential value of improved matching provided by unbundling is suggested by the observation that in the 1998-2003 period, over 60% of US licensing deals including two or more indications spanned multiple therapeutic areas.

Figure 2-5 illustrates our key empirical insight in this regard. This figure takes the total number of indications licensed in the US each year, and dissects it in terms of the total number of licensed indications by each contract. In particular, it presents the share of indications licensed each year by contracts granting rights for a total of I indications, with $I = 1$ (solid line), 2+ (dashed line). Panel A focuses on Medicare-oriented indications, while panel B on non Medicare-oriented ones. In the former, there is a marked spike in the share of single indication contracts following the program's passage. This pattern is absent from panel B. This suggests that indications with higher shock exposure were more often unbundled than those with lower exposure.

Table 2.4 crystalizes these insights by presenting results from probit models that estimate the probability of an indication being licensed by each type of contract. We first estimate an ordered probit model (column 1) with a dependent variable $I = 1, 2, 3+$, and a probit model (column 2) with a dependent variable the equals one if $I > 1$ (and zero otherwise). Both models use the same triple-diff specification employed before, except for that in this case each observation represents an indication.[26] In this context, negative-signed coefficients for the interactions of DMMS, US and the time period indicators reflect US commercialization rights to Medicare-oriented indications were more likely to be granted through licensing contracts extending rights for a smaller set of indications.

Both sets of results point to a relative increase in the unbundling of Medicare-oriented indications licensed for the US during the time period following the passage of Part D. In particular, the estimates of column 2 suggest that, all else constant, the probability that these indications were licensed through single-indication deals increased by 0.14 in 2004-2008 and by 0.13 in 2009-2014. According to the standard errors, however, the latter increase is not statistically significant.

2.7 Other sources of cooperation gains

Our rhetoric and theoretical model justify the post-2003 increase in Medicare-oriented deals in the US based on the realization innovative productivity gains. However, as described in section 2.2, gains from cooperation can also stem from the avoidance of duplicative investment in complementary assets (such as distribution channels and brand equity), or the preservation of downstream

[26] Therefore, in this case DMMS=1 if an indication is Medicare-oriented.

market power. Here we provide some evidence that counters the idea of either these sources being the primary driver of the documented cooperation surge.

Market power

To analyze the role of downstream market power preservation we invoke the framework of Gans et al. (2002), who conceptualize the incidence of cooperation as a function of the relative returns of competition to those of cooperation. Since in more concentrated markets new competitors will have a larger negative impact on incumbent profits, the relative returns to cooperation are expected to increase with market concentration. Thus, if preventing the erosion of market profits were a primary factor behind the licensing surge in Medicare-oriented markets, we would expect to see a more pronounced effect among more concentrated markets. That is, this view conceptualizes the surge in licensing-based cooperation as method to pre-empt competitive entry in those markets whose size was magnified by the Part D shock.

To test this idea we construct a measure of market concentration from the 2003 MEPS prescription drug consumption data. In addition to the purported use by each patient (as indicated by ICD9 codes), these data provides us with the National Drug Code associated to each consumed drug. The structure of this code allows us to retrieve an index for the labeler (distributing firm) of each drug. With this, we can compute the share of sales associated to each of these firms within each market, as defined by ICD9 3-digit categories. Using these, we construct the Herfindahl-Hirschman Index (HHI) of market concentration.[27] The resulting HHI variable has a low correlation to MMS (about 0.07 and not statistically significant) and does not exhibit significant correlations with baseline demand (as revealed by an analysis analogous to that of figure 3b).

Since we cannot compute a concentration measure for ex-US markets, we are unable to specify a regression model that exploits differences in concentration across markets. For this reason, our test is based on replicating our main estimates of table 2.3 (column 1) on a sample that does not include candidates licensed in the US that target the most concentrated markets. If pre-emptive motives were at the hart of the cooperation surge, then estimates from this sample should point to a significantly weaker increase in Medicare-oriented licensing following the program's enactment. Results from columns 1 and 2 of table 2.7 suggest otherwise.

Columns 1 and 2 respectively present the estimates obtained when we remove the 25% and 50% of US deals with higher associated HHI values.[28] As expected, the triple interaction parameters associated to US Medicare-oriented licensing in 2004-2008 are smaller than the 0.52 value of table 2.3 (column 1), but the differences are relatively small and the inferred Medicare-oriented licensing surge remains strong among the set of deals targeting the markets with lower levels of concentration.

[27] Recall that ICD9 3-digit categories represent the highest granularity at which we observe expenditure. Hence this measure represents the tightest market definition we can implement using the MEPS data. As stated in a previous footnote, the average number of conditions by ICD9 category is 2.6 in our data.

[28] To compute a deal's overall HHI we average all included indications' individual HHI values. When we instead remove the 25% and 50% indications licensed in the US with higher HHI values, results are qualitatively similar.

It should be noted that these results couldn't be used to argue that the preservation of downstream market power is an unimportant determinant of the incidence of cooperation. Gains derived from securing market power may large enough to warrant a high rate of cooperation among more concentrated markets even before the downstream demand shock came into place. If this were the case, there would be few infra-marginal deals in concentrated markets to sustain the large documented Medicare-oriented cooperation surge. Nevertheless, we take the evidence presented here as suggesting that the increased licensing activity was not articulated by pre-emptive motives.

Complementary assets

To assess the role of complementary assets as a driver for the documented licensing surge we implement a test that adapts the logic used above. It is based on a measure constructed with the goal of capturing the variation in the importance of complementary assets as they relate to market distribution. We re-estimate our main specification on samples from which we remove the US deals associated to larger cooperation gains as given by this metric. Our results suggest that the Part D-fueled surge in cooperation did not focus on those deals thatUS could benefit the most from avoiding duplicative investment on complementary assets.

Following Levine (2009), we develop a variable that counts the number of actively practicing physicians in the US that are likely to prescribe treatments for a given indication. We label this variable as the "number of prescribing physicians" (NPP). We exploit this type of variation because NPP can be thought proxy for the total expenditure in detailing (personal selling through sales representatives). This type of promotional expenditure represents the largest cost component among direct-to-physician promotional activities (Wittink, 2002),[29] which in turn represent the largest cost component of the overall promotional costs associated to the commercialization of new drugs (Donohue et al., 2007).

If commercialized in absence of cooperation, upstream innovators developing compounds associated to high-NPP indications will require larger investments in order to assemble and train representatives. Furthermore, the promotional activities carried out by these sales forces will not enjoy the lower average detailing costs derived from scope economies (i.e., promotion of multiple compounds in a single visit) or the benefits of branded reputation germane to established commercializers. It follows that gains of cooperation are higher for deals including indications with larger NPP values. If this type of gains were the primary reason behind the Medicare-oriented licensing surge, removing high-NPP US deals from the sample should identify a significantly milder increase in Medicare-oriented licensing.

To construct NPP we obtained data on the number of actively practicing physicians by specialty from the Association of American Medical Colleges "Physician Specialty Data Book" (Erikson et al., 2012). This source lists 36 specialties with their respective number of physicians actively providing patient care in the US (as of 2012). We then asked an expert to generate a mapping between targeted conditions and physician specialties that are likely to prescribe treatments for each of them in the

[29]See Manchanda and Honka (2005) for an excellent review on the extent, role of and attitudes towards detailing.

US[30] Using this indication/specialty mapping, we identified the set of unique specialties associated to each indication and generated NPP for each of these by summing the numbers of prescribing physicians among specialties within the set. The correlation between NPP and MMS (at the indication level) is about 0.02 and not statistically significant.

Estimates from columns 3 and 4 in table 2.7 are inconsistent with the hypothesis of a licensing surge driven by the avoidance of this type of duplicative investment.[31] Much to the contrary, the key triple interaction parameters for the period immediately following the program's enactment are respectively 0.62 and 0.61, larger than that obtained from the full sample (0.52). With the same caveat issued before (i.e., these results should not be taken to argue that complementary assets related to market distribution are an unimportant source of cooperation gains), and together with our results pertaining the role of market power, these results reinforce the idea that cooperation gains based stemming from superior innovative productivity may have been the key articulating force behind the documented licensing surge.

2.8 Robustness checks

Contract terminations and adverse selection

As we mentioned in section 2.2, the potential aggravation of an adverse selection problem could undermine the realization of productivity gains derived from cooperation. This concern is particularly accentuated given the immediacy of the rise in Medicare-oriented licensing since âone may presumeâ commercializers' drive to secure market presence in Medicare-oriented markets may have lead to rushed negotiations at the expense of rigorous due diligence, therefore leading to poorer selection standards and/or contract design. Here we provide some evidence that discredits this hypothesis.

To do so, we turn to contract termination data. Cortellis reports the status of each deal (i.e., active/completed or terminated) at the time the data was downloaded, as well as the termination dates for contracts encountering this outcome. It is well known that one leading reason for contract terminations is the observation of poor clinical testing outcomes. Thus, one repercussion of a potential strengthening of the adverse selection problem could be an increase in contract termination rates among deals including Medicare-oriented indications.[32]

[30] The mean number of physicians per specialty is about 17,000. The average number of specialties per indication is 3.2.

[31] As before, these are obtained by respectively removing 25% and 50% of deals with higher NPP. A deal's overall NPP value is computed by summing the number of prescribing physicians across the set of unique specialties across all included indications (i.e., no double counting of specialties). Similar results are obtained by summing indications' individual NPP values (i.e., allowing double counting) and by instead removing the 25% and 50% of indications with higher NPP values.

[32] A second reason for terminations is the unilateral exercise of a termination clause (Lerner and Malmendier, 2010), typically triggered by updated information regarding the candidate's economic prospects. However, since Part D improved market size expectations for Medicare-oriented indications, this factor is hard may partly offset the observed effect. With the data at hand, we are unable to tease these two effects apart.

Rates of terminations within 1, 2, 3, and 4 years of the contract signing date are reported in table 2.5 for contracts including the US and signed in 1998-2010.[33] Termination rates are increasing in time frame used to measure them, reflecting the gradual arrival of new information, but also the possibility that our data provider records terminations with a lag. Comparing rates across time frames, it is observed that roughly half of 4-year terminations unfold over the first two years.

In order to assess whether Part D-fueled deals were associated with higher termination rates we estimate probit models (at the licensing deal level) using as dependent variable an indicator that is activated if a contract is terminated within each time frame. For conciseness, we focus on terminations observed within 2 and 4 years. As before, we use triple difference specifications that include year and licensing stage fixed effects and clusters standard errors within DMMS/t cells. Results are presented in table 2.6.

A positive coefficient for the triple interacted variables would suggest that Part D-fueled deals were more likely to be terminated. While positive, estimated coefficients are small (particularly for the 4-year timeframe) and estimated imprecisely. Based on this evidence, we are reluctant to support the idea that a deepening of a potential adverse selection problem eroded the productivity gains associated to the surge in cooperation.

Alternative exposure measurements

To assess the robustness of our results to the measurement of exposure to the shock we consider two alternative proxies. Before describing their computation, recall that the baseline MMS measure is computed as the percentage of drug prescriptions within each therapeutic category that were issued to individuals enrolled in Medicare. The first alternative measure (MMS_EXP) is computed as the share of prescription drug expenditure paid for by Medicare within each 3-digit ICD9 therapeutic category; the second (MMS_PAT), as the percentage of patients purchasing medications within each category during each year that were enrolled in Medicare. Both measures are computed using 2003 MEPS prescription drug consumption and insurance data, and code an individual as enrolled in Medicare if he/she was enrolled during at least one month during that year.[34] All measures are significantly and highly correlated (i.e., correlations of 0.7 or more)

Columns 1-3 of table 2.8 reproduce our main result of Table 3 (column 1) using all each of these measures. For ease of comparison, we also reproduce the original results in column 1. Results are very similar across versions of the MMS variable, particularly in what regards the coefficient associated to the triple interaction for period 2004-2008.

Non-prescribability

The coverage benefits granted by Part D to Medicare enrollees apply to prescription drug expenditures. However, upon regulatory approval, many biotech compounds are live organisms, typically administered intravenously in in-patient settings. This type of drug expenditure is covered by a

[33] We restrict the sample to 2010 in order to compute 4-year terminations for all indications in the sample.

[34] Over 90% of individuals coded as enrollees in this way were enrolled full year.

separate insurance program (Medicare Part B), which implies that our estimates may suffer from an attenuation bias.

Our data does not allow us to ascertain the planned route of administration for each compound. Thus, in order to ascertain the magnitude of this bias we conduct two indirect checks. First, we use the MEPS pharmacy type information, which reports the types of pharmacy providers from which the person's prescribed medicines were purchased (drugstore, mail, HMO/clinic/hospital, etc.). For each 3-digit ICD9 category we compute the percentage of prescriptions purchase in in-patient settings (HMOs, clinics, hospitals) and correlate it with MMS. No significant correlation is observed.

Secondly, we re-estimate our main specification without including alliances including cancer indications, as in-patient administration is by and large more common among cancer treatments. Results are presented by column 4 of table 2.8. As expected, the coefficient for the triple interaction is slightly larger, but of overall similar magnitude.

Non-dichotomic shock exposure

Our last check pertains the dichotomic measurement of Medicare orientation. Here we employ an alternative empirical strategy to show that our main result holds when we instead consider a more continuous measure.

Using the baseline MMS measure, we define an deal's degree of Medicare orientation based on the MMS quintile associated with the highest-MMS indication included in the deal. With this, deals with the lowest exposure to the shock are those that only include indications in the lowest quintile of the MMS distribution. Alliances with the second highest level of exposure include at least one indication in the second quintile of the distribution but no indication in upper quintile, and so on.

For each of these 5 levels of Medicare-orientation we create an independent subsample including both deals including and not including the US among licensed territories, and estimate a diff-in-diff model with the following specification

$$N^q_{US,t,s} = \alpha^q_1 + \alpha^q_2 \cdot US \cdot DMMS + \sum_{j=2,3} \sigma^q_j \cdot US \cdot PER_j + \lambda^q_t + \mu^q_s + + \varepsilon^q_{US,t,s},$$

where q indexes each quintile and $N^q_{US,t,s}$ in analogous form to our analysis of section 2.6. For each subsample, coefficients represent the differential licensing patterns for development and commercialization including and not including the US territories. A Part D-fueled surge of US licensing would be consistent with estimates for σ^q_2 that increase across subsamples associated to progressively higher Medicare orientation.

Results (presented in table 2.9) are largely consistent with this idea. Starting from a small and imprecisely estimated σ^q_2 parameter in the first quintile subsample, estimates become progressively larger and more precise through the fourth quintile. For deals with in the fifth quintile, the effect is slightly smaller than in the third and fourth quintiles. A possible explanation for this stems

from the higher share of cancer targets in this subsample. Since these are more often administered intravenously than other therapies, for the reasons explained in the previous section, the Part D shock may have been less relevant in their case.

2.9 Concluding remarks

Based on the early insights of Schumpeter (1942) and Nordhaus (1969), the idea that a larger size of the downstream consumer market may increase the rate of technological innovation has received a considerable amount of attention by researchers. The underlying model in this stream of research is based on the *pull* effect exerted by market size: larger potential rewards to commercialization justify larger amounts of R&D investment, which in turn translate into an increased availability of new technology products for consumers. As embodied by strictly positive estimates for the elasticity of innovation to market size, numerous studies provide empirical evidence to back this effect. In this paper we argue that, in addition to *pulling* new technologies into the market, larger consumer markets may *catalyze* the development process of technologies in embryonic state, increasing the probability that they will reach the market and/or do so in a shorter period of time.

Our argument has two parts. First, inter-firm cooperation aimed at the development and commercialization of new technologies enhances productivity by pooling complementary capabilities. Second, a larger market size increases the extent and match quality of cooperation because it reduces the relative importance of transaction costs in Markets for Technology. By analyzing the impacts of Medicare Part D on the international drug candidate licensing market, we provide compelling evidence to the latter point. Although we cannot provide conclusive evidence for the existence of productivity gains derived from cooperation, a "ruling-out of alternatives" strategy provides some reassurance. Research from the fields of Strategy, Organizations, and Health Economics also supports the claim.

Our theoretical analysis shows that the impacts of Medicare Part D on drug candidate licensing can only be rationalized if transaction costs are meaningfully large. This suggests that policies encouraging the timely and properly-matched cooperation between upstream innovators and downstream commercializers could have a positive impact on overall innovative productivity. Our results from the pharmaceutical industry suggest that one important friction may be encountered in the negotiation room, as firms try to "agree to fair terms." To address these, as discussed by Agrawal et al. (2014), governmental intervention could take the form of "low cost, timely and predictable dispute resolution mechanisms, and insurance against certain types of risks."

A natural way to further this research agenda would be to decompose the aggregate surge in cooperation into firm-level responses and outcomes. As suggested by the insights of Ceccagnoli et al. (2010), Ceccagnoli and Jiang (2013), and Ceccagnoli et al. (2014), organizational capabilities and the nature and extent of downstream firms' internal R&D may be associated to varying marginal returns to cooperation, leading to heterogeneous licensing responses. Such analysis would be useful in order to evaluate the extent to which cooperation gains were realized, as well as the implications

for the industrial organization of the downstream markets.

Figure 2-1: Bundled and unbundled licensing upon the shock arrival

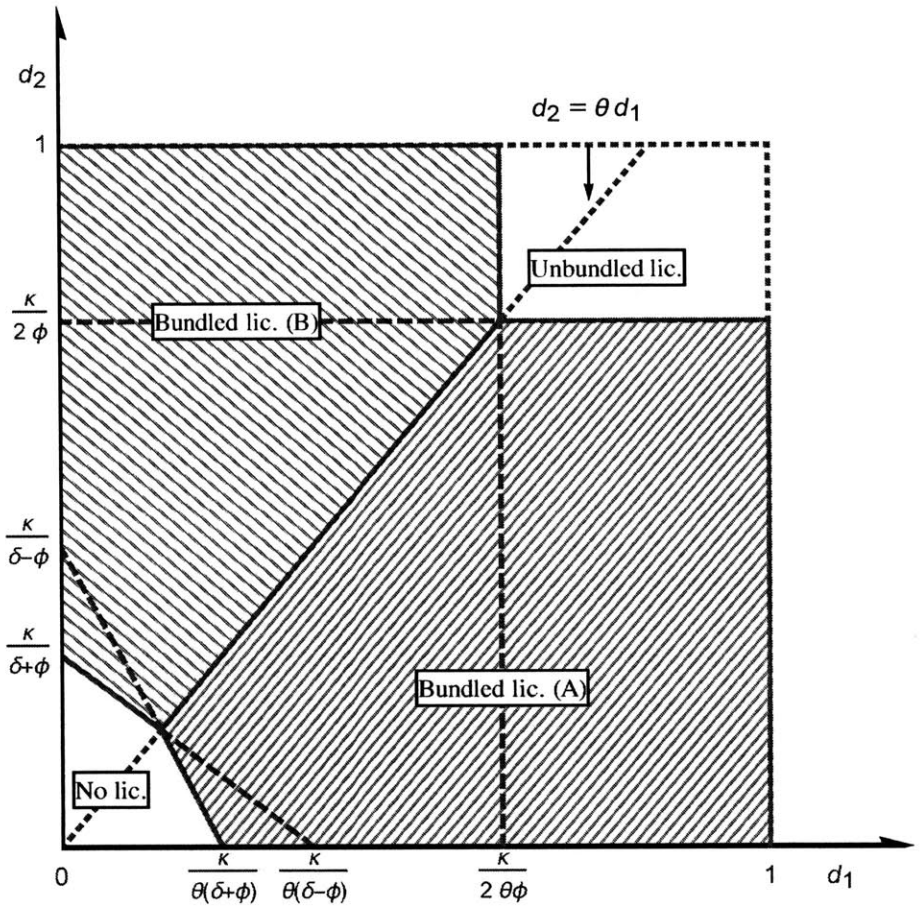


Figure 2-2: Kernel density of MMS scores for conditions targeted by indications in the data

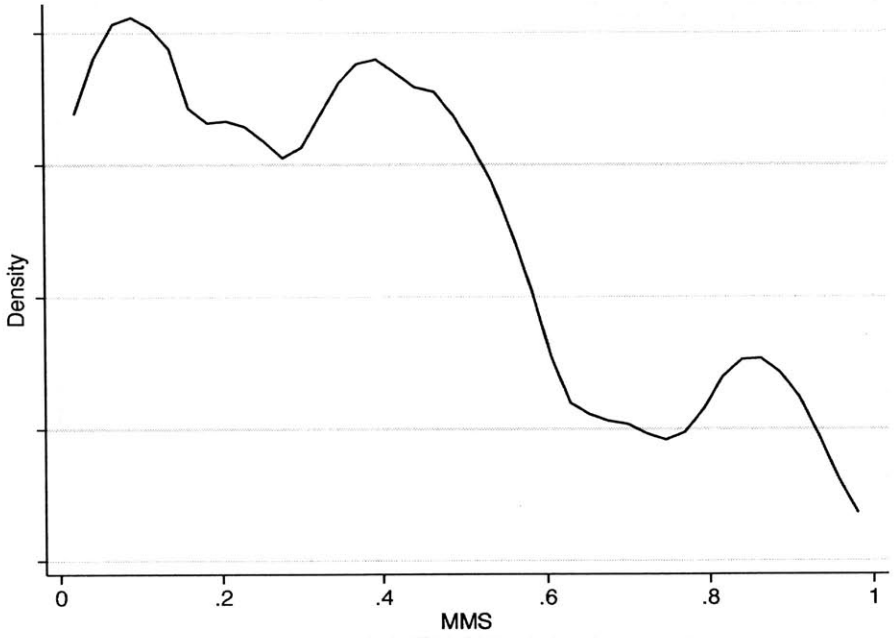
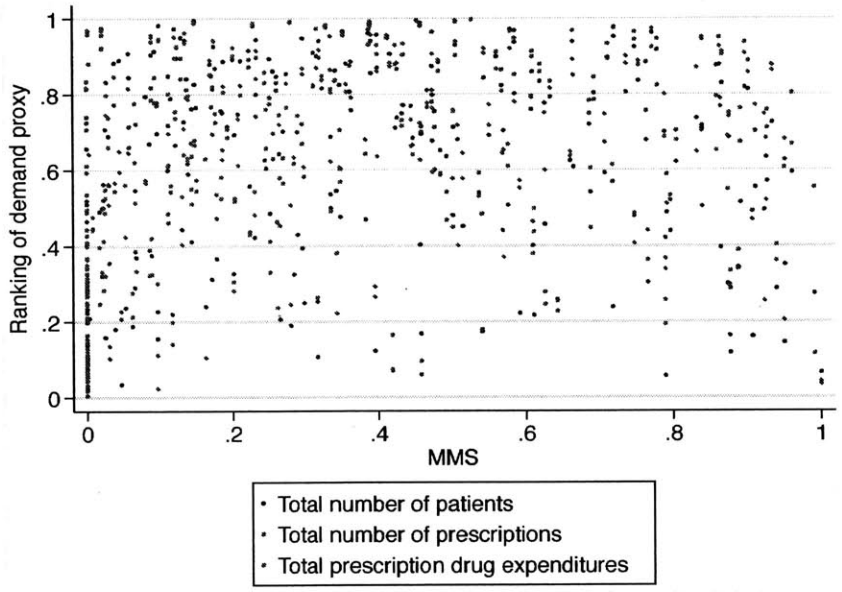
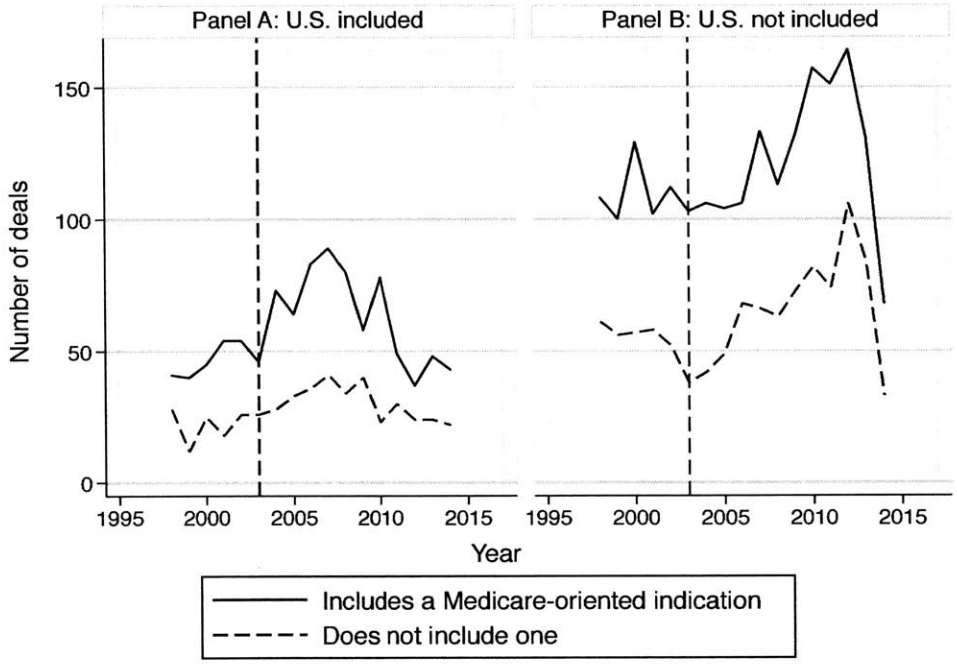


Figure 2-3: MMS and proxies for baseline demand



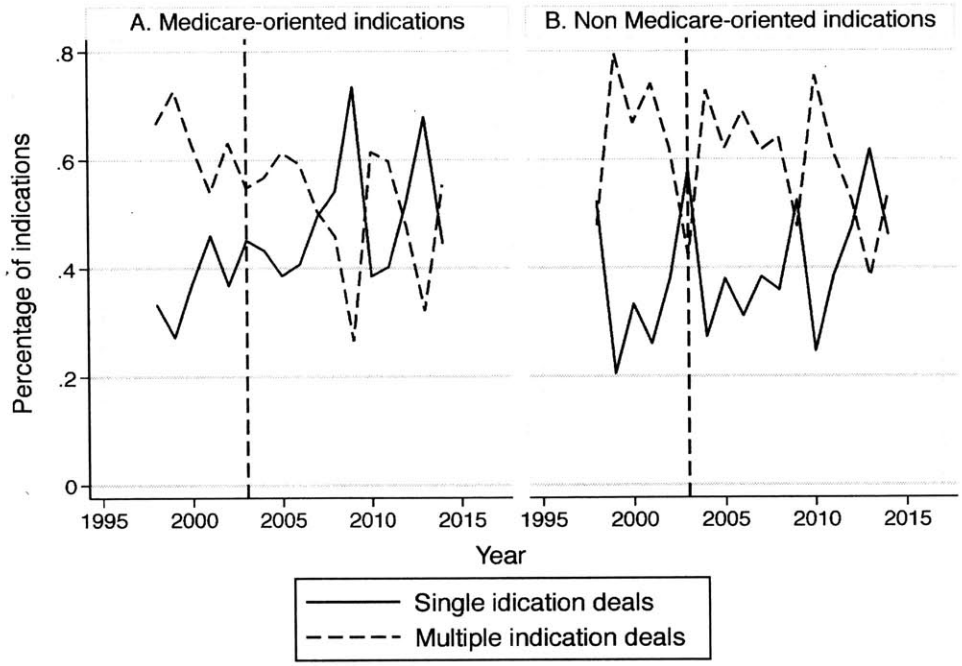
The vertical axis displays the ranking (re-scaled to the unitary interval) of each proxy within the total sample.

Figure 2-4: Number of licensing contracts by degree of Medicare orientation (1998-2014)



Medicare-oriented indications correspond to those associated with above-median MMS score.

Figure 2-5: Medicare orientation and number of licensed indications (contracts including the US)



Medicare-oriented indications are those associated with MMS scores above median. Conversely, non Medicare-oriented indications with MMS scores equal or lower than the median.. The vertical axis corresponds to the percentage of indications (of each Medicare orientation level) that were alliances for the development and commercialization of a total of 1, and 2+ indications.

Table 2.1: Commercialization probabilities, gains, and the number of new technology products (given d_1, d_2).

	(1) Probability of commercialization x_1	(2) Probability of commercialization x_2	(3) Expected total gains	(4) Expected number of new technology products
	Panel A: Bundled licensing only			
No licensing	p	p	$p(d_1+d_2)$	$2p$
Bundled licensing	$p+\delta$	$p+\delta$	$(p+\delta)(d_1+d_2)-\kappa$	$2(p+\delta)$
	Panel B: Bundled and unbundled licensing			
No licensing	p	p	$p(d_1+d_2)$	$2p$
Bundled licensing ($d_1 > d_2$)	$p+\delta+\phi$	$p+\delta-\phi$	$(p+\delta)(d_1+d_2)$ $+\phi(d_1-d_2)-\kappa$	$2(p+\delta)$
Bundled licensing ($d_1 < d_2$)	$p+\delta-\phi$	$p+\delta+\phi$	$(p+\delta)(d_1+d_2)$ $+\phi(d_2-d_1)-\kappa$	$2(p+\delta)$
Unbundled licensing	$p+\delta+\phi$	$p+\delta+\phi$	$(p+\delta+\phi)(d_1+d_2)-\kappa$	$2(p+\delta+\phi)$

Table 2.2: Main descriptive statistics.

	(1)	(2)	(3)	(4)
	Compounds		Indications	
	US included	US not included	US included	US not included
	Panel A: Distribution across time periods			
1998-2003	69.2	162.7	116.2	276.8
2004-2008	112.2	170.0	180.6	250.8
2009-2014	79.3	209.2	114.7	289.8
	Panel B: Distribution across development stages			
Discovery	39.6	104.6	59.1	146.2
Phase I	13.1	25.5	19.4	38.0
Phase II	18.6	28.0	32.5	49.1
Phase III	10.2	15.8	17.1	26.8
Pre-registered	4.2	7.4	6.9	13.7
Total	1,452	3,081	2,288	4,654

The development stage at licensing corresponds to the most advanced development stage across the set of licensed indications at the time the contract takes place. For a subset of licensed indications (1.3% of the total) Cortellis was not able to elucidate the specific development stage and thus reported it as âclinical.â We treat them as Phase I indications. In our analysis, results do not qualitatively change when we drop the contracts through which they were licensed.

Table 2.3: Main results.

	(1)	(2)	(3)	(4)
DMMS	0.76***	0.75***	0.66***	0.84***
	(0.06)	(0.06)	(0.06)	(0.12)
US	-0.78***	-0.74***	-0.96***	-0.42***
	(0.07)	(0.10)	(0.08)	(0.10)
US×DMMS	-0.05	-0.1	-0.1	-0.22
	(0.09)	(0.11)	(0.09)	(0.14)
DMMS×PER04_08	-0.20**	-0.04	-0.09	-0.04
	(0.08)	(0.12)	(0.09)	(0.18)
DMMS×PER09_14	-0.12		-0.03	-0.12
	(0.09)		(0.11)	(0.14)
US×DMMS×PER04_08	0.52***	0.43***	0.44***	0.56***
	(0.07)	(0.08)	(0.10)	(0.12)
US×DMMS×PER09_14	-0.05		-0.07	-0.21
	(0.12)		(0.14)	(0.21)
Sample	All	1998-2005	Licensed at discovery or phase I	Licensed at phase II or phase III
N	340	160	136	136

Negative binomial estimates using as dependent variable the number of licensing agreements, aggregated at the DMMS/US/year/licensing stage level. DMMS is an indicator activated if the licensing contract includes at least one indication with above-median MMS score. All models include year and licensing stage fixed effects. Standard errors (clustered within DMMS/US/year cells) are presented in parentheses. Legend: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Table 2.4: Unbundling of indications.

	(1)	(2)
Model:	Ordered Probit	Probit
Dependent variable:	$I = 1, 2, 3+$	1 if $I > 1$, 0 otherwise
DMMS	-0.22***	-0.23***
	(0.07)	(0.06)
US	-0.14	-0.17
	(0.14)	(0.13)
US×DMMS	0.18	0.22
	(0.16)	(0.14)
DMMS×PER04_08	0.08	0.09
	(0.09)	(0.09)
DMMS×PER09_14	0.12	0.14
	(0.10)	(0.09)
US×DMMS×PER04_08	-0.33*	-0.38**
	(0.19)	(0.17)
US×DMMS×PER09_14	-0.26	-0.33
	(0.23)	(0.22)
N	6,942	6,942

Models estimated at the licensed indication level. DMMS is an indicator activated if the indication has above-median MMS score. All models include year and licensing stage fixed effects, as well as interactions between the indicators for time periods and the inclusion of the US among licensed territories. Standard errors (clustered within DMMS/US/year cells) are presented in parentheses. Legend: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Table 2.5: Termination rates for contracts including the US.

Year of contract	(1) Years from contract signing 1	(2) 2	(3) 3	(4) 4
1998	0.04	0.08	0.10	0.12
1999	0.02	0.07	0.12	0.23
2000	0.01	0.08	0.16	0.21
2001	0.02	0.07	0.14	0.17
2002	0.02	0.07	0.09	0.12
2003	0.00	0.04	0.06	0.08
2004	0.01	0.02	0.05	0.07
2005	0.00	0.03	0.04	0.08
2006	0.00	0.02	0.03	0.04
2007	0.01	0.04	0.05	0.07
2008	0.00	0.01	0.02	0.06
2009	0.01	0.02	0.02	0.04
2010	0.00	0.01	0.01	0.01

Table 2.6: Termination rates and Medicare Orientation

Observed within:	(1) 2 years	(2) 4 years
DMMS	0.09	0.09
	(0.07)	(0.07)
US	0.43***	0.33***
	(0.10)	(0.08)
US×DMMS	-0.16	-0.02
	(0.12)	(0.13)
DMMS×PER04_08	-0.09	-0.11
	(0.22)	(0.14)
DMMS×PER09_14	-0.18	0.06
	(0.15)	(0.12)
US×DMMS×PER04_08	0.25	0.06
	(0.27)	(0.21)
US×DMMS×PER09_14	0.08	-0.13
	(0.22)	(0.16)
N	3,445	3,445

Probit estimates using as dependent variable an indicator activated if the licensing contract is terminated within each time frame from the contract signing date. Estimated at the licensing contract level. DMMS is an indicator activated if the licensing contract includes at least one indication with above-median MMS score. All models include year and licensing stage fixed effects, as well as interactions between the indicators for time periods and the inclusion of the US among licensed territories. Standard errors (clustered within DMMS/US/year cells) are presented in parentheses. Legend: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table 2.7: Market power and complementary assets.

	(1)	(2)	(3)	(4)
DMMS	0.73***	0.69***	0.73***	0.72***
	(0.05)	(0.05)	(0.06)	(0.06)
US	-1.15***	-1.77***	-1.18***	-1.45***
	(0.07)	(0.09)	(0.08)	(0.10)
US×DMMS	-0.13*	-0.05	-0.14*	-0.09
	(0.08)	(0.07)	(0.08)	(0.08)
DMMS×PER04_08	-0.08	-0.08	-0.10	-0.10
	(0.08)	(0.07)	(0.09)	(0.08)
DMMS×PER09_14	0.12	0.33***	0.00	-0.09
	(0.09)	(0.10)	(0.10)	(0.12)
US×DMMS×PER04_08	0.50***	0.46***	0.62***	0.61***
	(0.07)	(0.06)	(0.07)	(0.08)
US×DMMS×PER09_14	-0.09	-0.04	0.12	0.13
	(0.13)	(0.10)	(0.13)	(0.13)
Deals removed from full sample	25% highest HHI	50% highest HHI	25% highest NPP	50% highest NPP
N	340	340	340	340

Negative binomial estimates using as dependent variable the number of licensing agreements, aggregated at the DMMS/US/year/licensing stage level. DMMS is an indicator activated if the licensing contract includes at least one indication with above-median MMS score, implemented the median of each measureâs distribution. The construction of HHI and NPP is described in the text. All models include year and licensing stage fixed effects. Standard errors (clustered within DMMS/US/year cells) are presented in parentheses. Legend: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Table 2.8: Robustness (shock exposure and non-prescribability).

	(1)	(2)	(3)	(4)
DMMS	0.76***	1.30***	0.75***	0.42***
	(0.06)	(0.08)	(0.05)	(0.07)
US	-0.78***	-0.85***	-0.76***	-0.75***
	(0.07)	(0.08)	(0.07)	(0.07)
US×DMMS	-0.05	0.03	-0.10	-0.09
	(0.09)	(0.09)	(0.09)	(0.11)
DMMS×PER04_08	-0.20**	-0.41***	-0.19**	-0.30***
	(0.08)	(0.09)	(0.08)	(0.09)
DMMS×PER09_14	-0.12	-0.32***	-0.14*	-0.21*
	(0.09)	(0.10)	(0.08)	(0.11)
US×DMMS×PER04_08	0.52***	0.53***	0.53***	0.66***
	(0.07)	(0.06)	(0.07)	(0.09)
US×DMMS×PER09_14	-0.05	-0.05	0.00	0.03
	(0.12)	(0.11)	(0.12)	(0.16)
Medicare orientation meassure	Baseline	MMS_EXP	MMS_PAT	Baseline
Deals removed from the full sample	None	None	None	Deals including cancer indications
N	340	340	340	340

Negative binomial estimates using as dependent variable the number of licensing agreements, aggregated at the DMMS/US/year/licensing stage level. DMMS is an indicator activated if the licensing contract includes at least one indication with above-median MMS score, implemented the median of each measureâs distribution. All models include year and licensing stage fixed effects. Standard errors (clustered within DMMS/US/year cells) are presented in parentheses. Legend: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Table 2.9: Robustness (dichotomic exposure).

	(1)	(2)	(3)	(4)	(5)
US	-0.86***	-0.48**	-0.91***	-0.80***	-0.87***
	(0.13)	(0.22)	(0.09)	(0.11)	(0.09)
US×PER04_08	0.21	0.40*	0.51***	0.52***	0.44***
	(0.18)	(0.23)	(0.12)	(0.15)	(0.11)
US×PER09_14	-0.22	-0.29	-0.01	-0.05	-0.16
	(0.20)	(0.30)	(0.14)	(0.21)	(0.18)
Quintile of indication with highest MMS	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}
N	170	170	170	170	170

Negative binomial estimates using as dependent variable the number of licensing agreements, aggregated at the US/year/licensing stage level. All models include year and licensing stage fixed effects. Standard errors (clustered by year) are presented in parentheses. Legend: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Chapter 3

Regional Impacts of New Medical School Entries on the Supply of Physicians

3.1 Introduction

Do increases in the supply of physicians lead to changes in health expenditures and patientsâ health outcomes? This question relates to the controversial supplier-inducement hypothesis as an explanation for the increase in health expenditure (Cromwell and Mitchell, 1986; Evans Robert, 1974; McGuire et al., 1988; Newhouse, 1992; Rice, 1983). Earlier empirical studies used cross-national comparisons and took the supply of physicians as exogenous. Countries with more doctors had lower health expenditure, but this result is very sensitive to the specification. In contrast, a more robust result is that the number of doctors increased health expenditure in fee-for-service reimbursement systems.

As a first step to investigating the effects of the supply of physicians on the quality and effectiveness of medical services, this study exploits regional variations in the supply of physicians resulting from the large expansion of medical schools in the US during 1960-1980. Due to the lack of data on health expenditures and other second stage outcomes dating back to before the 1980s, this chapter focuses on uncovering and understanding the first stage, i.e., the differential effects of medical schools on the supply of physician across geographical areas.

Previous studies, such as Fox and Richards Jr (1977), have shown large geographical variation in the supply of physicians in the US. There have also been studies on what determines the geographical distribution of physicians. Standard location theory has been used to predict and explain choices of practice location by health professionals (Newhouse et al., 1982a,b; Rosenthal et al., 2005). They incorporate location preferences into utility maximization and assume that a number of factors affect the relative attractiveness of a certain location. Such factors vary across studies, and often include income, quality of leisure, distance to central cities and workload. As Dussault and Franceschini

(2006) point out, one implication of such models is that the distribution of health professionals depends on amenities as well as demand.

Intuitively, we expect medical school entries to have an impact on the geographical distribution of physicians and we expect the impact to be larger on physician supply in areas nearby. First of all, expansion of medical schools brought in new research facilities and new residency positions, and therefore had an immedicate and local effect in the region hosting the new medical school upon the entry. Second, in the framework of location theory, entry of a new medical school changes physicians' preference for working near vs. further away. Being closer to a teaching and research facility is an amenity that makes a location more attractive. Being closer to other physicians and health workers is another amenity, which, however, can work in both directions – physicians may prefer working closer to peers for reasons including human capital spillovers, but they may also prefer to work in areas with fewer competitors. Such effects took place upon the entry of a medical school and could have affected areas nearby besides the county hosting the entrant. Finally, there were smaller programs accompanying the expansion of medical schools, which are described in Section 3. Such programs often offered financial support to medical students in return for service in rural areas. Such programs could have resulted in faster growth in physicians in rural parts near medical schools.

Since new medical schools were not randomly located, I adopt a difference-in-difference strategy and try to detect breaks in any pre-existing differences across regions in the trend of physician supply around the time of new medical school entries. The identifying assumption is that absent the medical school expansion, any pre-existing differences would have continued on the same trends. I use the regional variation in the proximity to new medical schools and estimate the effect of a new medical school on physician supply in nearby areas. I first look at the effect in seven years to capture both the immediate and lagged responses. Then I investigate how such regional impacts of a new entry evolve over time (over 20 years), and also conduct a falsification test which confirms that the medical schools did not have an impact on physician supply in a 20-year pre-period.

Two event studies show the proximity to new entries to be linked with faster growth in county-level supply of physicians. Among five states which never had a medical school before 1967, two that opened a first medical school in 1967 experienced faster growth in physician levels thereafter. Among counties more than 300 miles from any medical school in 1960, those witnessing entries of new medical schools within 100 miles experienced faster growth in physician supply after the entries.

Using difference-in-difference regressions, I find that a new medical school increased physician supply within that county by about 500 after seven years, relative to other counties. This was more than 3 times the sample average. As for the number of physicians per 100,000 people, the increase was around 100 and above the sample average. Using an alternative geographic unit for analysis, I find that a new medical school increased physician supply by around 100 in counties within 50 miles, relative to counties farther away, and this increase is 2/3 of the sample average. Considering the number of physicians per 100,000 people, the increase was around 20 and about ÂŒ of the sample

average. These estimates suggest a strong positive regional effect on the supply of physicians. The regional estimates are smaller than the local (specific to county) effects as expected.

I also investigate how such regional impacts evolved over time. I find that a new medical school had similar impacts in the year of entry and in the next 20 years, which indicates that most of the impacts could be attributed to the immediate responses. The new job openings and new facilities brought by the new medical schools, and the attractiveness of working closer to a teaching and research facilities and peers are the main reasons why physician supply grew faster in areas closer to the newly established medical school, as opposed to the correlated location choices for education and professional practice of medical students. I find new medical schools to have no effect on the physician supply in most pre-entry years, which supports the hypothesis that locations of new medical schools were not correlated with other underlying determinants of physician supply.

This study is devoted to uncovering and understanding the differential effects of medical schools on the supply of physician across geographical units, as opposed to estimating the aggregate effects. Caution should be taken in interpreting the results as the overall effects of medical school expansion â it is plausible physician stock grew in all areas after the expansion, though the increase differed across areas. Therefore, my analysis probably underestimates the full impact of medical school expansion on the supply of physicians, but this does not harm the effectiveness and relevance of my analysis in terms of understanding the impact of the medical school expansion on the geographical variation of physicians.

The remainder of this paper is organized as follows. Section 3.2 describes the data. Section 3.3 introduces the institutional background and presents time-series evidence of the correlation between medical school expansion and growth in physician supply. Section 3.4 illustrates two event studies, which illustrate the regional impacts of medical schools. Section 3.5 discusses the empirical strategies motivated by previous work and the event study results. Section 3.6 reports estimation results, robustness checks and falsification checks. Section 3.7 concludes the study.

3.2 Institutional background

Due to publicly criticized shortage of physicians, there was a big federal push and authorized funding for new medical schools during 1960-1980. In the 1940s, the public began to demand a larger supply of physicians for several reasons - shortage of physicians, especially in rural and underserved areas, growing health consciousness, and swelling college enrollments and thus more applicants. The issue achieved center stage upon the publication of the Bane Report by the Surgeon Generalâs Consultant Group on Medical Education in 1959, which shocked the public by projecting a shortfall of 40, 000 physicians by 1975 and offered a blueprint for expanding existing medical schools and creating new medical schools.

The immediate legislative consequence was the Health Professions Educational Assistance Act of 1963, which funded establishment of new medical schools and larger enrollment size of existing health programs, and also made federal loans available to medical students. This legislation was

followed by revised acts in 1965 and more generous acts in 1968 and 1971 (Ludmerer, 1999). There is no direct evidence about how the funding was distributed and how locations of new medical schools were decided, and there are two important facts based on historical government reports and media reports. First, the timing of these medical school constructions was relatively concentrated, from 1960 to1980, and mostly from 1970 to 1980. Second, the federal government decided which applicants got the funding for establishing new medical schools or expanding pre-existing medical schools, and no evidence suggests any preference for location of new schools from the federal governmentâs side.

Some schools opened in areas where no medical schools ever existed, yet others located closer to pre-existing medical schools. Figures 3-1, 3-2, and 3-3 illustrate the geographical distribution of medical schools within the US in 1940, 1960 and 1980, respectively. Comparing these maps, there was little change from 1940 to 1960, but there was noticeable development of new medical schools from 1960 to 1980. As one example, between 1960 and 1980, Nevada and Arizona welcomed their first medical schools, while CA and the east saw more medical schools entering in addition to their previous medical schools. Figure 3-4 illustrates the geographical distribution of new medical schools established between 1960 and 1980. We can see that they were more sparsely located in the west and more spread out in the east.

3.3 Data

I use data on county-level number of active Non Federal M.D. physicians and population from the 2006 release of the Area Resource File by U.S. Department of Health and Human Services. Such data is available on a decennial basis for earlier periods and on an annual basis for more recent years. I use data from 1940, 1950, 1960, 1970, 1975, 1980, 1981-1986, 1988-1990, 1992-2000.

Census data from IPUMS provides an alternative source of data for physicians and population. I use the 1940-2000 Census, combined with the sample weights, to compute county-level number of physicians and population. The advantage about the census data is that it provides more potential controls, such as the gender and age composition of the population, but the census data covers even fewer years than the Area Resource File data and that the computed number of physicians contains more measurement error. Worse still, its county code has not been matched with medical school locations so can only be used for state-level regressions. Therefore, I report the main results using the Area Resource File data, but also include results from IPUMS data as a robustness check. In some of the analyses, I distinguish rural and urban areas, where the classification comes from the earliest SEER Rural-Urban Continuum Codes, i.e., the Codes as of 1974.

Information about former and current medical schools in the US comes from online sources and historical records such as of Medical Colleges of the United States and of foreign countries published by the American Medical Association. This data includes information on the year of establishment and year of exit for most medical schools that ever existed in the US. Unfortunately, I have no information on the size of each medical school, so I can only use the establishment of new medical

schools, but not the expansion in sizes of pre-existing medical schools as my source of variation. Among all these medical schools, I exclude those not existing during my period of study (exiting before 1930), those not recognized by the medical society, those that were forced to close due to fraud, and those that only produced one class of graduates. I also dropped a few schools which no longer exist and whose entering or exiting years are unknown. After these sample restrictions, we are left with 173 medical schools existing after 1930. Out of these 173 medical schools, 170 are still operating, 96 were established before 1960 and 47 were established between 1960 and 1980.

Federal legislation had an impact on both the number of medical schools and the number of graduates. Figure 3-5, from ?, shows the supply of national medical school graduates increased slowly during 1961-1970, sped up during the 1970s, slowed down in the early 1980s and stabilized eventually. Figure 3-6 plots the time series trend for medical schools and physician stock (using Area Resource File data), both showing a pattern of sharp increase between 1960 and 1980 and stabilization afterward. There seems to be a lag in the growth of physician supply in response to the expansion of medical schools. The number of medical schools increased drastically from 1960 to 1980 and stabilized after that, while the number of physicians increased slightly 1960-1970, more abruptly after 1970 and stabilized by 2000.

Notably, the Health Professions Educational Assistance Act of 1976 claimed there was no longer a shortage in the total supply of physicians in the United States, but geographical and specialty maldistribution– ânot enough doctors in rural and inner-city areas, and there is a continuing decline in the number of doctors practicing primary care.â The Act extended the National Health Service Corps (NHSC) program, which aimed at recruiting and placing health professionals in critical shortage areas. Scholarships were provided for students who agreed to serve in such areas on completion of their education. The size of the project was small but it was projected to expand fast. Such program designs aimed at increasing supply of physicians and other health workers in rural and underserved parts of the US and might have led to differential results across rural (medically underserved) and urban areas in the effects of medical school expansion, however, as discussed in Section 3.6.4, I have not detested this in the data.

3.4 Identifying the Impact of Medical School Expansion: Exploiting Geographical Variation of New Medical Schools

Inspired by Figures 3-1, 3-2, 3-3, and 3-4, I conduct two event studies below to investigate the relation between the proximity to newly established medical schools and growth in physician supply. These are not the most rigorous form of event studies, since I have less frequent data in earlier years and around the entries.

Figures 3-7 and 3-8 report statistics for five states, Arizona, Nevada, Idano, Montana and Wyoming, all of which never had any medical schools before 1967. Arizona and Nevada each welcomed their first medical school in 1967, while their close neighbors, Idaho, Montana and Wyoming, have never had any medical schools till the present. Figure 3-7 shows the growth in number of physi-

cians over time in these states around 1967. The dots are the data points from the available years, with straight lines connecting these dots over time. The red vertical line marks the year 1967. The lines stayed flat before 1967. There was an obvious upward trend for Arizona past 1967, a slight upward trend for Nevada and almost flat trends for the three states that have never had a medical school.

When I draw the same graph with the number of physicians per 100,000 people instead, the pre-trends were different for these states â it was declining slightly for Arizona and Nevada before 1967, but increasing slightly for the other states. The number grew after 1967 for all states, while the positive change in slope past 1967 was higher for Arizona. A simple difference-in-difference interpretation of Figure 3-8 also suggests faster growth in physician supply in states after their first medical school entries.

Since a state is probably too large a geographical unit to examine the regional or local impacts of medical school entries, I conduct another event study on the county level. Define the control group as the 108 counties more than 300 miles away from any medical school from 1960 to 1980, and the treatment group as the 23 counties more than 300 miles away from any medical school in 1960, but within 100 miles to the closest medical school in 1980. Within the treatment group, define the year when the distance declined as t = 0. For the control group, take 1965 as t = 0. I have tried other definitions and the patterns in Figures 3-9 and 3-10 are robust to choices of $t = 0$ for the control group.

Figures 3-9 and 3-10 present the growths in county-level physician supply for counties in both the treatment and control group. The average level of physicians for the treatment counties is noisier, since there are fewer counties in the treatment group than the control group. This, however, does not prevent us from discovering obvious and interesting patterns. Physician supply (both number and number per 100,000 individuals) in the control counties stayed almost flat over time. The pre-entry trend for the treatment group was not much different from the control, but the supply of physicians grew faster after t=0 and stayed at a higher level.

3.5 Empirical Strategy

Sharing the same nature as the event study exercise, I will compare changes in physician supply in areas where new medical schools opened with other areas. An alternative strategy is to compare changes in physician supply in regions closer to where a new medical school opened with regions further away. The difference between the two is the choice of geographical unit.

In an ideal experiment, one would like to have the new medical schools randomly distributed. There is no evidence that this really happened in reality. Areas differing in demographic and socio-economic factors that might affect location choices of physicians (and health expenditures) might also differed in the probability of establishing a new medical school. Therefore, I adopt a difference-in-difference strategy is detect breaks in any pre-existing differences in the trend of physician supply around the time of new medical school entries across regions. The identifying assumption is that

absent medical school expansion, any pre-existing differences would have continued on the same trends.

Finkelstein (2005) adopts a difference-in-difference approach in studying the aggregate effects of health insurance, using the nationwide introduction of Medicare in 1965. Regional variation comes from the fact that different fractions of elderly people had private health insurance across regions before the introduction of Medicare. As the author points out, the empirical strategy is to compare changes in outcomes in regions of the country where Medicare had a larger effect on the percentage of the elderly with health insurance to areas where it had less of an effect. Duflo (2000) and Duflo (2004) use a major shift in education policy in Indonesia to test the hypothesis that spending more money on infrastructure can increase human capital and reduce poverty. The author evaluates the effect of Indonesiaâs large school construction program by combining regional variation in program intensity with differences across cohorts, and using a difference in difference estimator that controls for (additive) systematic variation of education both across regions and across cohorts. My strategy is similar to the abovementioned papers.

I first exploit the regional variation in the proximity to new medical school entries in estimating Equation refeq:FS below. Following Duflo (2000)âs falsification checks using cohorts, I also check the pattern of effects over time (from 20 years into the past till 20 years into the future) in Equation 3.2 below.

3.5.1 Econometric Model

The basic regression equation for the impact of the medical school expansion on regional physician supply is as follows.

$$Y_{jt} = \alpha_j + \delta_t + \beta * (\text{Proximity to New Medical School})_{j,t-7} + \gamma X_{jt}, \tag{3.1}$$

This regression is either run on the state level, in which case j refers to state j, or on the county level, in which case j refers to county j. The dependent variable is either the number of physicians or the scaled number of physicians (i.e., per 100,000 people) in area j as of year t. α_j are area fixed effects. δ_t are year fixed effects, which control for any nationwide trends in the growth of physicians.

There are two alternative measures used as $(\text{Proximity to New Medical School})_{j,t-7}$, the number of medical schools in area j as of year (t-7), or the number of medical schools within 50 miles of county j as of year (t-7). Here we choose the seven-year lag of ($ProximitytoNewMedicalSchool$) to capture both the immediate effects and the potential lagged effects which first took place when the first class graduated. I will use $(\text{Proximity to New Medical School})_{j,t-\tau}$ instead later to track the effects over time to help interpret these effects separately, where the values of τ are taken discretely from -20 to 20.

Equation 3.1, although in the form of cumulative numbers, captures the impact of the medical school expansion because first, only changes in the number of medical schools should have an effect

on the supply of physicians by construction of this equation; and second, most of the variation (change in number of medical schools) comes from the expansion period 1960-1980. To account for potential serial correlation over time within areas, I allow for an arbitrary variance-covariance matrix in the error structure within each area. To alleviate concerns that other factors might also have been changing differentially across different areas over time, I also try to include a series of time-varying state or county-level co-variates (Xjt). Constrained by the richness of historical data, I control for population, gender composition and age distribution when using the census data and control for population when using the Area Resource File Data.

To check the robustness of the results, I report in Section 3.6 the regression results using alternative data sources (Area Resource File and IPUMS), using fewer or more years of the data (1940-2000, or 1960-1990 to focus more narrowly on the expansion period), using alternative measures of physician supply(level or scaled level), and using alternative specifications (with or without controls). The identification assumption, shared with other difference-in-difference estimations, should not be taken for granted – the pattern of changes in physician supply could vary systematically across regions and over time. Consequently, I want to test implications of the identification assumption, for example, if the parallel-trends hold, we should expect a zero coefficient if we run the regression of current physician stock on medical schools 20 years into the future, because it should not affect the present unless future medical school locations are correlated with other underlying factors. Inspired by such an idea, I also estimate the following equation, which is a slight variant of Equation 3.1.

$$Y_{jt} = \alpha_j + \delta_t + \beta * (\text{Proximity to New Medical School})_{j,t-\tau} + \gamma X_{jt}, \tag{3.2}$$

Here the values of τ are taken discretely from -20 to 20, where $\tau = 0, 1, ..., 20$ are used to check how the impacts of medical school expansion changed over time, while $\tau = -20, -19, ..., -1$ serve as Placebo tests. I plot the coefficients β_τ for τ ranging from -20 to 20 in Section 3.6.4.

It is important to highlight several limitations of this research strategy and its implementation, stated below and some are discussed again in combination with regression results in Section 3.6. First, availability of data restricts my analysis. For example, my strategy does not capture the size increase of pre-existing medical schools, but only the establishment of new medical schools. Furthermore, physician data for earlier years (1940-1980) is available on a very infrequent basis (every 10 years or 5 years), so we have a relatively small number of observations compared with the number of controls (FE for each county and age distributions, for example).

Second, I want to re-emphasize that this study is intended as the first stage for examining the effect of physician supply on health expenditures, instead of an attempt to estimate the full effects of medical school expansion on physician supply. It is plausible physician stock grew in all areas after the expansion, though the increase differed across areas. Therefore, my analysis probably underestimates the full impact of medical school expansion on the supply of physicians, but this does not harm the effectiveness and relevance of my exercise in providing a valid first stage for the geographical variation of physicians.

Third, it is not clear ex ante what makes the right geographical unit to use for measuring the regional impacts of medical school expansion. As Pong and Pitblado (2001) point out, geopolitical or administrative areas, such as counties and states, are artificial and not necessarily the most appropriate units of analysis. States are usually too large and tend to hide regional variations, while counties are usually too small to capture regional spillovers. Neither a state nor county might be the right geographical division to run the regressions on. There is much discretion in the choice, but I will describe results using three different units â state, county and counties within 50 miles to new medical schools, and discuss my preference. Related to the second point, this does not harm the effectiveness of the first-stage exercise as long as we detect differential regional effects on the supply of physicians caused by geographical variation of medical schools unrelated to other time-varying factors.

Finally, one would probably suspect that locations of new medical schools were endogenous, and my DD strategy might fail if their locations were correlated with other time-varying determinants of physician supply and second-stage outcomes. I will test for the hypothesis that locations of medical schools were not correlated with other time-varying factors using the falsification tests in Section 3.6.

3.6 Regression Results

Tables 3-1 to 3-8 report results from regressions estimating Equation 3.1. In every table, I investigate the robustness of the results to several alternative specifications, each of which is reported in a separate column. The results I report include the estimated coefficient β, the standard variation, F-statistics from the joint test of the reported coefficients equating zero, number of observations used in the regression, and the mean of the dependent variable. All standard errors are clustered at the area level, i.e., state-level for Tables 3-1, 3-2, 3-11, 3-12, and county-level for Tables 3-3-3-10.

Tables 3-1 and 3-2 report state-level regression results, i.e., effects of medical schools on the physician stock within the state. The results are not consistent across physician measures â Table 3-1 shows that the number of physicians increased significantly with the number of medical schools, however, Table 3-2 shows that the effect on number of physicians per 100,000 people is not statistically different from 0. This is not too surprising, given our earlier discussion that a state is usually too large a geographical unit to use for examining regional variation.

As a result, I focus on county-level regressions. The empirical results from estimating Equation 3.1 on county-level data are readily apparent in Tables 3-3 to 3-8. Tables 3-3 and 3-4 report the estimated local effects of new medical schools on the physician stock within the county, while also controlling for the number of medical schools within the state. The first two rows report the estimated coefficient on the number of medical schools within the state and its standard error, and the next two rows report the estimated coefficient on the number of medical schools within the county and its standard error. Taking the coefficients at face value, a new medical school increases the physician supply within that county, relative to other counties, by about 500 (from preferred

specifications 2 and 4) in seven years, which is above three times the average county-level physician stock (reported in the last row).

As for the number of physicians per 100,000 people in Table 4, the coefficient is around 100 and above the average across counties. The results are reasonably robust to alternative specifications in both tables. After controlling for number of medical schools within the county, the number of medical school within the state has no impact on physician supply. This also confirms our view that state is not the appropriate geographical unit to use – it is too large to detect local effects.

Motivated by the second event study, and trying to detect broader regional effects, I report the results using number of medical schools within 50 miles[1] as the variable of interest in Tables 5 and 6, while also controlling for medical schools within 100 and 200 miles. The results are robust across specifications and similar to those in Tables 3-3 and 3-4, except that the magnitudes are smaller. A new medical school increases physician supply by around 100 in counties within 50 miles, relative to counties farther away, and this increase is two thirds of the sample average. Considering the number of physicians per 100,000 people, the increase is around 20 and about one fourth of the sample average.

Tables 3-3 to 3-6 suggest a strong positive effect on the supply of physicians. These regional estimates in Tables 3-5 and 3-6 are smaller than the local (specific to county) effects in Tables 3-3 and 3-4. This is as expected, since the mechanical effect that more facilities and more positions opened at the county where the medical school operates, did not affect other nearby areas.

I prefer the results in Tables 3-5 and 3-6 over Tables 3-3 and 3-4 for two reasons: 1) The regional impacts are more interesting compared to the more local impacts. 2) One may argue that locations of new medical schools were potentially correlated with other time-varying determinants of physician supply, but being more or less than 50 miles away from a new medical school was arguably exogenous. I will test the validity of both approaches using falsification tests below.

One might suspect that, mechanically, being within 50 miles to a new medical school suffered from the same endogeneity problem as the location of a medical school, since this sample included locations of medical schools. In an attempt to address this challenge, I rerun the regressions on the sub-sample excluding counties starting with medical schools. The results, as presented in Tables 3-7 and 3-8, are almost identical to Tables 3-5 and 3-6 â having a new medical school within 50 miles raised county-level number of physicians by 100-200, which was of the same magnitude as the sample average; it raised the number of medical schools per 100,000 people by around 20, which was around a quarter of county-level average. My interpretation of these results is that the endogeneity issue is negligible in this case.

As discussed in the introduction and more elaborately in Section 3.3, there were programs during the expansion period to attract new MDs to serve in rural areas upon graduation. Such programs were small in size but it is of policy interest to know whether they made a difference. To test this in my empirical framework, I add the interactions of rural counties with the medical school measures to the regression in Tables 3-5 and 3-6, and report the results in Tables 3-9 and

[1]The distance is defined as the distance from this county to the nearest county that has a medical school.

3-10.

The results show that the impacts in rural parts (the sum of the first two coefficients reported) were not statistically different from 0, or that the regional effects were mostly on urban areas. However, I hesitate to reach further conclusions about the effectiveness of the programs in rural parts. The results show that there were no differential impacts across different geographical units of the rural areas, which doesn't mean physician supply didn't grow in rural regions – it is likely that all rural parts enjoyed similar gains in physician supply.

The county-level regressions using Area Resource File data shows a consistent pattern, while state-level regressions show nothing. As a separate robustness check, I also run the state-level regression using IPUMS data. The results are not very different from Tables 3-1 and 3-2, but the additional controls of gender composition and age distribution make the effect on number of physicians insignificant. Reading Tables 3-1, 3-2, 3-11 and 3-12 together, it is clear that a state is too large a geographical unit to use for examining regional effects.

We are mostly interested in the regression results using county-level Area Resource File data. As a falsification test, I estimate Equation 3.2. The values of τ are taken discretely from -20 to 20, where $\tau = 0$ to 20 is used to check how the impacts of medical school expansion changed over time, while τ = -20 to -1 serve as Placebo tests. I plot the coefficients β_τ for β_τ ranging from -20 to 20 in Figures 11 to 12. I report the results on the level of physician supply only, since the regression results have been qualitatively similar for number of physicians per 100,000 people.

The time pattern of β_τ is important for two reasons. First, it can help us check the validity of our measure of medical school expansion â the number of medical schools in the future should not have any impact on the supply of physicians now if the location of medical schools was not correlated with other time-varying determinants of physician supply. Second, the pattern of β_τ helps us understand how the effect evolves over time, and help us compare the relative magnitudes of the immediate effects from lagged effects.

Figures 3-11 displays the estimated coefficients of Equation 3.2 using the number of medical schools within the county and within the state as the dependent variables. The top panel of Figure 3-11 shows the time pattern for the differential local effect of medical school on the physician supply within that county, relative to other counties, controlling for county fixed effects, year fixed effects and population. The bottom panel of Figure 3-11 plots the time pattern for the effect of a new medical school within the state, controlling for county fixed effects, year fixed effects and population. The grey areas indicate the 95 percent confidence interval for each coefficient.

Consider first the top panel of Figure 3-11, the effect of medical school within the county. The coefficient stayed stable for t = -20 to 0, i.e., the number of medical schools 20 years back had the same effect as the number of medical schools now, or equivalently speaking, a new medical school had the same impact on the physician supply within that county in the entry year and in the following 20 years. The trend sloped downward past t=0 and the coefficients became not statistically different from 0 past t=5. The stable coefficients over time suggest that the differential regional effects mostly came from the immediate responses, i.e., the more positions and more facilities brought in

by the medical school and more physicians attracted by such amenities and locate nearby. The zero coefficients for t >= 5 supports the hypothesis that locations of new medical schools were not correlated with other underlying determinants of physician supply. This is reassuring and supports the validity of equation 3.1 as our estimating equation and also a valid first stage when one examines the impact of physician supply on medical spending and health outcomes.

The positive yet declining coefficients between t=0 and t=5 are a bit puzzling. I hesitate to make further interpretations and instead provide my speculation as follows. Since I am using the year of establishment, not the year of announcement of an upcoming medical school, one may suspect some anticipation effect to take place, i.e., physicians might adjust their location choices in anticipation of new medical schools being built in the near future. Unfortunately, I do not have year of announcement in the data so this is only a speculation. Figure 3-11 shows that the physician supply responded partially a bit earlier than the year of establishment, and it only came into full blown after the year establishment.

The bottom panel of Figure 3-11 is consistent with our estimation results from Tables 3-3 and 3-4 – after controlling for the number of medical schools within the county, the number of medical schools within the state had no effect on county-level physician supply. The estimated coefficient with controls was not statistically distinguishable from zero and stayed flat for all years. Figure 3-12 demonstrates this falsification exercise using the alternative geographical unit â counties within 50 miles to a new medical school. The estimates are smaller in magnitudes than estimates in the top panel of Figure 3-11, but the patterns over time are reassuringly similar. The coefficients stayed stable throughout t = -20 to t = -1, declined gradually from t = 0 forward and became not separable from 0 around t = 8. Equivalently speaking, a new medical school had the same impact on the physician supply within that county in the entry year and in the following 20 years, had some effect a few years before the actual entry and had no effect in earlier periods. Again I speculate the positive yet small coefficients between t=0 and t=10 to be the anticipation effects following the announcement of the upcoming medical school, but I do not have the data to test this.

Time patterns in Figures 3-11 and 3-12 have lead to some conclusions, or at least partially founded speculations, about the mechanisms underlying the effects. First, the stable coefficients until t = 0 shows that a new medical school had the same impact in the year of entry and in the next 20 years. This indicates that the most of the impacts could be attributed to the immediate effects, and the lagged effects did not play an important role. The new job opening and new facilities brought in by the new medical schools, and the attractiveness of working closer to a teaching and research facilities and peers were the main reasons why physician supply grew faster in areas closer to the newly established medical school, rather than the correlated location choices of education and professional practice by students at the new medical schools. Second, the zero coefficients for t large enough supports the hypothesis that locations of new medical schools were not correlated with other underlying determinants of physician supply and supports the validity of equation (1) as our estimating equation. Third, the subdued yet still positive coefficients slightly past t = 0 indicate some anticipation effects, that physicians began to adjust their location choices before the actually

establishment of new medical schools. Whether and why this was the case needs testing with additional data, such as data on when the establishment of new medical schools was announced.

As mentioned in Section 3.5, because of the low frequency of physician data for earlier years (1940-1980), we have a relatively small number of observations compared with the number of control variables. However, the results change little with and without the controls, using more or less years of the data, and using alternative data sources. This is reassuring and confirms that the results do not come from arbitrary specifications or data mining. Another concern is the potential endogeneity in the locations of new medical schools, I test for the hypothesis that locations of medical schools are not correlated with other time-varying factors in Section 3.6.4 above, and most of the results do not reject this hypothesis.

3.7 Conclusion

Adopting a difference-in-difference style regression framework, I examine the differential impacts of medical school expansion on the supply of physicians across geographical areas by comparing changes in physician supply in areas closer to new medical school entries with regions further away. I find that the entry of a new medical school increased the supply of physicians substantially more in nearby areas. It increased the physician supply by one to three times (depending on which measure of physician supply to use) the county average level in the county where the medical school was located, compared with other counties. The broader regional effect was smaller but still substantial – a new medical school increased physician supply by one fourth to two thirds of the sample average in counties within 50 miles, relative to counties farther away. The slightly subdued estimates still suggest a strong positive regional effect on the supply of physicians.

I also investigate how such regional impacts of a new medical school entry evolved over time. I find that a new medical school had the same impact in the year of entry and in the following 20 years, which indicates that the most of the impacts took place immediately. In other words, the new job opening and new facilities brought in by the new medical schools, and physicians attracted by research facilities and peers nearby were the main reasons why physician supply grew faster in areas closer to the newly established medical school. I find no effect on the physician supply in most of the pre-entry years, which suggests that the locations of new medical schools are not correlated with other underlying determinants of physician supply. Therefore, the expansion of medical schools could potentially be used as an exogeneous source of variation for physician supply.

This study is devoted to understanding the differential effects of medical schools across geographical units on the supply of physicians, and caution should be taken in interpreting the results as the overall effects of medical school expansion. It is plausible that all areas enjoyed growth in physician supply after the expansion, though the magnitude of the effect differed across areas. Therefore, my estimates could understate the full impact of medical school expansion on the supply of physicians, which, however, does not affect the effectiveness and relevance of this exercise in providing a valid first stage for the cross-sectional variation of physicians. With historical data on

county-level health expenditures and other second-stage outcomes, we can further investigate the impact of physician supply on these outcomes.

Figure 3-1: Distribution of Medical Schools by county, 1940

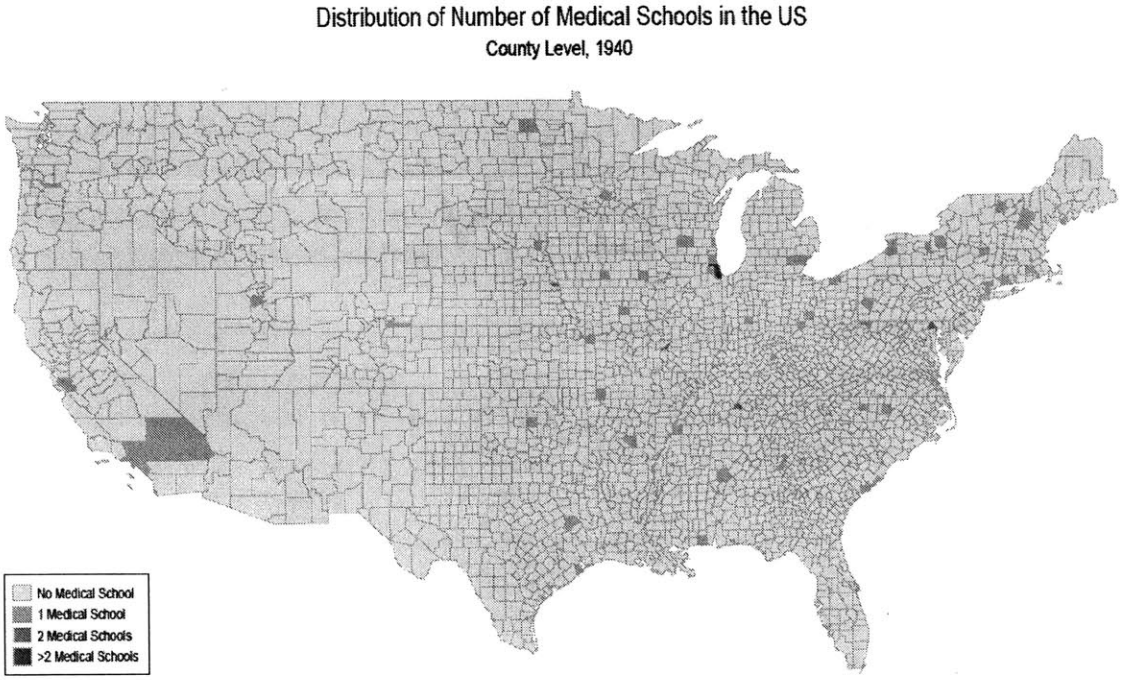


Figure 2: Distribution of Medical Schools by county, 1960
Distribution of Number of Medical Schools in the US
County Level, 1960

Figure 3-2: Distribution of Medical Schools by county, 1960

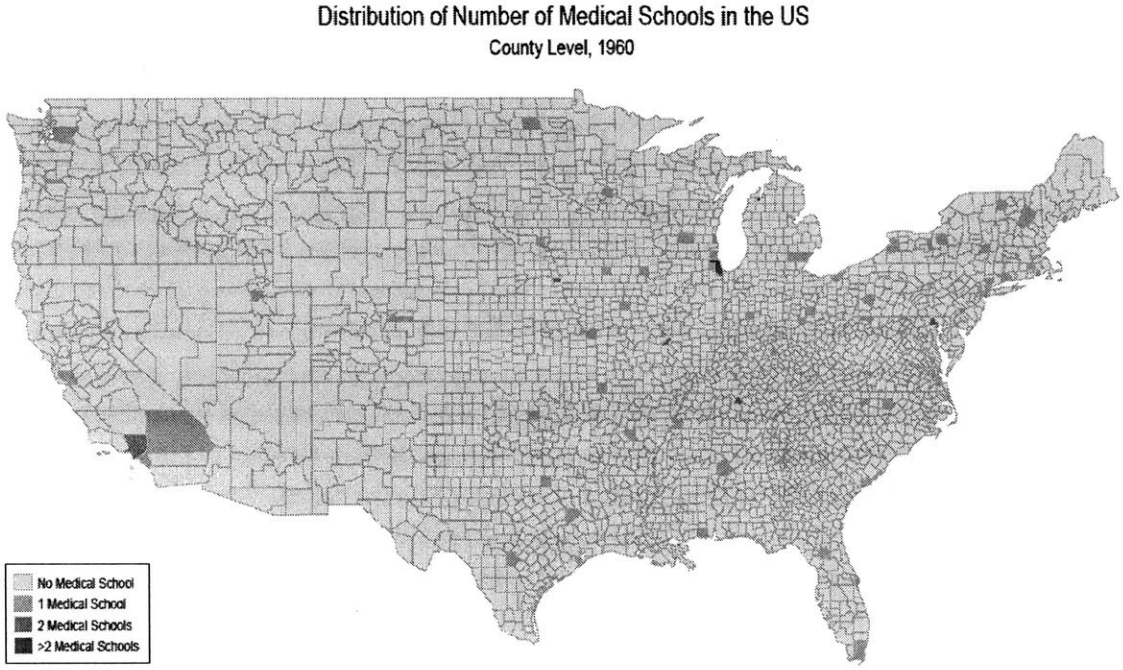


Figure 3-3: Distribution of Medical Schools by county, 1980

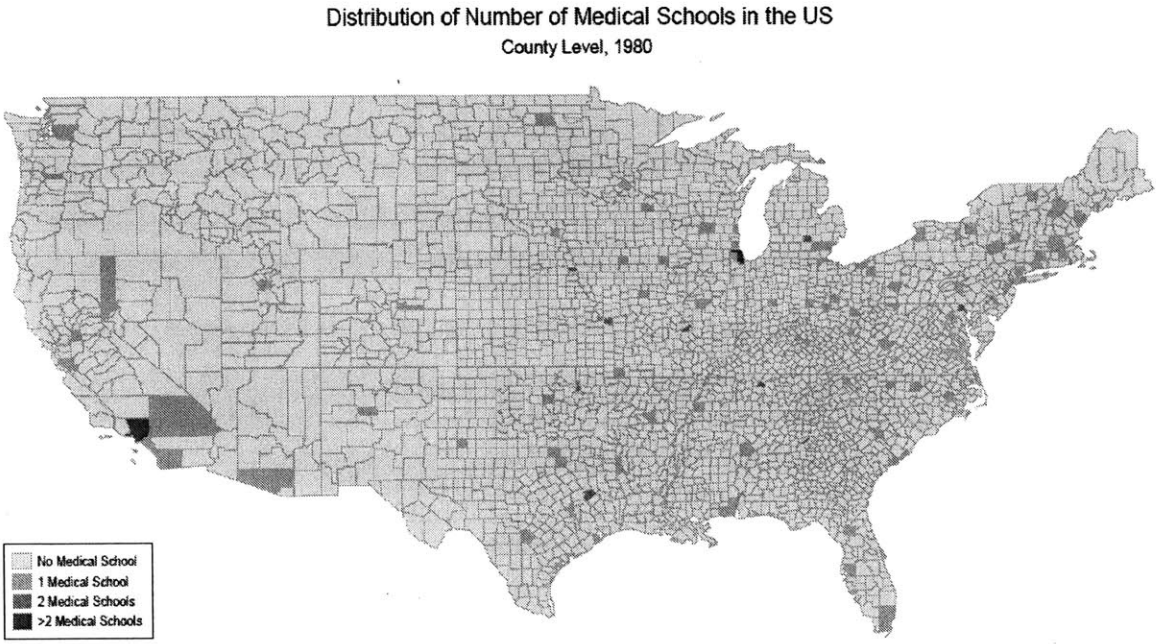
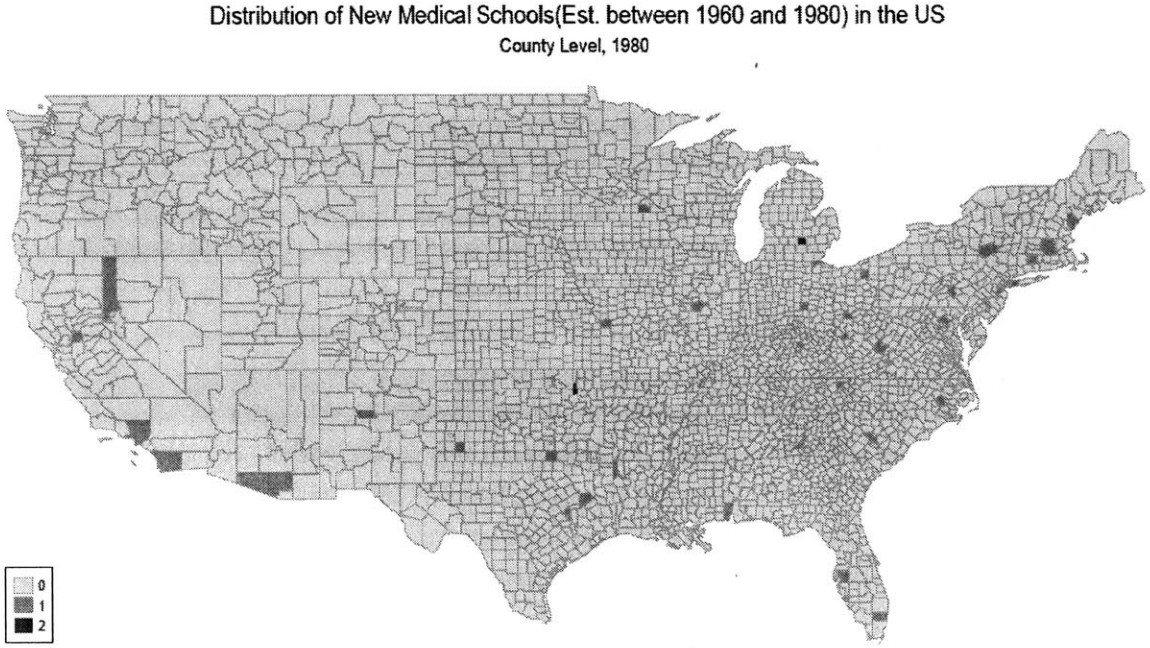
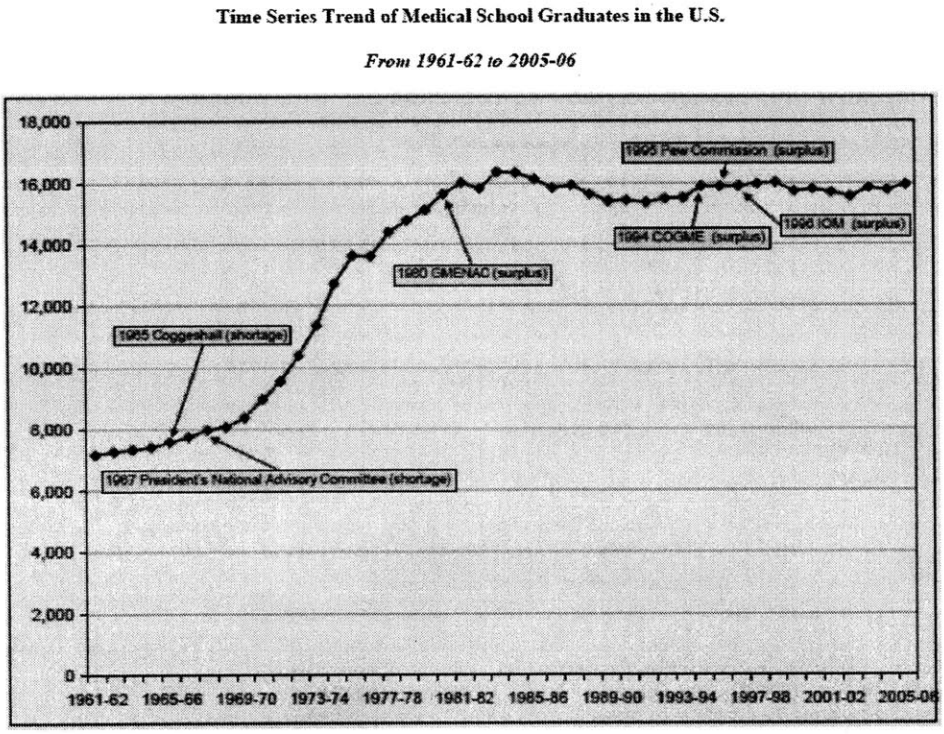


Figure 3-4: Distribution of new medical schools by counties, established between 1960 and 1980



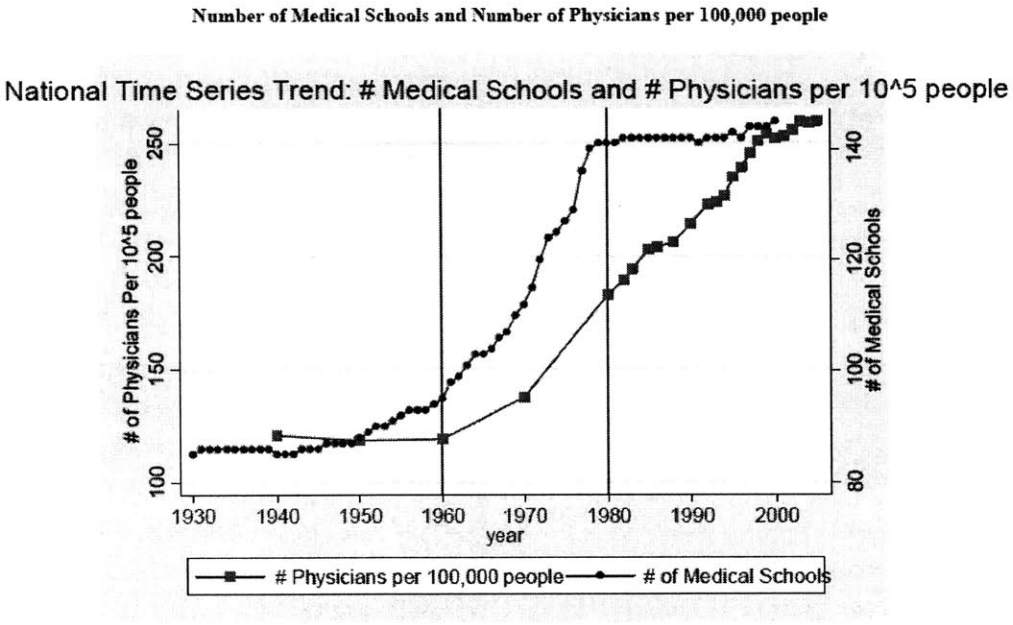
Note: Figure 4 illustrates the geographical distribution of new medical schools established between 1960 and 1980. We can see that they are more sparsely located in the west and more spread out in the east.

Figure 3-5: Time Series Trend of Medical School Graduates in the U.S.: From 1961-62 to 2005-06



Note: This graph is Figure 3 from Dill and Salsberg (2008), using data from the Association of American Medical Colleges (AAMC). It shows that national number of medical graduates grew slowly before 1970 and increased drastically from 1969 to 1978, grew slightly slower from 1978 to 1985 and stabilized thereafter.

Figure 3-6: Number of Medical Schools and Number of Physicians per 100,000 people



Note: Figure 6 shows the national time series patterns for both the total number of medical schools and the number of physicians per 100000 people (from the ARF data). There seems to be a lag in the growth of physician supply in response to the expansion of medical schools. The number of medical schools increased drastically from 1960 to 1980 and stabilized after, while the number of physicians increased slightly 1960-1970, more abruptly after 1970 and stabilized by 2000.

Figure 3-7: Event Study: States with First V.S. No Medical School Entry

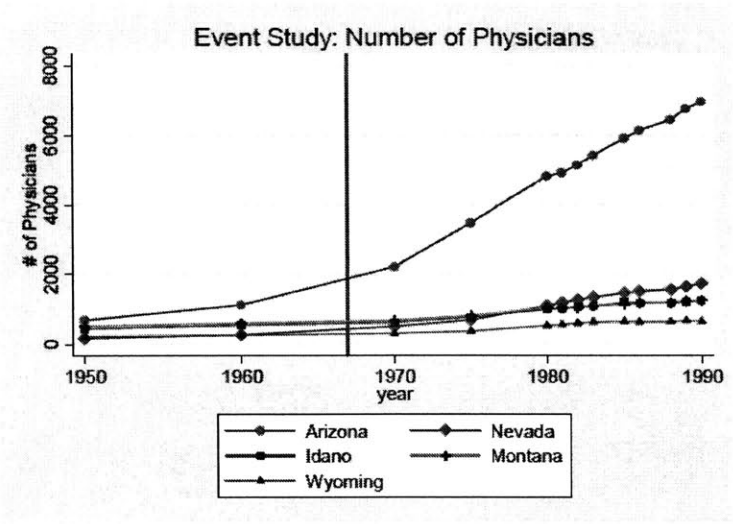


Figure 3-8: Event Study: States with First V.S. No Medical School Entry

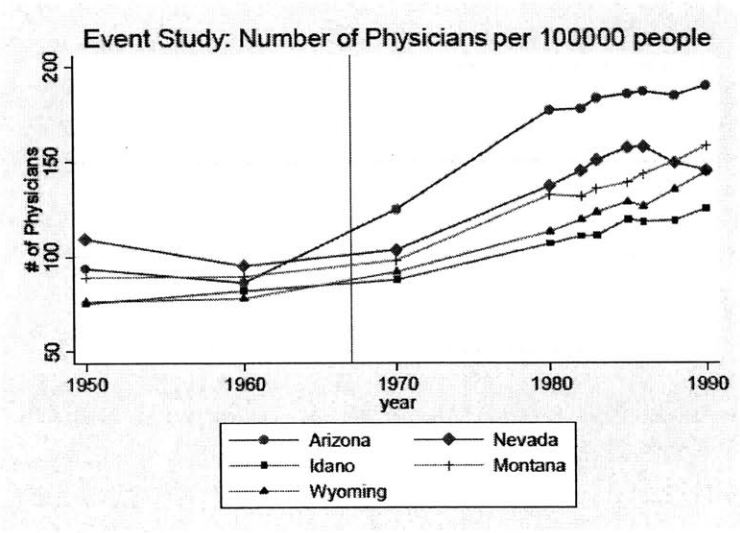


Figure 3-9: Event Study: Counties Near V.S. Far From Medical School Entry

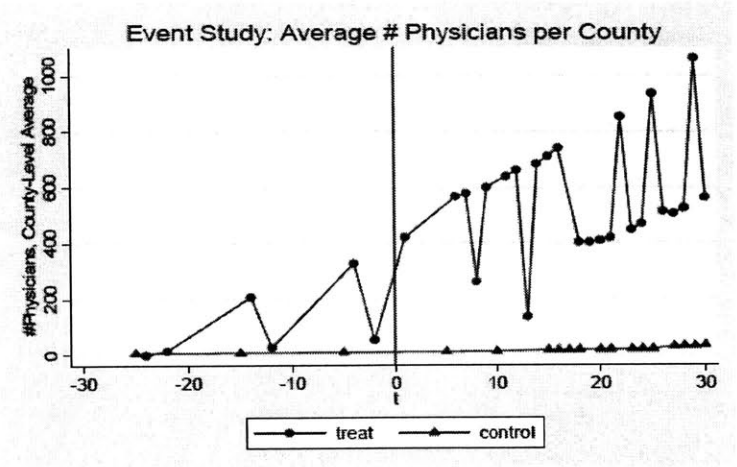


Figure 3-10: Event Study: Counties Near V.S. Far From Medical School Entry

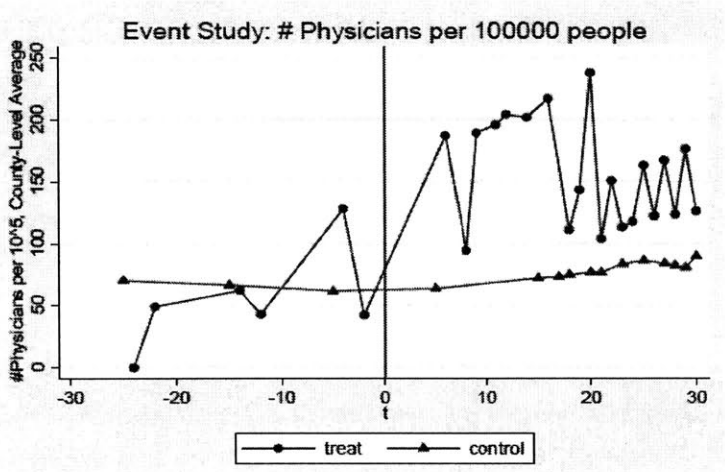
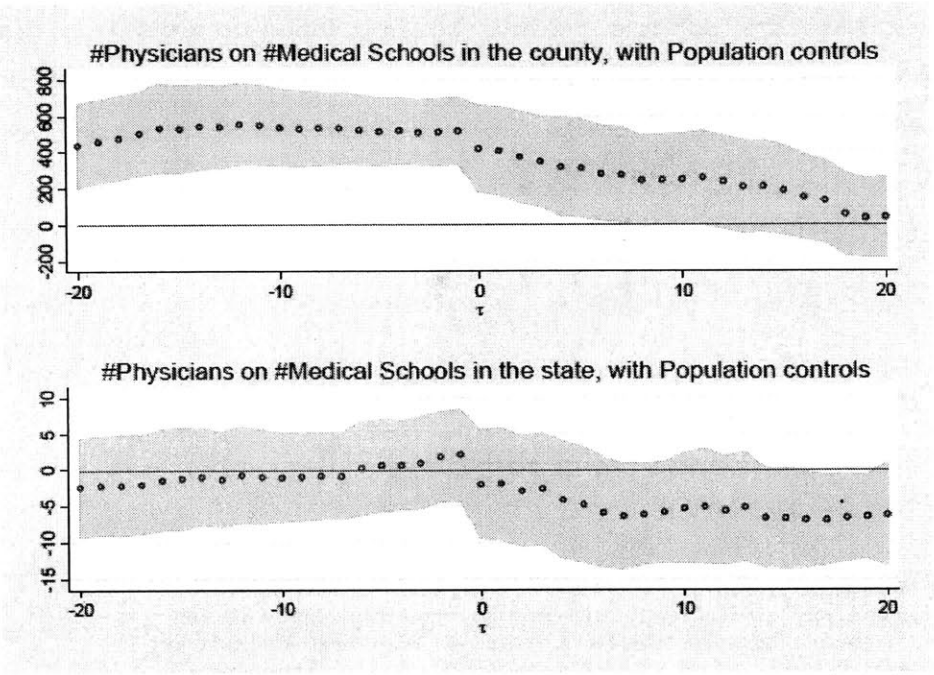


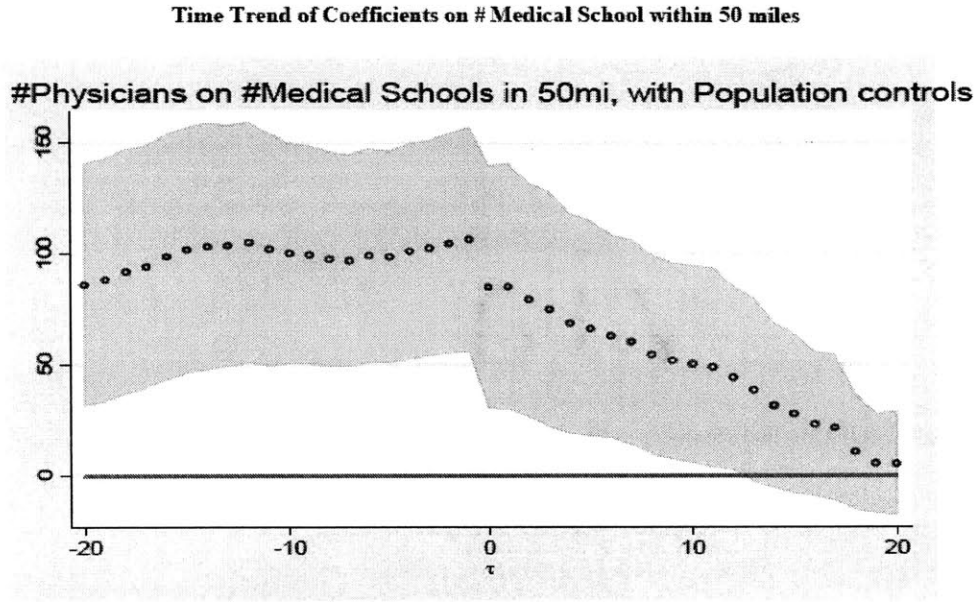
Figure 3-11: Time Trend of Coefficients on Medical School within the County and on Medical School within the State

Time Trend of Coefficients on # Medical School within the County and on # Medical School within the State



Note: This figure plots β_τ the coefficients of county-level # of physicians in year t regressed on both # of medical schools within the county in year (t-τ) and # of medical schools within the state in year (t-τ), where τ takes values from -20 to 20. The grey areas indicate the 95 percent confidence interval for each coefficient, where the standard errors are clustered at the county level. For $\tau > 0$, we are checking the regional effects of a future medical school on current physician supply, which serves as a falsification test. For $\tau < 0$, we are tracking the effect of a new medical school over time.

Figure 3-12: Time Trend of Coefficients on Medical School within 50 miles



Note: This figure plots β_τ the coefficients of county-level # of physicians in year t regressed on # of medical schools within 50 miles in year (t-τ), where τ takes values from -20 to 20. The grey areas indicate the 95 percent confidence interval for each coefficient, where the standard errors are clustered at the county level. For $\tau > 0$, we are checking the regional effects of a future medical school on current physician supply, which serves as a falsification test. For $\tau < 0$, we are tracking the effect of a new medical school over time.

Table 3.1: Number of physicians on number of medical schools by state

VARIABLES	(1) 1940-2000 State&Year FE	(2) 1940-2000 +population	(3) 1960-1990 State&Year FE	(4) 1960-1990 +population
# Medical Schools 7 years ago within the state	5,176 (1,366)	3,832 (1,041)	3,147 (833.6)	3,037 (789.1)
F	14.35	13.55	14.25	14.81
N	3966	3966	1702	1702
Mean: # physicians within the state	13590	13590	11949	11949

Notes: Standard Errors clustering at the state-level are reported in the parentheses. F-statistics for the test of the reported coefficient = 0 is reported in the third last row. Number of observations used in each regression is reported in the second last row. Sample average of the dependent variable is reported in the last row.

Column (1) uses 1940-2000 ARF data and includes state and year FE in the regression. Column (2) uses 1940-2000 ARF data and controls for time-varying population in addition to state and year FE in the regression. Columns (3) and (4) replicate Columns (1) and (2) using 1960-1990 data to focus on the expansion period.

Notes: Standard Errors clustering at the state-level are reported in the parentheses. F-statistics for the test of the reported coefficient = 0 is reported in the third last row. Number of observations used in each regression is reported in the second last row. Sample average of the dependent variable is reported in the last row. Column (1) uses 1940-2000 ARF data and includes state and year FE in the regression. Column (2) uses 1940-2000 ARF data and controls for time-varying population in addition to state and year FE in the regression. Columns (3) and (4) replicate Columns (1) and (2) using 1960-1990 data to focus on the expansion period.

Table 3.2: Number of physicians per 10^5 people on number of medical schools by state

VARIABLES	(1) 1940-2000 State&Year FE	(2) 1960-1990 State&Year FE
# Medical Schools 7 years ago within the state	0.271	0.711
	(3.190)	(2.114)
F	0.00721	0.113
N	3527	1265
Mean: # physicians per 100, 000 individuals within the state	200.8	173.8

Notes: Standard Errors clustering at the state-level are reported in the parentheses. F-statistics for the test of the reported coefficient = 0 is reported in the third last row. Number of observations used in each regression is reported in the second last row. Sample average of the dependent variable is reported in the last row.

Column (1) uses 1940-2000 ARF data and includes state and year FE in the regression. Column (2) replicates Column (1) using 1960-1990 ARF data to focus on the expansion period.

Notes: Standard Errors clustering at the state-level are reported in the parentheses. F-statistics for the test of the reported coefficient = 0 is reported in the third last row. Number of observations used in each regression is reported in the second last row. Sample average of the dependent variable is reported in the last row. Column (1) uses 1940-2000 ARF data and includes state and year FE in the regression. Column (2) replicates Column (1) using 1960-1990 ARF data to focus on the expansion period.

Table 3.3: Number of physicians on number of medical schools by state and county

VARIABLES	(1) 1940-2000 county&year FE	(2) 1940-2000 +population	(3) 1960-1990 county&year FE	(4) 1960-1990 +population
# Medical Schools 7 years ago within the state	21.97	-0.964	14.46	2.316
	(5.464)	(3.374)	(4.133)	(3.112)
# Medical Schools 7 years ago within the county	1,208	541.8	676.1	444.1
	(362.0)	(119.7)	(185.1)	(96.98)
F	11.64	10.77	10.19	10.49
N	72243	62820	37692	28269
Mean: # physicians within the county	154.3	154.3	134.9	134.9

Notes: Standard Errors clustering at the county level are reported in the parentheses. F-statistics for the joint test of the two reported coefficients = 0 is reported in the third last row. Number of observations used in each regression is reported in the second last row. Sample average of the dependent variable is reported in the last row.

Column (1) uses 1940-2000 ARF data and includes county and year FE in the regression. Column (2) uses 1940-2000 ARF data and controls for time-varying population in addition to county and year FE in the regression. Columns (3) and (4) replicate Columns (1) and (2) using 1960-1990 data to focus on the expansion period.

Notes: Standard Errors clustering at the county level are reported in the parentheses. F-statistics for the joint test of the two reported coefficients = 0 is reported in the third last row. Number of observations used in each regression is reported in the second last row. Sample average of the dependent variable is reported in the last row. Column (1) uses 1940-2000 ARF data and includes county and year FE in the regression. Column (2) uses 1940-2000 ARF data and controls for time-varying population in addition to county and year FE in the regression. Columns (3) and (4) replicate Columns (1) and (2) using 1960-1990 data to focus on the expansion period.

Table 3.4: Number of physicians per 10^5 people on number of medical schools by state and county

VARIABLES	(1) 1940-2000 county and year FE	(2) 1960-1990 county and year FE
# Medical Schools 7 years ago within the state	-0.610	-0.376
	(0.668)	(0.599)
# Medical Schools 7 years ago within the county	134.7	93.92
	(17.62)	(13.21)
F	29.32	25.70
N	62094	27699
Mean: # physicians per 100, 000 individuals within the county	95.63	84.96

Notes: Standard Errors clustering at the county level are reported in the parentheses. F-statistics for the joint test of the two reported coefficients = 0 is reported in the third last row. Number of observations used in each regression is reported in the second last row. Sample average of the dependent variable is reported in the last row.

Column (1) uses 1940-2000 ARF data and includes state and year FE in the regression. Column (2) replicates Column (1) using 1960-1990 ARF data to focus on the expansion period.

Notes: Standard Errors clustering at the county level are reported in the parentheses. F-statistics for the joint test of the two reported coefficients = 0 is reported in the third last row. Number of observations used in each regression is reported in the second last row. Sample average of the dependent variable is reported in the last row. Column (1) uses 1940-2000 ARF data and includes state and year FE in the regression. Column (2) replicates Column (1) using 1960-1990 ARF data to focus on the expansion period.

Table 3.5: Number of physicians on number of nearby medical schools

VARIABLES	(1) 1940-2000 county&year FE	(2) 1940-2000 +population	(3) 1960-1990 county&year FE	(4) 1960-1990 +population
# Medical Schools 7 years ago within 50 miles	191.3	109.9	109.4	81.48
	(41.66)	(33.93)	(22.45)	(22.97)
# Medical Schools 7 years ago within 100 miles	43.73	18.20	23.19	11.24
	(12.48)	(12.74)	(8.107)	(9.838)
# Medical Schools 7 years ago within 200 miles	-3.291	5.125	-0.160	5.517
	(4.365)	(3.346)	(2.795)	(2.720)
F	20.43	9.692	19.27	10.72
N	72243	62820	37692	28269
Mean: # physicians within the county	154.3	154.3	134.9	134.9

Notes: Standard Errors clustering at the county level are reported in the parentheses. F-statistics for the joint test of the three reported coefficients = 0 is reported in the third last row. Number of observations used in each regression is reported in the second last row. Sample average of the dependent variable is reported in the last row. Column (1) uses 1940-2000 ARF data and includes county and year FE in the regression. Column (2) uses 1940-2000 ARF data and controls for time-varying population in addition to county and year FE in the regression. Columns (3) and (4) replicate Columns (1) and (2) using 1960-1990 data to focus on the expansion period.

Notes: Standard Errors clustering at the county level are reported in the parentheses. F-statistics for the joint test of the three reported coefficients = 0 is reported in the third last row. Number of observations used in each regression is reported in the second last row. Sample average of the dependent variable is reported in the last row. Column (1) uses 1940-2000 ARF data and includes county and year FE in the regression. Column (2) uses 1940-2000 ARF data and controls for time-varying population in addition to county and year FE in the regression. Columns (3) and (4) replicate Columns (1) and (2) using 1960-1990 data to focus on the expansion period.

Table 3.6: Number of physicians per 10^5 people on number of nearby medical schools

VARIABLES	(1) 1940-2000 county and year FE	(2) 1960-1990 county and year FE
# Medical Schools 7 years ago within 50 miles	22.49	21.01
	(4.438)	(7.346)
# Medical Schools 7 years ago within 100 miles	-2.255	-8.053
	(3.832)	(7.606)
# Medical Schools 7 years ago within 200 miles	4.573	5.384
	(1.568)	(2.681)
F	32.45	24.12
N	62094	27699
Mean: # physicians per 100,000 individuals within the county	95.63	84.96

Notes: Standard Errors clustering at the county level are reported in the parentheses. F-statistics for the joint test of the three reported coefficients = 0 is reported in the third last row. Number of observations used in each regression is reported in the second last row. Sample average of the dependent variable is reported in the last row. Column (1) uses 1940-2000 ARF data and includes state and year FE in the regression. Column (2) replicates Column (1) using 1960-1990 ARF data.

Notes: Standard Errors clustering at the county level are reported in the parentheses. F-statistics for the joint test of the three reported coefficients = 0 is reported in the third last row. Number of observations used in each regression is reported in the second last row. Sample average of the dependent variable is reported in the last row. Column (1) uses 1940-2000 ARF data and includes state and year FE in the regression. Column (2) replicates Column (1) using 1960-1990 ARF data.

Table 3.7: Number of physicians on number of nearby medical schools, excluding counties already with a medical school in 1940

VARIABLES	(1) 1940-2000 county&year FE	(2) 1940-2000 +population	(3) 1960-1990 county&year FE	(4) 1960-1990 +population
# Medical Schools 7 years ago	150.8	83.54	88.04	64.92
within 50 miles	(30.27)	(26.77)	(17.93)	(20.88)
# Medical Schools 7 years ago	40.57	16.30	22.09	11.26
within 100 miles	(10.18)	(8.094)	(6.232)	(5.853)
# Medical Schools 7 years ago	-4.117	0.715	-1.295	1.667
within 200 miles	(2.931)	(2.298)	(1.796)	(1.616)
F	16.30	6.567	14.65	6.969
N	70748	61520	36912	27684
Mean: # physicians within the county	104.4	104.4	89.83	89.83

Notes: Standard Errors clustering at the county level are reported in the parentheses. F-statistics for the joint test of the three reported coefficients = 0 is reported in the third last row. Number of observations used in each regression is reported in the second last row. Sample average of the dependent variable is reported in the last row. Column (1) uses 1940-2000 ARF data and includes county and year FE in the regression. Column (2) uses 1940-2000 ARF data and controls for time-varying population in addition to county and year FE in the regression. Columns (3) and (4) replicate Columns (1) and (2) using 1960-1990 data to focus on the expansion period.

Table 3.8: Number of physicians per 10^5 people on number of nearby medical schools, excluding counties already with a medical school in 1940

VARIABLES	(1) 1940-2000 county and year FE	(2) 1960-1990 county and year FE
# Medical Schools 7 years ago within 50 miles	21.12	19.99
	(4.387)	(7.513)
# Medical Schools 7 years ago within 100 miles	-1.064	-7.389
	(3.802)	(7.844)
# Medical Schools 7 years ago within 200 miles	3.457	4.738
	(1.523)	(2.730)
F	29.01	21.43
N	60805	27123
Mean: # physicians per 10^5 people within the county	89.69	79.90

Notes: Standard Errors clustering at the county level are reported in the parentheses. F-statistics for the joint test of the three reported coefficients = 0 is reported in the third last row. Number of observations used in each regression is reported in the second last row. Sample average of the dependent variable is reported in the last row. Column (1) uses 1940-2000 ARF data and includes state and year FE in the regression. Column (2) replicates Column (1) using 1960-1990 ARF data.

Notes: Standard Errors clustering at the county level are reported in the parentheses. F-statistics for the joint test of the three reported coefficients = 0 is reported in the third last row. Number of observations used in each regression is reported in the second last row. Sample average of the dependent variable is reported in the last row. Column (1) uses 1940-2000 ARF data and includes state and year FE in the regression. Column (2) replicates Column (1) using 1960-1990 ARF data.

Table 3.9: Number of physicians on number of nearby medical schools in rural vs. urban areas

VARIABLES	(1) 1940-2000 county&year FE	(2) 1940-2000 +population	(3) 1960-1990 county&year FE	(4) 1960-1990 +population
# Medical Schools 7 years ago	159.5	90.84	92.99	70.22
within 50 miles	(33.79)	(29.83)	(20.11)	(23.26)
Rural * # Medical Schools 7 years	-163.2	-90.31	-93.87	-69.19
ago within 50 miles	(33.89)	(29.87)	(20.18)	(23.30)
# Medical Schools 7 years ago	49.04	19.79	27.40	14.09
within 100 miles	(12.54)	(9.961)	(7.790)	(7.290)
Rural * # Medical Schools 7 years	-47.50	-21.19	-26.27	-15.25
ago within 100 miles	(12.64)	(10.11)	(7.850)	(7.352)
# Medical Schools 7 years ago	-2.666	0.483	0.233	1.918
within 200 miles	(3.392)	(2.756)	(2.087)	(1.952)
Rural * # Medical Schools 7 years	-7.902	1.044	-8.083	-1.851
ago within 200 miles	(3.256)	(2.920)	(1.960)	(1.977)
F	30.33	4.576	26.69	5.694
N	70748	61520	36912	27684
Mean: #physicians within the county	104.4	104.4	89.83	89.83

Notes: Standard Errors clustering at the county level are reported in the parentheses. F-statistics for the joint test of the six reported coefficients = 0 is reported in the third last row. Number of observations used in each regression is reported in the second last row. Sample average of the dependent variable is reported in the last row. Column (1) uses 1940-2000 ARF data and includes county and year FE in the regression. Column (2) uses 1940-2000 ARF data and controls for time-varying population in addition to county and year FE in the regression. Columns (3) and (4) replicate Columns (1) and (2) using 1960-1990 data to focus on the expansion period.

Notes: Standard Errors clustering at the county level are reported in the parentheses. F-statistics for the joint test of the six reported coefficients = 0 is reported in the third last row. Number of observations used in each regression is reported in the second last row. Sample average of the dependent variable is reported in the last row. Column (1) uses 1940-2000 ARF data and includes county and year FE in the regression. Column (2) uses 1940-2000 ARF data and controls for time-varying population in addition to county and year FE in the regression. Columns (3) and (4) replicate Columns (1) and (2) using 1960-1990 data to focus on the expansion period.

Table 3.10: Number of physicians per 10^5 people on number of nearby medical schools in rural vs. urban areas

VARIABLES	(1) 1940-2000 county and year FE	(2) 1960-1990 county and year FE
# Medical Schools 7 years ago	21.47	21.58
within 50 miles	(5.195)	(9.515)
Rural * # Medical Schools 7 years ago	-21.84	-20.69
within 50 miles	(6.276)	(9.504)
# Medical Schools 7 years ago	-1.328	-9.051
within 100 miles	(4.691)	(9.873)
Rural * # Medical Schools 7 years ago	0.710	7.691
within 100 miles	(4.910)	(9.792)
# Medical Schools 7 years ago	4.819	6.093
within 200 miles	(1.717)	(3.039)
Rural * # Medical Schools 7 years ago	-7.476	-7.045
within 200 miles	(1.392)	(1.673)
F	25.21	18.97
N	60805	27123
Mean: # physicians per 10^5 people within the county	89.69	79.90

Notes: Standard Errors clustering at the county level are reported in the parentheses. F-statistics for the joint test of the six reported coefficients = 0 is reported in the third last row. Number of observations used in each regression is reported in the second last row. Sample average of the dependent variable is reported in the last row. Column (1) uses 1940-2000 ARF data and includes state and year FE in the regression. Column (2) replicates Column (1) using 1960-1990 ARF data.

Notes: Standard Errors clustering at the county level are reported in the parentheses. F-statistics for the joint test of the six reported coefficients = 0 is reported in the third last row. Number of observations used in each regression is reported in the second last row. Sample average of the dependent variable is reported in the last row. Column (1) uses 1940-2000 ARF data and includes state and year FE in the regression. Column (2) replicates Column (1) using 1960-1990 ARF data.

Table 3.11: Robustness check: number of physicians on number of medical schools by state

VARIABLES	(1) 1940-2000 State&year FE	(2) 1940-2000 +populatio n	(3) 1940-2000 +sex & age distribution	(4) 1960-1990 State&year FE	(5) 1960-1990 +populatio n	(6) 1960-1990 +sex & age distribution
# Medical Schools 7 years ago within the state	6,359 (1,604)	2,110 (1,040)	371.5 (390.4)	4,452 (1,466)	1,856 (1,178)	-213.9 (353.8)
F	15.71	4.111	0.905	9.225	2.483	0.366
N	317	317	317	181	181	181
Mean: # physicians within the state	8499	8499	8499	8563	8563	8563

Notes: Standard Errors clustering at the state-level are reported in the parentheses. F-statistics for the test of the reported coefficient = 0 is reported in the third last row. Number of observations used in each regression is reported in the second last row. Sample average of the dependent variable is reported in the last row.

Column (1) uses 1940-2000 IPUMS data and includes state and year FE in the regression. Column (2) uses 1940-2000 IPUMS data and controls for time-varying population in addition to state and year FE in the regression. Column (3) uses 1940-2000 IPUMS data and controls for time-varying population, sex composition and age distribution in addition to state and year FE in the regression. Columns (4)(5)(6) replicate Columns (1) (2)(3) using 1960-1990 IUMS data to focus on the expansion period.

Notes: Standard Errors clustering at the state-level are reported in the parentheses. F-statistics for the test of the reported coefficient = 0 is reported in the third last row. Number of observations used in each regression is reported in the second last row. Sample average of the dependent variable is reported in the last row. Column (1) uses 1940-2000 IPUMS data and includes state and year FE in the regression. Column (2) uses 1940-2000 IPUMS data and controls for time-varying population in addition to state and year FE in the regression. Column (3) uses 1940-2000 IPUMS data and controls for time-varying population, sex composition and age distribution in addition to state and year FE in the regression. Columns (4)(5)(6) replicate Columns (1) (2)(3) using 1960-1990 IUMS data to focus on the expansion period.

Table 3.12: Robustness check: number of physicians per 10^5 people on number of medical schools by state

VARIABLES	(1) 1940-2000 State&year FE	(2) 1940-2000 +sex & age distribution	(3) 1960-1990 State&year FE	(4) 1960-1990 +sex & age distribution
# Medical Schools 7 years ago within the state	7.129	8.734	4.387	4.732
	(4.180)	(4.805)	(5.600)	(5.940)
F	2.909	3.304	0.614	0.635
N	317	317	181	181
Mean: # physicians per 100, 000 individuals within the state	172.0	172.0	173.5	173.5

Notes: Standard Errors clustering at the state-level are reported in the parentheses. F-statistics for the test of the reported coefficient = 0 is reported in the third last row. Number of observations used in each regression is reported in the second last row. Sample average of the dependent variable is reported in the last row.

Column (1) uses 1940-2000 IPUMS data and includes state and year FE in the regression. Column (2) uses 1940-2000 IPUMS data and controls for time-varying sex composition and age distribution in addition to state and year FE in the regression. Columns (3) and (4) replicate Columns (1) and (2) using 1960-1990 IUMS data to focus on the expansion period.

Notes: Standard Errors clustering at the state-level are reported in the parentheses. F-statistics for the test of the reported coefficient = 0 is reported in the third last row. Number of observations used in each regression is reported in the second last row. Sample average of the dependent variable is reported in the last row. Column (1) uses 1940-2000 IPUMS data and includes state and year FE in the regression. Column (2) uses 1940-2000 IPUMS data and controls for time-varying sex composition and age distribution in addition to state and year FE in the regression. Columns (3) and (4) replicate Columns (1) and (2) using 1960-1990 IUMS data to focus on the expansion period.

Bibliography

Abaluck, J., Gruber, J., June 2011. Choice inconsistencies among the elderly: Evidence from plan choice in the medicare part D program. The American Economic Review 101, 1180–1210.

Abaluck, J., Gruber, J., June 2013. Evolving choice inconsistencies in choice of prescription drug insurance. Working Paper 19163, National Bureau of Economic Research.

Acemoglu, D., 2002. Directed technical change. The Review of Economic Studies 69 (4), 781–809.

Acemoglu, D., Cutler, D., Finkelstein, A., Linn, J., 2006. Did medicare induce pharmaceutical innovation? The American economic review 96 (2), 103–107.

Acemoglu, D., Linn, J., 2004. Market size in innovation: Theory and evidence from the pharmaceutical industry. The Quarterly Journal of Economics, 1049–1090.

Aghion, P., Howitt, P., 1992. A model of growth through creative destruction. Econometrica 60, 323–351.

Agrawal, A., Cockburn, I., Zhang, L., 2014. Deals not done: Sources of failure in the market for ideas. Strategic Management Journal.

Akcigit, U., Celik, M. A., Greenwood, J., 2013. Buy, keep or sell: Economic growth and the market for ideas. Tech. rep., National Bureau of Economic Research.

Arie, G., Grieco, P. L., 2014. Who pays for switching costs? Quantitative Marketing and Economics 12 (4), 379–419.

Arora, A., Fosfuri, A., 2003. Licensing the market for technology. Journal of Economic Behavior & Organization 52 (2), 277–295.

Arora, A., Fosfuri, A., Gambardella, A., 2001. Markets for technology and their implications for corporate strategy. Industrial and corporate change 10 (2), 419–451.

Arora, A., Gambardella, A., 2010. Ideas for rent: an overview of markets for technology. Industrial and corporate change 19 (3), 775–803.

Arora, A., Gambardella, A., Magazzini, L., Pammolli, F., 2009. A breath of fresh air? firm type, scale, scope, and selection effects in drug development. Management Science 55 (10), 1638–1653.

Arrow, K., 1962. Economic welfare and the allocation of resources for invention. In: The rate and direction of inventive activity: Economic and social factors. Princeton University Press, pp. 609–626.

Bajari, P., Benkard, C. L., Levin, J., 2007. Estimating dynamic models of imperfect competition. Econometrica 75 (5), 1331–1370.

Bajari, P., Hong, H., Ryan, S. P., 2010. Identification and estimation of discrete game of complete information. Econometrica 78 (5), 1529–1568.

Beggs, A., Klemperer, P., 1992. Multi-period competition with switching costs. Econometrica 60 (3), 651–666.

Benkard, C. L., 2004. A dynamic analysis of the market for widebodied commercial aircraft. Review of Economic Studies 71, 581–611.

Bessen, J., Meurer, M. J., 2008. Patent failure: How judges, bureaucrats, and lawyers put innovators at risk. Princeton University Press.

Blume-Kohout, M. E., Sood, N., 2013. Market size and innovation: Effects of medicare part d on pharmaceutical research and development. Journal of public economics 97, 327–336.

Boards of Trustees, 2012. 2012 annual report of the boards of trustees of the federal hospital insurance and federal supplementary insurance trust funds.

Bosse, D. A., Alvarez, S. A., 2010. Bargaining power in alliance governance negotiations: evidence from the biotechnology industry. Technovation 30 (5), 367–375.

Cassiman, B., Veugelers, R., 2006. In search of complementarity in innovation strategy: Internal r&d and external knowledge acquisition. Management science 52 (1), 68–82.

Ceccagnoli, M., Graham, S. J., Higgins, M. J., Lee, J., 2010. Productivity and the role of complementary assets in firms demand for technology innovations. Industrial and Corporate Change 19 (3), 839–869.

Ceccagnoli, M., Higgins, M. J., Palermo, V., 2014. Behind the scenes: Sources of complementarity in r&d. Journal of Economics & Management Strategy 23 (1), 125–148.

Ceccagnoli, M., Jiang, L., 2013. The cost of integrating external technologies: Supply and demand drivers of value creation in the markets for technology. Strategic Management Journal 34 (4), 404–425.

Cerda, R. A., 2007. Endogenous innovations in the pharmaceutical industry. Journal of Evolutionary Economics 17 (4), 473–515.

Civan, A., Maloney, M. T., 2009. The effect of price on pharmaceutical r&d. The BE Journal of Economic Analysis & Policy 9 (1).

Cockburn, I. M., Henderson, R. M., 2001. Scale and scope in drug development: unpacking the advantages of size in pharmaceutical research. Journal of Health Economics 20 (6), 1033–1057.

Cromwell, J., Mitchell, J. B., 1986. Physician-induced demand for surgery. Journal of health economics 5 (4), 293–313.

Danzon, P. M., Nicholson, S., Pereira, N. S., 2005. Productivity in pharmaceutical–biotechnology r&d: the role of experience and alliances. Journal of health economics 24 (2), 317–339.

De Mouzon, O., Dubois, P., Scott Morton, F. M., Seabright, P., 2015. Market size and pharmaceutical innovation. The RAND Journal of Economics, 844–871.

Decarolis, F., 2015. Medicare part D: Are insurers gaming the low income subsidy design? American Economic Review 105 (4), 1547–80.

Decarolis, F., Polyakova, M., Ryan, S. P., June 2015. The welfare effects of supply-side regulations in medicare part d. Working Paper 21298, National Bureau of Economic Research.

DiMasi, J. A., Hansen, R. W., Grabowski, H. G., 2003. The price of innovation: new estimates of drug development costs. Journal of health economics 22 (2), 151–185.

Donohue, J. M., Cevasco, M., Rosenthal, M. B., 2007. A decade of direct-to-consumer advertising of prescription drugs. New England Journal of Medicine 357 (7), 673–681.

Drandakis, E. M., Phelps, E. S., 1966. A model of induced invention, growth and distribution. The Economic Journal, 823–840.

Dranove, D., Garthwaite, C., Hermosilla, M., 2014. Pharmaceutical profits and the social value of innovation. Tech. rep., National Bureau of Economic Research.

Dranove, D., Meltzer, D., 1994. Do important drugs reach the market sooner? The RAND Journal of Economics, 402–423.

Dubé, J.-P., Hitsch, G. J., Rossi, P. E., 2009. Do switching costs make markets less competitive? Journal of Marketing Research 46 (4), 435–445.

Duflo, E., 2000. Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. Tech. rep., National Bureau of Economic Research.

Duflo, E., 2004. The medium run effects of educational expansion: Evidence from a large school construction program in indonesia. Journal of Development Economics 74 (1), 163–197.

Duggan, M., Scott-Morton, F., 2010. The effect of medicare part d on pharmaceutical prices and utilization. The American Economic Review, 590–607.

Dussault, G., Franceschini, M. C., 2006. Not enough there, too many here: understanding geographical imbalances in the distribution of the health workforce. Human resources for health 4 (1), 12.

Einav, L., Finkelstein, A., Schrimpf, P., forthcoming. The response of drug expenditures to nonlinear contract design: Evidence from medicare part d. The Quarterly Journal of Economics.

Ericson, K. M., 2014. Consumer inertia and firm pricing in the medicare part D prescription drug insurance exchange. American Economic Journal: Economic Policy 6 (1), 38–64.

Ericson, K. M., Starc, A., 2015. Pricing regulation and imperfect competition on the massachusetts health insurance exchange. Review of Economics and Statistics 97 (3), 667–682.

Erikson, C., Jones, K., Tilton, C., 2012. Physician specialty data book. Association of American Medical Colleges.

Evans Robert, G., 1974. Supplier-induced demand: Some empirical evidence and implications in the economics of health and medical care.

Farrell, J., Klemperer, P., 2007. Coordination and lock-in: Competition with switching costs and network eects. In: Armstrong, M., Porter, R. (Eds.), Handbook of Industrial Organization. Vol. 3. Elsevier B.V., pp. 1967–2072.

Finkelstein, A., 2004. Static and dynamic effects of health policy: Evidence from the vaccine industry. The Quarterly Journal of Economics 119 (2), 527–564.

Finkelstein, A., 2005. The aggregate effects of health insurance: Evidence from the introduction of medicare. Tech. rep., National Bureau of Economic Research.

Forman, C., Goldfarb, A., Greenstein, S., 2008. Understanding the inputs into innovation: Do cities substitute for internal firm resources? Journal of Economics & Management Strategy 17 (2), 295–316.

Fox, J. G., Richards Jr, J. M., 1977. Physician dominance and location of foreign and us trained physicians. Journal of health and social behavior, 366–375.

Gans, J. S., Hsu, D. H., Stern, S., 2002. When does start-up innovation spur the gale of creative destruction? RAND Journal of Economics 33 (4), 571–586.

Gans, J. S., Hsu, D. H., Stern, S., 2008. The impact of uncertain intellectual property rights on the market for ideas: Evidence from patent grant delays. Management Science 54 (5), 982–997.

Gans, J. S., Stern, S., 2003. The product market and the market for ideas: commercialization strategies for technology entrepreneurs. Research policy 32 (2), 333–350.

Giaccotto, C., Santerre, R. E., Vernon, J. A., 2005. Drug prices and research and development investment behavior in the pharmaceutical industry*. Journal of Law and Economics 48 (1), 195–214.

Giovanetti, G. T., Jaggi, G., 2013. Beyond borders: Global biotechnology report 2013. Ernst & Young.

Gold, M., March 2006. The landscape of private firms offering medicare prescription drug coverage in 2006.

Grabowski, H. G., Vernon, J. M., 2000. Effective patent life in pharmaceuticals. International Journal of Technology Management 19 (1-2), 98–120.

Grossman, G. M., Helpman, E., 1991. Trade, knowledge spillovers, and growth. European Economic Review 35 (2), 517–526.

Handel, B. R., 2014. Adverse selection and inertia in health insurance markets: When nudging hurts. American Economic Review 103 (7), 2643–82.

Hausman, J. A., Poterba, J. M., Summer 1987. Household Behavior and the Tax Reform Act of 1986. Journal of Economic Perspectives 1 (1), 101–19.

Heiss, F., Leive, A., Mcfadden, D., Winter, J., 2013. Plan selection in medicare part D: Evidence from administrative data. Journal of Health Economics 32 (6), 1325–1344.

Heiss, F., McFadden, D., Winter, J., 2008. Mind the gap! consumer perceptions and choices. Tech. rep., Munich Center for the Economics of Aging (MEA) at the Max Planck Institute for Social Law and Social Policy.

Hellmann, T., 2007. The role of patents for bridging the science to market gap. Journal of Economic Behavior & Organization 63 (4), 624–647.

Henderson, R., Cockburn, I., 1996. Scale, scope, and spillovers: the determinants of research productivity in drug discovery. The RAND Journal of Economics, 32–59.

Hermosilla, M., 2015. Imperfect outsourcing of technological innovations. Working Paper, Johns Hopkins University.

Ho, K., Hogan, J., Scott Morton, F., March 2015. The Impact of Consumer Inattention on Insurer Pricing in the Medicare Part D Program. Working Paper 21028, National Bureau of Economic Research.

Hoadley, J., Cubanski, J., Summer, L., Huang, J., October 2012. Medicare policy data spotlight medicare part D: A first look at part D plan offerings in 2013Kaiser Family Foundation Report.

Hoadley, J., Summer, L., Hargrave, E., Cubanski, J., Neuman, T., August 2014. Medicare part D in its ninth year: The 2014 marketplace and key trends, 2006–2014Kaiser Family Foundation Report.

Kessel, M., Frank, F., 2007. A better prescription for drug-development financing. Nature biotechnology 25 (8), 859–866.

Kesternich, I., Heiss, F., Mcfadden, D., Winter, J., 2013. Suit the action to the word, the word to the action: Hypothetical choices and real decisions in medicare part D. Journal of Health Economics 32 (6), 1313–1324.

Ketcham, J. D., Lucarelli, C., Miravete, E. J., Roebuck, M. C., 2012. Sinking, swimming, or learning to swim in medicare part D. American Economic Review 102 (6), 2639–73.

Ketcham, J. D., Lucarelli, C., Powers, C. A., 2015. Paying attention or paying too much in medicare part d. American Economic Review 105 (1), 204–33.

Ketcham, J. D., Simon, K. I., 2008. Medicare part d's effects on elderly patients' drug costs and utilization. The American journal of managed care 14 (11).

Klemperer, P., 1987. Markets with consumer switching costs. The Quarterly Journal of Economics 102 (2), 375–94.

Kling, J. R., Mullainathan, S., Shafir, E., Vermeulen, L. C., Wrobel, M. V., 2012. Comparison friction experimental evidence from medicare drug plans. The Quarterly Journal of Economics 127 (1), 199–235.

Knittel, C., 1997. Interstate long distance rates: Search costs, switching costs, and market power. Review of Industrial Organization 12 (4), 519–536.

Krasner, J., December 31 2006. Insurer hits millions of seniors with drug cost hike. Retrieved from `http://www.boston.com`.

Kyle, M. K., 2006. The role of firm characteristics in pharmaceutical product launches. The RAND Journal of Economics 37 (3), 602–618.

Kyle, M. K., McGahan, A. M., 2012. Investments in pharmaceuticals before and after trips. Review of Economics and Statistics 94 (4), 1157–1172.

Larkin, I., 2008. Bargains-then-ripoffs: Innovation, pricing and lock-in in enterprise software. Academy of Management Proceedings.

Lerner, J., Malmendier, U., 2010. Contractibility and the design of research agreements. The American Economic Review, 214–246.

Lerner, J., Merges, R. P., 1998. The control of technology alliances: An empirical analysis of the biotechnology industry. The Journal of Industrial Economics 46 (2), 125–156.

Lerner, J., Shane, H., Tsai, A., 2003. Do equity financing cycles matter? evidence from biotechnology alliances. Journal of Financial Economics 67 (3), 411–446.

Levine, A., 2009. Licensing and scale economies in the biotechnology pharmaceutical industryWorking paper.

Lichtenberg, F. R., 1996. Do (more and better) drugs keep people out of hospitals? The American economic review, 384–388.

Lichtenberg, F. R., 2010. Pharmaceutical innovation, mortality reduction, and economic growth. Measuring the Gains from Medical Research: An Economic Approach 1980, 74.

Lichtenberg, F. R., Waldfogel, J., 2008. Does misery love company-evidence from pharmaceutical markets before and after the orphan drug act. Mich. Telecomm. & Tech. L. Rev. 15, 335.

Longman, R., 2006. Bigger deal values by splitting indications: Pdl's experience. Pharma MedTech Business Intelligence: In Vivo. Accessed at https://www.pharmamedtechbi.com/publications/in-vivo/24/2/bigger-deal-values-by-splitting-indications-pdls-experience.

Ludmerer, K. M., 1999. Time to heal: American medical education from the turn of the century to the era of managed care. Oxford University Press.

Malerba, F., Orsenigo, L., 2002. Innovation and market structure in the dynamics of the pharmaceutical industry and biotechnology: towards a history-friendly model. Industrial and corporate change 11 (4), 667–703.

Manchanda, P., Honka, E., 2005. Effects and role of direct-to-physician marketing in the pharmaceutical industry: An integrative review, the. Yale J. Health Pol'y L. & Ethics 5, 785.

McGuire, A., Henderson, J., Mooney, G., et al., 1988. The economics of health care: an introductory text. Routledge & Kegan Paul.

Miller, D. P., 2015. Risk adjustment and low income subsidy distortions in medicare part DMimeo, Clemson University.

Miller, D. P., Yeo, J., 2015a. Discrete Choice Models to Estimate Switching Costs: An Application to Medicare Part DMimeo, Clemson University and Singapore Management University.

Miller, D. P., Yeo, J., 2015b. The Consequences of a Public Health Insurance Option: Evidence From Medicare Part DMimeo, Clemson University and Singapore Management University.

Miller, K., December, 9 2014. Do private medicare firms have lower costs? Job market paper, Department of Economics, University of Minnesota.

Mossinghoff, G. J., 1999. Overview of the hatch-waxman act and its impact on the drug development process. Food & Drug LJ 54, 187.

Murphy, K. M., Topel, R. H., 2003. The economic value of medical research. Measuring the gains from medical research: An economic approach 15 (30), 125–146.

Murphy, K. M., Topel, R. H., 2006. The value of health and longevity. Journal of Political Economy 114 (5).

Newhouse, J. P., 1992. Medical care costs: how much welfare loss? The Journal of Economic Perspectives 6 (3), 3–21.

Newhouse, J. P., Williams, A. P., Bennett, B. W., Schwartz, W. B., 1982a. Does the geographical distribution of physicians reflect market failure? The Bell Journal of Economics, 493–505.

Newhouse, J. P., Williams, A. P., Bennett, B. W., Schwartz, W. B., 1982b. The geographic distribution of physicians: Is the conventional wisdom correct? Rand.

Nordhaus, W. D., 1969. Invention, growth, and welfare: A theoretical treatment of technological change. MIT press Cambridge, MA.

Nosal, K., 2012. Estimating switching costs for medicare advantage plansMimeo, University of Mannheim.

Office, C. B., 2014. Estimated effects on direct spending and revenues for health care programs of proposals in the president's 2015 budget. Tech. rep.
URL `http://www.cbo.gov/sites/default/files/cbofiles/attachments/45250-Health_Programs_Proposals.pdf`

Park, M., 2010. The economic impact of wireless number portabilityWorking Paper.

Patrick, G. L., 2013. An introduction to medicinal chemistry. Oxford university press.

Pisano, G. P., 1997. R&d performance, collaborative arrangements and the market for know-how: A test of the lemons' hypothesis. Unpublished Working Paper, Harvard Business School.

Pisano, G. P., 2006. Science business: The promise, the reality, and the future of biotech. Harvard Business Press.

Polyakova, M., September 2015. Regulation of insurance with adverse selection and switching costs: Evidence from medicare part D. Working Paper 21541, National Bureau of Economic Research.

Pong, R. W., Pitblado, J. R., 2001. Don't take" geography" for granted! some methodological issues in measuring geographic distribution of physicians. Canadian Journal of Rural Medicine 6 (2), 103.

Powell, W. W., 1996. Inter-organizational collaboration in the biotechnology industry. Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft, 197–215.

Powell, W. W., Brantley, P., 1992. Competitive cooperation in biotechnology: Learning through networks. Networks and organizations, 366–394.

Rice, T. H., 1983. The impact of changing medicare reimbursement rates on physician-induced demand. Medical care 21 (8), 803–815.

Romer, P. M., 1990. Endogenous technological change. Journal of Political Economy 98 (5 pt 2).

Rosenthal, M. B., Zaslavsky, A., Newhouse, J. P., 2005. The geographic distribution of physicians revisited. Health Services Research 40 (6p1), 1931–1952.

Samuelson, P. A., 1965. A theory of induced innovation along kennedy-weisäcker lines. The Review of Economics and Statistics, 343–356.

Schumpeter, J. A., 1942. Socialism, capitalism and democracy. Harper and Brothers.

Sharpe, S. A., 1997. The effect of consumer switching costs on prices: A theory and its application to the bank deposit market. Review of Industrial Organization 12, 79–94.

Shepard, M., 2015. Hospital Network Competition and Adverse Selection: Evidence from the Massachusetts Health Insurance Exchange. Job market paper, Harvard University.

Shi, M., Chiang, J., Rhee, B., 2006. Price competition with reduced consumer switching costs: The case of "wireless number portability" in the cellular phone industry. Management Science 52 (1), 27–38.

Spulber, D. F., 2014. The Innovative Entrepreneur. Cambridge University Press.

Stango, V., 2002. Pricing with consumer switching costs: Evidence from the credit card market. The Journal of Industrial Economics 50 (4), 475–492.

Starc, A., 2014. Insurer pricing and consumer welfare: evidence from medigap. The RAND Journal of Economics 45 (1), 198–220.

Summer, L., Hoadley, J., Hargrave, E., September 2010. The medicare part d low-income subsidy program: Experience to date and policy issues for considerationKFF Report.

Teece, D. J., 1986. Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. Research policy 15 (6), 285–305.

U.S. Department of Health and Human Services, September 2014. U.S. Department of Health and Human Services Rate Review Annual Report`http://aspe.hhs.gov/health/reports/2014/RateReview/rpt_RateReview.pdf`.

Viard, V. B., 2007. Do switching costs make markets more or less competitive? the case of 800-number portability. The RAND Journal of Economics 38 (1), 146–163.

Von Hippel, E., 1994. sticky information and the locus of problem solving: implications for innovation. Management science 40 (4), 429–439.

Waterson, M., 2003. The role of consumers in competition and competition policy. International Journal of Industrial Organization 21 (2), 129–150.

Weintraub, G. Y., Benkard, C. L., Van Roy, B., 2008. Markov perfect industry dynamics with many firms. Econometrica 76 (6), 1375–1411.

Wittink, D. R., 2002. Analysis of roi for pharmaceutical promotion (arpp). http://www.rxpromoroi.org/arpp/media/arpp$_h$*andout*$_0$927.*pdf*.

Yin, W., 2008. Market incentives and pharmaceutical innovation. Journal of Health Economics 27 (4), 1060–1077.

Yin, W., 2009. R&d policy, agency costs and innovation in personalized medicine. Journal of health economics 28 (5), 950–962.

Yin, W., Basu, A., Zhang, J. X., Rabbani, A., Meltzer, D. O., Alexander, G. C., 2008. The effect of the medicare part d prescription benefit on drug utilization and expenditures. Annals of internal medicine 148 (3), 169–177.