
DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY


Price Regulation in the Pharmaceutical Market

Link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:40050065

Terms of use
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at
https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFlOWI2M2EwYTg

Accessibility
https://accessibility.huit.harvard.edu/digital-accessibility-policy

Share Your Story
The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story

Price Regulation in the Pharmaceutical Market

A dissertation presented

by

Luca Maini

to

The Department of Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Economics

Harvard University

Cambridge, Massachusetts

May 2018

© 2018 Luca Maini

All rights reserved.

Dissertation Advisors:
Professor Ariel Pakes
Professor David Cutler

Author:
Luca Maini

Price Regulation in the Pharmaceutical Market

Abstract

This dissertation explores the economic impact of price regulation in the pharmaceutical market. The first chapter is methodological in nature. I show that in the presence of unobserved regulatory constraints traditional structural estimation techniques fail to deliver reliable moment restrictions. I then develop a general methodology that can be used for estimation in models where regulation generates unobserved restrictions on firms' strategies. The second chapter (co-authored with Fabio Pammolli) uses this novel methodology to calculate the impact of External Reference Pricing (ERP) on launch delays of new drugs in the European pharmaceutical market. Governments that implement ERP use prices in other countries as negotiation benchmarks in an effort to bring down the cost of prescription drugs. By doing so, they limit the ability of firms to price discriminate across countries, and create an incentive to withhold drugs from countries with lower willingness to pay. We find that removing ERP would reduce delays in Eastern Europe by up to 14 months per drug. The third chapter (co-authored with Josh Feng), analyzes the impact of value-based pricing (VBP) on overall welfare in the pharmaceutical market. Advocates of VBP promote it as a potential solution to concerns that drug prices are increasingly misaligned with their therapeutic benefits. We find that even though VBP can help realign R&D incentives, it does not maximize welfare for a wide range of economic models. Moreover, implementing VBP requires accurate measures of value. Using a set of disease-modifying treatments for multiple sclerosis we show that different data and methodologies yield measures of drug value that are only weakly correlated, suggesting that precisely estimating drug value may be a challenging proposition.

Contents

List of Tables

List of Figures

Preface

The pharmaceutical market is one of the largest in the world. Total sales of prescription drugs amounted to almost $750 billion in 2013, with $300 billion coming from the United States alone. The net present value of the industry's R&D pipeline is estimated to be in the $500 billion range. In recent years, drug spending has quickly increased, spurred by the introduction of several groundbreaking therapies for hepatitis, diabetes, and cancer. Many of these new drugs arrived on the market with considerable initial price tags, bringing drug prices to the forefront of the policy debate. This dissertation is organized in three chapters that examine the implication of price regulation in the pharmaceutical market from a methodological, empirical, and theoretical perspective.

The presence of regulation can create issues in the identification and estimation of model parameters. Since the constraints imposed by regulation are often unobserved by the econometrician, estimation techniques that do not explicitly take into account the presence of regulation can result in biased estimates. In the first chapter of this dissertation I provide a general method to estimate parameters in behavioral choice models where agents face unobserved constraints that may prevent them from choosing the ex-post optimal action. I extend previous methodologies that rely on best-response conditions to allow for shocks that affect the agent's intended strategy instead of his payoff. I characterize the assumptions needed to obtain identification and apply the methodology to two examples: a pricing problem where firms face unobserved price ceilings, and a single-agent dynamic model of entry with unobserved time delays.

The second chapter of this dissertation (coauthored with Fabio Pammolli) applies the methodology derived in the first chapter to analyze the impact of external reference pricing (ERP) on launch delays in the market for pharmaceutical products. Governments that implement ERP use prices in other countries as negotiation benchmarks in an effort to bring down the cost of prescription drugs. By doing so, they limit the ability of firms to price discriminate across countries, and create an incentive to withhold drugs from countries with lower willingness to pay. Using data on pharmaceutical sales in European countries from

2002 to 2012, we document the presence of widespread launch delays across Europe — up to three years on average in Eastern Europe. To distinguish between strategic delays caused by ERP and delays that arise for other reasons, we develop a dynamic structural model of entry that allows for externalities in price, which we estimate using a novel moment inequality approach. We find that removing ERP would reduce delays in Eastern Europe by up to 14 months per drug. At the same time, ERP has a small impact on firm revenue, so it would be theoretically possible to compensate firms for their profit loss in exchange for forgoing strategic delays.

The third chapter of this dissertation (coauthored with Josh Feng) analyzes the impact of value-based pricing (VBP) on overall welfare in the pharmaceutical market. We show that VBP does not maximize welfare for a wide range of economic models. In static models that disregard R&D incentives, the optimal policy is actually the opposite of VBP: the most effective drugs should be priced at the lowest possible level to maximize patient access. In models that include R&D incentives, VBP is only optimal if patients are fully insured. However, we also show that VBP has a desirable property: it maximizes the probability that an R&D projects with a positive net expected value will reach the market. This property however, depends on whether the social planner can accurately measure the "value" of a drug. We turn to this question in the second part of the chapter, using both clinical trial data and a database of patient claims for a set of disease-modifying treatments for multiple sclerosis. We find that different data and methodologies yield only weakly correlated measures of drug value, suggesting that precisely estimating drug value may be a challenging proposition.

Acknowledgments

It's hard to begin an acknowledgment section without making some trite, obvious statement. And yet, obvious things tend to have a redeeming quality. They are true. So it is true, though obvious, that I could not have gotten where I am right now without help. Lots of it.

Everyone finds inspiration in different places. My inspiration came from my teachers. It would be easy to say that there are too many of them to name — spending three fourths of your life as a student will do that — but that would be a disservice to me and them alike. And so: Fratel Marco, Marina Truant, Giannina Silva, Paola Pannuti, Achille Maffini, Gianna Raschi, Silvano Quomori, Paul Sally, Derek Neal, Hugo Sonnenschein, Victor Lima, Grace Tsiang, David Laibson, Brigitte Madrian. Each one of them left something of theirs with me. An impression, a parting thought, a quip, and, on occasion, a well-placed scolding. They showed me the beauty in discovery, and they dared me to drive further. They had the discipline to demand perfection, and the patience to let me catch up. They pushed me until I reached my limit, and then they hurled me past it.

My advisors Ariel Pakes, David Cutler, Robin Lee, and Ernie Berndt deserve a separate mention. I could say so many things about the support they showed me over the past few years, but perhaps only one really matters: they pointed the way by being themselves. They are the scholars I aspire to be.

I would also be remiss if I didn't mention the people who shared this journey with me. If teachers were my guides, they were my travelling companions: Marko, Dan, Nikhil, Pete, Molly, Tony, Jung, Annetta, Max, Xavier, Andy, Elaine, Wentao, Filippo, Matteo, Edo A. And Edo T, who kept me sane with coffee and executive lunch; and Josh, who is not only my coauthor, but also the best sparring partner I could wish for countless office chats.

Finally, my family. My parents, who were my first teachers; my sister, who was my first student; and Kemeng, who picks up the pieces of the glasses I break and calls me Harvard when I forget the stove on and the keys on the door. They are the ones who have made it all possible in the first place. I may never know just how much I owe them.

Thank you. It's been one hell of a ride.

Ai miei maestri

Chapter 1

Identification and Estimation in Models with Unobserved Strategy Constraints

1.1 Introduction

Estimation of many behavioral choice models relies on the traditional assumption that observed actions reflect ex-post optimizing behavior (Pakes, 2010; Pakes, Porter, Ho and Ishii, 2015). When agents maximize payoffs, the econometrician can derive restrictions on the unobserved parameters of the model by building moment conditions rooted in revealed preference: the expected value of any feasible action profile cannot exceed the expected payoff of the agent's observed behavior. In certain situations however, agents may face constraints that prevent them from choosing the payoff-maximizing action. These restrictions could take the form of regulatory constraints, such as price ceilings (Dubois and Lasio, 2017), or entry delay shocks (Maini and Pammolli, 2017); or they may be attributable to bounded rationality. In these instances the traditional estimation strategy might not deliver reliable restrictions on the parameters.

In this paper I extend the revealed-preference estimation framework to allow for an

element of randomness in agent choice. Throughout the rest of the paper I will refer to this stochastic component as *action shocks*. In the framework I consider, observed actions are a result of the agent's strategy plus the action shock component. Agents choose strategies that maximize expected payoffs over the distribution of the action shocks. The actions that the econometrician observes in the data therefore represent only one of many possible outcomes, which might not satisfy the ex-post optimality condition.

The primary challenge of this setup is that without observing strategies, it is impossible to simulate the expected payoff of the agent for arbitrary values of the unknown parameter. That is how traditional moment inequalities generate restrictions: by finding out the values of the unknown parameter for which the agent's strategy yields a higher payoff than all other strategies. These restrictions cannot be calculated if the econometrician does not know what the agent intended to do.

I show that in this setting one can construct restrictions based on aggregate moments that average payoffs across units of observation (usually markets, but also agents). The intuition behind this methodology is simply: even without observing the strategy of the agent, the realized action that appears in the data is a result of optimal play; hence, it is possible to obtain a sample analog of the average expected payoffs across markets (or agents) by observing data on repeated choices. The resulting moment conditions compare the average expected payoff of the agent (or of various agents) across markets to the counterfactual expected payoff earned by playing an arbitrary strategy for an arbitrary value of the parameter.

The main assumptions needed for this result is that the econometrician can recover a consistent approximation of the payoff function, i.e. an approximation that does not include a structural disturbance. At its core, the assumption that the structural error is zero implies that the econometrician and the agent observe the same information set. This assumption can be restrictive in many settings, but it is necessary for identification in this context. A model that incorporates a flexible enough structural disturbance will be observationally equivalent to a model with action shocks, since it will usually be possible to find a value of the structural

error that rationalizes the choice of the observed action.[1]

This estimation strategy is most likely to be helpful in situations where the econometrician cannot recover the full solution to the model. When a full solution can be recovered, the fact that the econometrician cannot observe the strategy played by the agent is irrelevant, because the agent is assumed to play the optimal strategy. In these cases the estimation technique will still be valid, but the unknown parameters can generally be recovered using more traditional solution methods, like maximum likelihood estimation. When the solution cannot be recovered however, the traditional approach will usually fail.

A key difference of this approach is that the expected payoff is not simulated, but rather calculated as a sample average. Therefore, the econometrician cannot recover expected payoffs for arbitrary values of the unknown parameter, but only for the value of the parameter that pertains to the data — that is, the true value of the parameter. As a consequence, my estimator will have different properties than traditional moment inequality estimators. In particular, I show that in many cases the estimator will only recover a one-directional bound on the unknown parameter.

The main environment that the model in the paper is trying to describe is an environment where firms face constraints that do not directly affect payoffs. For example, a price ceiling prevents the firm from choosing a higher price, thus indirectly affecting profits, while a fixed cost of entry does not restrict the options available to the firm, but directly changes the payoff associated with some of them. While models that include regulatory constraint of this kind may be estimated with ad-hoc methodologies (i.e. Dubois and Lasio, 2017), the approach in this paper provides a general solution.

Though this is not the main focus, one could also interpret the constraints of the model as being generated by bounded rationality. I briefly explore this possibility at the end of the paper. Several recent papers have considered models where agents deviate from the optimal strategy due to behavioral biases, such as rational inattention (Sallee, 2014), inertia (Ericson,

[1]The two models may however deliver different counterfactuals, so it will be up to the econometrician to determine what assumptions is a better fit for the data generating process.

2014), reference-dependence (Crawford and Meng, 2011), and cognitive hierarchy (Brown, Camerer and Lovallo, 2013). Ellison (2006) provides a good summary of earlier work on bounded rationality in IO models.

This paper's main contribution is to the theory of estimation from revealed preference conditions. The most general framework for this type of estimation is in Pakes *et al.* (2015) and Pakes (2010), who generalize and extend previous work by Hansen and Singleton (1982) and Luttmer (1996). This paper proposes a further generalization of this framework to allow for instances when the strategy of the agent is subject to unobserved stochastic shocks. The approach however, is only partially nesting the work of Pakes *et al.* (2015) and Pakes (2010) for the case of models without an added structural disturbance.

The rest of the paper proceeds as follows. In Section 1.2 I describe the setup of the general problem I consider, and lay out the assumptions needed for identification. I derive the main identification results and characterize some properties of the estimator in Section 1.3. Section 1.4 provides two practical examples of problems that can be analyzed with these methodological tools. The first problem considers a firm choosing an optimal pricing strategy when facing stochastic price ceilings. This problem is a generalized version of the problem analyzed in Dubois and Lasio (2017). The second problem is a firm choosing an optimal launch sequence for its products across many markets, and facing stochastic time-to-entry delays. The second chapter of this thesis (Maini and Pammolli, 2017) is an empirical analysis of this latter example. Finally, in Section 1.5 I discuss how this framework may be adapted to allow for the analysis of bounded rational play by agents who are choosing ex-post suboptimal actions not because of stochastic shocks, but because they are unable to fully solve the problem. Section 3.5 provides some conclusive remarks.

1.2 Theoretical Framework

1.2.1 Setup

The initial setup of the problem is identical to the moment inequality framework from Pakes *et al.* (2015), also discussed in Pakes (2010). For simplicity of comparison, I use a similar notation here as well. Consider a model with n agents, indexed by $i = 1, \ldots, n$. Let $\mathcal{J}_i$ denote the information set of agent i at the time actions are chosen, and let $\mathcal{A}_i$ be the set of possible actions an agent could take.

Agent i's payoff is determined by his action $a_i \in \mathcal{A}_i$, the actions of other agents $a_{-i} \in \mathcal{A}_{-i}$, and an additional set of variables $\mathbf{y}_i \in \mathbf{Y}_i$. The payoff function maps these variables to the set of the real numbers:

$$\pi : \mathcal{A}_i \times \mathcal{A}_{-i} \times \mathbf{Y}_i \to \mathbb{R}$$

The function $\pi(\cdot)$ and the probability distribution of $\mathcal{J}_i$ and $\mathbf{Y}_i$ are primitives of this game.

1.2.2 Stochastic Strategies

I refer to the observed behavior of agents in the data as *actions*, and instead define *strategies* as what agents choose when they maximize payoffs. Strategies and actions may not be linked by a bijective relationship. Instead, actions are the result of strategies plus a stochastic element. Formally, let $\mathcal{S}_i$ denote the set of all possible strategies an agent could choose. Strategies can be contingent on the information contained in $\mathcal{J}_i$ (i.e. $\mathbf{s}_i(\mathcal{J}_i)$), though to ease notational burden I will simply denote them as $\mathbf{s}_i$ in what follows. Then, the realized action a_i will be a random variable with conditional probability distribution $f_{\mathbf{s}_i}(a_i | \theta^s)$. I assume that the agent has perfect information on $f_{\mathbf{s}_i}(a_i | \theta^s)$, while the econometrician only knows $f(\cdot)$ up to a finite parameter vector θ^s, whose true value I denote as θ_0^s.[2]

[2]This assumption is analogous to distributional assumptions that are common in the IO literature, and necessary for identification in dynamic games (Rust, 1994).

Assumption 1 (Stochastic Best Response Condition)**.** *If $\mathbf{s}_i'$ is the strategy played by agent i, then*

$$\sup_{\mathbf{s}_i} \Pi\left(\mathbf{s}_i, \mathbf{s}_{-i}, \mathbf{y}_i\right) \leq \Pi\left(\mathbf{s}_i', \mathbf{s}_{-i}, \mathbf{y}_i\right) \tag{1.1}$$

for all $i = 1, \ldots, n$, where

$$\Pi\left(\mathbf{s}_i, \mathbf{s}_{-i}\mathbf{y}_i\right) = \int_{\mathcal{A}_i} \int_{\mathcal{A}_{-i}} \mathbb{E}\left[\pi\left(a_i, a_{-i}, \mathbf{y}_i\right) \middle| \mathcal{J}_i\right] \cdot f_{\mathbf{s}_{-i}}\left(a_{-i} \middle| \theta^s\right) da_{-i}\, f_{\mathbf{s}_i}\left(a_i \middle| \theta^s\right) da_i \tag{1.2}$$

If $f_{\mathbf{s}_i}\left(\cdot \middle| \theta^s\right)$ is a degenerate distribution, then this assumption is identical to the best-response assumption in Pakes (2010) and Pakes *et al.* (2015). In all other cases, the difference is that agents calculate expected payoffs over all possible realizations of $\mathbf{s}_i$ and $\mathbf{s}_{-i}$.

Assumption 2 (Counterfactual Conditions)**.** $\mathbf{y}_i = y\left(\mathbf{z}_i, a_i, a_{-i}\right)$, $\mathbf{s}_{-i} = s_{-i}\left(\mathbf{s}_i, \mathbf{z}_i\right)$ *and the distribution of $\mathbf{z}_i$ conditional on $\mathcal{J}_i$ and $\mathbf{s}_i = a_i$ does not depend on a_i.*

This assumption is necessary to account for situations in which $\mathbf{y}_i$ may respond endogenously to changes in actions, and situations where the strategies of other agents respond to changes in the strategy of agent i.

I note two things about this assumption. First, I have implicitly assumed that variables in $\mathbf{y}_i$ respond to actions a_i, and not to strategies $\mathbf{s}_i$ per se. This simply reflects a belief that in most real world applications it will be the realization of a strategy (i.e. the action a_i) that will affect payoff-relevant random variables, and not the strategy itself. However, the assumption can be modified to incorporate situations where $\mathbf{y}_i$ may depend on $\mathbf{s}_i$ without significantly altering any of the core results. Second, I am assuming that agents can observe each other's strategies. This may not be the case in certain applications, and in those cases one will need a model that explains how agents react after observing certain actions (i.e. $\mathbf{s}_{-i} = s_{-i}\left(a_i, \mathbf{z}_i\right)$).

Together, Assumptions 1 and 2 form the basis for the estimation strategy. Define

$$\Delta\Pi\left(\mathbf{s}_i, \mathbf{s}_i', \mathbf{s}_{-i}, z_i\right) \equiv \Pi\left(\mathbf{s}_i, \mathbf{s}_{-i}, z_i\right) - \Pi\left(\mathbf{s}_i', \mathbf{s}_{-i}, z_i\right)$$

Then, it follows from these two assumptions that

$$\mathbb{E}\left[\Delta\Pi\left(\mathbf{s}_i, \mathbf{s}_i', \mathbf{s}_{-i}, \mathbf{z}_i\right)\right] \geq 0 \qquad \forall\, \mathbf{s}_i' \in \mathcal{S}_i \tag{1.3}$$

Written out in full, this condition implies that deviating from the strategy played by the agent (after accounting for endogenous responses in other player's actions as well as in the payoff-relevant variables $\mathbf{y}_i$) results in lower expected payoffs:

$$\int_{\mathcal{A}_i}\int_{\mathcal{A}_{-i}} \mathbb{E}\left[\pi\left(a_i, a_{-i}, \mathbf{y}_i\right)\middle|\, \mathcal{J}_i\right] \cdot f_{s_{-i}(\mathbf{s}_i,\cdot)}\left(a_{-i}\middle|\, \theta^s\right) \mathrm{d}a_{-i}\, f_{\mathbf{s}_i}\left(a_i\middle|\, \theta^s\right) \mathrm{d}a_i -$$
$$\int_{\mathcal{A}_i}\int_{\mathcal{A}_{-i}} \mathbb{E}\left[\pi\left(a_i, a_{-i}, \mathbf{y}_i\right)\middle|\, \mathcal{J}_i\right] \cdot f_{s_{-i}(\mathbf{s}'_i,\cdot)}\left(a_{-i}\middle|\, \theta^s\right) \mathrm{d}a_{-i}\, f_{\mathbf{s}'_i}\left(a_i\middle|\, \theta^s\right) \mathrm{d}a_i \geq 0 \qquad \forall \mathbf{s}'_i \in \mathcal{S}_i$$

1.2.3 Measurement Model

In order to conduct estimation the model requires an additional assumption, which rules out the presence of structural error in the measurement of the payoff function. Let $r\left(a_i, a_{-i}, z_i^o; \theta^r\right)$ be a parametric function that approximates the payoff function $\pi(\cdot)$. This function depends on actions, an observable vector of variables z_i^o, and a vector of parameters θ^r whose true value I denote as θ_0^r.

Assumption 3 (Observability of payoff-relevant variables)**.** *Let z_{ij} be such that $z_{ij} \in z_i$ and $z_{ij} \notin z_i^o$. Then $z_{ij} \notin \mathcal{J}_i$.*

Here I simply rule out the possibility that the agent can condition its choice over a larger set of variables than the one observed by the econometrician. Hence, the econometrician knows all the payoff-relevant random variables.

A consequence of this assumption is that the difference between the parametric approximation of the payoff function and its realization conditional on the agent's action will be a random variable with mean zero: define

$$r\left(a_i, a_{-i}, z_i^o; \theta_0^r\right) \equiv \pi\left(a_i, a_{-i}, z_i\right) + \nu_{i,a_i,a_{-i}} \tag{1.4}$$

then

$$\mathbb{E}\left[\nu_{i,a_i,a_{-i}}\middle|\, \mathcal{J}_i\right] = 0 \quad \Longleftrightarrow \quad \mathbb{E}\left[r\left(a_i, a_{-i}, z_i^o; \theta_0^r\right)\middle|\, \mathcal{J}_i\right] = \mathbb{E}\left[\pi\left(a_i, a_{-i}, z_i\right)\middle|\, \mathcal{J}_i\right]$$

This assumption, which is the most onerous of the model, is nonetheless necessary to ensure identification. Adding to the payoff function a structural error known to the agent, but

not to the econometrician, implies that deviations from what the econometrician estimates to be the optimal strategy could be attributed to this error.[3] Therefore, these two unobserved sources of error would be indistinguishable.[4]

It is important to note that this assumption does not imply that the econometrician can recover the agent's *expected* payoff, which does not only depends on z_i, but also on $\mathbf{s}_{-i}$, which is unobserved. The estimation strategy I present in the next section deals with this limitation.

Note that Assumption 3 also ensures that one could recover a consistent estimate of θ^r by minimizing the distance between $r(\cdot)$ and observed payoffs:

$$\hat{\theta^r} = \arg\min_{\theta_r} \left\| r\left(a_i^o, a_{-i}^o, z_i^o; \theta^r\right) - \pi_i^o \right\|$$

for some measure of distance $\|\cdot\|$.

The measurement of the expected payoff as conditional on strategies $\mathbf{s}_i$ and $\mathbf{s}_{-i}$ is

$$R\left(\mathbf{s}_i, \mathbf{s}_{-i}, z_i^o; \theta^s\right) = \int_{\mathcal{A}_i} \int_{\mathcal{A}_{-i}} r\left(a_i, a_{-i} z_i^o; \hat{\theta^r}\right) \cdot f_{\mathbf{s}_{-i}}\left(a_{-i} \middle| \theta^s\right) \mathrm{d}a_{-i} \, f_{\mathbf{s}_i}\left(a_i \middle| \theta^s\right) \mathrm{d}a_i \tag{1.5}$$

By Assumption 3

$$\mathbb{E}\left[\Pi\left(\mathbf{s}_i, \mathbf{s}_{-i}, z_i\right)\right] = \mathbb{E}\left[R\left(\mathbf{s}_i, \mathbf{s}_{-i}, z_i^o; \theta_0^s\right)\right]$$

Hence, it is possible to rewrite the condition from equation 1.3 as

$$\mathbb{E}\left[\Delta R\left(\mathbf{s}_i, \mathbf{s}_i', \mathbf{s}_{-i}, z_i; \theta_0^s\right)\right] \geq 0 \qquad \forall \mathbf{s}_i' \in \mathcal{S}_i \tag{1.6}$$

where

$$\Delta R\left(\mathbf{s}_i, \mathbf{s}_i', \mathbf{s}_{-i}, z_i; \theta^s\right) \equiv R\left(\mathbf{s}_i, \mathbf{s}_{-i}, z_i^o; \theta^s\right) - R\left(\mathbf{s}_i', \mathbf{s}_{-i}, z_i^o; \theta^s\right)$$

[3] It may be possible to allow for some structural error in the payoff function, as long as this error can be recovered through static estimation. For example, if payoffs depend on a demand system, it may be possible to allow for an unobserved heterogeneity component, as long as the econometrician also has an instrument to account for the selection that this unobserved component generates. However, certain types of structural error, such as additive payoff shocks will generally be collinear with the unobserved parameter θ, and therefore must be ruled out a priori.

[4] Being able to observe behavior of the same agent in situations where the econometrician can assume that the strategy chosen always results in the optimal action may help disentangle the structural error from the strategy uncertainty. For example, Doraszelski *et al.* (2018) extrapolate unobserved costs from a market in equilibrium and use those estimates to fit models of learning on earlier data.

Equation 1.6 will form the basis of the estimation strategy.

1.3 Main Identification Results

I now turn to the task of generating sample moment conditions for estimation. The challenge that needs to be overcome is that the econometrician only observes the realization of the strategy, and not the strategy itself. Thus, the observed outcome may not be payoff-maximizing: per Assumption 1 agents play strategies that maximize the *expected* payoff, but the *realization* of the strategies need not result in ex-post optimality. Therefore the question is: is it possible to generate restrictions on the distribution of $f_{\mathbf{s}_i}\left(a_i|\,\theta^s\right)$ using only observed data on a_i and $r\left(\cdot\right)$?

Before considering the answer, note that identification in moment inequality is generally based on the idea that the agent's action contains some information about the true value of the unknown parameter of the model. In particular, the strategy of the agent will only be optimal for a certain range of values of the parameter. For values outside that range the econometrician should be able to find strategies that would earn higher revenue, which in turn allows him to rule those values out.

In this setting, the agent's action may not necessarily be optimal at the true value of the parameter. This generates a challenge, since it implies that moment conditions based on comparing observed outcomes to simulated counterfactuals may not always hold. In order to solve this problem, I show that aggregation across markets can generate a consistent sample analogue to the expected payoff.

Imagine that the econometrician observes data from K markets, indexed by $k = 1, \ldots, K$. For each market, denote the optimal strategy of each agent as $\mathbf{s}_{ik}^{\star} \equiv \mathbf{s}_i^{\star}\left(\mathcal{J}_{ik}, \theta^s\right)$. Assume that data contains actions a_{ik}^o for all agents, and that the econometrician can recover the payoff measurement function $r\left(\cdot\right)$. The observations in each market are assumed to be independent of one another.

Since equation 1.6 will hold in each market, it follows that the aggregate condition across

markets will also hold:

$$\mathbb{E}\left[\frac{1}{K}\sum_{k=1}^{K} R\left(\mathbf{s}_{ik}, \mathbf{s}_{-ik}, z_{ik}^{o}; \theta^{s}\right) - \frac{1}{K}\sum_{k=1}^{K} R\left(\mathbf{s}'_{ik}, \mathbf{s}_{-ik}, z_{ik}^{o}; \theta^{s}\right)\right] \geq 0 \qquad \forall \left\{\mathbf{s}'_{ik}\right\}_{k=1}^{K}, \mathbf{s}'_{ik} \in \mathcal{S}_{ik} \quad (1.7)$$

The following theorem shows that if the payoffs of the agent are well-behaved, it is possible to recover a sample analog of the first term of equation 1.7.

Theorem 1. *Assume that*

$$\int_{\mathcal{A}_i}\int_{\mathcal{A}_{-i}} r^2\left(a_i, a_{-i} z_i^{o}; \hat{\theta}^{r}\right) \cdot f_{\mathbf{s}_{-i}}\left(a_{-i} \middle| \theta^{s}\right) da_{-i}\, f_{\mathbf{s}_i}\left(a_i \middle| \theta^{s}\right) da_i < \infty$$

Then, for any $\varepsilon > 0$, there exists M' such that

$$\frac{1}{M}\sum_{k=1}^{M} r\left(a_{ik}^{o}, a_{-ik}^{o}, z_{ik}^{o}; \hat{\theta}^{r}\right) - \frac{1}{M}\sum_{k=1}^{M} R\left(\mathbf{s}_{ik}^{\star}, \mathbf{s}_{-ik}, z_{ik}^{o}; \theta_0^{s}\right) < \varepsilon$$

for all $M > M'$.

The proof of this theorem is in Appendix A.1.1. Intuitively, the theorem follows from the assumptions made in the model. Because of Assumption 1, the agent must choose the optimal strategy $\mathbf{s}_{ik}^{\star}$. Hence, each observed action a_{ik}^{o} is a random draw from $f_{\mathbf{s}_{ik}^{\star}}\left(\cdot \middle| \theta^{s}\right)$. This implies that $r\left(a_{ik}^{o}, a_{-ik}^{o}, z_{ik}^{o}; \hat{\theta}^{r}\right)$ is a random realization from the conditional distribution of the payoff of firm i, whose expectation is $R\left(\mathbf{s}_{ik}^{\star}, \mathbf{s}_{-ik} z_i^{o}; \theta_0^{s}\right)$ (by Assumption 3). Then, the result of the theorem follows from a generalized version of the law of large numbers for independent, non-identical random variables, which applies as long as the variance of the payoffs is bounded.

The second element of equation 1.7 depends on simulated counterfactual expected payoffs $R\left(\mathbf{s}'_{ik}, \mathbf{s}_{-ik}, z_{ik}^{o}; \theta^{s}\right)$. Assumption 2 ensures that we have a model for how other agents will react to deviations. Hence, $R\left(\mathbf{s}'_{ik}, \mathbf{s}_{-ik}, z_{ik}^{o}; \theta^{s}\right)$ can be simulated. In instances when we do not have a model for $\mathbf{s}_{-i}$, assuming that the behavior of other agents does not change would also suffice.[5]

[5]This may be the case in single-agent problems and in static games for example.

Theorem 2. *Assume that*

$$\int_{\mathcal{A}_i}\int_{\mathcal{A}_{-i}} r^2\left(a_i, a_{-i} z_i^o; \hat{\theta}^r\right)\cdot f_{\mathbf{s}_{-i}}\left(a_{-i}\middle|\,\theta^s\right) da_{-i}\, f_{\mathbf{s}_i}\left(a_i\middle|\,\theta^s\right) da_i < \infty$$

Then, for any θ^s, and for any $\varepsilon > 0$, there exists M' such that

$$\frac{1}{M}\sum_{k=1}^{M}\left(\int_{\mathcal{A}_i} r\left(a_{ik}, a_{-ik}^o, z_{ik}^o; \hat{\theta}^r\right)\cdot f_{\mathbf{s}'_{ik}}\left(a_{ik}\middle|\,\theta^s\right) da_{ik}\right) - \frac{1}{M}\sum_{k=1}^{M} R\left(\mathbf{s}'_{ik}, \mathbf{s}_{-ik}, z_{ik}^o; \theta^s\right) < \varepsilon$$

for all $M > M'$.

The proof of Theorem 2 is analogous to that of Theorem 1 (see Appendix A.1.2). Combining the result of the two theorems, the moment conditions are

$$\mathcal{M}\left(\left\{\mathbf{s}'_{ik}\right\}_{k=1}^{K}, \theta^s\right) = \min\left\{0, \frac{1}{K}\sum_{k=1}^{K} r\left(a_{ik}^o, a_{-ik}^o, z_{ik}^o, \hat{\theta}^r\right) - \frac{1}{K}\sum_{k=1}^{K} R\left(\mathbf{s}'_{ik}, \mathbf{s}_{-ik}, z_{ik}^o, \hat{\theta}^r\right)\right\} = 0 \quad (1.8)$$

and must hold for $\theta^s = \theta_0^s$ and for any set of strategies $\left\{\mathbf{s}'_{ik}\right\}_{k=1}^{K}$. In this case, the identified set is

$$\Theta^s = \left\{\theta^s : \sum_{\mathbf{s}_i}\mathcal{M}\left(\mathbf{s}_i, \theta^s\right) = 0\right\}$$

Before discussing the properties of the estimator, I note that the result of Theorem 1 could be extended to allow for some correlation between observations across markets, simply by relying on a law of large numbers for correlated random variables. The cost of this extension is that the number of markets needed to obtain an accurate approximation would be larger.

1.3.1 Properties of the Moment Inequality Estimator

I now discuss the properties of the estimator. In doing so, I elucidate some conceptual differences that differentiate other moment inequality approaches from the one presented in this paper.

The basic identifying principle of moment inequalities is rooted in a revealed preference axiom. The approach in Pakes (2010) and Pakes *et al.* (2015) exploits information contained in the agent's strategy. By assumption, the firm knows the true value of the unknown parameter, and chooses a strategy to maximize expected payoffs. Therefore, the econometrician can

generate restrictions by showing that for some values of the unknown parameter there are other strategies that generate a higher expected payoff.

In the setting I presented in the previous section, the strategy is unobserved: the econometrician sees a^o_{ik}, but not the underlying s_{ik}. Hence, restrictions are based on expected payoffs, which can be recovered in the aggregate. Moment inequalities rule out values of the unknown parameter for which it is possible to earn a higher expected payoff than the agent earned in the data.

Formally, the key distinction between the two approaches is that when the strategy is observed, the econometrician can simulate the expected payoff for *any arbitrary value* of the unknown parameter θ^s, whereas in my setting the econometrician can only recover the expected payoff for the true value θ^s_0. As a result, the difference that forms the basis for the inequality condition compares the agent's expected revenue for the true value of the parameter θ^s_0 to a simulated counterfactual that uses an arbitrary value of θ. This contrasts with the traditional approach, which compares the observed strategy with a counterfactual strategy, both evaluated for the same value of θ.

This difference has implications for the properties of the estimator. In particular, whenever $R\left(\cdot;\theta^s\right)$ is monotonic in θ^s, this implies that equation 1.8 will provide a one-directional bound (proof in Appendix A.1.3).

Theorem 3. *Let $R\left(\cdot;\theta^s\right)$ be defined as in equation 1.5 and $r\left(\cdot\right)$ be defined as in equation 1.4. The following two conditions hold:*

1. *If $\frac{\partial R(\cdot;\theta^s)}{\partial \theta^s} \geq 0$ for all θ^s, then*

$$\mathcal{M}\left(\left\{\mathbf{s}'_{ik}\right\}_{k=1}^{K}, \theta^s\right) = 0$$

   for all $\left\{\mathbf{s}'_{ik}\right\}_{k=1}^{K}$, $\theta^s \leq \theta^s_0$;

2. *if $\frac{\partial R(\cdot;\theta^s)}{\partial \theta^s} \leq 0$ for all θ^s, then*

$$\mathcal{M}\left(\left\{\mathbf{s}'_{ik}\right\}_{k=1}^{K}, \theta^s\right) = 0$$

   for all $\left\{\mathbf{s}'_{ik}\right\}_{k=1}^{K}$, $\theta^s \geq \theta^s_0$.

For an example, consider a model where $a_i \sim \mathcal{N}\left(s_i;\theta^s\right)$, i.e. θ^s regulates the variance of

the observed action. Then, a higher θ^s implies that the realized action is more likely to be further away from the desired action s_i, which in turn will generally imply a lower expected payoff.

Theorem 3 suggests that in many empirical applications it may be impossible to generate bounds on the parameter of interest in both directions by simply relying on the condition in equation 1.8. In these cases, I propose an alternative approach (proof in Appendix A.1.4).

Theorem 4. *Let $R(\cdot;\theta^s)$ be defined as in equation 1.5 and $r(\cdot)$ be defined as in equation 1.4. Furthermore, let $F^{ub}\left(\mathbf{s}_{ik}, \mathbf{s}_{-ik}, z^o_{ik}; \theta^s\right)$ be a function such that*

$$F^{ub}\left(\mathbf{s}_{ik}, \mathbf{s}_{-ik}, z^o_{ik}; \theta^s\right) \geq R\left(\mathbf{s}_{ik}, \mathbf{s}_{-ik}, z^o_{ik}; \theta^s\right)$$

for all θ^s. If

$$F^{ub}\left(\mathbf{s}_{ik}, \mathbf{s}_{-ik}, z^o_{ik}, \tilde{\theta^s}\right) < \frac{1}{K}\sum_{k=1}^{K} r\left(a^o_{ik}, a^o_{-ik}, z^o_{ik}, \hat{\theta^r}\right)$$

for some $\tilde{\theta^s}$, then $\tilde{\theta^s}$ is rejected from the identified set (i.e. $\tilde{\theta^s} \notin \Theta$).

The intuition behind the theorem is simple. One way to rule out a potential value $\tilde{\theta^s}$ is to show that the agent cannot achieve as high a payoff as it did in the data if $\theta = \tilde{\theta^s}$. In general, one could do so by testing all the strategies available to the agent. This approach is generally unfeasible if the action space is too large. Instead, one could find a function of θ^s that is always greater than the payoff of the firm, and then simply test that function.

Finding the right function could be tricky. In general, many functions will satisfy the premise of the theorem, but functions that do so, and still deliver meaningful restrictions, may be hard to come by. An approach that could be successful is to build functions $F^{ub}(\cdot;\theta)$ that recreate an environment similar to what the agent is facing, but with more favorable conditions. This may be a situation where a firm is facing reduced competition, or where the payoff function is modified to shut down some channels that have a negative impact on payoffs (for example, if the payoff function includes negative spillover effects). I provide explicit examples for the two applications I consider in the next section.

1.4 Applications and Discussion

1.4.1 A pricing game with unobserved constraints

In this example I consider the problem of a firm that is setting prices in an environment where prices are subject to an unobserved price cap. This is a generalized version of the problem considered by Dubois and Lasio (2017), who study price regulation on pharmaceuticals in France.

There are n firms indexed by $i = 1, \ldots, n$. Each firm sets prices for a set of m_i products, indexed by $j = 1, \ldots, m_i$. Prices are subject to a price cap $\bar{p}_{ij}$, which is unobserved by the econometrician. Assume the price cap can be modeled as a draw from a distribution $g\left(\bar{p}_{ij} \middle| \theta^s\right)$. I assume that firms do not observe price caps in advance.[6] Instead, firms submit bids to the government, and then prices that are above the price cap are adjusted downward. Formally, the set of possible strategies chosen by the firm is $\mathbb{R}^{m_i}_{\geq 0}$ (i.e. the positive real n-space). Hence, a strategy $\mathbf{s}_i = \left\{s_{ij}\right\}_{j=1}^{m_i}$ is a positive, real-valued, m_i-dimensional vector. The realized price $p_i = \left\{p_{ij}\right\}_{j=1}^{m_i}$ is defined such that

$$p_{ij} = \min\left\{s_{ij}, \bar{p}_{ij}\right\}_{j=1}^{m_i}$$

The payoff function is simply the profit function of the firm. For each market, the maximization problem of the firm is

$$\max_{\mathbf{s}_i} \Pi\left(\mathbf{s}_i, \mathbf{s}_{-i}, z, c_i\right) \equiv \sum_{j=1}^{m_i}\left[\int_{\mathbb{R}^{m_i}_{\geq 0}} \int_{\mathbb{R}^{M_{-i}}_{\geq 0}} D_{ij}\left(p_{ij}, p_{-i}, z\right) \cdot \left(p_{ij} - c_{ij}\right) \mathrm{d}F_{\mathbf{s}_{-i}}\left(p_{-i} \middle| \theta^s\right) \mathrm{d}F_{\mathbf{s}_i}\left(p_{ij} \middle| \theta^s\right)\right]$$

where $M_{-i} = \sum_{l \neq i} m_l$ is the total number of products sold by other firms, $D_{ij}(\cdot)$ denotes demand for product j, sold by firm i, $F_{\mathbf{s}_i}\left(p_i \middle| \theta^s\right)$ and $F_{\mathbf{s}_{-i}}\left(p_{-i} \middle| \theta^s\right)$ represent the cumulative distribution functions for p_i and p_{-i} given strategies $\mathbf{s}_i$ and $\mathbf{s}_{-i}$, and I include a vector z of demand controls, as well as a vector of marginal costs of production $c_i = \left\{c_{ij}\right\}_{j=1}^{m_i}$. Note that demand depends on the prices charged by other firms, which in turn may be constrained.

[6] This assumption is the key difference between this example, and the problem considered by Dubois and Lasio (2017). In their model, firms know price caps in advance. That allows them to use first-order conditions of the unconstrained problem (together with a few additional assumptions) to recover marginal cost.

The fact that the price of competitors may be constrained means that the firm cannot easily solve this problem: the best response to the unconstrained price may not be the best response to the constrained price. Hence, the choice of $\mathbf{s}_i$ will reflect the firm's perceived probability that their competitors will be price constrained, which makes this a hard problem to solve analytically.[7]

This problem cannot be solved using the usual estimation methods – e.g. derive the first order condition and find the value of θ^s that minimizes it. The first-order condition will hold for the firm's choice of $\mathbf{s}_i$ (which is unobserved), but not necessarily for the realized price p_i.[8] Instead, one can generate restrictions on θ^s by building moment conditions based on expected payoffs.

Assume that the econometrician can observe data from K markets, indexed by $k = 1, \ldots, K$. Denote market-specific variables with a k subscript and observed data with an o-superscript (i.e. observed market prices are p^o_{ijk}). Also assume that the econometrician can recover consistent estimates $\hat{D}_{ij}(\cdot)$ and $\hat{c}_{ijk}$ for the demand function and the marginal cost of production. Finally, since this is a static model, I evaluate deviations while holding the strategies of competitors fixed.

Let θ^s_0 denote the true value of θ^s. By Theorems 1 and 2, the following must hold for $\theta^s = \theta^s_0$ and any given pricing strategy $\mathbf{s}'_i$:

$$\frac{1}{K}\sum_{k=1}^{K}\sum_{j=1}^{m_i}\hat{D}_{ij}\left(p^o_{ijk}, p^o_{-ik}, z^o_k\right)\cdot\left(p^o_{ijk}-\hat{c}_{ijk}\right) - \frac{1}{K}\sum_{k=1}^{K}\sum_{j=1}^{m_i}\left(\int_{\mathbb{R}^{m_i}_{\geq 0}} D_{ij}\left(p_{ijk}, p^o_{-ik}, z^o_k\right)\cdot\left(p_{ijk}-\hat{c}_{ijk}\right)\mathrm{d}F_{\mathbf{s}'_{ik}}\left(p_{ijk}\middle|\,\theta^s\right)\right) \geq 0 \quad (1.9)$$

Notice that the strategies of competitors are held fixed in this setting (to satisfy the require-

[7] For example, imagine the pricing problem of a generic manufacturer who is competing with a price-controlled brand. When announcing prices for the following period, the optimal price will depend on the price cap of the brand, which in turn may not be known in advance.

[8] In some cases, it may be possible to recover the optimal strategy of the firm. For example, in a single-product, single-agent problem, the unconstrained first-order condition (where $p_{ijk} = s_{ijk}$) can be used to identify the profit-maximizing price $p^\star_{ijk}$. Then, the optimal strategy will be to set $s_{ijk} = p^\star_{ijk}$. Things become more complicated with multiple products and multiple agents, because the optimal price for any one product will also depend on the likelihood that the prices of other products will be constrained by the price cap.

ments of Assumption 2). Since this is a static game, the deviations are equivalent to the conditions of a Nash equilibrium.

Equation 1.9 compares the average expected payoff of firm i across markets under the current strategy, to the average expected payoff using an arbitrary strategy $\mathbf{s}_i'$. With more than one firm (and each firm in a sufficiently large number of markets), it is possible to build moment conditions for each firm. Alternatively, with a small number of markets, one could aggregate expected payoffs across firms.

In this example, whether or not the second element of the equation is monotonic in θ^s will depend on the distribution $g\left(\bar{p}_{ij}\middle|\,\theta^s\right)$. For example, if $\bar{p}_{ij}$ follows a an exponential distribution with parameter θ^s, the profits of the firm will be monotonically decreasing in θ^s: higher θ^s means that the expected price cap is more likely to be binding.

There are two approaches to obtain bounds in both directions when expected profits are monotonically decreasing in θ^s. First, notice that 1.9 will yield a lower bound on θ^s. If θ^s is very low, the firm's pricing strategy is less likely to be constrained, which in turn should give it more room to generate higher profits. Hence, it is possible that the econometrician may find an alternative pricing strategy that yields higher expected profits.

To obtain an upper bound, there are two possibilities. First, the lowest possible value for θ^s is the value implied by simple maximum likelihood combined with the assumption that the price ceiling is binding in all markets. Hence

$$\theta^s_{UB1} = \arg\max \prod_{k=1}^{K} \mathcal{P}\left(\bar{p}_{ij} = p^o_{ikj}\middle|\,\theta^s\right)$$

Second, it may be possible to obtain a better bound using the method described in Theorem 4. One potential candidate function for $F^{ub}\left(\cdot;\theta^s\right)$ could be the profit of the firm when there are no competitors. In this case, the firm will definitely earn higher profits, and the optimal price can be derived easily by solving the maximization problem of the firm (which is now easily solvable because there are no competitors). Then, if for a given value of $\theta^s > \theta^s_{UB1}$ the firm in this counterfactual environment earns lower profits than in the data, θ^s can be rejected.

1.4.2 A single-agent dynamic entry model with time delays

In this second example consider the problem of a firm that is deciding when to launch its product across a set of different markets, and faces time delays that may affect its ability to enter in each market in the desired period. The second chapter of this thesis focuses on the empirical analysis of this problem. Here, I simply provide a brief theoretical overview.

The firm operates in a discrete-time, finite-horizon environment, where periods are indexed by $t = 0, 1, \ldots, T$. It owns a set of m products indexed by $i = 1, \ldots, m$, which it can sell in a set of n markets, indexed by $j = 1, \ldots, n$. The observed action of the firm is a launch sequence $S = \{s_{ij}\}$, where s_{ij} denotes the launch period of product i in market j.

The state variable of the problem is $S_t = \{s_{ijt}\}_{i,j}$ where

$$s_{ijt} = \begin{cases} s_{ij} & \text{if } s_{ij} \leq t \\ 0 & \text{otherwise} \end{cases}$$

The profits from selling product i in market j may depend on the state variable, and is given by $\pi(S_t, z_{ij})$. Specifically, profits may be affected by the launch of competing products owned by the same firm, as well as the launch of the same product in other markets, through channels like parallel trade or reference pricing.

The firm faces stochastic shocks to the launch sequence in the form of random time delays. Formally, a delay shock is a binary random variable ρ_{ijt}. If $\rho_{ijt} = 1$, then drug i cannot enter in country j until period $t + 1$ (when a new shock is drawn). Denote the parameter that governs the probability that a shock comes up positive as ψ, i.e. let $\psi_{ijt} = f(\psi)$ denote the probability that $\rho_{ijt} = 1$.

The firm maximizes profits by choosing where to send entry applications at the beginning of each period. Hence, a strategy is a map

$$\mathcal{A} : \mathcal{S} \to \{0, 1\}^{n \times T}$$

where $\mathcal{S}$ is the set of all possible launch sequences, and $\mathcal{A}$ generates a set of binary vectors A_{lt} with an action profile for each period until T_l.

The firm solves

$$\max_{\mathcal{A}} \sum_{S} \left(\sum_{t=0}^{T} \beta^t \sum_{j=1}^{n} \pi \left(S_t, z_{ij} \right) \cdot \mathcal{P} \left(S_t | \, \mathcal{A} \right) \right)$$

where $\mathcal{P} \left(S_t | \, \mathcal{A} \right)$ depends on ψ_{ijt} and the strategy of the agent.

Since $\pi_{ij} \left(S_t \right)$ is state-space specific, this problem will not have a general analytic solution. Moreover, due to the size of the state space ($(n \cdot m)^T$ possible states), computing the solution through numerical methods is unfeasible. However, one can build moment conditions based on the expected profit of the firm.

Assume that the econometrician observe from multiple firms, indexed by $k = 1, \ldots, K$; each operating in the same markets and facing the same delay shocks. Assume that the econometrician can recover a function $r \left(S_t, z_{ij} \right)$ that approximates $\pi \left(S_t, z_{ij} \right)$. Denote observed entry sequences as S_k^o. Then, the following moment conditions will hold for $\psi = \psi^0$, and can be used to generate restrictions:

$$\frac{1}{K} \sum_{k=1}^{K} \left(\sum_{t=0}^{T} \beta^t \sum_{j=1}^{n} \pi \left(S_{kt}^o, z_{ij} \right) \right) - \frac{1}{K} \sum_{k=1}^{K} \left(\sum_{S} \beta^t \sum_{j=1}^{n} \pi \left(S_t, z_{ij} \right) \cdot \mathcal{P} \left(S_t | \, \mathcal{A}', \psi \right) \right) \geq 0 \qquad (1.10)$$

The expected profit of the firm will be decreasing in ψ_{ijt}, hence, if $f \left(\psi \right)$ is monotonic, so will the profits of the firm with respect to ψ. As in the previous example, equation 1.10 will provide a lower bound: for smaller values of ψ, one can find strategies that allow the firm to enter more quickly than it was able to in the data, and in doing so may able to identify strategies that earn more profits and reject those values of ψ.

As in the previous example, there are two ways to obtain an upper bound. First, a simple upper bound can once again be estimated by maximum likelihood:

$$\psi_{UB1} = \arg\max \prod_{k=1}^{K} \mathcal{P} \left(S_k^o | \, \psi, \mathcal{A}_0 \right)$$

where $\mathcal{A}_0$ is the strategy whereby the firm always applies for entry in all markets right away. This generates an upper bound on the parameter ψ because it assumes that all delays are due to the delay shocks.

Second, to obtain a tighter bound using the method described in Theorem 4 one could

simulate expected profits using $\pi'\left(z_{ij}\right) = \max_{S_t} \pi\left(S_t, z_{ij}\right)$ in each market j and for each product i. In this scenario, the firm always earns more money, and the solution to the model should be easy to derive (this follows from the fact that the new profit function does not incorporate spillover effects, hence each market can be analyzed separately).

1.5 Another potential application: bounded rationality

While the main objective of this paper is to describe an environment where firms are facing unobserved constraints generated by government regulation, the methodology developed in the previous sections could also be applied to situations where the agent is unable to solve the maximization problem specified in the model. In these instances, the shocks that prevent strategies from realizing the ex-post optimal outcome would be interpreted as a sign that the agent is held up by constraints generated by bounded rationality.

The setup presented previously retains a best-response assumption that assumes the agent can fully solve the problem. Hence, it appears conceptually inconsistent with the idea that agents exhibit bounded rationality. However, I show in the next section how to reformulate the framework of the model using a weaker assumption: agents maximize from a restricted action set. This setup is more clearly related to the concept of bounded rationality, but observationally equivalent to the previous one, and therefore can be analyzed using the same tools.

1.5.1 An alternative formulation: restricted action sets

I keep the same notation and setup as before, but I assume that agents maximize payoffs by choosing strategies from a subset $\mathcal{A}_i^f$ of the choice set $\mathcal{A}_i$. I call this subset the *feasible action set*. The econometrician knows $\mathcal{A}_i$, but cannot observe $\mathcal{A}_i^f$. One possible interpretation of this model is that the agent has limited processing power, and therefore only evaluates a subset of all the possible actions (for example, most chess players can generally only think a handful of moves ahead, and can only compare a few possible alternatives).

I replace Assumptions 1 and 2 with two closely related assumptions and keep Assumption 3. I drop the distinction between actions and strategies, and in the notation write the actions of other players as $\mathbf{a}_{-i}$ (in bold to denote the fact that the agent is treating it as a random variable).

Assumption 4 (Limited Best Response)**.** *If a_i' is the action chosen by agent i, then*

$$\sup_{a_i} \pi\left(a_i, \mathbf{a}_{-i}\mathbf{y}_i\right) \leq \pi\left(a_i', \mathbf{a}_{-i}\mathbf{y}_i\right) \tag{1.11}$$

for all $a_i \in \mathcal{A}_i^f \subseteq \mathcal{A}_i$, and $i = 1, \ldots, n$.

Assumption 5 (Counterfactual Conditions II)**.** $\mathbf{y}_i = y\left(\mathbf{z}_i, a_i, \mathbf{a}_{-i}\right)$, $\mathbf{a}_{-i} = a_{-i}\left(a_i, \mathbf{z}_i\right)$ *and the distribution of $\mathbf{z}_i$ conditional on $\mathcal{J}_i$ and a_i does not depend on a_i.*

From these three assumptions, and given observed action a_i^o in the data, it is straightforward to derive the the following restriction on $\mathcal{A}_i^f$:

$$\mathbb{E}\left[r\left(a_i', a_{-i}, z_i^o, \hat{\theta}_r\right)\right] \geq \mathbb{E}\left[r\left(a_i^o, a_{-i}, z_i^o, \hat{\theta}_r\right)\right] \implies a_i' \notin \mathcal{A}_i^f \tag{1.12}$$

The basic idea of this inequality is that the econometrician can rule out certain actions from the feasible action set by checking that the agent would have been better off by choosing them.

Both this model and the previous one share a common characteristic: the observed action may not satisfy global ex-post optimality. Both will be able to rationalize the same data: if one assumes that there are no frictions, then actions that would earn a higher payoff in counterfactual simulations are simply excluded from the feasible set. Therefore the two models are observationally equivalent.

The condition in equation 1.12 will set bounds for the set $\mathcal{A}_i^f$. In particular, the econometrician will be able to characterize a set

$$\overline{\mathcal{A}}_i^f = \left\{a_i' : \mathbb{E}\left[r\left(a_i', a_{-i}, z_i^o, \hat{\theta}_r\right)\right] \leq \mathbb{E}\left[r\left(a_i^o, a_{-i}, z_i^o, \hat{\theta}_r\right)\right]\right\}$$

Note that it will generally be the case that $\mathcal{A}_i^f \subseteq \overline{\mathcal{A}}_i^f \subseteq \mathcal{A}_i$. In particular, the econometrician

will not be able to exclude from the feasible set actions that yield lower payoffs than observed in the data. This asymmetry mimics the uni-directional bound issue discussed in Section 1.3.1.

1.6 Conclusion

This paper provides a general framework for estimation and identification of parameters in models where observed behavior may not satisfy ex-post optimality restrictions. I build moment conditions by adapting and extending the standard best-response inequalities to allow for shocks that affects the strategies of agents, under the assumption of ex-ante optimizing behavior.

In many models, agents may be subject to regulatory constraints that prevent them from choosing their preferred strategy. Unless the econometrician knows the exact nature of these constraints, the observed behavior of agents will appear suboptimal in these situations. To obtain information on model parameters in these settings, I show how to construct moment conditions that take into account the impact of constraints to the strategy modeled as action shocks: disturbances that affects the realization of the agent's strategy.

Identification depends on the assumption that the econometrician and the agent share the same information set. This assumption is restrictive in many settings, and therefore may limit the practical applications of this methodology. Generally speaking, the approach will work well in situations where the econometrician has excellent information on payoffs. This could be the case in the analysis of consumer choices when agents purchase goods that don't have an inherent utility value, such as financial objects, certain types of insurance, and potentially even health services. It will also be the case with firm choices when the objective function can be assumed to match the profit function, and accurate data on costs and revenue is available. Conversely, the assumptions will almost certainly fail when considering consumer choices for goods where match-specific utility is an important component, and in any situation where data on important variables is missing.

The main objective and contribution of this paper is to provide a flexible estimation

method for environments where regulation generates unobserved restrictions on the set of actions that agents can choose. These restrictions will not affect payoffs directly, but will generate behavior that would appear suboptimal in an unconstrained model. The methodology described in the paper provides a simple way to incorporate these types of restrictions in the model and still generate meaningful restrictions on the relevant parameters. The two examples characterized in the paper demonstrate the range of environments that this methodology can successfully tackle.

As a second contribution, I also show that the same setup can be easily adapted to describe the behavior of a bounded rational agent. In this case, the restrictions on feasible strategies do not come from a regulator, but rather from the agent's rationality constraints. Given that models where the econometrician himself is unable to find the full solution are becoming more and more common, it may be important to allow for suboptimal agent behavior as well. The tools derived in this paper can be applied to these problems.

Chapter 2

Reference Pricing as a Deterrent to Entry: Evidence from the European Pharmaceutical Market[1]

2.1 Introduction

The rise in prescription drug spending is a growing concern for many countries around the world. Outside the US, governments try to control spending by directly negotiating prices with manufacturers, and implementing a variety of policy tools to limit markups on patent-protected drugs. A widely adopted practice, usually referred to as External Reference Pricing (ERP), consists of benchmarking drug prices by using prices in other countries. ERP has an obvious appeal for governments because it is simple, it guarantees that prices will be in line with other countries, and it can reduce spending. Firms however, worry about the limitations that ERP imposes on their ability to optimally price discriminate across countries with different willingness to pay.

The debate over the effects of reference pricing has important implications for policy. Using ERP carries virtually no negative repercussions for the home country, but it can impose

[1]Co-authored with Fabio Pammolli

an externality on foreign countries. In particular, firms will have an incentive to delay entry in low-income countries whose prices are referenced by high-income countries whenever the externality imposed through reference pricing outweighs the revenue earned by expanding into additional markets. Drugs often enter in many markets several years after receiving marketing approval, so the welfare loss generated by ERP could be very large (Reich, 2000).

In this paper, we develop a structural model of entry that allows for externalities in price across markets, and use it to provide an estimate of the impact of ERP on launch delays. In the model, firms maximize profits by choosing an optimal entry sequence, conditional on demand and price conditions in each country. To isolate the impact of reference pricing we exploit the fact that ERP operates through a very specific channel. On the margin, ERP generates delays when the change in expected profits from launching in an additional country is outweighed by the expected loss from the externality generated by reference pricing.

We estimate the model using data on sales of pharmaceutical products across all Member States of the European Economic Area (EEA). Most European countries adopt ERP among the criteria used to set prices. Moreover, the EEA includes countries with highly heterogeneous income, which creates the potential for strategic delays. Empirically, widespread launch delays occur in almost all European countries, though some of these delays are likely due to factors other than reference pricing. Before launching, governments and drug manufacturers engage in negotiations that can last several months, and countries can also delay the entry of products they consider unsafe. On average, relative to their marketing approval date, products are delayed by about 3 years in Eastern European countries, and 1 year elsewhere.

We begin our exercise by estimating demand and price in each country. We use a random utility nested logit model for demand, and estimate prices using a flexible parametric function that tries to capture the decision process of the government. The function predicts equilibrium prices as a combination of the reference price and an internal *government price* — which represents the price that would have been granted in the absence of reference pricing. The degree to which the reference price affects the equilibrium is a country-specific parameter in the pricing model. We find that allowing for this additional degree of heterogeneity is

important, as many countries do not follow their stated ERP guidelines perfectly.

Our estimated demand and price primitives suggest that the externality generated by reference pricing is large enough to incentivize strategic delays. In simulations that compare the expected revenue of various entry sequences we find that firms earn higher revenue when delaying entry in some countries. In particular, delaying entry in at least some Eastern European countries yields higher revenue for 70% of drugs. In roughly 20% of cases, withholding the product from *all* Eastern European countries would be preferable to launching everywhere at the same time. However, we find that virtually no drugs would earn higher revenue by delaying entry in countries outside Eastern Europe.

To quantify the impact of ERP we must isolate delays due to reference pricing from idiosyncratic delays generated by sources outside the control of the firm, such as the time needed to negotiate the terms of entry. We model this residual source of delay using binary shocks. In each period, firms that apply for entry in a given country draw a shock. If it comes up negative, entry has to be postponed until the next period, when a new shock is drawn.[2]

We estimate the probability of experiencing an idiosyncratic delay using the novel moment inequality estimator derived in the first chapter of this dissertation.[3] While moment inequalities have been applied to single-agent models of entry with spillover effects or externalities (Holmes, 2011; Morales *et al.*, 2017), what makes this approach novel is that our empirical framework presents an additional complication: firm strategies are partially unobserved. The entry sequence we observe only reveals when firms were able to enter. Situations in which the firm applied and was delayed however, are observationally equivalent to situations in which the firm did not apply.

The inequalities rely on a revealed preference argument. We assume that firms are

[2]We abstract from fixed entry costs. In a finite-horizon setting with fixed entry costs, delaying is only justified if costs are declining or highly fluctuating. Neither of these seems likely in the context we study. In general, fixed costs of entry of should be small for drugs that have already received marketing approval.

[3]Estimation techniques for dynamic entry models that rely on solving the firm's problem are generally unfeasible in settings similar to ours due to the cardinality of the action space of the firm (for a set of N countries over a T-period horizon, the firm can choose between T^N possible strategies). Instead, our approach does not require us to identify the optimal strategy of the firm or compute the value function, though it can only provide bounds on the parameters of the model.

maximizing expected profits and compare the expected profits of the observed entry sequence to the counterfactual profits of playing a different strategy. These inequalities will not always hold for individual firms: the realization of the random delay shocks in the data might prevent the firm from achieving the desired entry sequence. However, the first chapter of this dissertation shows that these differences will disappear when considering average payoffs across many firms. In this empirical application we find that these moment conditions can only provide a lower bound on the parameter of interest. We calculate an upper bound by exploiting the fact that the approval date is the earliest time at which the firm could have sent an entry application.

Over the period from 2002 to 2012, our estimates imply that replacing ERP with a pricing mechanism that does not generate externalities in price across countries would reduce delays in each Eastern European country by up to 63%, or 14 months per drug on average.[4] Several possible alternatives to ERP have been proposed in the policy literature, from transitioning to a centralized European cost-effectiveness evaluation system (Drummond, 2003), to implementing two-part pricing systems where products are supplied at cost and governments make transfers to firms in order to reach static and dynamic efficiency, to creating barriers preventing reference pricing and import-export of pharmaceutical products across countries (Towse *et al.*, 2015). The exact policy would affect firm profits, but not the implications for strategic delays, so our counterfactual has broad external validity.

Removing ERP would get rid of strategic delays, but would be hard to implement politically, so we also suggest a way to eliminate strategic delays while leaving ERP in place. Under our proposal, a central European authority would offer each firm a lump-sum subsidy in exchange for sending entry applications to all countries simultaneously. We estimate that the size of the subsidy would be small — around €20 million for the average drug. This is a consequence of the fact that in the current equilibrium ERP does not have a large impact on

[4]By distinguishing between Eastern Europe and Western Europe we do not mean to associate any value to the geographic location of these sets of countries. Rather, we draw this demarcation for convenience. Countries in Eastern Europe share certain traits that make their bundling convenient for our purpose: they have lower income (and prices), and smaller market size than virtually all countries in Western Europe.

firm revenue (our estimates indicate that Eastern European prices are not much lower than prices in Western European countries that use ERP more aggressively).

Our analysis has some limitations. The complexity of the problem we consider forces us to make several simplifying assumptions, which are necessary to perform the empirical analysis, but not necessarily desirable. First, we assume that firms act as single-agents. Even though the firms we consider are monopolists with regards to the specific molecule they produce, virtually all therapeutic classes we consider contain at least a few different molecules, which are presumably substitutable with one another to some degree. Second, we assume that there is no structural error in either our demand or price model. This assumption is necessary to build the moment inequalities given that firm strategies are unobserved. Finally, we do not explicitly model the government's choice of a reference pricing function, opting instead to treat it as an exogenous feature. We discuss these limitations more in depth in the relevant sections of the paper and in the conclusion.

Our paper contributes to four main strands of economic literature. First, it contributes to a growing body of work, both empirical and theoretical, studying the impact of price regulation on the access to pharmaceutical products. The empirical side of this literature usually analyzes the impact of government policy on launches using a reduced-form framework (Danzon *et al.*, 2005; Danzon and Epstein, 2012; Kyle, 2007; Kyle and Qian, 2013; Cockburn *et al.*, 2016). A notable exception is Duso *et al.* (2014), which examines the welfare impact of parallel trade in Germany.[5] On the theory side, this literature has focused on simulating the impact of reference pricing on firm strategy (e.g. Borja, 2014; Toumi *et al.*, 2013; Stargardt and Schreyögg, 2006; Houy and Jelovac, 2015), or establishing conditions under which regulation that limits price discrimination is beneficial or harmful to welfare (e.g. Birg, 2016; Brekke *et al.*, 2007, 2015, 2016; Matteucci and Reverberi, 2017). Our contribution is that we explicitly model the impact of reference pricing on firm incentives and develop an estimation strategy to isolate the effect of this policy on launch delays.

[5] Another methodologically related paper is Chaudhuri *et al.* (2006), which uses structural techniques to estimate the impact of patent policy on patient welfare in the Indian market for quinolones.

Second, our paper is related to a series of studies on the impact of regulation that links prices to endogenous market benchmarks. For example, both Medicare Part B and Medicaid tie drug reimbursements to the average of reported private market prices. Duggan and Scott Morton (2006) show that in the case of Medicaid this regulation creates a distortion that leads to higher prices in the private market. Another set of policies with a similar effect are so-called "price-linked" subsidies, i.e. subsidies that are linked to market prices. Jaffe and Shepard (2017) and Decarolis (2015) show that these types of subsidies can distort premiums in health exchanges and Medicare Part D respectively. More generally, price externalities across firms have been detected in the absence of government intervention. Grennan (2013) and Grennan and Swanson (2016) show that knowing how much rival hospitals paid for medical devices can affect future prices. Our paper shows that if pricing strategies are constrained, firms can also respond along different margins (i.e. by manipulating their entry strategy).

Third, we contribute to the vast empirical Industrial Organization literature on entry models, which originated with Bresnahan and Reiss (1991) (for an overview of this literature see Berry and Reiss, 2007). Most papers in this literature use stochastic fixed costs of entry (e.g. Seim, 2006). Since these costs are less relevant in our setting, we take a different approach and replace them with stochastic delay shocks. These shocks have different implications for estimation: they do not directly affect the profit of the firm, but rather impose stochastic constraints on the action space of the firm. Moreover, when the realization of delay shocks is unobserved, the firm's strategy is also unobserved, which creates additional challenges in estimation.

Our final contribution is to the literature on partial identification started by Manski (2003). We develop a novel approach to deal with the challenges introduced by delay shocks. The empirical literature on partial identification is growing and includes several papers (Dickstein and Morales, 2018; Eizenberg, 2014; Holmes, 2011; Illanes, 2016; Katz, 2007; Morales *et al.*, 2017; Pakes *et al.*, 2015). Our approach is closest to that of Holmes (2011) and Morales *et al.* (2017), but differs in the way that identification is obtained. While their approach identifies the set of parameter values for which the firm's observed strategy is optimal, our approach

identifies the set of parameters consistent with the revenue earned by the firm in the data.

The rest of the paper proceeds as follows. Section 2.2 introduces the institutional environment of the European pharmaceutical market. Section 2.3 describes the data and discusses preliminary evidence that supports the hypothesis that firms are delaying launches in countries with low prices in order to avoid the impact of external reference pricing. We present our theoretical model of entry in Section 2.4. The estimation is then divided in two parts. We present our empirical model and estimation results for demand and price in Section 2.5, while Section 2.6 contains the dynamic analysis. We discuss the implications of our results for counterfactuals and policy analysis in Section 2.7. Finally, in Section 3.5 we provide some concluding remarks, a discussion of the paper's limitations, and a roadmap for future research.

2.2 The European Pharmaceutical Market

Europe represents roughly 22% of the world's pharmaceutical market in terms of ex-factory sales; half of the US and Canada combined (EFPIA, 2016). Much like in the United States, the debate over pharmaceutical spending and how to best control it is a topical issue in many European countries. Even though prices tend to be lower than US prices (Danzon, 2003; Danzon and Furukawa, 2006), governments are concerned about rising healthcare expenditure, and increasingly look to pharmaceuticals as a potential area for savings (Deloitte, 2013).

In this section we outline the main characteristics of the European pharmaceutical market, focusing on the aspects of regulation that are most often associated with launch delays.

2.2.1 Marketing Approval of Pharmaceutical Products in Europe

In all countries around the world new drugs can only be sold after being reviewed for efficacy and safety. The European Medicines Agency (EMA) oversees this process in the European Economic Area. While in other parts of the world marketing approval for new drugs is generally granted by a regulatory authority whose jurisdiction is limited to one country

(i.e. the FDA in the United States, or the PMDA in Japan), member states of the European Economic Area have been relying on a shared approval process since 1995 – the year the EMA was founded.[6] Though national drug agencies still exist, their effort is now organized and regulated by the EMA.

Pharmaceutical companies seeking approval for their products can choose between three possible procedures. The *centralized procedure* is administered by the EMA itself, and grants automatic approval in all EEA Member States. It is available to all drugs, and compulsory for certain classes of drugs, including biologics.[7] Drugs for which the centralized procedure is not mandatory can also go through two additional channels. If the drug already has a marketing authorization from any EEA member state, the firm can use the *mutual recognition procedure* to extend it to any other member state using a fast-track procedure taking no longer than 90 days (European Parliament, 2001).[8] The other alternative is the *decentralized procedure*. In this case, the firm submits an application to multiple countries at the same time and designs one as the Reference Member State in charge of reviewing it (European Parliament, 2004).

The centralized marketing approval process ensures that the cost of seeking additional marketing approvals all but disappears as soon as firms receive marketing approval from any country in Europe (or from the EMA). This eliminates one of the most common explanations for launch delays. However, firms may still incur into delays because individual countries retain the ability to regulate prices independently from one another.

---

[6]The European Economic Area consists of all Member States of the European Union, plus Norway, Iceland, and Liechtenstein. Switzerland (also in our data), is a member of the European Free Trade Area, but not of the EEA. However, it has a series of bilateral trade agreements that allow it to take part in the common European market.

[7]The full list of drugs that must receive approval by the EMA includes: human medicines containing a new active substance to treat HIV or AIDS, cancer, diabetes, neurodegenerative diseases, auto-immune and other immune dysfunctions, and viral diseases; medicines derived from biotechnology processes; advanced-therapy medicines, such as gene-therapy, somatic cell-therapy or tissue-engineered medicines; and medicines seeking orphan designation.

[8]Other countries can to refuse the extension by claiming that doing so would create significant risks for public health, although this does not happen very frequently.

2.2.2 Price Regulation of Pharmaceutical Products in Europe

Most European countries provide some form of single-payer coverage, meaning the government bears the vast majority of prescription drug costs. Even in countries where the government does not directly insure patients (e.g. Germany, Netherlands, Switzerland), the government negotiates a price cap with manufacturers. Thus, all governments impose strict restrictions on drug prices, with the primary goal of controlling spending.

Pricing restrictions typically only apply to drugs that are paid for by the government through the public health insurance system. However, since European citizens overwhelmingly access health care through government-funded programs, the exclusion of a product from public formularies results in its de-facto exclusion from the national market (European Commission, 2012).[9] The ability to effectively deny entry provides governments with the necessary leverage to demand lower prices.

Firms petition for reimbursement status by submitting pricing and reimbursement applications to each government. The time required to review an application and negotiate a price can vary significantly across countries. In theory, Directive 89/105/EEC (informally known as the Transparency Directive) states that governments can take no longer than 180 days to review a pricing and reimbursement application (Council of European Communities, 1988). In practice however, this limit is often surpassed, both because of enforceability issues, and because governments can stop the clock by asking for additional information. Data on turnaround times for applications is scarce, but survey evidence from the late 1990s indicates that the average varies substantially, from 0 days in UK and Germany, to over two years in Belgium (OECD Health Policy Studies, 2008; PICTF, 2006).

The requirements of pricing and reimbursement applications vary across countries, though firms must generally include both a clinical dossier detailing the medical benefits of the drug, as well as an economic report with projected sales and a proposed price. The government then uses this information as inputs into the pricing decision. In theory, the

[9]The only exceptions to this rule tend to be generic drugs, whose low price tag makes them a cost-effective option even in the absence of government coverage.

government strives to set prices that reflect the value of each drug and reward the firm's innovative efforts, while at the same time keeping spending under control. In practice, estimating the value of a drug is a complicated and costly exercise. Most countries require firms to disclose the price charged for the drug in other countries, and then try to set prices that are approximately consistent with what other governments are paying.

It is important to stress that ERP is not the only (or even the main) policy instrument used by countries in setting prices. Countries often rely on a variety of other methods, including, but not limited to Health Technology Assessments, internal reference pricing (which links prices of molecules within a pre-specified equivalence class), and price freezes/cuts. These policies will be especially important in situations when a reference price cannot be observed, or when the reference price is higher than what the government is able to pay.

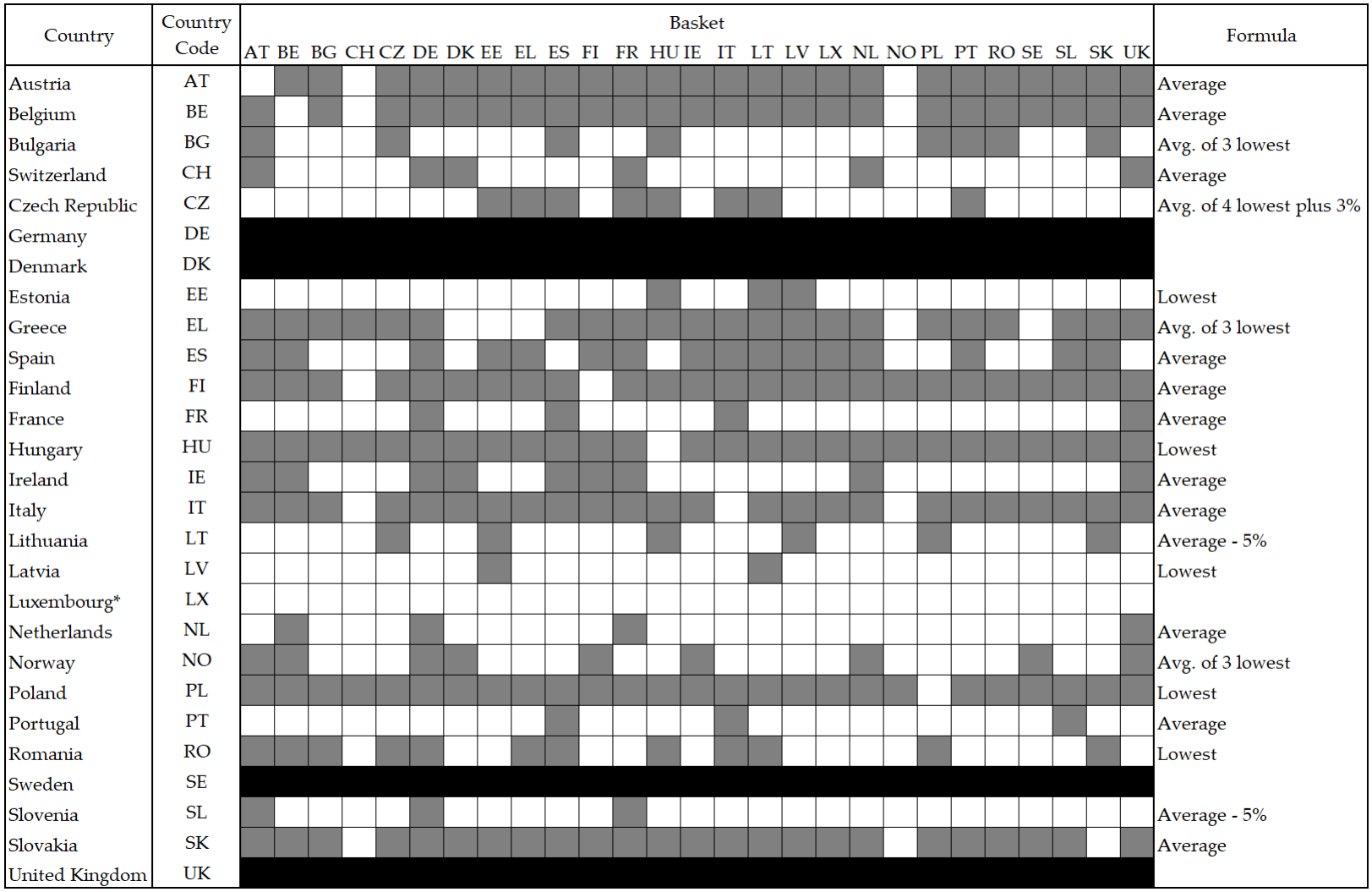
Finally, governments ultimately care about overall pharmaceutical spending rather than prices. Price-volume agreements that protect spending such as clawback policies are common, meaning that the outcome of the negotiation between the government and the firm is usually not a set list price, but rather a price schedule that depends on volume. Prices and volumes will also be correlated if the government is willing to provide more favorable coverage to firms that give better pricing conditions.

2.2.3 Overview of External Reference Pricing

In 2012 all European countries indicated ERP as one of the criteria used in setting prices except Denmark, Germany, Sweden, and the UK.[10] Both the pharmaceutical industry and policymakers have acknowledged the externality generated by ERP and its role in producing launch delays for new pharmaceutical products (EFPIA, 2014; Carone *et al.*, 2012). However, governments remain reluctant to abandon the policy, because of the savings they claim it generates.

The two most important aspects of ERP policies are the reference basket (i.e. the basket of countries whose prices are sampled), and the formula used to compute the reference

[10]Since then, Denmark and Germany have adopted ERP as well.



* Luxembourg only references the drug's country of origin.

Figure 2.1: *Reference Price Baskets and Formulas for EEA countries*

price. For both, there is significant variation across countries.[11] Some governments (e.g. Austria, Belgium, Finland, Hungary, and Poland) require firms to submit prices from all other countries in the European Union. Others only reference similar countries, both in terms of geographical proximity, size, and income level — for example, Estonia references Hungary, Latvia, and Lithuania, while France references Germany, Italy, Spain, and the UK. In terms of the reference formula, most countries use the average across the reference basket, but a few (e.g. Latvia, Poland, and Romania) use the lowest price, while others still use slight variations: Bulgaria, Greece, and Norway use the average of the three lowest prices in the basket. Figure 2.1 offers an overview of cross-country variation in reference baskets and formulas.[12]

The stringency with which each country adheres to their stated ERP guidelines may vary across countries. Some governments state that ERP is only used to "inform" the pricing decision, meaning that we might expect prices to be affected by ERP but not necessarily to be perfectly aligned with the reference pricing benchmark. In other instances, governments may push for prices that are below the benchmark if they expect to sell higher volumes than the referenced countries.[13] Countries whose governments claim to only use ERP informally include Belgium, Finland, France, Italy, Poland, and Spain.[14]

We use the reference basket and formulas to estimate reference prices in our model, but

[11]Contrary to cross-country variation, over time variation in ERP policies is much more scarce. The biggest policy change happened in 2010 when Greece switched its reference function in an effort to lower prices. Other changes are tied to the entry of new Member States in the European Union or in the Eurozone. These occurred in Austria, Belgium, Finland, and Italy (all of whom reference EU member states); and Spain (which references countries in the Eurozone). Finally, Portugal, Hungary, and Poland made small adjustments to their reference baskets at various points.

[12]The table shown in the paper represents a snapshot of reference baskets and formulas in 2012. It was built by combining several published sources (Carone *et al.*, 2012; European Federation of Pharmaceutical Industries and Associations, 2014; Kanavos *et al.*, 2011; Leopold *et al.*, 2012; Wilsdon *et al.*, 2013) with unpublished IMS reports. For the analysis used in the paper we generated yearly tables to capture some small changes that happened in a few countries with regard to their reference basket and the formula used.

[13]According to several informal conversations of the authors with industry insiders, these sort of arguments appear to be popular in countries such as Italy, where the prices of smaller markets are included in the reference function.

[14]Sources are in disagreement as to whether France and Italy use ERP formally or informally. EFPIA (2014) classifies both as Informal/Formal.

we exclude a few additional characteristics of ERP policies that can also vary across countries. We briefly list them here for completeness. First, countries update the reference prices with varying frequency, from as little as every 6 months (e.g. Greece, and Slovenia) , to as many as 60 (Finland). Second, countries can use raw ex-factory prices, or apply a PPP adjustment (all Scandinavian countries do so). Third, not all countries apply ERP to the same set of drugs. Most countries apply ERP only to drugs that are reimbursed through the national health insurance system, but some apply it to all new innovative drugs (e.g. France), and others to all drugs, regardless of reimbursement status (e.g. Greece).

2.2.4 Parallel Trade

ERP affects firm's incentives by limiting their ability to price discriminate across countries. Another force that operates through the same channel is parallel trade. The European Economic Area is a free trade area, with no limits over the flow of goods across borders. This includes patent-protected products, such as brand prescription drugs. Parallel importers purchase drugs in countries with low prices and sell them abroad where prices are higher.

In a world without frictions, parallel trade would drive the price of all drugs to the minimum available, and act as an extreme version of reference pricing. In practice, this does not happen, because firms have a variety of tools that they employ to fight the impact of parallel trade. For example, Kyle (2011) shows that firms differentiate the products they sell across countries by changing strength, dosage, or packaging. Another important tool is the ability to manage supply quotas, which was affirmed in a European Court decision in 2008.[15] Moreover, Kanavos and Costa-Font (2005) suggest that the difference in prices between the parallel traded product and the original is negligible, with gains from trade accruing mostly

[15]The case pitted GlaxoSmithKline against a group of Greek wholesalers and parallel importers (Joined Cases C-468/06 to C-478/06, court opinion and proceedings available at http://curia.europa.eu/juris/liste.jsf?language=en&num=C-474/06). In November 2000 GlaxoSmithKline stopped supplying certain drugs to Greek wholesalers, opting instead for direct delivery to pharmacies and hospitals. Their goal was to prevent the sale of medicines destined to the Greek market to parallel importers. The resulting lawsuit ended with a decision on September 18th, 2008, which prohibited the direct restriction of parallel trade by altering the traditional distribution channels. However, the decision also carved out the possibility for the firm, to take reasonable measures to protect itself, such as establishing supply quotas to wholesalers based on current market demand.

to intermediaries in the distribution chain, perhaps as a result of weak substitution incentives for pharmacies and physicians.

Even if it does not affect prices, however, parallel trade can still undermine firm revenue. Using our data we are unable to separate the effect of reference pricing from that of parallel trade. That being said, our model will generally apply to any policy that limits price discrimination. In counterfactual exercises where we lift pricing restrictions we likewise assume that both reference pricing and parallel trade are unavailable (together with any other policy that could potentially affect price discrimination).

2.3 Preliminary Evidence of the Impact of ERP

We now present some preliminary evidence of the presence of strategic delays. Our results confirm empirical patterns that several researchers have documented using similar data (Cockburn *et al.*, 2016; Danzon *et al.*, 2005; Danzon and Epstein, 2012; Kyle, 2007), but we also highlight novel facts on the effect of reference pricing. We show that delay patterns are consistent with a strategic response to reference pricing rules, but that some delays are more likely to be generated by idiosyncratic mechanisms, such as the turnaround time of pricing and reimbursement applications. We also find that over time prices fall in countries who apply ERP, relative to countries who do not.

The section is divided in three parts. We start by describing our data in 2.3.1. Next, in 2.3.2, we characterize the extent of launch delays in Europe, and the differences we observe across countries. Finally, we look at price trends and their correlation with delays in 2.3.3.

2.3.1 Data

The main source of data for the empirical analysis is the MIDAS database maintained by Quintiles-IMS, a global information company specializing in the health care sector. The data covers sales of all pharmaceutical products for European countries from 2002 to 2012.[16] It

[16]We are missing data entirely for Cyprus, Iceland, Lichtenstein, and Malta. A few other countries have partially missing data. See Appendix B.2 for more details.

consists of a quarterly panel of volume and revenue sales divided by country. Products are defined by a combination of molecule, firm, product name, form, strength, and package. Quintiles-IMS collects this information by surveying pharmacies and hospitals.

To the best of our knowledge, this database represents the most comprehensive source of data on sales in the European pharmaceutical market. Nonetheless, it has a few important limitations, which we discuss below.

First, the data does not provide any information on the approval dates of drugs. We collect approval dates for all EMA-approved medications from the EMA's website, and the approval date of all mutual recognition applications from an internet database maintained by the Heads of Medicines Agencies (HMA).[17]

Second, IMS reports ex-factory revenue sales, which do not usually incorporate rebates and discounts that are sometimes granted by individual payers. While in the US estimates of discounts for brand drugs oscillate between 20-40% during the period we consider (see Congressional Budget Office, 2005; Aitken *et al.*, 2016), discounts tend to be much lower in Europe (Danzon, 2003; Danzon and Furukawa, 2006). According to industry insiders, average rebates for patent-protected brand drugs are rarely above 10%.[18] Unfortunately, there is currently no available data on pharmaceutical rebates in Europe, so in our paper we simply use the prices implied by the IMS data.[19]

Third, the data contains some missing information for certain countries and years. Because of the externalities generated by reference pricing, missing data points can have an impact on non-missing observations as well. To minimize the impact of missing data we resort to

[17] Both datasets are publicly available. The HMA is a network of the heads of the European national authorities in charge of the regulation of medicinal products for human and veterinary use in the European Economic Area. The data can be found at http://mri.cts-mrp.eu/Human

[18] Author's own estimation based on several personal conversations.

[19] As an aside, contracts that combine list prices with hidden discounts could circumvent ERP and restore the ability to use price discrimination. However, this does not seem to be the case even though these contracts are theoretically available. One possible reason why rebates are underutilized in Europe is that governments have higher transparency requirements than private firms, due to the necessity to account for their spending. Hence hiding discounts might be more difficult. Another possibility is that firms know that if they were to avoid ERP by using large discounts, government might find ways to force them to reveal those discounts.

imputation using a variety of techniques.[20]

We integrate the IMS data with a few additional sources. On top of the aforementioned EMA and HMA data on approval dates, we collect GDP and population data from Eurostat, as well as data on the incidence of diseases in each European country from the Global Burden of Disease Study. We use that information to build the market size variable for the demand estimation. We also use quarterly exchange rates from the European Central Bank to convert sales data from countries that use currencies other than the Euro.

Sample Selection We observe around 6,000 molecules and 3,000 firms in our data. Most of these molecules and firms are old off-patent and generic products with negligible yearly sales, and many are only available in one or two countries. We are interested in new, on-patent products whose potential market spans multiple European countries. Hence, we select a subsample of drugs that satisfy the following three criteria:

1. The drug was first launched on or after January 1st, 1995.
2. The drug had at least one new launch in a European country on or after January 1st, 2002
3. One of the three following conditions is satisfied:
   (a) The drug was approved by the EMA using the centralized procedure between 1995 and 2012,
   (b) The drug successfully completed at least one Mutual Recognition Application between 1995 and 2012,
   (c) The drug is sold in at least 10 countries by 2012, and is classified as a patent-protected brand drug for at least some part of our sample period.

We end up selecting 481 drugs (we define a drug as the combination of a molecule, a firm,

[20] See Appendix B.2 for more details.

Table 2.1: *Summary Statistics*

		Full Sample	Main Sample	Dynamic Sample
# therapeutic classes		241	109	44
# firms		2,944	168	47
# molecules		6,354	475	86
Class-firm-mol combinations		55,134	481	87
Class-firm combinations			375	84
Diffusion	mean	2.1	20.1	21.3
	median	1	22	24
Yearly sales	mean	€3,547,665	€115,427,475	€121,977,298
	median	€92,489	€39,214,276	€46,340,438
Approval Method	EMA		312	24
	MRP		127	46
	Other		42	17

This table reports summary statistics for the IMS MIDAS database. The full sample includes all prescription drugs in the data. The main sample includes prescription drugs that satisfy the criteria laid out in section 2.3.1. The dynamic sample includes a subset of main sample drugs whose patent expired by December 31st, 2012. See Section 2.2.1 for an overview of the marketing approval methods. Yearly sales refers to the entire EEA territory.

and a therapeutic class).[21] Most of the products we select received approval from the EMA through the centralized procedure or applied for mutual recognition. We also include a few drugs that we were not able to match with the EMA and HMA data on approval dates, but that we observe being sold in many European countries.[22] Unsurprisingly, drugs in our main sample experience much greater sales and diffusion relative to the average drug in the data (see Table 2.1). The median drug in our main sample is available in 22 countries (by the end of 2012), and collects yearly sales of €38.6 million across all European markets. For comparison, the median product in the full sample is sold in only 1 country, and earns less than €100,000 every year.

We also select a subsample of drugs within our main sample whose patent expired prior to

[21] All molecules in our main sample are sold by a single firm, but can sometimes be available in different therapeutic classes. In these cases, IMS reports separate observations for each therapeutic class. We keep this distinction since our demand estimation relies on therapeutic classes to define markets.

[22] For these products we impute the European approval date as the date of the fifth launch. We do not use the first launch because in many instances products that apply through the mutual recognition procedure start selling in the reference member state a year or two in advance relative to the rest of Europe.

December 31st, 2012. This smaller group is used in dynamic analysis, when our methodology requires that we are able to compute the overall expected payoff of a drug until the time its patent expires.[23] This smaller sample consists of 87 drugs and has similar characteristics relative to the main sample of drugs in terms of sales and diffusion across European countries. The only main difference is in how these drugs were approved. Most of the drugs in our main sample were approved by the EMA using the centralized procedure. However, a majority of the drugs in the dynamic sample chose the Mutual Recognition Procedure.

2.3.2 Launch Delays in the EEA

This section documents the extent and variation of launch delays across European countries. We uncover two main patterns. First, virtually all countries experience meaningful launch delays on average. Second, there is substantial heterogeneity in average delays across countries. Some of this heterogeneity is explained by income: low-income Eastern European countries experience on average 2 additional years of delay relative to the rest of Europe. However, we also find significant variation in average delays within high- and middle-income countries.

Figure 2.2 reports delays by country using the full unbalanced sample of 481 main products.[24] The graph plots average launch delay in months, conditional on entry and calculated starting from the approval date of each product. The average delay ranges from as little as 3 months in the Netherlands, to as much as 3 and a half years in Romania. We distinguish between short- and long-term delays to highlight the different nature of delays in low-income countries (we consider delays of less than 4 years as "short-term"). Delays

[23] Since patent expiration dates can vary slightly across countries, we set period T as the latest expiration date among those of France, Italy, and Spain. Patents generally expire roughly at the same time in most countries, since they are administered by the European Patent Office. However, some countries can choose to grant extensions to individual patents. In our data we also occasionally observe earlier than expected patent expiration dates for some Eastern European countries. We choose these three countries because they are the three largest markets that use ERP. Therefore, when their patent protection expires, the strategic incentives to delay launches should become close to zero. Empirically, we observe only a total of 11 country launches occurring after period T, which suggests that our approximation is accurate.

[24] Note that this means that the sample is truncated. For a similar look at delays in a balanced (but smaller) sample, see Appendix B.4.

in high- and middle-income countries are almost exclusively short-term in nature, and potentially explained by slowdowns in the bureaucratic process, or other idiosyncratic factors outside the control of the firm. Delays in low-income countries however, are very likely to be long-term in nature, which is more consistent with an explanation based on strategic delays: Eastern European countries cannot afford to pay high-enough prices, so firms deliberately withhold their products until later.

2.3.3 Price Trends Over Time and Across Countries

This section examines the relationship between prices and delays, and describes price variation across countries and over time. We document a strong inverse relationship between price level and delays, and show that prices in countries that use ERP fall over time relative to prices in countries that do not use ERP.

We calculate the price level in each country using a fixed effect regression, and plot the coefficients against average delays (Figure 2.3).[25] There are three main takeaways from this figure. First, price levels and delays are negatively correlated, as we would expect if firms delayed entry in countries that cannot afford to pay high prices. Second, countries with a large population tend to experience shorter delays relative to the benchmark set by the best-fit line, while countries with low population tend to experience longer delays. This is also consistent with an explanation for delays that is based on reference pricing: firms will be more willing to surrender a low price in large countries since the larger market size will make up for the potential loss from the externality on price. Third, the effect of market size is asymmetric: low-income countries are consistently penalized for market size, but small high-income countries are equally likely to fall above or below the best-fit line. This is once again consistent with a strategic reaction to ERP: small market countries where prices are

[25] Average delay comes from Figure 2.2. Price level variable is calculated as a country-specific fixed effect γ_j in a regression of log price on various fixed effects:

$$\ln\left(p_{ijt}\right) = \theta_i + \gamma_j + \delta_t + \varepsilon_{ijt}$$

In the regression, i indexes drugs, j indexes countries, and t indexes years.

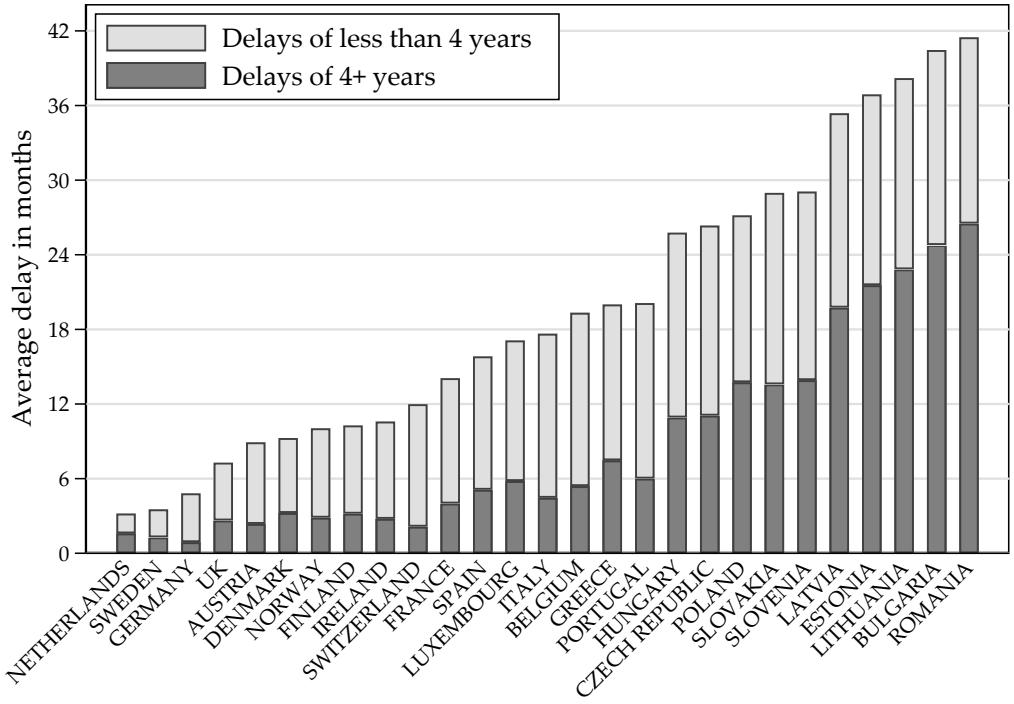


Figure 2.2: *Average delay from marketing approval by country*

This graph shows the average launch delay in months, from the approval date, conditional on entry, for all European countries. Dark bands represent the overall impact of long-term delays (defined as delays of 4 years or longer). We calculate average delays using all 481 drugs in our main sample. See Table 2.1 for summary statistics of this group of drugs, and section 2.3.1 for a description of how this sample was created.

high do not generate a negative effect on prices, and therefore should not be penalized.[26]

In Figure 2.4 we look at the price of drugs over their life-cycle. To do so, we run the following regression:

$$\ln\left(p_{ijt}\right) = \theta_i + \gamma_{ja} + \delta_t + \varepsilon_{ijt}$$

where γ_{ja} is a fixed effect for country and drug age, measured in years starting from the approval year. We also include drug fixed effects θ_i and year fixed effects δ_t.

[26]Once again, Netherlands is an outlier on this graph. The price level in the Netherlands is probably underestimated for three reasons. First, the data is missing hospital drugs, which tend to be more expensive. Second, oftentimes in our data the same product is more expensive when sold through the hospital channel, which means that if we only observe retail sales we will underestimate the price of a product that is also sold in hospitals. Third, the fact that the sample starts in 2007 means that most of the years occur during the Great Recession, when many countries cut prices.

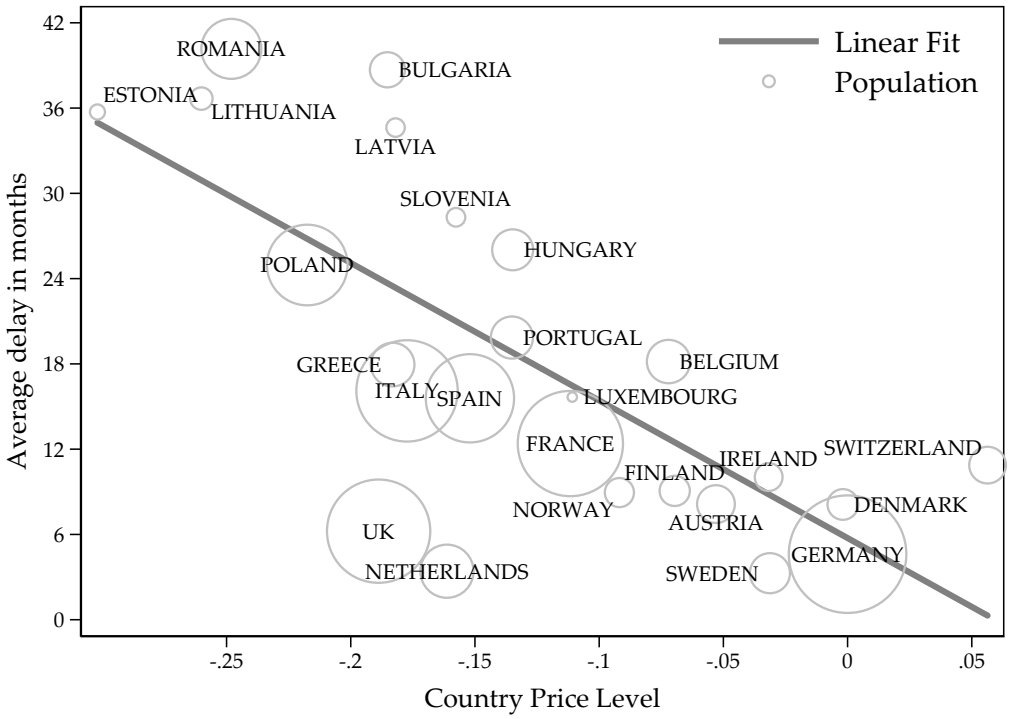


Figure 2.3: *Correlation between Price Level and Average Delays by Country*

This figure plots the relationship between price level average delay in months. Price level is calculated as the country fixed effect from a regression of log price on drug, year, and country fixed effects. The coefficient for Germany was normalized to 0 in this regression. Average delays are the same as in Figure 2.2. The plot is weighted by population (we use data from 2012).

We compare the evolution of drug prices across two groups of similar countries, some of which do not use ERP. The first is the group of the four largest countries in the Eurozone: France, Germany, Italy, and Spain. The second is the block of Scandinavian countries: Denmark, Finland, Norway, and Sweden. Both groups contain countries that do not use ERP (Germany in the first group; Denmark and Sweden in the second). We are not interested in the price level differences among these countries, but rather in whether price trends are divergent. To check whether that is the case we plot the difference between the γ_{ja} coefficients for Germany and Denmark and those of the other countries in the respective groups. In both cases we see that relative prices fall over time in countries that use ERP relative to the benchmark provided by countries that do not use ERP.[27] This is consistent with the additional downward pressure that we would expect to see through the external reference pricing channel: as the product is launched in more countries, prices fall wherever ERP is used relative to countries where ERP is not used.

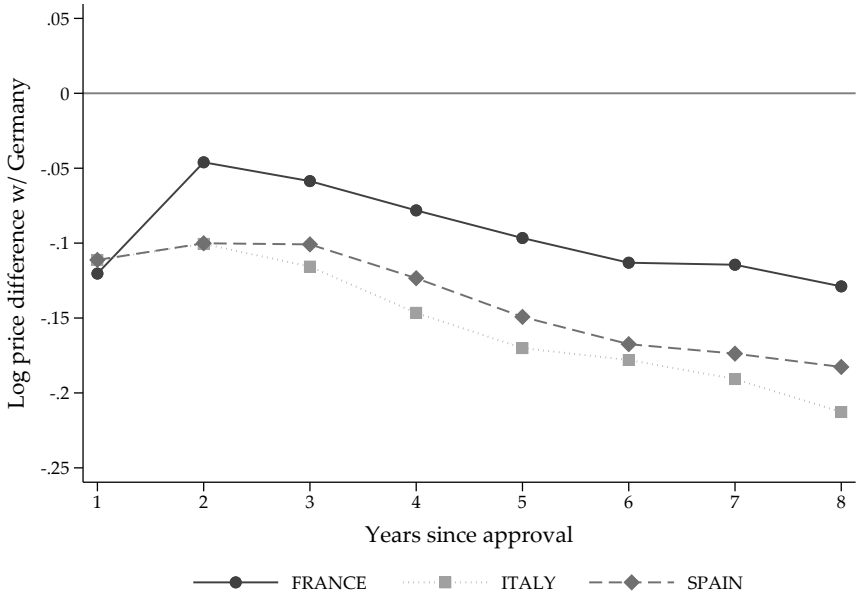
2.4 A Dynamic Model of Entry with Externalities in Price

A pharmaceutical firm l owns a set $\mathcal{I}_l$ of patent-protected molecules (indexed by i) with a marketing authorization for sale in a finite set $\mathcal{N}_i \subseteq \mathcal{N} = \{1, \ldots, N\}$ of markets (European countries), indexed using the subscript j.[28] The patent on each molecule has an expiration date, T_i periods into the future, at which point generic alternatives are allowed to enter and profits are driven to zero.[29] The firm's objective is to maximize profits over the life-cycle of

[27]The only potentially unexpected pattern is that prices fall in Sweden after about 5 years, almost to the level of Finland, even though Sweden does not use External Reference Pricing. However, prices can fall over time for many reason, and countries can use many other tools outside of Reference Pricing to reduce spending.

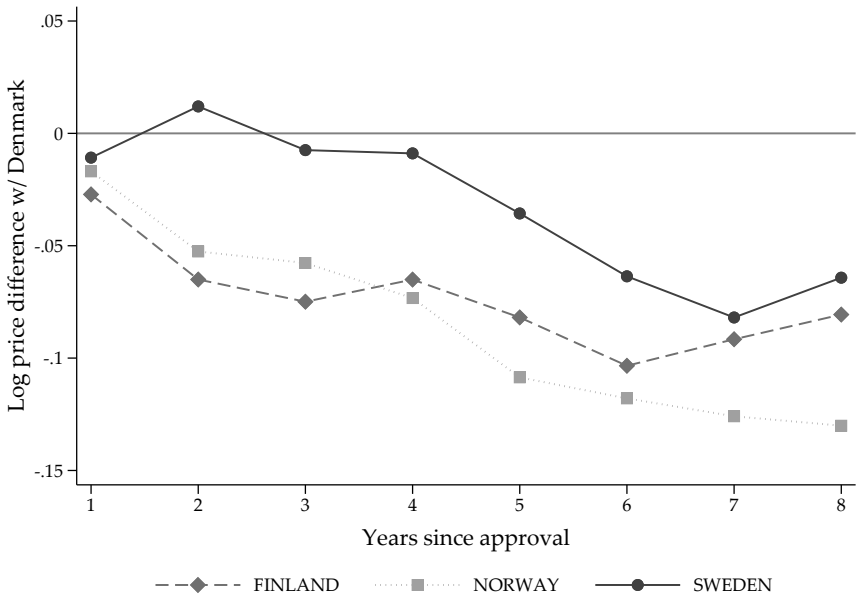
[28]Even though in theory all drugs can easily be approved for all countries in the EEA, we allow for the possibility that the drug might not be able to enter in all countries. In some occasions, governments can ban drugs if they are concerned about side effects. Hence, we only assume that a drug can enter in a country if we observe sales in the data.

[29]This assumption can be relaxed in many ways without significantly altering the model. The key result that has to hold is that there are no more strategic incentives to delay once the product has lost patent protection. This is almost certainly the case because once a product loses patent protection, governments can rely on much more effective price-cutting measures, and no longer need to resort to external reference pricing.



(a) *Prices in France, Germany, Italy, and Spain*

This figure plots average price trends over the life-cycle of a drug in France, Italy, and Spain, relative to Germany. We use a specification that controls for drug fixed effects. For details, please consult section 2.3.3.



(b) *Prices in Denmark, Finland, Norway, and Sweden*

This figure plots average price trends over the life-cycle of a drug in Finland, Norway, and Sweden, relative to Denmark. We use a specification that controls for drug fixed effects. For details, please consult section 2.3.3.

Figure 2.4: *Change in prices over time and across countries*

their products. We denote the last period of the firm as $T_l = \max_{i \in \mathcal{I}_l} T_i$.

In each period, the firm is solving a two-part problem:

1. In what countries should the products be launched?

2. What prices should be set in each country?

We are interested in understanding strategic launch delays, which are the outcome of the first part of the problem. Of course, the optimal launch strategy will depend on the equilibrium prices that are set in each country. Firms, however, have limited agency in determining these prices, because in Europe drug spending is subject to strict government regulation. Therefore, we do not explicitly model the price-setting stage, but instead use a flexible parametric function to predict equilibrium prices.[30]

Start by introducing some notation. Denote the *launch sequence* of firm l as $S_l = \left\{s_{ij}\right\}_{j \in \mathcal{N}_i, i \in \mathcal{I}_l}$, where s_{ij} denotes the period of entry of product i in country j. Furthermore, we denote the launch sequence at the end of a given period t as $S_{lt} = \left\{s_{ijt}\right\}_{j \in \mathcal{N}_i, i \in \mathcal{I}_l}$ where

$$s_{ijt} = \begin{cases} s_{ij} & \text{if } s_{ij} \leq t \\ 0 & \text{otherwise} \end{cases}$$

Once a product has entered, we assume that it cannot be voluntarily withdrawn by the firm.[31] Notice that under these assumptions, for each value of S_{lt} there is exactly one possible path of previous realizations of the state variable that leads to S_{lt}. This implies that knowing S_{lt} is enough to know $S_{l\tau}$ for all $\tau < t$. We similarly denote the launch sequence of other firms as S_{-l}. Occasionally, we will also use the shorthand S or S_t to indicate the launch sequences of all firms.

[30] We remain agnostic with respect to possible micro-foundations of this function, which do not matter for the results of this paper. We show in Appendix B.1.3 that the equation can be derived from a static Nash Bargaining model.

[31] Empirically, we only observe products being withdrawn because the EMA has decided to revoke the marketing authorization upon reviewing post-clinical evidence, or after demand falls for several periods, suggesting that the product is no longer economically or therapeutically viable. In both cases we assume that the choice is not taken by the firm.

The model considers the possibility that the price of a product in a given market may affect the price of that same product in other markets through the reference pricing channel. When the firm launches a product in multiple countries, governments can observe prices abroad and calculate a reference pricing benchmark that affects their own price. The extent to which this happens is captured by a country-specific parameter in the price control function.

The optimal launch strategy will also depend on demand. The demand system and the price-setting equation are primitives that the firm takes as given when making entry decisions. We describe each in turn before specifying the dynamic entry model.

2.4.1 Demand System

We base demand on the logit random utility model.[32] Markets are defined by a country, year, and therapeutic class.[33] We aggregate products within a therapeutic class at the molecule-brand status level. We define three possible brand statuses: originator product (i.e. the brand sold by the patent-holder or main manufacturer), non-originator brand (usually a parallel traded product), and generic. All the products in the main sample of analysis are originator brands. The utility of consumer ℓ, in country j, from consuming drug i (molecule m), belonging to therapeutic class κ, in year t is given by

$$u_{i(m,\kappa)\ell(j)t} = \delta_{ijt} + \nu_{i\ell t} \tag{2.1}$$

To obtain more realistic substitution patterns we also add a nesting structure at the molecule level. The error term $\nu_{i\ell t}$ is parametrized as

$$\nu_{i\ell t} = \left(\zeta_{m,\kappa} + (1-\sigma_\kappa)\,\varepsilon_{i\ell t}\right)$$

[32] Variations on the logit model are commonly used to describe the pharmaceutical market. Duso *et al.* (2014) and Stern (1996) use a two-level nested logit model to model demand for oral anti-diabetics and a set of four therapeutic classes (gout, sedatives, minor tranquilizers, and oral anti-diabetics) respectively. Dunn (2012) uses a random-effect logit model (micro-BLP) to describe the market for anti-cholesterol drugs using individual data. In our case, we found that adding a more sophisticated nesting structure did not substantially improve model fit (we experimented with adding an upper level at the ATC4 level or at the market level, separating the outside option from all other products). We did not have enough data to implement a random-effect logit model.

[33] For details on the definition and construction of therapeutic classes, see Appendix B.2.

where m indicates the molecule of drug i, σ_κ lies on the unit interval, $\varepsilon_{i\ell t}$ is distributed according to a standard Extreme Value Type 1 distribution and $\zeta_{m,\kappa}$ is an error term whose distribution satisfies the property that $\nu_{i\ell t}$ is distributed according to an Extreme Value Type 1 distribution as long as $\varepsilon_{i\ell t}$ is also EV1 (Cardell, 1997).[34]

We parametrize δ_{ijt} to incorporate two important empirical features of drug demand in our data: heterogeneity in preferences across countries, and growing demand over time. Demand for drugs can differ substantially across countries because of heterogeneity in prescribing guidelines, incidence of disease, and patient preferences. Moreover, possibly because drugs are generally considered experience goods (Crawford and Shum, 2005), most products experience increasing demand over their life-cycle as both patients and providers learn about new therapies.

We do not include a coefficient for price, since we do not observe the price that patients pay. Instead, we include realized demand as a control in the price function, implicitly assuming that any relationship between price and volume sold is mediated by the government. In general, patients in European countries only pay a fraction of the cost of prescription drugs, so any degree of price elasticity that is picked up in the data is likely driven by the government.[35]

Our specification for δ_{ijt} is

$$\delta_{ijt} = \alpha_{ij} + \beta_i \text{age}_{it} + \eta_i NF_{ijt} + \xi_{ijt} \tag{2.2}$$

α_{ij} captures a country-specific preference for each drug, which could reflect differences in

[34]For the nested logit model to make sense, σ_κ must lie on the unit interval. We do not implement this restriction in the estimation, but instead opt to abandon the nesting structure in favor of a simple logit whenever the parameter falls outside the limits set by the theory.

[35]All European governments provide universal health insurance coverage. Cost-sharing for drugs tends to be very low. Drugs administered in an inpatient setting are usually completely free. Cost-sharing of outpatient prescription drugs is disciplined by a variety of regulations that weaken the relationship between the price paid by the government and that paid by the patient. In a few countries outpatient drugs are also completely free (Netherlands, Scotland). Some countries use copays that are common to all drugs, regardless of price (Austria, Germany, Ireland, Italy, UK). Others have coinsurances but relatively low caps on the amount that each patient can spend each year/month (Belgium, Finland, Spain, Sweden). Finally, countries that have coinsurances without caps either have low coinsurances for the most valuable products (France, Greece, Poland), or allow for a variety of exemptions to protect sick and low-income individuals (Denmark, Portugal) (Barnieh *et al.*, 2014; Panteli *et al.*, 2016; Thomson and Mossialos, 2010).

prescribing guidelines or disease burden.[36] The β_i coefficient allows for the possibility that patients and physicians might learn about new drugs over time. We measure age starting with the approval date of the drug.[37] For non-originator products we also keep track of the number of selling firms as a separate control variable NF_{ijt}.[38] Finally, we add a drug-country-year random shock, ξ_{ijt}, so that the model can fit the data.

Inverting market shares (and normalizing the utility of the outside option to 0) yields the standard estimating equation for nested logit models:

$$\ln\left(\frac{s_{ijt}}{s_{0jt}}\right) = \alpha_{ij} + \beta_i \text{age}_{it} + \eta_i NF_{ijt} + \sigma_m \ln\left(\frac{s_{ijt}}{s_{mjt}}\right) + \xi_{ijt} \tag{2.3}$$

where s_{0jt} is the share of the outside good, and s_{ijt} and s_{mjt} are the market share of the product, and the overall market share of the molecule nest respectively.[39]

We denote the demand function generated by this model as $D_{ijt}\left(S_t, \xi_{jt}^{\kappa}\right)$, where $\xi_{jt}^{\kappa} = \left\{\xi_{ijt}\right\}_{i\in\kappa}$ is the vector of shocks for all products in therapeutic class κ.

2.4.2 Price-Setting Equation

Drug prices are set in negotiations between firms and governments. The exact form of the these negotiations is hard to capture in an explicit model. Presumably, the government is trying to reconcile several goals, such as providing access to valuable medications and rewarding costly innovation, while at the same time facing spending constraints. Since we do not have any information on the government's objective function we opt for a more agnostic

[36] Even though the impact of disease burden is mostly reflected through market size, therapeutic classes can sometimes encompass large clusters of disease. For example, oncologics are divided in three large therapeutic classes (targeted therapies, cytotoxics, and hormonal therapies).

[37] An alternative definition of age could start counting from the year in which a product was launched in a specific country. Our prior is that information about a drug can spread to countries even if that product is not yet available. As a robustness check, we estimated the model using country-specific age, obtaining very similar results.

[38] Virtually all originator products are sold by a single firm in each country, though the firm is not necessarily the same across countries. However, most molecules face multiple brand and generic competitors, which we aggregate in order to avoid excessive entry and exit.

[39] See Appendix B.1.1 for the derivation. We think of the outside good as an aggregate of non-drug therapies (doctor visits, surgery, etc.) or drugs in other classes.

approach, and model prices using a flexible control function.

Our price-setting equation includes two components. The first component is what we call *government price*, p_{ijt}^{gov}. This is the price that is agreed upon between the firm and the government in the absence of reference pricing. We write the government price as a function of product and country fixed effects, as well as three additional control variables that are meant to capture the potential effect of some of the other price-control policies implemented by the government. As a first control we include an indicator for whether the firm has headquarters in country j. Kyle (2006) shows that this variable is important to determine probability of launch; we include it to check whether we can detect a significant effect on price. As a second control we include a flexible function of the number of other molecules available in the same market. There are several reasons why this variable should matter, all of which suggest it should have negative sign: the availability of alternatives should decrease the additional welfare generated by a drug, competitive pressure could bring prices down, and finally, governments sometimes benchmark prices to the lowest price available within a group of substitutable drugs. As a third control we include total realized demand for the drug. Intuitively, the government might react to high prices by steering patients towards cheaper options. Governments also make widespread use of payback policies aimed at preventing budget overshooting (Carone *et al.*, 2012). Demand for drugs has an idiosyncratic component (in the real world as in the model), but prices have to be agreed upon in advance, which can lead to costly unexpected spending. Hence, we conjecture that governments may react by establishing price volume schedules that include downward adjustments to prices for drugs with higher demand.

The specification of the government price is

$$p_{ijt}^{\text{gov}}\left(D_{ijt}\left(S_t, \xi_{jt}^{\kappa}\right)\right) = \theta_i \cdot \gamma_j \cdot \exp\left(\beta_Z Z_{ijt} + \beta_D \ln\left(D_{ijt}\left(S_t, \xi_{jt}^{\kappa}\right)\right)\right) \tag{2.4}$$

where θ_i and γ_j are the product and country fixed effects, Z_{ijt} is the matrix of controls, and $D_{ijt}\left(S_t, \xi_{jt}^{\kappa}\right)$ the realized demand, which depends on the random shocks of the products in class κ: $\xi_{jt}^{\kappa} = \left\{\xi_{ijt}\right\}_{i \in \kappa}$. We interpret this equation as a price-schedule, rather than a set list

price.

The second component of the price-setting equation is the *reference price*. The reference price is not directly observed, but reference price functions F_{jt}^{ref} and baskets R_{jt} are reported by various sources.[40] Nonetheless, some details regarding the implementation of these functions require additional assumptions. In particular, we need to establish how soon governments can see the price schedules that have been set in other countries, and what is the right volume at which to evaluate the price schedules when calculating the reference. We assume that governments see prices with a 1-period lag, and apply the volume from the home country in the current period to the price schedules used in the reference formula.[41] Therefore, the reference pricing function that we implement empirically is given by

$$p_{ijt}^{\text{ref}}\left(S_t, D_{ijt}\left(\cdot\right)\right) = F_{jt}^{\text{ref}}\left(\left\{p_{ikt-1}\left(S_t, D_{ijt}\left(\cdot\right)\right)\right\}_{k\in\left(R_{jt}\cap E_{it-1}\right)}\right) \tag{2.5}$$

where E_{it-1} is the set of countries where product i is sold as of time $t-1$ (this set can be obtained from information in S_t).

To combine these two components we assume that whenever the governments observes a reference price that is inferior to the government price, the equilibrium price is set as a weighted average of the government price and the reference price. We let the weight be country-specific in order to capture eventual heterogeneity in the application of reference

[40] Figure 2.1 shows the reference baskets and prices for 2012.

[41] For the volume adjustment, there are two natural options: the price schedules could be evaluated at the volume sold in the referenced country, or in the home country in the current period. The first option implies that each country references the *observed list prices* from the previous period. The second option is to evaluate the price schedules at the volume sold in the home country in the current period. This option implies that each country references the *price schedule*. We experiment with both options and find the second one to be more accurate. It provides a better fit with the data, and is more aligned with the way reference pricing is applied in practice according to informal conversations with industry insiders.

pricing guidelines. The overall price-setting equation is then given by

$$p_{ijt}\left(S_t, D_{ijt}\left(\cdot\right)\right) = \begin{cases} p_{ijt}^{\text{gov}}\left(S_t, D_{ijt}\left(\cdot\right)\right) & \text{if } p_{ijt}^{\text{ref}}\left(\cdot\right) \geq p_{ijt}^{\text{gov}}\left(\cdot\right) \\ \left(1-\mu_j\right) p_{ijt}^{\text{gov}}\left(S_t, D_{ijt}\left(\cdot\right)\right) + \mu_j p_{ijt}^{\text{ref}}\left(S_t, D_{ijt}\left(\cdot\right)\right) & \text{if } p_{ijt}^{\text{ref}}\left(\cdot\right) < p_{ijt}^{\text{gov}}\left(\cdot\right) \end{cases} \tag{2.6}$$

Identifying the Impact of Reference Pricing Before moving on to the dynamic model, we briefly discuss the variation in price that allows us to identify the impact of reference pricing. Our identification strategy is based on two key elements: the functions that each country uses to calculate the reference pricing benchmark, and the asymmetry in the application of reference pricing that comes from the assumption that reference price only matters if it is below the government price.

We illustrate how these assumptions are reflected in the data by looking at a specific example: the drug Abacavir (brand name Ziagen, marketed by ViiV Healthcare, a UK-based company). Abacavir is an important antiretroviral therapy used to treat HIV-positive patients. It was approved in July 1999, and lost patent protection in 2014. It is currently included in the WHO List of Essential Medications.

Figure 2.5 displays the order of entry observed in the data. The launch sequence is fairly typical: shortly after the approval date the product is launched in a large set of high- and middle-income countries. Poland, a low-income country with a large population, is part of this initial group as well.[42] Subsequent launches are scattered, but occasionally grouped together. Bulgaria and Czech Republic receive the product in 2003, about four years after the initial approval, followed by Lithuania and Romania in 2005, Hungary in 2006, and Slovakia in 2008. As of December 2012, the product was still not available in Latvia and Estonia.[43]

We look at price trends for Abacavir in various countries in Figure 2.6. We compare prices

[42]This is also not uncommon. Very often the product will be launched in a couple of low-income countries during the first wave of launches. Those countries tend to be the ones with the largest demand for the product.

[43]Two drugs that use Abacavir in combination with other molecules were launched in Latvia in 2004 and 2006. Estonia does not have any recorded sales of Abacavir or associated combination therapies in our data.

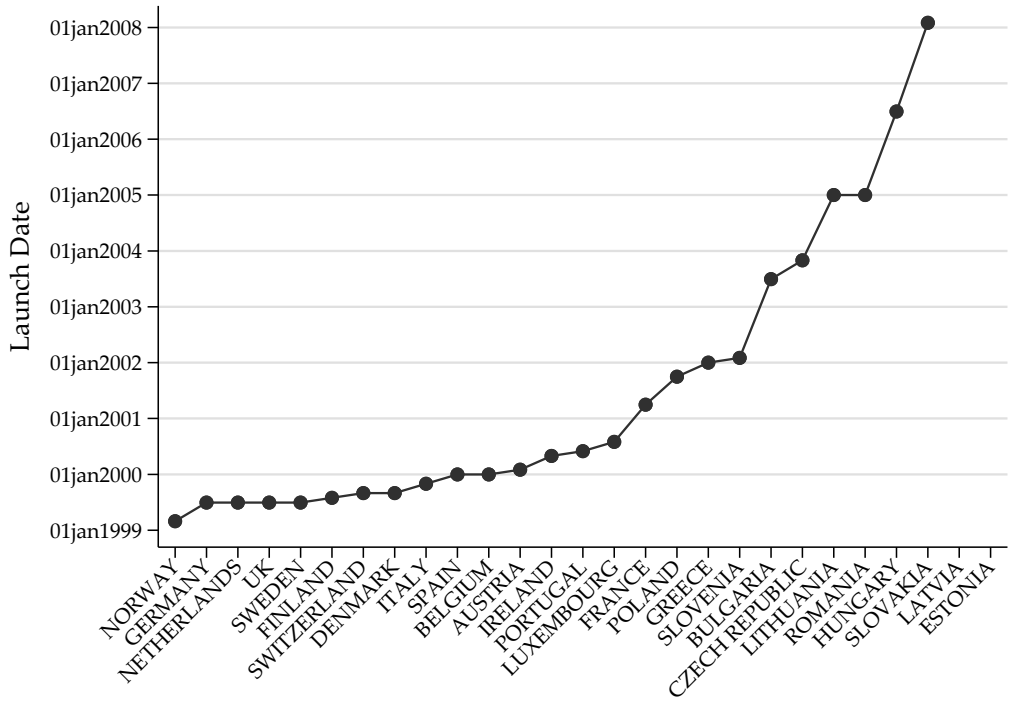
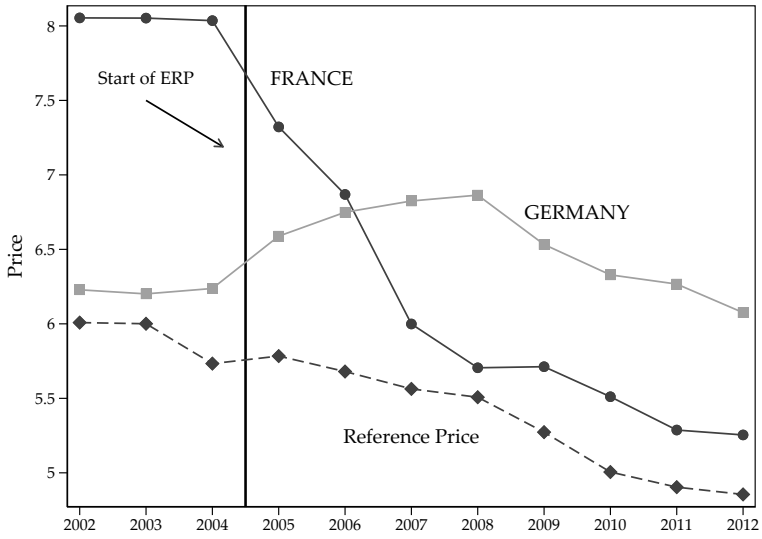


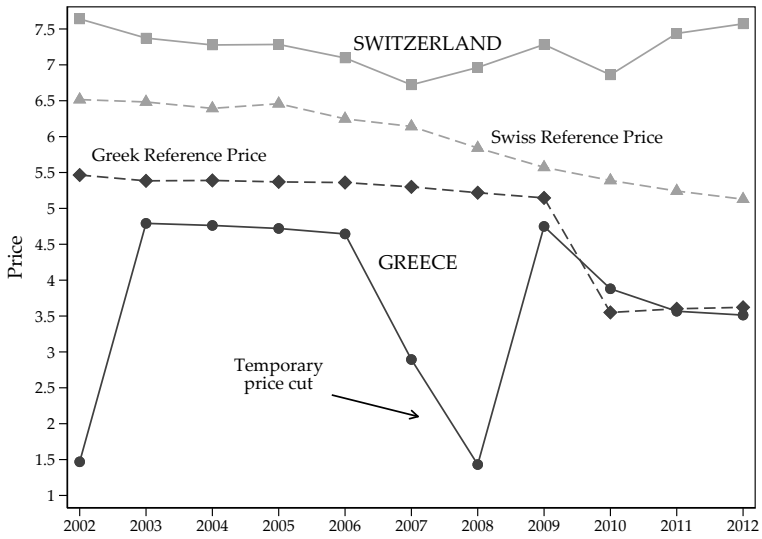
Figure 2.5: *Abacavir entry sequence*

Note: This figure plots the entry sequence of the drug for Abacavir (brand name Ziagen, developed and sold by ViiV Healthcare), an antiretroviral therapy used to treat HIV-positive patients. The drug was originally approved in July 1999, using the centralized procedure administered by the EMA. The drug was not available in Latvia and Estonia as of the fourth quarter of 2012, which is the last quarter of available data.



(a) *France and Germany*

This figure plots the yearly prices and reference prices for Abacavir in France and Germany (Germany does not use reference pricing). France applies ERP only starting 5 years after the product is launched, which in this case happened in early 2001, meaning that ERP would start first in 2005.



(b) *Greece and Switzerland*

This figure plots the yearly prices and reference prices for Abacavir in Switzerland and Greece. There is some noise in the series because Greece implemented price cuts at various points after the financial crisis of 2007-2008. The price charged in 2002 is also presumably an introductory price, potentially generated by noise or measurement error in the data.

Figure 2.6: *Abacavir Price Trends*

in France and Germany first (2.6a). ERP rules in France state that reference pricing should be applied at launch, and then updated on a yearly basis, but only starting five years after the original launch. This discontinuity allows us to observe price differences in a quasi-event study setting. The price in Germany is fairly constant over time. The price in France is also constant for a couple of years, but quickly falls to the level of the reference price between 2005 and 2007 (the product was launched in the first months of 2001, which means that the price was probably set in late 2000, and the reference price would first be updated at some point in 2005).

To compare two countries that use ERP policies, but appear to put different emphasis on them, we look at Greece and Switzerland (2.6b). The graph shows that the Swiss price is above the reference price benchmark, so our model will predict that reference price should have an effect on the price. However, there is no apparent correlation between the two lines. In fact the Swiss price even seems to diverge from its reference price after 2007. In contrast, Greece comes in below the reference price until 2009. Hence, the model would interpret prices between 2002 and 2009 as measures of the government price. Those prices are constant, with the exception of the first observed price in 2002, and a brief interlude of approximately 6 quarters between 2007 and 2008 in which the price was momentarily cut. However, in 2010, when the reference price falls, so does the Greek price, suggesting that reference pricing guidelines are followed closely, at least in this case.

These example illustrates how the model determines the impact of reference pricing. The model will predict a government price for Abacavir for each country to match instances when the reference price does not bind. When the reference binds, the degree to which prices move in the same direction as the reference price will determine whether μ_j is close to 1 (as in the case of Greece), close to 0 (as in the case of Switzerland), or somewhere in between (as in the case of France).

2.4.3 Dynamics

We now turn to the dynamic aspect of the model. The firm operates in a single-agent, discrete-time, finite-horizon environment. Its goal is to maximize profits by choosing the order and timing of entry in each country, conditional on demand and price primitives (over which it has perfect information).[44]

Firms face stochastic shocks in their launch sequence in the form of random entry delays. Formally, a delay shock is a binary Bernoulli random variable ρ_{ijt}, with country-specific parameters ψ_j, independently distributed across countries, years, and drugs. If $\rho_{ijt} = 1$, then drug i cannot enter in country j until period $t + 1$ (when a new shock is drawn). These shocks help capture variation in delays that cannot be explained through the reference pricing channel. Delays caused by ERP will arise when firms voluntarily decide to withhold one of their products because they expect that doing so will result in higher overall revenues. Launch delays that occur for any other reason will be soaked up by delay shocks.[45]

Each period unfolds as follows. At the beginning, the firm chooses a set of countries where it will send entry applications. We represent this action as a binary vector $A_{lt} = \left\{a_{ijt}\right\}_{j\in\mathcal{N}_i, i\in\mathcal{I}_l}$ where $a_{ijt} = 1$ whenever the firm chooses to launch product i in country j. More generally, we denote *a strategy* for firm l in the extended-form game as a map

$$\mathcal{A}_{lt} : \mathcal{S} \to \{0,1\}^{T_l - t + 1}$$

where $\mathcal{S}$ is the set of all possible values of the launch sequence S_l, and $\mathcal{A}_{lt}$ generates a set of binary vectors A_{lt} with an action profile for each period until T_l.[46] The launch sequence of the firm as of period t represents the state variable of the problem. While the launch sequence of other firms has the potential to affect expected revenue (by stealing market share and

[44] Notice that revenue and profits are equivalent, since the model assumes that all costs are zero.

[45] The main source of delays other than ERP is probably the time required to review price and reimbursement applications and to complete price negotiations. However, other sources may exist as well: firms may need to wait for certain negotiations to be resolved before engaging in additional launching because of capacity constraints to their negotiating workforce; or countries may block the entry of products they consider potentially dangerous for a number of years.

[46] We call this a game because we treat Nature as a player whose actions (i.e. the delay shocks) are stochastic.

potentially affecting prices), we assume that strategies are not conditional on the actions of other firms.[47] Afterward, the vector of binary shocks for the current period $\{\rho_{ijt}\}$ is realized, determining the new value of the state variable. After observing the shocks, governments set price schedules, and finally, products are sold and profits are realized.

The firm's problem at time t is to pick a strategy to maximize

$$V_t\left(S_{lt-1}, S_{-lt-1}\right) = \max_{\mathcal{A}_{lt}=\{A_{l\tau}(S_{l\tau-1})\}_{\tau=t}^{T}} \sum_{S_l}\left(\sum_{S_{-l\tau}}\left(\sum_{\tau=t}^{T}\beta^{\tau-t}\Pi_\tau\left(S_l, S_{-l}\right)\right)\cdot\mathcal{P}\left(S_{-l}\middle|\, S_{-lt-1}, \mathcal{A}_{-lt}\right)\right)\cdot\mathcal{P}\left(S_l\middle|\, S_{lt-1}, \mathcal{A}_{lt}\right) \tag{2.7}$$

where β is the discount factor (which we set equal to 0.98), $\Pi_\tau\left(S_l, S_{-l}\right)$ is the expected period profit of the firm, for a given realization of the launch sequence (both the own sequence and the sequence of competitors), and $\mathcal{P}\left(S_l\middle|\, S_{lt-1}, \mathcal{A}_{lt}\right)$ and $\mathcal{P}\left(S_{-l}\middle|\, S_{-lt-1}, \mathcal{A}_{-lt}\right)$ are the probabilities of S_l and S_{-l} conditional on S_l and S_{lt-1}, for given strategies $\mathcal{A}_{lt}$ and $\mathcal{A}_{-lt}$ of the firm and its competitors.

The expected period payoff is defined as

$$\Pi_\tau\left(S_l, S_{-l}\right) = \sum_{i\in\mathcal{I}_l}\mathbb{E}_{\zeta_{jt}^{xr}}\left[\sum_{j\in S_{l\tau}} p_{ij\tau}\left(S_t, D_{ijt}\left(\cdot\right)\right) D_{ij\tau}\left(\cdot\right)\right] \tag{2.8}$$

The expectation is taken over the possible realizations of the stochastic error in the demand system.

2.5 Estimates of Demand and Price Primitives

In this section we describe how we carry out the estimation of demand and price primitives, and briefly discuss results and model fit. Demand estimates are in 2.5.1, while price estimates are in 2.5.2. Finally, in 2.5.3 we use these estimates to simulate and compare the expected revenue of entry sequences where products are launched everywhere right away to sequences

[47] This assumption is necessary to obtain identification from the model. Allowing firms to react to each other is desirable, but makes counterfactual scenarios hard to compute, since our data does not allow us to make inferences about reactions off the equilibrium paths.

that delay launches in certain countries. We find that firms earn higher revenue when their products are delayed in Eastern European countries, but find no evidence that delays in Western Europe lead to higher revenue.

2.5.1 Demand Estimation

Equation 2.3 provides the basis for estimation. The independent variables can be easily constructed from the IMS data, with the exception of age, for which we use approval date from the EMA or the Heads of Medicines Agencies.[48] To measure market size we use data from the Global Burden of Disease Study. We use a map from ATC4 to GBD indication constructed by Costinot *et al.* (2016) to calculate the number of patients that might potentially use drugs in a certain therapeutic class.[49] We then scale up the number of patients to obtain a number of standard units and construct market shares from data on sales volumes.[50]

Two potential issues arise. First, $\ln\left(\frac{s_{ijt}}{s_{mjt}}\right)$ (i.e. the within-molecule market share of product i) is correlated with the error term ξ_{ijt}, so instruments are needed in order to recover a consistent value of the associated coefficients. We use the number of years since the patent on molecule m expired, the average market share of parallel traded products for other molecules within the same country, and the number of other firms selling the same molecule.[51] These variables are meant to capture the impact of the two potential channels of within-molecule competition: generic products (for off-patent molecules), and parallel traders (for on-patent products).

The second issue is that firms might be able to observe ξ_{ijt} prior to entry. This leads to

[48] We manage to match over 90% of molecules in our main sample to an approval date. In the handful of cases for which we do not have a match, we use the fifth-earliest launch date from the IMS MIDAS database as a proxy for the approval date. In the sample we were able to match, the fifth-earliest launch occurs after the approval date in 95% of the cases.

[49] We thank the authors of the paper for sharing the map with us ahead of publication.

[50] The scaling number is chosen to be the smallest number such that the outside option has at least 1% market share in all countries and years. In that sense, our estimate can be thought of as a lower bound on the actual market size. We pick a different scaling number for each therapeutic class.

[51] We identify as parallel traders all non-originator firms who sell a product sharing the same molecule and product name as the original product.

the classic selection problem that is common to many IO settings: countries where entry is recorded would have unobservably high values of ξ_{ijt}, leading to a biased estimator. In practice however, there are several attenuating circumstances that suggest selection is a second-order concern in this case. First, firms never exit voluntarily, so we do not need to worry about exit selection. Second, since we control for drug-country-specific preferences, our model will pick up the average preference of each country. The only remaining possibility is that demand prior to entry could be lower than our model prediction, because firms wait until demand reaches a certain threshold before entering. If that were the case, we would expect our α_{ij} coefficient to be biased upwards. In the dynamic estimation, a higher α_{ij} coefficient makes firm i more likely to enter in country j. Therefore, this type of selection bias would lead us to underestimate the amount of strategic delays.[52] We conclude that if unobserved selection does exist, it should act in the direction of preventing us from finding a result.

Results We estimate a separate equation for each of the 109 therapeutic classes that contain at least one of the products in our main sample using linear regressions with instruments.

The coefficients of the model itself are not very informative, so we discuss how the model fits the data. Figure 2.7 plots the distribution of R^2 coefficients from each regression. Virtually all regressions achieve a coefficient of determination of at least 0.8. This is because demand for all drugs tends to be fairly well-behaved, without many fluctuations. The main exception is flu medications ($R^2 = 0.6$). Since the flu is a highly seasonal illness, with stronger strains in specific years, demand for these products tends to vary a lot year-to-year in a way that is hard for our model to capture.[53]

Figure 2.8 plots predicted shares against actual shares for products in the main sample.

[52] If potential demand were higher before entry, that might lead us to overestimate delays due to strategic behavior. However, this type of selection runs counter to the main intuition for the selection mechanism in this case.

[53] The other outlier in the graph is a class called "N7D9|OTHER ANTI-ALZHEIMERS" which contains a single molecule, Memantine (brand name Namenda), which is used to treat Alzheimer's disease in people who are intolerant to cholinesterase inhibitors.

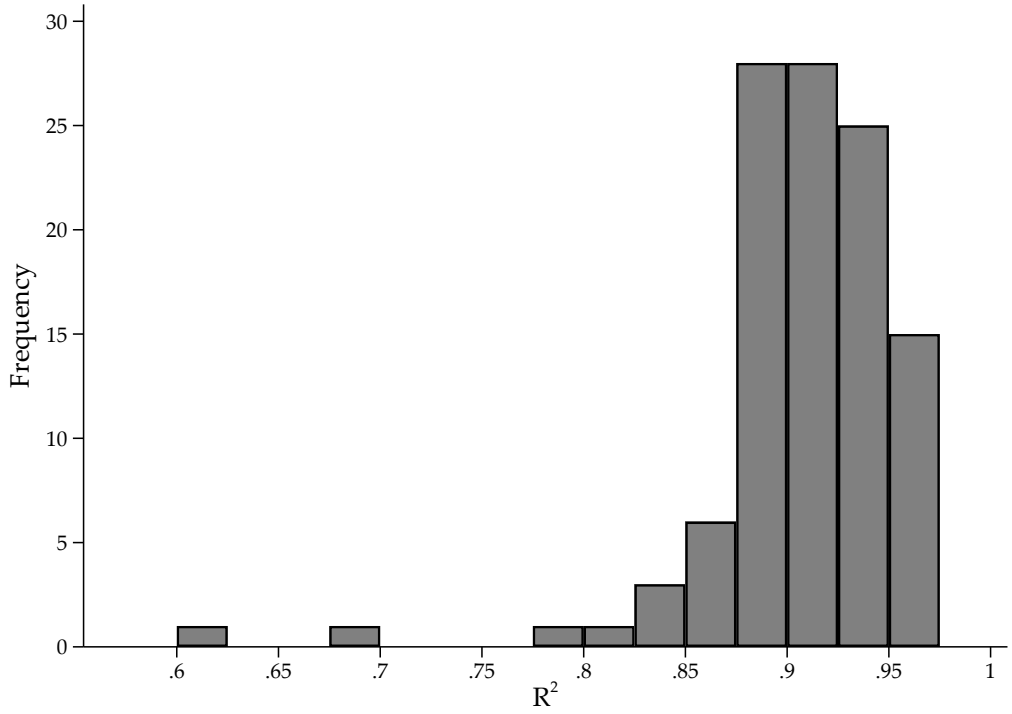


Figure 2.7: *R^2 coefficients from demand regressions*

This figure plots the R^2 coefficients from the demand regressions. We estimate 109 regressions, one for each therapeutic class that contains at least one of the products in our main sample.

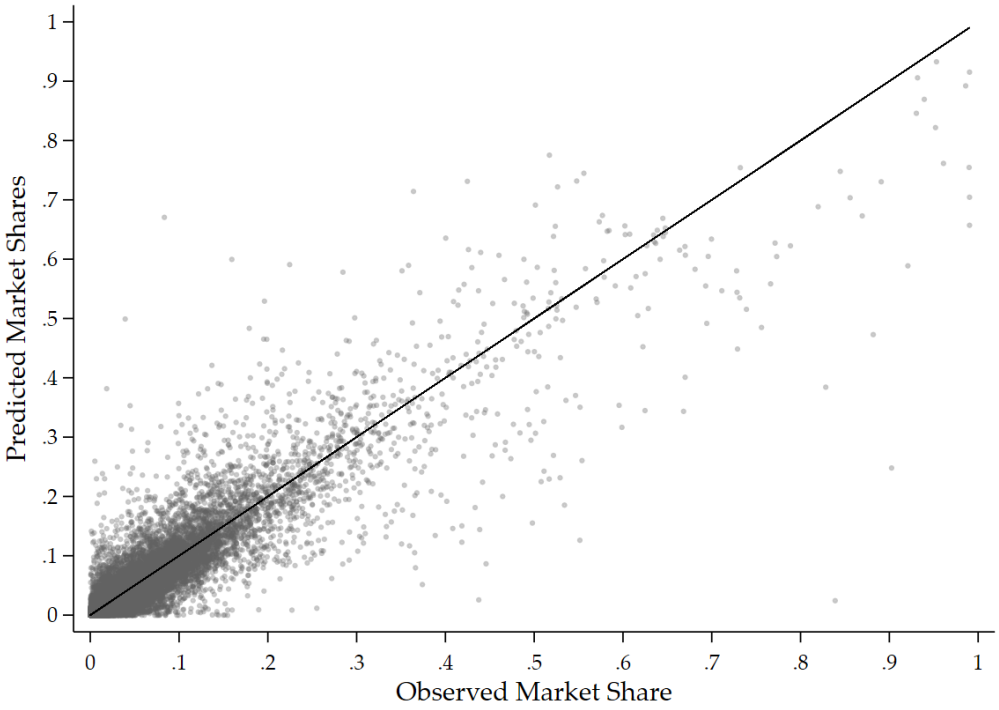


Figure 2.8: *Demand Model Fit*

This figure plots the observed market share of main sample products, against the market share we predict using our demand model.

The predicted shares of dominant products (i.e. products whose market share is greater than 0.6) tend to be underestimated by the model. However, the overall average market share ends up being quite close to the overall average (it is in fact slightly overestimated, by 0.02%).

2.5.2 Price and ERP parameters

Equation 2.6 provides the basis for price estimation. To fit the model we assume that prices are subject to measurement error. There almost certainly is some amount of measurement error in our data: IMS collects their data from surveys of pharmacies and hospitals, and the prices we match are yearly averages. For simplicity, we do not include any additional sources of error.[54] Since our price model has a multiplicative function, we also assume that the measurement error is multiplicative (i.e. additive in logs). Let p_{ijt} be the model-predicted price, and p^o_{ijt} the observed price. Then

$$\ln\left(p^o_{ijt}\right) = \begin{cases} \ln\left(p^{\text{gov}}_{ijt}\left(\cdot\right)\right) + \eta_{ijt} & \text{if } p^{\text{ref}}_{ijt}\left(\cdot\right) \geq p^{\text{gov}}_{ijt}\left(\cdot\right) \\ \ln\left(\left(1-\mu_j\right) p^{\text{gov}}_{ijt}\left(\cdot\right) + \mu_j p^{\text{ref}}_{ijt}\left(\cdot\right)\right) + \eta_{ijt} & \text{if } p^{\text{ref}}_{ijt}\left(\cdot\right) < p^{\text{gov}}_{ijt}\left(\cdot\right) \end{cases}$$

We assume that η_{ijt} is i.i.d. across countries, drugs, and years.

Our estimation routine searches the vector of parameters that minimizes the difference between the model prediction and the data. Since the price function includes a fixed effect for each product and country (there are 481 distinct products in our main sample), the total number of parameters is quite large, and estimation matching log price levels would be quite slow. To improve the speed and efficiency of the procedure we match log differences in price. The key property of this ratio is that it does not depend on the product fixed effect θ_i. We state the result here as a Theorem, and provide a formal proof in Appendix B.1.2).

[54] Introducing any other source of error in the price would potentially lead to insurmountable estimation issues due to the propagation of this error through the reference pricing channel.

Theorem 5. *Let $p_{ijt}(\cdot)$ be defined as in equation 2.6. Then, for any $j, k \in \mathcal{N}_i$*

$$\ln\left(\frac{p_{ijt}(\cdot)}{p_{ikt+1}(\cdot)}\right) = \begin{cases} \ln\left(\frac{\gamma_j \cdot \exp\left(\beta_Z Z_{ijt} + \beta_D \ln\left(D_{ijt}\right)\right)}{\gamma_k \cdot \exp(\beta_Z Z_{ikt+1} + \beta_D \ln(D_{ikt+1}))}\right) & \text{if } p^{ref}_{ijt}(\cdot) \geq p^{gov}_{ijt}(\cdot) \wedge \\ & p^{ref}_{ikt+1}(\cdot) \geq p^{gov}_{ikt+1}(\cdot) \\ \ln\left(\frac{\gamma_j \cdot \exp\left(\beta_Z Z_{ijt} + \beta_D \ln\left(D_{ijt}\right)\right)}{(1-\mu_k)\gamma_k \cdot \exp(\beta_Z Z_{ikt+1} + \beta_D \ln(D_{ikt+1})) + \mu_k \tilde{p}^{ref}_{ikt+1}(\cdot)}\right) & \text{if } p^{ref}_{ijt}(\cdot) \geq p^{gov}_{ijt}(\cdot) \wedge \\ & p^{ref}_{ikt+1}(\cdot) < p^{gov}_{ikt+1}(\cdot) \\ \ln\left(\frac{(1-\mu_j)\gamma_j \cdot \exp\left(\beta_Z Z_{ijt} + \beta_D \ln\left(D_{ijt}\right)\right) + \mu_j \tilde{p}^{ref}_{ijt}(\cdot)}{\gamma_k \cdot \exp(\beta_Z Z_{ikt+1} + \beta_D \ln(D_{ikt+1}))}\right) & \text{if } p^{ref}_{ijt}(\cdot) < p^{gov}_{ijt}(\cdot) \wedge \\ & p^{ref}_{ikt+1}(\cdot) \geq p^{gov}_{ikt+1}(\cdot) \\ \ln\left(\frac{(1-\mu_j)\gamma_j \cdot \exp\left(\beta_Z Z_{ijt} + \beta_D \ln\left(D_{ijt}\right)\right) + \mu_j \tilde{p}^{ref}_{ijt}(\cdot)}{(1-\mu_k)\gamma_k \cdot \exp(\beta_Z Z_{ikt+1} + \beta_D \ln(D_{ikt+1})) + \mu_k \tilde{p}^{ref}_{ikt+1}(\cdot)}\right) & \text{if } p^{ref}_{ijt}(\cdot) < p^{gov}_{ijt}(\cdot) \wedge \\ & p^{ref}_{ikt+1}(\cdot) < p^{gov}_{ikt+1}(\cdot) \end{cases} \tag{2.9}$$

where $\tilde{p}^{ref}_{ijt}(\cdot)$ is such that $p^{ref}_{ijt}(\cdot) = \tilde{p}^{ref}_{ijt}(\cdot) \cdot \theta_i$, and $\tilde{p}^{ref}_{ijt}(\cdot)$ is not a function of θ_i.

The proof of this Theorem hinges on showing that the reference price can be written as a linear function of the drug fixed effect. Intuitively, this result follows from the fact that the reference price can be written as a weighted average of government prices, which are all linear functions of the drug fixed effect.

Since log differences do not depend on the product fixed effect θ_i, this strategy drastically cuts down on computation time. We match $\ln\left(\frac{p_{ijt}}{p_{ikt+1}}\right)$, where j and k are two (randomly selected) countries, and we look at the difference in the price of product i in these two countries over consecutive periods.[55] We compute the model predictions using equation 2.9. The estimating equation in differences is

$$\ln\left(\frac{p^o_{ijt}}{p^o_{ikt+1}}\right) = \ln\left(\frac{p_{ijt}(\cdot)}{p_{ikt+1}(\cdot)}\right) + \eta_{ijt} - \eta_{ikt+1}$$

[55] We compare prices of different countries to avoid differencing out the country fixed effect. In order to minimize the number of observations lost through the differencing process, we occasionally consider $\frac{p_{ij2012}}{p_{ik2002}}$ as a moment. By doing so we only lose one observation per drug, instead of one observation per drug-country.

Our routine minimizes the sum of the two error terms:

$$O\left(\gamma_j, \mu_j, \beta_Z, \beta_D\right) = \sum_{i,j,k,t} \left[\ln\left(\frac{p_{ijt}\left(\cdot\right)}{p_{ikt+1}\left(\cdot\right)}\right) - \ln\left(\frac{p^o_{ijt}}{p^o_{ikt+1}}\right)\right]^2 \tag{2.10}$$

We restrict μ_j to the unit interval. A negative value of μ_j does not make sense; a value of μ_j greater than 1 is theoretically possible, but raises the possibility of negative prices in counterfactual predictions. Once we have obtained estimates for $\left(\hat{\gamma}_j, \hat{\mu}_j, \hat{\beta}_Z, \hat{\beta}_D\right)$, we plug them into the price equation to calculate $\hat{\theta}_i$.

Results We report price estimation results for the vector of parameters $\left(\hat{\gamma}_j, \hat{\mu}_j, \hat{\beta}_Z, \hat{\beta}_D\right)$. Since our estimation is in logs, we report coefficients for $\ln\left(\gamma_j\right)$, which are more easily interpretable as proportional changes in relative terms with respect to a benchmark (in this case, the omitted coefficient used as a benchmark is $\ln\left(\gamma_{\text{GERMANY}}\right)$).

The first column of Table 2.2 shows the coefficients for $\ln\left(\gamma_j\right)$. The point estimates roughly match our intuition: countries with lower income tend to pay lower prices, with a couple of exceptions. Poland and Hungary have higher coefficients than many other countries with higher income. However, both countries use the minimum price in Europe as their reference, and the coefficients on μ_j for both are very close to 1. This suggests that the government is occasionally willing to grant higher prices, knowing that reference rules will bring them down very quickly.[56]

The second columns shows estimates for μ_j, which is the coefficient measuring how strictly each country adheres to its own ERP guidelines. We observe significant heterogeneity across countries in this respect. Seven countries have coefficients above 0.9, suggesting that reference pricing guidelines are followed closely. Four other countries (including Italy and Spain) have coefficients above 0.8. The remaining 8 countries who use ERP all have coefficients below one-third, which suggests that they either do not follow their guidelines

[56]Luxembourg and Norway (the two richest countries in Europe) are also outliers. Norway is relatively isolated in the reference pricing matrix, because it is only referenced by Finland. Hence granting a lower price to Norway might carry relatively little consequences for firms. This effect is not captured explicitly by our model, but is incorporated in the country fixed effect. Luxembourg is harder to explain, though its status as a relatively small countries might give rise to all sort of irregularities and exceptions. Many drugs do not even record sales in Luxembourg, so it is possible that selection might play a role here.

Table 2.2: *Price Estimation Results*

Country	$\ln(\gamma_j)$		μ_j	
Austria	-0.099	[-0.134,-0.071]	0.199	[0.022,0.444]
Belgium	-0.123	[-0.156,-0.093]	0.119	[0.010,0.446]
Bulgaria	-0.230	[-0.317,-0.145]	0.987	[0.432,0.998]
Denmark[b]	-0.068	[-0.102,-0.046]	0	
Estonia	-0.170	[-0.380,-0.128]	0.994	[0.249,0.998]
Finland	-0.126	[-0.165,-0.098]	0.332	[0.018,735]
France	-0.098	[-0.130,-0.061]	0.098	[0.008,841]
Germany[a,b]	0		0	
Greece	-0.186	[-0.216,-0.035]	0.987	[0.841,0.999]
Hungary	-0.145	[-0.240,-0.082]	0.991	[0.326,0.998]
Ireland	-0.089	[-0.127,-0.062]	0.229	[0.018,0.840]
Italy	-0.191	[-0.221,-0.110]	0.854	[0.219,0.994]
Latvia	-0.201	[-0.317,-0.151]	0.805	[0.013,0.941]
Lithuania	-0.204	[-0.359,-0.151]	0.996	[0.014,0.999]
Luxembourg[c]	-0.200	[-0.247,-0.176]	0	
Netherlands	-0.183	[-0.220,-0.151]	0.293	[0.073,0.826]
Norway	-0.143	[-0.186,-0.111]	0.846	[0.107,0.981]
Poland	-0.119	[-0.238,0.025]	0.904	[0.239,0.989]
Portugal	-0.176	[-0.220,-0.138]	0.093	[0.012,0.842]
Romania	-0.253	[-0.351,-0.105]	0.994	[0.204,0.999]
Slovenia[d]	-0.232	[-0.272,0.204]	/	
Spain	-0.159	[-0.186,-0.131]	0.874	[0.151,0.975]
Sweden[b]	-0.102	[-0.131,-0.080]	0	
Switzerland	0.007	[-0.024,0.032]	0.020	[0.003,0.080]
UK[b]	-0.201	[-0.233,-0.179]	0	
Controls				
Log quantity sold		-0.019		[-0.024,-0.017]
Home-Firm Indicator		0.041		[0.011,0.079]
At least 1 other molecule in class		-0.043		[-0.079,0.028]
At least 2 other molecules in class		0.013		[-0.024,0.045]
At least 5 other molecules in class		-0.029		[-0.063,0.007]
At least 10 other molecules in class		-0.003		[-0.031,0.016]

[a] The price level is normalized to Germany's.

[b] Denmark, Germany, Sweden, and the UK do not use reference pricing during the period 2002-2012, so we set μ_j to zero.

[c] Luxembourg references the price of the country of origin of the drug. We don't know country of origin, so we simply assume that μ_j is equal to 0.

[d] On occasion, a country's price level is such that its reference price never binds. In these cases the coefficient is undetermined.

all that closely, or that they apply them only to selected drugs. In particular Switzerland does not appear to use reference pricing at all, with a coefficient that is almost exactly zero. We are not able to identify a coefficient for Slovenia because the model estimates that its government price is always below the reference price (Slovenia references Austria, France, and Germany, which are all countries with much higher price levels).

Almost all the coefficients on the control variables in the welfare function behave as expected. Higher quantity sold is associated with lower prices. Prices tend to be approximately 4% higher in countries where the firm has headquarters. Finally, having a higher number of competitors in the same class is associated to lower prices, though the relationship appears to be nonlinear and noisy.

One key result that emerges from the analysis is that the price level (as indicated by $\ln(\gamma_j)$) in Western European countries that follow reference pricing closely is only marginally higher than the price level of Eastern European countries. This suggests that in the current equilibrium firms may be under pressure to keep prices higher in Eastern European countries to reduce the size of the externality generated by ERP.

2.5.3 Simulation-Based Evidence of Optimal Delays

Our price estimation results suggest that ERP does indeed affect equilibrium prices, but are the externalities strong enough to generate delays? We test whether this is the case by simulating firm revenue from a variety of strategies. We do not include delay shocks in this simulation, since all we are interested in is establishing whether or not the incentive for strategic delays exists.

We simulate the expected payoff of two sets of strategies, and compare it to the payoff obtained when launching everywhere right away. The first set of strategies, denoted as $\mathcal{A}^{kr}$, consists of withholding a single product from country k for r periods.[57] Formally,

[57] When a firm owns multiple products in the same therapeutic class, we test strategies where only one product is withheld and other products are launched immediately.

$\mathcal{A}_{lt}^{kr} = \left\{A_{l\tau}^{kr}\left(S_{l\tau-1}\right)\right\}$, with $A_{l\tau}^{kr}\left(S_{l\tau-1}\right) = \left\{a_{ij\tau}^{kr}\right\}_{i\in\mathcal{I}_l}$, and

$$a_{ijt}^{kr} = \begin{cases} 1 & \text{if } j \neq k \vee (j = k \wedge \tau < t + r) \\ 0 & \text{if } j = k \wedge \tau \geq t + r \end{cases}$$

We simulate the expected revenue of these strategies for all the 87 drugs in our dynamic estimation sample, and calculate the fraction of drugs for which delaying in country j is optimal for at least one value of r.[58] Figure 2.9 plots the results. We find that these types of delaying strategies are optimal for all Eastern European countries in at least some cases. Only three countries outside of Eastern Europe would experience delays according to this simulation: Austria, Belgium, and Luxembourg. Luxembourg does not have particularly low prices, but it's a very small market, so it's not surprising that for some drugs it would be optimal to exclude it. The model also predicts a delay for one drug in Austria, and one in Belgium. In both cases, the drug in question was indeed launched with a long delay in both countries, and earned a very small amount in sales. Even though both Belgium and Austria tend to have high price levels relative to most other countries, they can affect the prices of countries with higher price levels (for example Ireland).[59]

Figure 2.9 is remarkably consistent with the reduced-form patterns described in Section 2.3.2 (Figure 2.2). No information about delays or launch sequences was used in generating the figure, which is purely based on price and demand estimates. This confirms the intuition that firm behavior adheres to the incentives laid out by the regulatory environment, and supports the idea that our model can capture the relevant features of this market.

In the next set of strategies, we focus on Eastern Europe as a block, and simulate revenue

[58] In some cases, drug i is already in country j in 2002, which makes it impossible to simulate revenue under a counterfactual delay. We exclude those drugs from our calculations.

[59] Delaying entry in Romania and Bulgaria is optimal for a surprisingly low number of drugs, given our price and demand estimates. This is because prior to acquiring EU membership in 2007, Bulgaria and Romania where only referenced by a few other Eastern European countries. Thus, our model predicts that entering in Romania and Bulgaria only has a small effect on prices prior to 2007. In the data, we do not find a significant difference in the average delays before and after EU entry in these two countries. This is not too surprising however, as entry in the EU also removes several bureaucratic obstacles that might have generated idiosyncratic delays. Hence, the net effect of EU membership on launch delays could be close to 0.

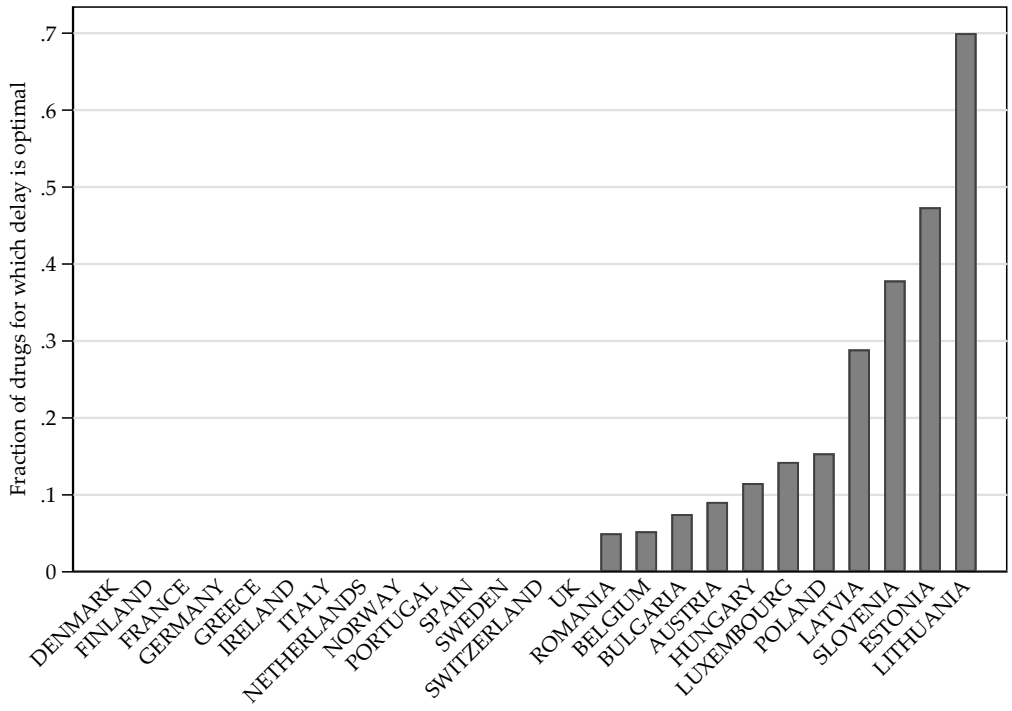


Figure 2.9: *Optimality of strategic delays by country*

This figure plots the fraction of drugs for which delaying in each country is optimal. We use a sample of 87 drugs whose patent expires on or before December 31st, 2012. By the time our sample started, in 2002, launches in many countries had already occurred for some of these drugs. Hence, for each country j we estimated the y-axis variable using the subsample of drugs that had not already entered in country j before 2002. This implies that the sample used is slightly different for each country.

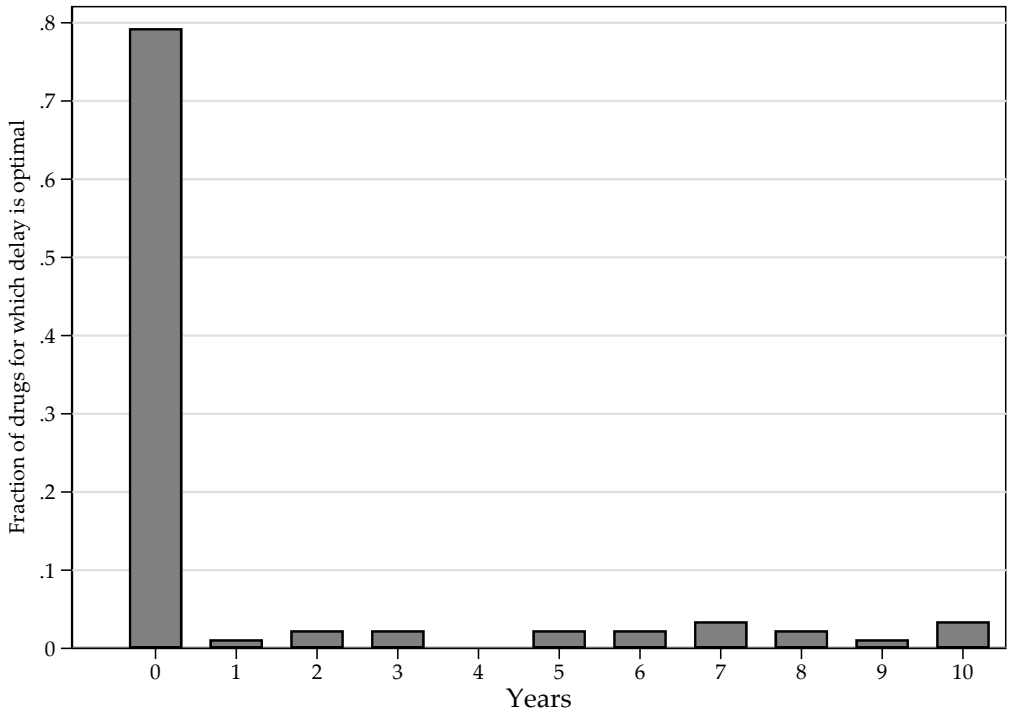


Figure 2.10: *Strategic Delays in Eastern European Countries*

In this figure we consider a set of strategies where entry in Eastern European countries is delayed for t periods, where t is between 0 and 10. We count the number of drugs for which each strategy is optimal, and plot it in a histogram (the y-axis displays the fraction of drug, rather than the level). The figures shows that for 80% of drugs no delay is preferable, but for 20% of drugs some delay is optimal.

from strategies that delay entry in all Eastern European countries.[60] Formally, let $\mathcal{N}_E$ denote the set of Eastern European countries. Then $\mathcal{A}_{lt}^{Er} = \left\{A_{l\tau}^{Er}\left(S_{l\tau-1}\right)\right\}$, with $A_{l\tau}^{Er}\left(S_{l\tau-1}\right) = \left\{a_{ij\tau}^{Er}\right\}_{i\in\mathcal{I}_l}$, and

$$a_{ijt}^{kr} = \begin{cases} 1 & \text{if } j \notin \mathcal{N}_E \vee \left(j \in \mathcal{N}_E \wedge \tau < t+r\right) \\ 0 & \text{if } j \in \mathcal{N}_E \wedge \tau \geq t+r \end{cases}$$

We plot the distribution of optimal delays for our 87 drugs in Figure 2.10. We find that for around 20% of drugs, delaying in all Eastern European countries as a block is preferable.

These simulations suggest that firms are better off when their products are not available in Eastern European countries right away. At the same time, we find almost no evidence that firms have any strategic incentive to delay entry in countries in Western Europe (this includes even countries with relatively lower income levels like Greece and Portugal).

2.6 Dynamic Analysis

With estimates for demand and price primitives in hand, we now turn our attention to the parameters governing idiosyncratic delays. These parameters, which may differ across countries, capture the probability that an application may be delayed. Traditional estimation methods for dynamic entry models are unfeasible in our case: the complicated net of price externalities prevents us from deriving an analytic solution, and the total number of countries and periods we consider means that the state space is too large to solve the problem numerically. Finally, the structure of delay shocks generates an additional identification challenge on top the curse of dimensionality: firm strategies are unobserved. In the data, we see firms entering, but not when applications are sent. In other words, in the observed launch sequences, an application that was not sent is observationally equivalent to an application that was rejected. To overcome these obstacles we approach the problem using a partial identification approach.

[60] As with the previous simulation exercise, sometimes drug are already available in some Eastern European countries. We delay entry only in countries where the product has not already entered.

We use two sets of moment inequalities to construct bounds on the probability of idiosyncratic delays. The first set of moment inequalities uses data on entry and approval dates and exploits the intuition that firms can only apply after they have received marketing approval for a product. Hence, the average probability of an idiosyncratic delay must be less than or equal the overall probability of delay implied by the data (notice that these inequalities can only recover an upper bound). The second set of inequalities uses information contained in the expected revenue of the observed entry sequence. The estimator rejects parameter values for which we can find strategies that yield higher expected revenue than what firms obtained in the data.

In constructing the second set of moment inequalities we rely on the novel methodological approach introduced in the first chapter of this thesis. In our setting we cannot rely on the model to compute the expected revenue of the firm's strategy, because the strategy is unobserved. At the same time, moment inequalities that use the expected revenue of the observed entry sequence may not hold: some firms may earn less (or more) than their expectation depending on the realization of the delay shocks. Hence we use aggregate outcomes across drugs to build inequalities based on aggregate moments. In other words, our moment conditions are based on the average expected revenue across all drugs in our sample.

The rest of this section is divided in three parts. First, in 2.6.1, we derive an upper bound using moment inequalities based on entry and approval data. Next, in 2.6.2 we show how our empirical model maps to the theoretical framework introduced in the first chapter of this dissertation. Finally, we discuss the empirical implementation of this latter methodology and our results in 2.6.2.

2.6.1 Moment Inequalities based on Entry Data

We use entry and approval data to recover an upper bound on the delay parameters. Our distributional assumption on the random delay shocks is that in each period the probability of an application for entry in country j being delayed is ψ_j. Let $\left(1-\bar{\psi}_j\right)$ be the probability

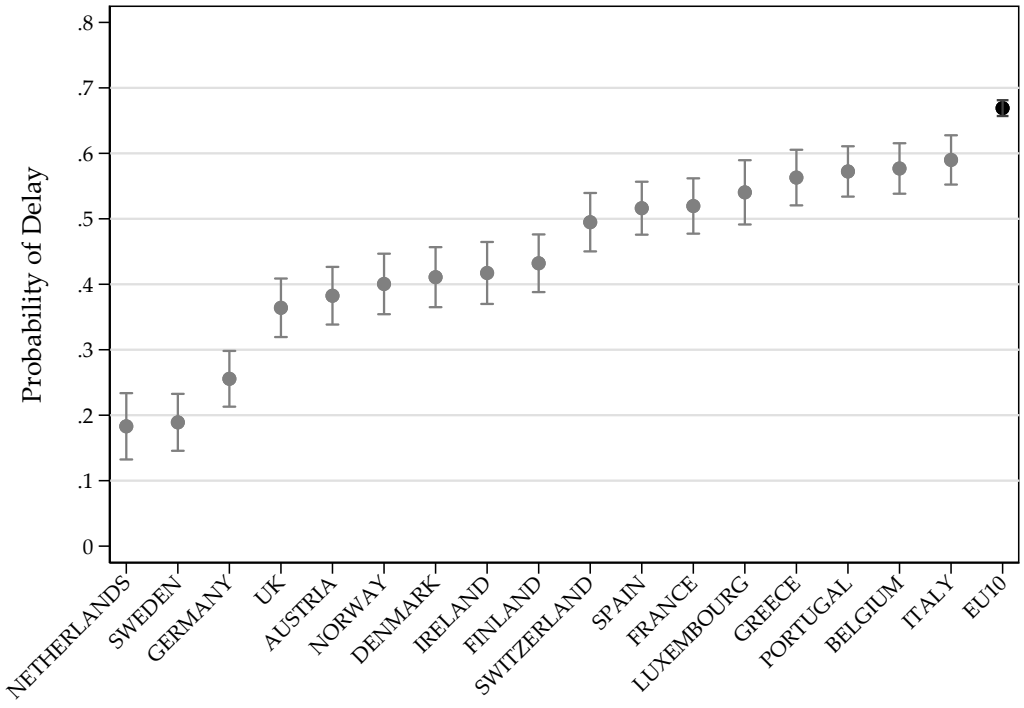


Figure 2.11: *Probability of Overall Delays by Country*

In this figure we plot the overall probability of a delay in each country in Europe. We aggregate Eastern European countries as a whole for consistency with the moment inequality estimation (details in Section 2.6.2). The plot includes 95% confidence interval bands for each parameter.

that product i will enter in country j in a given year. This is the combination of the probability that the firm will apply, times the probability of the application being accepted. Hence, for all j, $\bar{\psi}_j \geq \psi_j$.

To estimate $\bar{\psi}_t$ we can simply calculate the probability of a delay by using data on approval dates and launch dates: suppose that a product approved in year 0 enters France in year 2. Then we are going to assign an expected probability of entry of $\frac{1}{3}$: the product had three opportunities to enter — in years 0, 1, and 2 — and registered one success, in year 2.

Figure 2.11 plots the coefficients $\bar{\psi}_j$ for all European countries, together with confidence intervals.[61] We estimate a single coefficient for Eastern Europe — denoted as ψ_{EU10} — for consistency with the moment inequality estimation (see the next section).[62] The range of these parameters varies between 0.2 and 0.6 for Western European countries, while the coefficient for Eastern Europe is 0.669.[63]

2.6.2 Moment Inequalities based on Revenue Data

Theoretical Derivation

Identification in moment inequality is generally based on the idea that the firm's strategy contains some information about the true value of the unknown parameter of a model. In particular, the strategy of the firm will only be optimal for a certain range of values of the parameter. For values outside that range we should be able to find strategies that would earn higher revenue, which in turn allows us to rule those values out.

In our setting firm strategies are unobserved (e.g. if we observe entry in France with a

[61] We follow instructions laid out by Brown *et al.* (2001) in building approximate confidence intervals for these parameters.

[62] EU10 is the shorthand for Eastern European countries that are part of the European Union. In our data we have 8 Eastern European countries (instead of 10), because we are missing sales data for Czech Republic and Slovakia.

[63] One could also obtain a tighter bound by choosing a partition P of the set of all drugs, calculating a subset-specific upper bound $\bar{\psi}_j^p$ for all $p \in P$, and choosing $\bar{\psi}_j^{p,\min} = \min_p \bar{\psi}_j^p$ to be the upper bound. We show in Appendix B.5 that doing so can yield a lower upper bound for $\bar{\psi}_j$. However, we only include the estimate from the full sample in our main results, as this exercise relies heavily on the assumption that ψ_j is constant across drugs.

2-year delay, we don't know whether the firm applied starting in year 0, year 1, or year 2). As a result, we are unable to compute the expected revenue of the firm's observed strategy for arbitrary values of the unknown parameter. However, in the data we can observe the revenue earned by the firm for the true value of the parameter. We use this information to obtain identification and build inequalities in a way that is similar to the rest of the literature: we compare the revenue earned by the firm in the observed data, and calculate counterfactual revenue under alternative strategies for arbitrary values of the parameter.

The loss of information that comes from not observing the optimal strategy means that our inequalities will not necessarily hold for individual firms: the revenue we observe depends not only on strategies, but also on the realization of the delay shocks. If a firm incurs in longer-than-expected delays, this will decrease their revenue and we may end up finding a violation of the inequality even at the true parameter value. To address this issue, we aggregate across drugs and instead match the average expected revenue across all firms. Doing so will remove the noise generated by the realization of the delay shocks.

Formally, let $\mathcal{A}_{lt}^{\star}$ denote the optimal strategy of firm l starting in period t. By definition, $\mathcal{A}_{lt}^{\star}$ is the solution to the dynamic programming problem of the agent as expressed in equation 2.7. In other words, for all possible strategies $\mathcal{A}_{lt}'$, $\mathcal{A}_{lt}^{\star}$ satisfies

$$\tilde{V}_t\left(\mathcal{A}_{lt}^{\star}; S_{lt-1}, S_{-lt-1}\right) \geq \tilde{V}_t\left(\mathcal{A}_{it}'; S_{lt-1}, S_{-lt-1}\right) \tag{2.11}$$

where

$$\tilde{V}_t\left(\mathcal{A}_{lt}; S_{lt-1}, S_{-lt-1}\right) = \mathbb{E}\left[\sum_{\tau=t}^{T} \beta^{\tau-t} \Pi_\tau\left(S_l, S_{-l}\right) \middle| \mathcal{A}_{lt}, S_{lt-1}, S_{-lt-1}\right] \tag{2.12}$$

is the expected payoff of the agent conditional on playing strategy profile $\mathcal{A}_{lt}$ (the expectation is taken over the possible realizations of the launch sequences of all products, as in equation 2.7).

Under the assumption that firms maximize expected returns we can use equation 2.11 to build "revealed preference" inequalities. In the data we observe a certain number of firms with molecules in specific therapeutic classes (each firm-class combination constitutes one observation for us, since molecules in different therapeutic classes do not affect each other).

The *expected* payoff of the observed launch sequence S_l^o of firm l starting at period t is

$$V_t\left(S_l^o, S_{-l}^o\right) = \sum_{\tau=t}^{T} \beta^{\tau-t} \Pi_\tau\left(S_l^o, S_{-l}^o\right)$$

Notice that $\Pi_\tau\left(S_l^o, S_{-l}^o\right)$ is an object that we can recover using simulation, since it only depends on the shock $\xi_t^\kappa = \left\{\xi_{jt}^\kappa\right\}$. This error is unobserved by both the firm and the econometrician by assumption.[64] We recover the distribution of ξ_t^κ from demand estimation, and use it to simulate ξ_t^κ. Then, $V_t\left(S_l^o, S_{-l}^o\right)$ converges to the average of the expected payoff of each firm when playing the optimal strategy.

Theorem 6. *For any $\varepsilon > 0$, we can find M' such that*

$$\frac{1}{M}\sum_{i=1}^{M}\left(V_t\left(S_l^o, S_{-l}^o\right) - \tilde{V}_t\left(\mathcal{A}_{lt}^\star; S_{lt-1}, S_{-lt-1}\right)\right) < \varepsilon$$

for all $M > M'$.

The proof of this Theorem essentially shows that $V\left(S_l^o, S_{-l}^o\right)$ is bounded, and therefore $V\left(S_l^o, S_{-l}^o\right)$ satisfies the premise of the framewoek introduced in the first chapter of this dissertation (see Appendix B.1.4). The basic intuition for the result is that if firms are playing the optimal strategy, then the observed launch sequences are draws from the probability distributions $\mathcal{P}\left(S_l \middle| S_{lt-1}, \mathcal{A}_{lt}^\star\right)$ and $\mathcal{P}\left(S_{-l} \middle| S_{-lt-1}, \mathcal{A}_{-lt}^\star\right)$. This distribution need not be the same for all firms: each may face a different initial state S_{lt-1}, and different demand or price primitives. However, since the random shocks used in the model are independently distributed across drugs, we can apply a generalized version of the law of large number for non-identical independent random variables to argue that the sample average across firms converges to the average expected payoff.

Theorem 6 suggests that we can write moment inequalities based on the average payoff obtained by the firms in our sample:

$$\frac{1}{M}\sum_{i=1}^{M}\tilde{V}_t\left(\mathcal{A}_{lt}^\star; S_{lt-1}, S_{-lt-1}\right) \geq \frac{1}{M}\sum_{i=1}^{M}\tilde{V}_t\left(\mathcal{A}_{lt}'; S_{lt-1}, S_{-lt-1}\right) \tag{2.13}$$

[64] There is no additional error from price since the residual is assumed to be measurement error.

The left-hand side of this inequality is approximated by $\frac{1}{M}\sum_{i=1}^{M} V_t\left(S_l^o, S_{-l}^o\right)$. To compute the right-hand side we use simulation. The expected payoff of any deviation $\mathcal{A}'_{lt}$ can be written as an integral over the possible realization of the launch sequence S_l, holding S_{-l} fixed [65]

$$\tilde{V}_t\left(\mathcal{A}'_{lt}; S_{lt-1}, S_{-lt-1}\right) = \sum_{S_{-l}}\sum_{S_l}\sum_{\tau=t}^{T}\beta^{\tau-t}\Pi\left(S_l, S_{-l}\right)\mathcal{P}\left(S_l|\, S_{lt-1}, \mathcal{A}'_{lt}\right)\mathcal{P}\left(S_{-l}|\, S_{-lt-1}, \mathcal{A}^{\star}_{-lt}\right)$$

We approximate $\tilde{V}_t\left(\mathcal{A}'_{lt}; S_{lt-1}, S_{-lt-1}\right)$ by simulating the distribution of $\mathcal{P}\left(S_l|\, S_{lt-1}, \mathcal{A}'_{lt}\right)$. For a given guess of the parameter vector ψ_{EU10}, and for each drug $i \in \mathcal{I}_l$ draw $\left\{\nu_{ijt}^r\right\}_{r=1}^{nsim}$ and use them to calculate simulated entry paths $\left\{S_l^r\right\}$. The average simulated payoff is

$$V_t^{sim}\left(\mathcal{A}'_{lt}; S_{lt-1}, S_{-lt-1}, \psi_{EU10}\right) = \frac{1}{nsim}\sum_{r=1}^{nsim}\left[\sum_{\tau=t}^{T}\beta^{\tau-t}\Pi\left(S_l^r, S_{-l}\right)\right]$$

The difference between $\tilde{V}_t\left(\mathcal{A}'_{lt}; S_{lt-1}, S_{-lt-1}\right)$ and $V_t^{sim}\left(\mathcal{A}'_{lt}; S_{lt-1}, S_{-lt-1}, \psi^0_{EU10}\right)$ — for the true parameter vector ψ^0_{EU10} — is simulation error, which can be eliminated by choosing *nsim* large enough, and the error in the realization of S_{-lt-1} (which is only one of many possible draws). When we aggregate across firms, the error in S_{-lt-1} will disappear for a large enough sample of drugs.

We define empirical moment moment conditions as

$$\mu\left(\mathcal{A}'_{lt}, \psi_{EU10}\right) = \min\left\{0, \frac{1}{M}\left(\sum_{i=1}^{M} V_t\left(S_l^o, S_{-l}^o\right) - V_t^{sim}\left(\mathcal{A}'_{lt}; S_{lt-1}, S_{-lt-1}, \psi_{EU10}\right)\right)\right\}$$

The estimation identifies a set Ψ^I of parameters that satisfy

$$\Psi^I = \left\{\psi_{EU10} \,:\, \sum_{\mathcal{A}'_t}\mu\left(\mathcal{A}'_{lt}, \psi_{EU10}\right) = 0\right\}$$

Empirical Implementation

Based on our simulation results, we assume that firms never optimally delay entry in Western Europe (i.e. $\bar{\psi}_j = \psi_j$ for all countries outside Eastern Europe). Since moment inequalities

[65] We cannot simulate S_{-l} because we do not observe strategies. However, each observation of S_{-l} is drawn from the probability distribution of $\mathcal{P}\left(S_{-l}|\, S_{-lt-1}, \mathcal{A}^{\star}_{-lt}\right)$, so when we average across firms, we will approximate the right distribution. See the proof of Theorem 6 in Appendix B.1.4.

work best with a small number of parameters we assume that the probability of delay is the same in all Eastern European countries. This parameter captures the probability of incurring a delay in Eastern European countries due to forces outside of the firm's control. We estimate an identified set for this parameter by building moment conditions based on equation 2.13, and checking what range of parameters satisfies it for a series of possible strategies $\mathcal{A}'_{lt}$.

Calculating $V_t\left(S_l^o, S_{-l}^o\right)$ and $V_t^{sim}\left(\mathcal{A}'_{lt}; S_{lt-1}, S_{-lt-1}, \psi_{EU10}\right)$ does not require observing drugs from their original launch, but it does require observing them until period T_l. Hence we perform this analysis on the dynamic sample (see Table 2.1 for details).

We let our simulation analysis guide our choice of counterfactual strategies. For each drug i, our counterfactual strategy delays entry in all countries where simulations suggest that doing so would result in higher revenue. For firms that have more than one drug we simulate multiple strategies where the entry sequences of all but one drugs are held fixed (as if they were owned by another firm).[66]

Ex-ante, we expect these inequalities to be more effective in finding a lower bound on the probability of an idiosyncratic delay ψ. To see why, assume that the expected revenue conditional on playing the optimal strategy are monotonically decreasing in ψ (i.e $\frac{\partial \tilde{V}(\mathcal{A}^\star(\psi),\psi;\cdot)}{\partial \psi} < 0$).[67] Then, it is impossible to find an alternative strategy A' such that

$$\tilde{V}_t\left(A', \psi'; S_{lt-1}, S_{-lt-1}\right) > \tilde{V}_t\left(\mathcal{A}^\star_{lt}\left(\psi^0\right), \psi^0; S_{lt-1}, S_{-lt-1}\right)$$

for $\psi' > \psi^0$.[68] It will still be possible to find A' such that $V\left(A', \psi'\right) > V\left(A^\star\left(\psi^0\right), \psi^0\right)$ for $\psi' < \psi^0$.[69]

[66] We also tested a variety of simpler strategies, such as applying right away in all countries, or applying a set number of period before observed entry. These strategies yielded looser bounds.

[67] We do not have a formal proof of this statement. However, this is intuitively true: firms should be better off when the probability of a delay is lower, as they have better control over which entry sequence will be realized. For example, if the probability of delay were 0, the firm would be able to choose the profit-maximizing entry strategy. An increase in the probability of delay would reduce the likelihood of achieving the profit-maximizing entry sequence, therefore expected revenue would fall.

[68] To see why, notice that by assumption $V\left(A^\star\left(\psi^0\right), \psi^0\right) > V\left(A^\star\left(\psi'\right), \psi'\right)$ for all $\psi' > \psi^0$. Moreover, by definition of $A^\star(\cdot)$, $V\left(A^\star\left(\psi'\right), \psi'\right) > V\left(A', \psi'\right)$ for all ψ'.

[69] We could obtain an upper bound from data on expected revenue by showing that for a given value of ψ, no strategy will ever yield expected revenue as high as what the firm obtained in the data. The problem is that the

Our empirical results confirm this intuition: we find that our moment inequalities only provide a lower bound on the probability of an idiosyncratic delay. The identified set Ψ^I that we generate using these inequalities rejects values of ψ_{EU10} lower than 0.416.

2.7 Counterfactuals and Policy Analysis

In many cases, estimation using moment inequality precludes the ability to run counterfactuals. In our case however, we are able to simulate an equilibrium strategy under an important alternative reference pricing scenario: the elimination of ERP. In this scenario, regardless of the pricing policy that is implemented as a replacement, the optimal solution is easy to derive: firms no longer have any incentive to delay entry in any country, so they will apply for entry everywhere right away. This scenario gives an idea of the welfare loss from reference pricing in terms of forgone drug-years. We examine this scenario in 2.7.1 and find that removing ERP could accelerate entry in Eastern European countries by up to 14 months.

The removal of ERP could be hard to implement in practice, as it would require all countries to surrender part of their sovereign power to the European Union. Historically, governments have been reluctant to do so. However we estimate that leaving ERP in place and paying firms to forgo strategic delays would require a relatively small amount of money (around €20 million for the average drug). This solution, which we briefly discuss in 2.7.2, could be easier to implement than a complete overhaul of the pricing system in each country.

2.7.1 Delays in the absence of ERP

In the absence of ERP (and assuming that no other policy that limits price discrimination across countries exists), the optimal strategy is for the firm to launch as soon as possible in all countries. Hence, the firm will send entry applications to all countries as soon as the

firm achieves the highest possible expected revenue when playing the optimal strategy, and the reason we resort to moment inequalities is precisely that we cannot compute this strategy. However, we can find functions of data and parameters that always returns a value greater than the expected revenue under the optimal strategy. We describe this approach in Appendix B.5, including an example of a function that satisfies these requirements, but that unfortunately only yields an upper bound that is already ruled out from the moment inequalities that use entry data.

product is approved, and the only delays in the model will arise through the bureaucratic delay channel.

This counterfactual is consistent not only with a situation in which ERP is ruled out by the European Union, but also with any policy that places the responsibility to set prices with a European authority, and out of the hands of individual countries. Many such policies have been indeed suggested as a way to circumvent the effect of reference pricing. For example Towse *et al.* (2015) suggest that various forms of tiered pricing could be implemented centrally by the European Union to avoid delays in low-income countries. They propose Ramsey pricing and Value-Based Pricing as possible candidates. The implication of our model is that the specific pricing rule does not matter, as long as cross-market externalities are eliminated (this would also involve imposing limits on parallel trade, for example by carving out an exclusion for patent-protected pharmaceuticals). Importantly this also implies that we do not need to rely on our pricing model to work out the implications of this counterfactual. This is crucial as it is quite likely that such a large change in the regulatory environment would also bring about changes in the level and variation of prices across countries that our model might be unable to predict accurately.

Figure 2.12 shows the range of possible delays in Eastern Europe implied by our identified set $[0.416, 0.669]$ for ψ_{EU10}. The parameter at the upper bound of the interval implies that idiosyncratic delays match observed delays. Under this scenario, the average drug should expect around 2 years of delay in each Eastern European country. At the lower bound of the interval, firms would instead expect an average delay of around 8-9 months in these countries, a reduction of approximately 63% in delays. A potentially useful comparison is what would happen if we assigned to Eastern European countries the same delay parameter as the slowest Western European country, which is Italy. Italy has a delay probability of 0.590. If ψ_{EU10} took on this value, delays in the absence of ERP would fall by 27%, or around 6 months in each country.

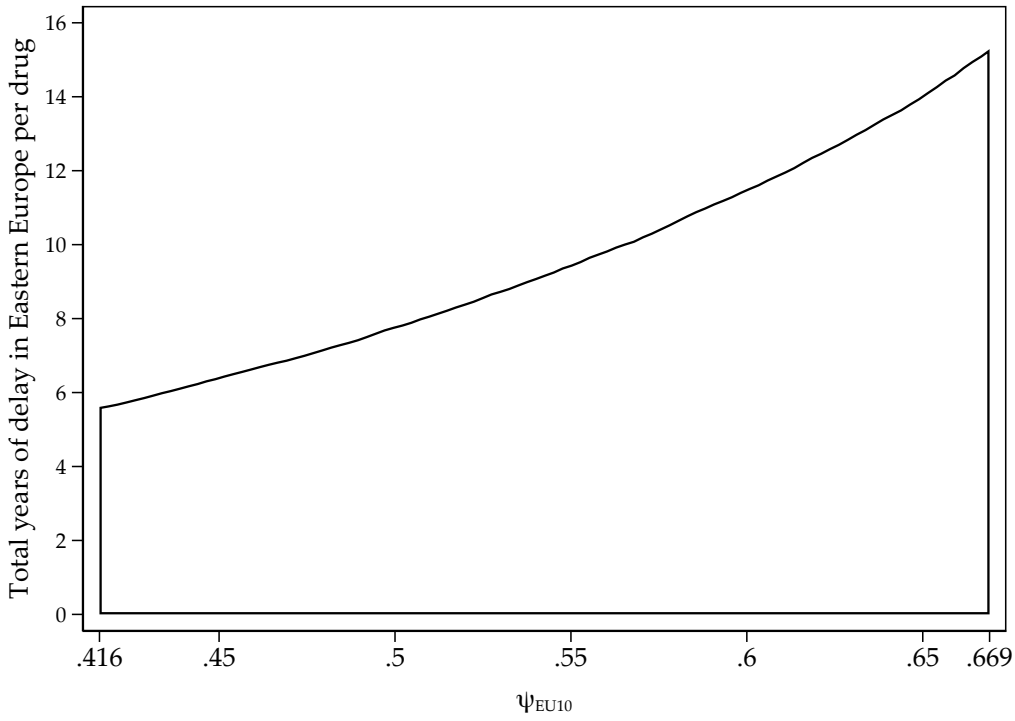


Figure 2.12: *Range of Potential Delays in the Absence of ERP*

In this figure we plot the range of possible delays in Eastern European countries for the identified set of possible delay parameters. The y-axis reports the total delay of the average drug among all Eastern European countries. The upper bound of the interval returns the same average delay that is observed in the data (approximately 16 years, or 2 years per country), because it is calculated to match the overall probability of a delay.

2.7.2 Policy Proposal: Pay for (no) Delay

One of the main obstacles that prevents the implementation of a pricing mechanism without reference pricing is the political will of EEA Member States. The prerogative of governments to manage drug pricing independently stems directly from the Treaty of Lisbon, meaning that any action to remove it would require a complicated bureaucratic process.[70] Moreover, governments tend to be quite protective of their independence from the central EU government, so efforts to move away from reference pricing are likely to be met with resistance.

A potential solution that would leave the current pricing system in place, could be to provide subsidies for firms that make their products available in peripheral European countries upon receiving marketing approval. These subsidies would be handed out by a centralized European agency upon confirmation that an entry application has been sent and approved for entry in all Eastern European countries (even though we do not observe this information in the data, we assume that this agency would have a way to verify the firm's action).

Our estimates suggest that the overall budget impact of this policy would be small. The average drug would receive an offer around €20 million in order to induce immediate entry.[71] On average, during the period between 1995 and 2017, around 27 new drugs received approval in the EEA. Hence, the overall impact of this subsidy would be around €550 million per year. This is not a negligible sum, but still represents a very small fraction of the overall budget of the EU, which in 2016 was around €150 billion. It is also worth noting that from the point of view of a social planner, the lump sum constitutes a transfer, so any gains from early access, however small, would improve the overall welfare in the system.[72]

---

[70] Article 168 of the Treaty on the Functioning of the European Union (i.e. the Treaty of Lisbon) explicitly states that "Union action in the field of public health shall fully respect the responsibilities of the Member States for the definition of their health policy and for the organization and delivery of health services and medical care and the allocation of the resources assigned to them."

[71] This number is the expected revenue of the average drug. It is obtained by aggregating across drugs similarly to the way that the expected profit for the moment inequality was built. The only difference is that we used the full sample of 481 drugs to build this estimate.

[72] This conclusion might not hold in a model where additional frictions or dynamic considerations exist (e.g a shadow cost of raising governments funds, or dynamic implications on the incentives to invest in R&D). However,

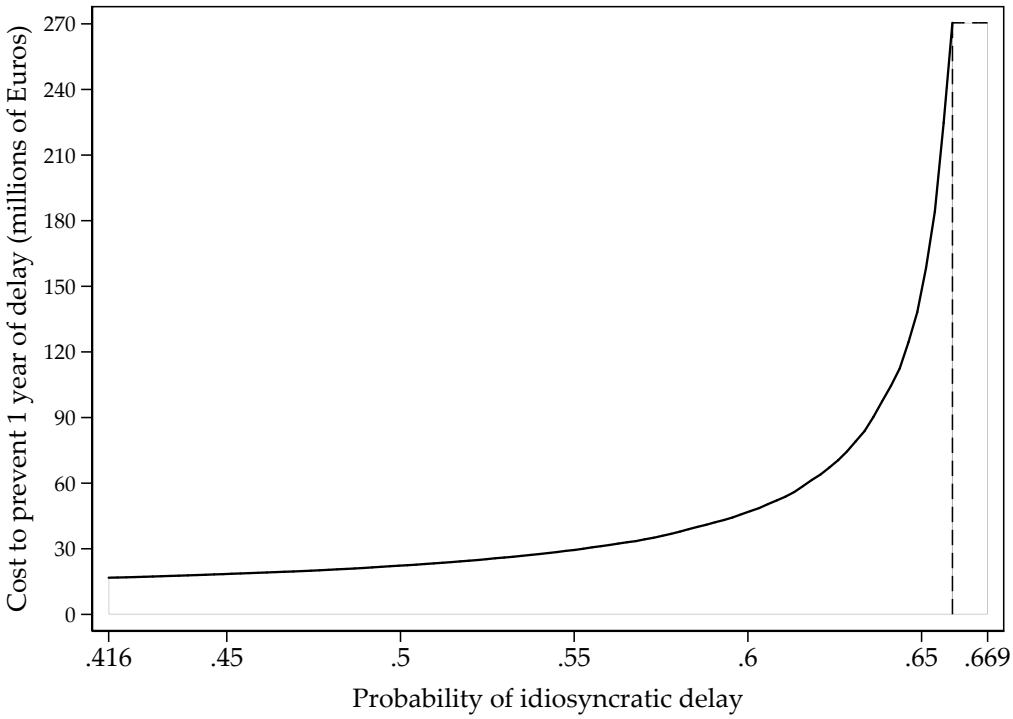


Figure 2.13: *Cost of Reducing Delays by 1 drug-year in Eastern European countries*

In this figure we plot the range estimate of the cost of reducing delays by 1 drug-year in all 8 Eastern European countries in our sample. The y-axis is truncated above because the curve tends to infinity as ψ_{EU10} approaches 0.669, which is the value for which all delays are idiosyncratic, and therefore subsidies cannot change the delay profile.

Figure 2.13 plots a distribution of the cost of reducing delays by 1 country-year for all countries in Eastern Europe over the range of the identified set of ψ_{EU10}. The purpose of this figure is not to argue whether paying the subsidy is worth it, but rather to exemplify the kind of tradeoff involved. A larger probability of idiosyncratic delays implies that subsidies have a smaller impact on overall delays, hence the cost increases. At the upper bound, all delays are idiosyncratic, and therefore subsidies cannot affect the entry sequence of any drug. However, for more reasonable values of the parameter, the cost to increase access is not prohibitive. If we again assumed that ψ_{EU10} is equal to ψ_{ITALY}, then it would cost approximately €43 million to reduce delays by one drug-year in all Eastern European countries. Given that Eastern Europe contains roughly 80 million individuals, this comes at the cost of 50 cents per person.

2.8 Conclusion and Next Steps

This paper tries to determine the extent to which external reference pricing policies contribute to the disparity in access to prescription drugs across countries. ERP generates complex incentives for firms who might benefit from strategically delaying entry in countries with low willingness to pay for drugs. Using a novel moment inequality approach, we characterize the impact of these policies on launch delays and on firm revenue. Our methodology allows us to obtain identification even though the firm's actions are unobserved, thus contributing to a growing body of literature showing how moment inequalities can make even the most complicated models tractable.

While generally speaking estimation using moment inequalities precludes the ability to run counterfactuals, we are able to simulate one important scenario: the removal of reference pricing. In this particular case, the solution to the problem is easy to derive: firms no longer have any reason to delay entry in any country, so the optimal strategy is simply to send applications everywhere right away. We find that if ERP were eliminated delays in Eastern Europe would fall by up to 14 months per drug in each Eastern European country.

the size of the subsidy is small enough that these additional considerations will likely be second-order.

Policymakers and industry insiders are both aware of the externality generated by ERP, and several proposals to replace reference pricing with alternative systems have been suggested (Kanavos *et al.*, 2011; OECD Health Policy Studies, 2008; Towse *et al.*, 2015; Vogler *et al.*, 2015). Our hope is that the analysis in this paper will contribute to the policy discussion on this important topic. One of our key results is that while the impact on delays is potentially large, revenue implications in the current equilibrium are relatively small. We exploit this insight to propose a centralized system of lump-sum subsidies that compensate firms for the profit loss of ERP in exchange for forgoing strategic delays. We believe that this approach is particularly promising, as it would not require any political transfer of power from individual countries to a central authority, and its budget impact would be minimal.

Our analysis also has some limitations. One of these limitations is that our model does not include a source of structural error, such as drug-country-year specific shocks in demand or price that are observed by the firm, but not by the econometrician. As we discuss in Section 2.5.1, unobservable demand shocks would likely lead us to underestimate the impact of strategic delays. Hence, this omission may lead us to underestimate the impact of reference pricing.

We also assume that our price model predicts prices exactly. Introducing a structural error in the price estimation would create an insurmountable econometric challenge, since this error would propagate through reference pricing channel in ways that would be difficult to account for. One possible avenue to relax this assumption might be to assume that the structural error on price is known to the firm, but not to foreign governments. This would prevent the error from propagating across countries. However, the intuition behind the model would not change much. This type of error could help justify earlier-than-predicted entry, but the opposite (later-than-predicted entry) is usually much more common in the data.

We assume that firms act as single agents even though our demand and price models imply that the actions of other firms can have an impact on revenue. Our demand model implies that entry of a competitor will negatively affect the market shares of all other products

within the therapeutic class, and our price function suggests that entry of a competitor may have a negative impact on price. This introduces a certain degree of internal inconsistency in the model, though empirically, these effects may not be large enough to elicit a strategic reaction — in particular, the effect of competitors on price is very noisy according to our estimation results. Allowing for multiple agents is unfortunately not feasible in the current environment, as the expected revenue of alternative strategies in the moment inequality can only be computed holding the strategy of other firms fixed. This is a more general problem in the moment inequality literature, which has usually been more successfully applied to single-agent problems, rather than games involving multiple agents.

Finally, there are also elements that we do not model explicitly. We take the regulatory environment as exogenous. Presumably, the government could choose the reference pricing function according to some optimal decision-making criterion. We ignore this possibility here. The question of why reference pricing is effective in lowering prices is also interesting. Experience suggests that governments do not need to resort to ERP if they want to implement price cuts (we observe several instances of temporary price cuts in the data that do not seem to be related to reference pricing). Indeed, there is some evidence that reference pricing is sometimes used only as a pretense for cuts that would have occurred anyway. For example, Greece changed its reference pricing function in 2010 to a formula that resulted in roughly 10-20% lower prices across the board. It is likely that these cuts would have been mandated regardless due to the financial situation of the country at the time. These considerations go beyond the scope of the paper (we believe it is reasonable to assume that these changes are exogenous from the point of view of the firm), but might provide fertile ground for future research.

Another area for future work is the translation of the delay loss generated by ERP into a welfare measure. Throughout the paper, we have focused on launch delays as our main outcome measure, because a reliable measure of welfare is hard to infer from the information available in our data. Translating delays into welfare might require a more sophisticated demand or price model. Alternatively, one could try to obtain data on health outcomes and

examine whether the introduction of specific drugs affected mortality rates (see for example Dubois and Kyle, 2017).

Chapter 3

Is Value-Based Pricing of Pharmaceuticals Optimal?[1]

3.1 Introduction

In October of 2012, just two months after ziv-aflibercept (brand name Zaltrap) had received FDA-approval for treating metastatic colorectal cancer, doctors at the Memorial Sloan-Kettering cancer center published an op-ed on the New York Times stating that they would not prescribe the new drug to their patients.[2] The reason? The drug cost about twice as much as its competitors, while offering roughly the same improvement in outcomes.

Over the past several years, excess spending on pharmaceutical products has loomed large over the healthcare industry. Many practitioners and policymakers are expressing a growing sentiment that drugs are priced above what they can offer in terms of therapeutic value. The may be right: shortly after the publication of the New York Times op-ed, Sanofi, the company who owns the marketing rights for Zaltrap, cut its price in half. In spite of this, the drug was still not deemed to be cost-effective by the UK's National Institute for Health and Care Excellence (NICE) in 2014, leading to its exclusion from the UK market.

[1] Co-authored with Josh Feng

[2] Bach *et al.* (2012), available at http://nyti.ms/OA0x88.

Value-based pricing has emerged as a potential solution to concerns that drug prices are increasingly misaligned with their therapeutic benefits. Several countries around the world, including the UK, Australia, and France, have incorporate cost-effectiveness evaluations in their pricing and reimbursement decisions. The US government has expressed an interest in these approaches: the Centers for Medicare and Medicaid Services proposed a Medicare Part B pilot in 2016 that would have involved the utilization of value-based purchasing tools (Centers for Medicare & Medicaid Services, 2016). The plan was abandoned later that year, but value-based pricing remains a frequent buzzword, most recently heard in Alex Azar's remarks to the Federation of American Hospitals in March of 2018.[3] On the private side of the market, many insurers have been pushing manufacturers to sign outcome-based contracts that tie reimbursements to the achievement of specific health benchmarks. While this is not necessarily the same as value-based pricing, it is nonetheless a step towards realigning value and price.

The goal of this paper is to critically evaluate the implications and feasibility of value-based pricing for pharmaceuticals, theoretically and empirically. From the point of view of theory our contribution is to provide a rigorous framework for thinking about optimal pricing policy in the pharmaceutical market. We first show that in a static model patient welfare is not maximized under value-based pricing, but rather through a pricing system that lowers the price of effective treatments and increases the price of ineffective ones. The intuition is fairly simple and holds across a broad range of models: lower prices maximize access, and therefore lowering the price of the most effective treatments is the best thing to do from the point of view of patients.

Our key theoretical result is that in a model where the social planner takes into account R&D incentives, value-based pricing has some desirable properties, though it is generally not optimal. When drugs are priced at value, all new innovations will generate a net positive impact on welfare. Moreover, this policy maximizes the probability of developing a new drug

[3]Full text available at https://www.hhs.gov/about/leadership/secretary/speeches/2018-speeches/remarks-on-value-based-transformation-to-the-federation-of-american-hospitals.html.

(subject to the constraint that the drug will have a net positive effect on welfare). Intuitively this occurs because value-based pricing aligns private incentives and public incentives. However, we also show that the optimal price is always lower than value-based pricing, unless patients are fully insured. Intuitively this happens because when patients are not fully insured, the social planner can expand access to drugs by lowering prices. Hence, it will usually be worth it to sacrifice a marginal innovation in order to expand access to all other ones.

This result hinges crucially on the ability of the government to measure drug value with accuracy and consistency. If estimates of drug value are noisy or biased, the implementation of value-based pricing is unlikely to generate any significant benefits.

Ultimately, this is an empirical question, and one that we turn to in the second part of the paper. We look at disease modifying treatments for multiple sclerosis and estimate the value of these therapies using three different methodologies: using clinical trial data, population health statistics, and finally data on consumer choices. We find that even though measures of value are loosely correlated with one another, there is significant heterogeneity across methodologies in the results we obtain. Moreover, the information about therapeutic value that is available upon FDA-approval is very noisy, particularly for what concerns the safety profile of new treatments.

Our paper builds upon and expands on previous research on value-based pricing. In our model, we attempt to formalize the intuition behind the two main motivations for value-based pricing: improving patient outcomes by ensuring a degree of consistency between value and price (Bach, 2015), and aligning firm's R&D incentives to the welfare generated by their products (Jayadev and Stiglitz, 2009). We find support only for the latter. Instead, our findings suggest that patient welfare may be better served by the use of value-based insurance design principles, which aim at lowering patient costs for undervalued therapies (and vice-versa for overvalued therapies).

Other papers have also attempted to establish theoretical frameworks for value-based pricing. Our model is loosely related to Danzon *et al.* (2015), who argue that value-based

pricing caps might approximate first-best outcome if the firm is able to use indication pricing (e.g. through the use of personalized medicine). Lakdawalla and Sood (2009) point out that public insurance can be welfare-enhancing when dynamic incentives of firms are taken into account. This is the same intuition we use in our model, though they do not apply to a study of value-based pricing. The interaction of value-based pricing with firm's R&D incentives in the context of the pharmaceutical market has also been studied in a variety of papers (see for example Hughes, 2011; Jena and Philipson, 2008; Mcguire *et al.*, 2008).

Our empirical section draws on methodologies that have been used both in economics and in the healthcare literature to estimate drug value. We discuss these more in depth in the main body of the paper, as well as in previous work (Feng and Maini, 2016). To the best of our knowledge, our paper is the first to directly compare different methodologies for estimating drug value.

The rest of the paper proceeds as follows. In Section 3.2 we provide a definition of drug value, and discuss ways in which it can be measured, as well as some practical approaches to incorporate value into pricing considerations. In Section 3.3 we outline our theoretical framework for value-based pricing. In Section 3.4 we test various methodologies for estimating drug value to disease-modifying treatments for multiple sclerosis to check how consistent results across methodologies are. Section 3.5 provides some conclusive remarks.

3.2 What Does "Value" Mean?

Defining drug value is not easy. At a basic level, value can be interpreted as the effect of a drug on patient health outcomes, but it can also include many additional considerations, such as budgetary impact, scientific novelty, R&D costs, rarity of the disease, and number of existing treatments available. In this paper, we focus on aspects of value that may be reflected in patient choices (i.e. demand). We do so not because we believe other factors to be unimportant, but because their importance is likely to be subjective, and their impact hard to

measure.[4]

In this section, we first describe a few ways to empirically measure drug value, and then discuss what types of value-based pricing strategies have been proposed, or implemented in the pharmaceutical market.

3.2.1 Measuring Drug Value

Broadly speaking, there are three approaches to measuring drug value. Each one comes with some advantages and disadvantages.

First, drug value can be measured directly by looking at experimental data from clinical trials. This method is common in the medical literature, and forms the basis of most currently existing value-based pricing systems. It has also been occasionally used in economics to calculate quality-adjusted price indexes (Berndt *et al.*, 1996; Lucarelli and Nicholson, 2009). The key advantages of this approach is that it only relies on data that is already available by the time a drug enters the market, and that it makes it easy to draw a causal link. The main drawback is that clinical trial data may be noisy and have limited external validity (Rothwell, 2005; 2006). Moreover, these measures generally do not include considerations that are not strictly related to efficacy and safety, but that may still be important for patients, such as ease of use (Lee *et al.*, 2011; Pawaskar *et al.*, 2007).

Second, drug value can be measured by analyzing how new treatments affect population health, usually by exploiting heterogeneity in the availability of new treatments across geographical areas and over time. For example, Dubois and Kyle (2017) exploit variation in the introduction of oncology drugs in Europe to look at changes in site-specific cancer mortality. Budish *et al.* (2015) - who study underinvestment in early-stage cancer R&D - use differential progress in five-year survival rates across cancer types to estimate the quality-adjusted life years gained from various R&D policies. This method provides a relatively complete assessment of the impact of a drug. However, it usually leaves out a few important

[4]The interested reader should consult Sussex *et al.* (2013) for a discussion of how to value these additional aspects.

aspects of quality, such as the impact of side-effects on the quality of life. It also requires a significant amount of time to track health improvements, making it difficult to evaluate new drugs.

Finally, researchers can derive measures of drug quality from patient choices by estimating a demand system, using a revealed preference approach. These techniques usually require some structural assumptions in the form of a demand system (e.g. Arcidiacono *et al.*, 2013; Dunn, 2012), unless the researcher can identify an exogenous source of price variation that fully identifies the demand curve (e.g. Goldman *et al.*, 2010). This approach has two main advantages. First, it considers the decisions of the individuals who are actually going to be using the products, helping to capture idiosyncratic match quality between patient and drug; and second, it provides a comprehensive measure that synthesizes all drug characteristics, from efficacy and side effects to ease of use and more. However, it is predicated on the assumption that a patient's welfare is reflected through her choices. Evidence on patient behavior seems to suggest that in some instances this might not be the case. Many patients tend to prefer brand to generic drugs, despite the fact that generics are engineered to exactly replicate the active principle contained in brand medications (Bronnenberg *et al.*, 2015). Underusing and overusing of drugs has also been reported in various circumstances. Compliance with maintenance drugs is usually imperfect (Cramer, 2004; Huser *et al.*, 2005), while overuse of opioids has been widely documented in both media and economic research. The latter example also highlights the potential role of physician incentives in producing suboptimal outcomes (Schnell and Currie, 2017).

3.2.2 Approaches to Value-Based Pricing

Proponents of value-based pricing broadly argue that the price of pharmaceuticals should reflect their therapeutic value to patients and to the health system as a whole. In practice though, there are several different approaches to achieving this goal.

The main example of a value-based pricing system is the one in use in England, where the National Institute for Health Evaluation (NICE) is responsible for evaluating every new drug

in terms of its cost-effectiveness. Every time a new drug applies to receive reimbursement through the national health insurance system, NICE produces an estimate of the maximum price at which the drug would achieve a predetermined cost-effectiveness threshold. Firms seeking prices that would put the drug beyond that threshold are rarely approved (Dakin *et al.*, 2015; Devlin and Parkin, 2004; McCabe *et al.*, 2008).

Pay-for-performance contracts are another possible approach to value-based pricing. These agreements tie prices to the achievement of certain health outcomes in patients. Many European governments have already adopted these approaches for some drugs (Adamski *et al.*, 2010), and in recent years these contracts have garnered more attention among US private payers.[5] The main obstacle preventing the widespread diffusion of these contracts is health outcomes are generally hard to measure. Another potential drawback is that drugs are only one of many inputs into the health function of a patient, so by signing these contracts manufacturers would be exposed to some idiosyncratic risk. Finally, some critics have noted that even though pay-for-performance would tie payments more closely to outcomes, if would not necessarily tie them to value (Kaltenboeck and Bach, 2017). Furthermore, empirical evidence from European system has found that these types of contracts do not result in substantial savings (Navarria *et al.*, 2015).

Lastly, closely related to value-based pricing is indication-based pricing, which would allow manufacturers to charge different prices for drugs depending on the disease they are used to treat (Bach, 2014). Current regulation in the US makes indication-based pricing virtually impossible, since doctors can prescribe drugs for off-label use, and do not necessarily need to disclose the diagnosis of the patient. However, as with outcome-based contracting, indication-based pricing would not necessarily tie prices to value. Economic theory only suggests that indication-based pricing would allow firms to engage in price discrimination, thus allowing them to extract a larger share of consumer surplus (Chandra and Garthwaite, 2017). Whether this is be beneficial or not is an open and important question: with the rise of

[5]For example, CIGNA signed a pay-for-performance contract for the new class of PCSK9 anti-cholesterol drugs: http://www.wsj.com/articles/health-insurers-push-to-tie-drug-prices-to-outcomes-1462939262.

personalized healthcare, the scope for indication-based pricing is likely to expand (Garrison and Towse, 2017).

3.3 A Theoretical Framework for Value-Based Pricing

From the perspective of economic theory, there are several reasons why government intervention in the pharmaceutical market may be desirable. On the demand side, patients and providers often have imperfect information on the efficacy of treatments (Arrow, 1963), and on top of that the presence of insurance inflates demand through moral hazard (Pauly, 1968; Arrow, 1968). On the supply side, the patent system ensures that new products enjoy a certain degree of market power, at least for a limited time, thus allowing firms to charge large markups. Monopoly power generates potentially harmful, but necessary static inefficiency. At the margin, the need to provide firms with the necessary incentives for innovation must be balanced against the desire to expand access to medication.

In this section, we examine whether value-based pricing is the appropriate tool to address some of the existing frictions in the pharmaceutical market. We focus on two in particular. First, we examine whether value-based pricing is the optimal policy in the presence of systematic discrepancies between the actual value of drugs, and the value patients perceive. Next, we consider whether value-based pricing can be used as a tool to ensure that innovation incentives of firms align with social welfare.

3.3.1 A Demand-Side Model for Value-Based Pricing

A common discussion point in the medical literature is that value-based pricing would eliminate inconsistencies between the price that is charged to payers and the value of a product (Bach, 2015; Stabile *et al.*, 2013). Hence, value-based pricing regulation may be appropriate to correct instances where patient choices deviate from what a social planner would consider a fully efficient allocation. These deviations may occur for a variety of reasons: moral hazard may lead patients to choose treatments that are more valuable, but less cost-effective; behavioral biases may lead patients to under-utilize certain types of drugs (Baicker

et al., 2015), or show preferences for brand drugs over their generic equivalent (Bronnenberg *et al.*, 2015); financial incentives may lead physicians to increase prescriptions (Jacobson *et al.*, 2010); and finally, direct-to-consumer advertising may lead to shifts in demand as well (Shapiro, 2018; Sinkinson and Starc, 2018).

We consider a model where consumer choices may not reflect social welfare. We impose a minimal amount of restrictions on the form and nature of the bias generated by this discrepancy, and consider the problem of a benevolent social planner whose goal is to maximize welfare. We focus on static welfare only, and temporarily abstract from dynamic incentives. Hence, we leave out R&D considerations for now.

Let I denote the set of potential patients in a therapeutic market, indexed by i; and let J denote the set of treatments available in this market, indexed by j, including the outside option (no treatment). Each treatment j provides net welfare benefit w_{ij} to patient i.[6] Patients choose treatments by maximizing a choice utility function, which we do not model explicitly. Instead, we simply denote $\pi_{ij}(p)$ as the choice probabilities that arise from the maximization problem of patient i as a function of drug prices $p = \{p_j\}_{j=0}^{J}$. We make two restrictions on the properties of $\pi_{ij}(p)$.

Assumption 6. *For all $i \in I$ and $j \in J$, $\frac{\partial \pi_{ij}(p)}{\partial p_j} \leq 0$.*

Assumption 7. *For all $i \in I$, and $j, k \in J$ such that $j \neq j$, $\frac{\partial \pi_{ij}(p)}{\partial p_k} \geq 0$.*

These two assumptions are always satisfied in discrete choice demand models as long as demand is downward sloping. The second assumption may be violated in circumstances where patients can choose more than one product, specifically in the case of complementary products. In these cases, the assumption would still hold as long as the set of products is re-defined to include combination therapies separately.

The problem of the social planner is:

[6] We implicitly assume marginal costs of production are zero, though one could easily interpret w_{ij} as welfare net of costs without changing the solution or the interpretation of our findings.

$$\max_{p} \sum_{i \in I} \left(\sum_{j \in J} w_{ij} \pi_{ij}(p) \right) \tag{3.1}$$

Notice that prices in this model are a transfer from the payer (either the patient, the government, or a private insurer) to the firm, so their only impact is through $\pi_{ij}(p)$. Moreover, $\pi_{ij}(p)$ is a sufficient statistic because the choice utility of the patient does not otherwise appear in the problem of the social planner.

A general property of this model is that therapies that generate high welfare should have lower prices. The intuition is simple: by Assumptions 1 and 2, lowering the price of valuable treatments will nudge patients towards those treatments and away from treatments that are less beneficial on average.

To drive home the intuition, consider first the simplified case where the problem has a representative agent, i.e. $w_{ij} \equiv w_j$, and $\pi_{ij}(p) \equiv \pi_j(p)$. The first-order conditions of the problem can be written as

$$[p_k] : \sum_{j \in J} w_j \frac{\partial \pi_j(p)}{\partial p_k} = 0 \quad \forall k \in J$$

We can manipulate the FOC to obtain

$$\begin{aligned}
[p_k] : \quad & w_k \frac{\partial \pi_k(p)}{\partial p_k} + \sum_{j \neq k} w_j \frac{\partial \pi_j(p)}{\partial p_k} \\
= \quad & w_k \frac{\partial \pi_k(p)}{\partial p_k} + \sum_{j \neq k} w_k \frac{\partial \pi_j(p)}{\partial p_k} - \sum_{j \neq k} w_k \frac{\partial \pi_j(p)}{\partial p_k} + \sum_{j \neq k} w_j \frac{\partial \pi_j(p)}{\partial p_k} \\
= \quad & w_k \sum_{j \in J} \frac{\partial \pi_j(p)}{\partial p_k} + \sum_{j \neq k} (w_j - w_k) \frac{\partial \pi_j(p)}{\partial p_k} \\
= \quad & \sum_{j \neq k} (w_j - w_k) \frac{\partial \pi_j(p)}{\partial p_k} = 0
\end{aligned}$$

where the last step is a consequence of the fact that choice probabilities must sum up to 1, and therefore the sum of the partial derivatives must sum up to 0:

$$\sum_{j \in J} \frac{\partial \pi_j(p)}{\partial p_k} = 0$$

Now consider the product k such that $w_k \geq w_j$ for all products $j \in J$. The first-order condition

for product k will be weakly negative by Assumption 2, which means that the optimal price for product k is the lowest possible price, which in most practical examples will be the marginal cost of production. Conversely, if we consider the product ℓ such that $w_\ell < w_j$ for all products $j \in J$, then the first-order condition will be positive, which in turn implies that the implied price should be the highest possible.

In the general case, the first-order condition reduces to

$$[p_k] : \sum_{i \in I} \sum_{j \neq k} \left(w_{ij} - w_{ik} \right) \frac{\partial \pi_{ij}(p)}{\partial p_k} = 0 \quad \forall k \in J$$

Since there may not be a strictly dominating product (i.e. a product k such that $w_{ik} \geq w_{ij}$ for all $i \in I$ and $j \in J$), the problem may not have a corner solution as the version with a representative agent discussed above. However, it will still be generally true that higher w_{ik} will translate into a lower price for product k, all else equal (this is easy to see since the derivative of the FOC with respect to w_{ik} is negative).

The implication of the model runs counter to the intuition of value-based pricing: effective treatments should be provided at cost, while ineffective ones should be priced at high markups. Instead, this principle more accurately reflects value-based insurance design, whose goal is to lower the cost-sharing of undervalued therapies (Chernew *et al.*, 2007).

3.3.2 Supply-Side Model for Value-based Pricing

We now examine what changes if we incorporate dynamic incentives into the model. The social planner now sets prices while also considering a firm's incentives to invest in R&D.

In order to streamline notation and focus on the relevant aspects of the model, we make three simplifying assumptions relative to our earlier setup. First, we assume that there are at most two treatments on the market: an outside option whose value is normalized to 0, and a new treatment to be developed by a firm. Second, we assume that all treatments are covered by insurance.[7] Patients are only responsible for a fraction α of the price of a new treatment,

[7]While private insurance companies generally do not have coverage requirements, Medicare and Medicaid must cover all FDA-approved drugs.

leading to the possibility that price may exceed value for some patients. Third, we assume that demand accurately represents patient welfare.[8]

Our model covers two periods. In period 0, a firm decides whether to spend C in order to develop a new drug, which they will be able to market in the following period. For simplicity, we assume that if the firm invests, the innovation is always successful.[9]

Let $D(\cdot)$ denote demand for the new product. In period 1, in the absence of price regulation, the firm will choose the profit-maximizing price

$$p^{firm} = \arg\max_{p} D(\alpha p)\, p$$

where we assume for simplicity that the marginal cost of production is zero.[10] The firm will invest in period 0 as long as

$$\Pi\left(p^{firm}\right) = D\left(\alpha p^{firm}\right) p^{firm} \geq C \tag{3.2}$$

From the point of view of the social planner however, the innovation is only valuable as long as the additional welfare generated is greater than C:

$$W\left(p^{firm}\right) = \int_0^{D\left(\alpha p^{firm}\right)} D^{-1}(r)\, \mathrm{d}r > C \tag{3.3}$$

Depending on the shape of the demand function, it is possible that there may be instances in which equation 3.2 is satisfied, but equation 3.3 isn't (Figure 3.1b).

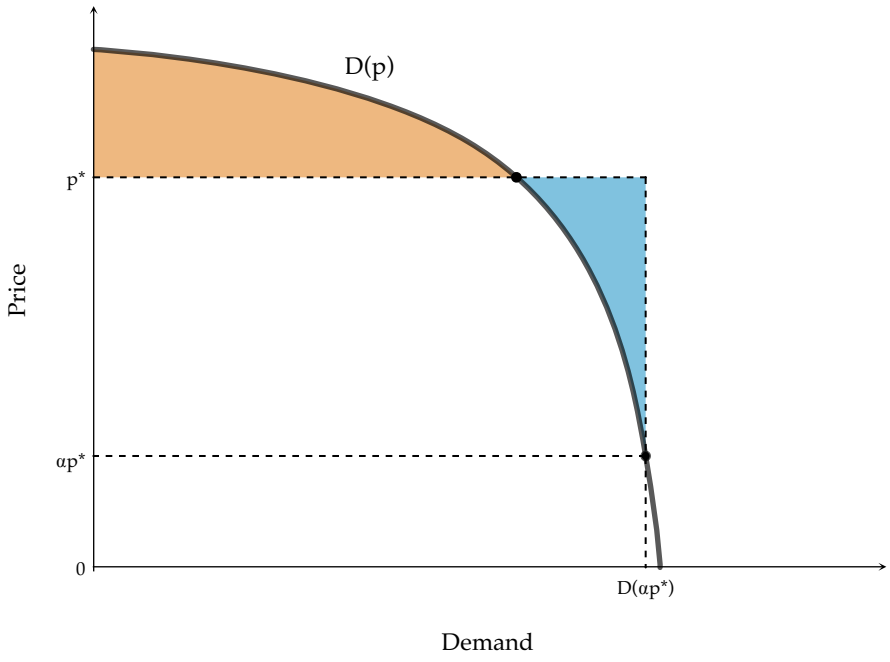
More generally, it is unlikely that p^{firm} would represent the optimal price from the point of view of the social planner, who maximizes a different objective function:

$$\max_{p} \int_0^{\Pi(p)} (W(p) - C)\, \mathrm{d}F(C) = W(p) \cdot F(\Pi(p)) - \int_0^{\Pi(p)} C\, \mathrm{d}F(C)$$

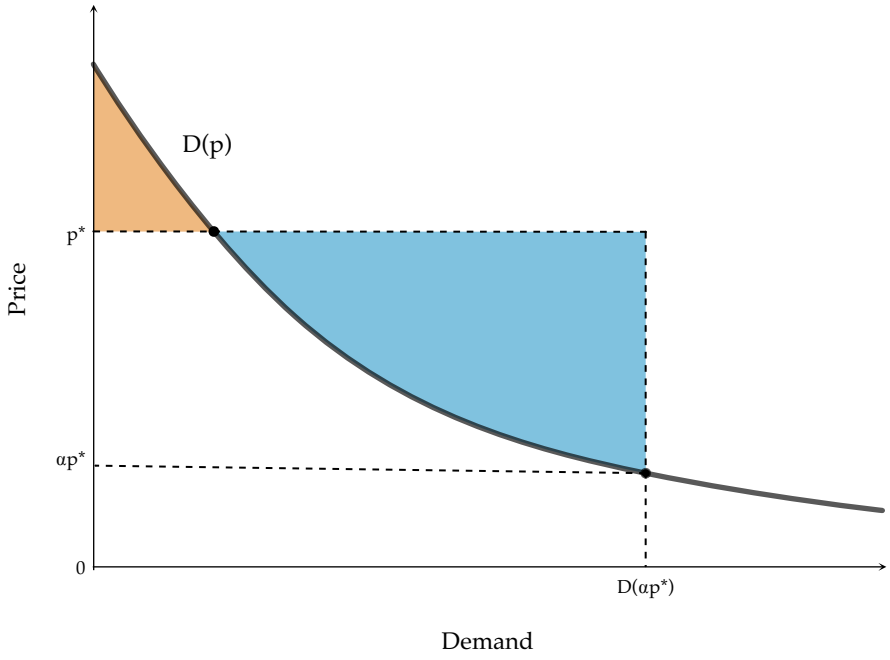
[8]This assumption is mostly out of convenience, as generalizing the model to an arbitrary welfare function $W(p)$ is straightforward. We do not do so for ease of exposition, and because moral hazard is sufficient to generate the incentive to price above value. Systematic biases in the evaluation of the clinical value of a drug may also lead to the same effect, but the implications would not be substantially different.

[9]Explicitly including a probability of success does not substantially change the analysis. C can easily be interpreted as the cost of innovation inclusive of eventual failed projects.

[10]This assumption is easy to relax as well.



(a) *Consumer welfare is greater than firm profits*



(b) *Firm profits are greater than consumer welfare*

Figure 3.1: *Firm profits with Moral Hazard*

where $F(C)$ is the cumulative distribution function of C. The first-order condition of the social planner's problem is

$$[p]: \underbrace{(W(p) - \Pi(p)) \cdot f(\Pi(p)) \cdot \frac{\partial \Pi(p)}{\partial p}}_{\text{innovation effect}} + \underbrace{\frac{\partial W(p)}{\partial p} \cdot F(\Pi(p))}_{\text{access effect}} = 0$$

The first term represents the increase in welfare from a marginal increase probability that the product is developed (*innovation effect*). In that case the potential gain is $W(p) - \Pi(p)$ (because for a product at the margin, $C = \Pi(p)$), and the increase in the probability is $f(\Pi(p)) \cdot \frac{\partial \Pi(p)}{\partial p}$ (i.e. the probability density at $\Pi(p)$ times the change in $\Pi(p)$ generated by a marginal increase in p). The second term represents the marginal loss in welfare times the probability that the product was going to be developed anyway (*access effect*). Intuitively, at the margin these two effects should be the same.

To see if this maps to value-based pricing we can rearrange the first-order condition to obtain

$$W(p) - \Pi(p) = -\frac{W'(p)}{\Pi'(p)} \cdot \frac{F(\Pi(p))}{f(\Pi(p))} \tag{3.4}$$

Value-based pricing implies $\Pi(p) = W(p)$. This will only be the case if

1. $W'(p) = 0$, i.e. welfare does not depend on price, which could be the case if, for example, patient reimbursement did not depend on price.

2. $F(\Pi(p)) = 0$, i.e. the probability that costs are below profits is zero (i.e. there are no profitable projects, in which case nothing is developed and both welfare and profits are equal to 0.

Hence, value-based pricing will generally *not* arise as the optimal solution to the social planner's problem. In fact, while the optimal solution is hard to characterize for the general case, one can easily see that it will always be *below* value-based pricing. Intuitively, when development costs equal total welfare, the net value of a marginal innovation is zero. At the same time, lowering the price marginally below value increases access (and welfare) for all products that would have been developed anyway. Hence, welfare should increase if prices

are lowered. Formally, one can see that this will be the case because the RHS of equation 3.4 is always greater than or equal to zero.[11]

Discussion

It is important to point out that this simple derivation hides some nuanced aspects. First, the notion of "value" in our model is endogenously related to price. The average value of a new innovation is tied to the number of patients who have access to the drug. This in turn, will depend on price. Under standard assumptions, a lower price will increase the number of patients who use the drug, which will increase the total contribution to welfare, but, since marginal patients should receive a lower benefit, decrease the value to the average patient.

Second, the results point to the fact that the characterization of the optimal pricing policy will depend on the choice of cost-sharing between the patient and the insurer. In particular, if the social planner can only affect pricing, but not cost-sharing, then the optimal pricing policy should take into account how an optimizing insurer would adjust coverage. In our treatment we have focused on the pricing problem alone, leaving a more complete treatment for future work.

Third, the implementation of the optimal policy (which is hard to characterize for the general case), requires the social planner to know the distribution of R&D costs for potential new projects. This is a primitive that is generally hard to know.[12] Furthermore, these costs are likely to differ substantially across project types. This may call for different pricing policies across different therapeutic areas, which could be hard to implement in practice. In this respect, value-based pricing has the advantage that its implementation does not require information about the distribution of R&D costs. It also has two properties that are likely to be perceived as desirable from a policy point of view. The first is that all new innovations

[11] The RHS could be negative if increasing p decreased firm profits (i.e. $\Pi'(p) < 0$). But that price would never be optimal anyway because if that were the case welfare could be unambiguously increased by lowering prices.

[12] A few papers have tried to estimate the cost of developing a new drug in the past. The most cited estimates come from a series of papers (DiMasi et al., 1991; 2003; 2016) that focus on a somewhat restricted set of drugs (Avorn, 2015).

will have a net positive effect on welfare. Value-based pricing equates welfare to profits, and firms will only develop a new drug if the expect its profits will exceed R&D costs. Second, it will maximize the number of new products that generate positive expected net welfare.

At the same time, correct implementation of value-based pricing still assumes that the social planner can accurately measure welfare. If the social planner can only observe a noisy signal of value, then value-based pricing will not satisfy the two properties mentioned above. Hence, the degree of accuracy with which value can be measured is an important empirical question. In the next section we test how well ex-ante measures of value (i.e. from clinical trials) predict drug quality in non-experimental data and revealed preferences.

3.4 Empirical Application: the Market for MS Drugs

The model developed in the previous section highlights

This is an empirical and a practical question. Given available data and information, how accurately can a government agency evaluate the quality of a drug?

3.4.1 Background

Multiple sclerosis is a disease of the brain and the spinal cord, in which the immune system attacks and disrupts the nerve fibers connecting the brain to the rest of the body. The disease is progressive, with widely varying symptoms across patients. Its most common form – Relapsing Remitting Multiple Sclerosis, or RRMS – is characterized by short episodes of increasing neurological symptoms (usually called relapses), followed by periods of partial or complete recovery (remissions).[13] The FDA has approved several so-called disease-modifying therapies (DMTs) for RRMS. The efficacy of these drugs is measured by the degree to which

[13]MS has four main types. Clinically Isolated Syndrome (CIS) consists of a first episode of MS-like symptoms, but not all patients who experience CIS go on to develop MS. Relapsing-Remitting MS (description in the main body of the text) affects roughly 85% of MS patients. Primary-Progressive MS is a more aggressive form of MS characterized by continuous worsening of neurological functions without the presence of relapses or remissions. Ocrelizumab (Ocrevus) is currently the only FDA-approved treatment for PPMS. Secondary Progressive MS is also characterized by a continuous worsening of neurological function, but only following an initial RRMS course. Most patients who are diagnosed with RRMS will eventually transition to SPMS.

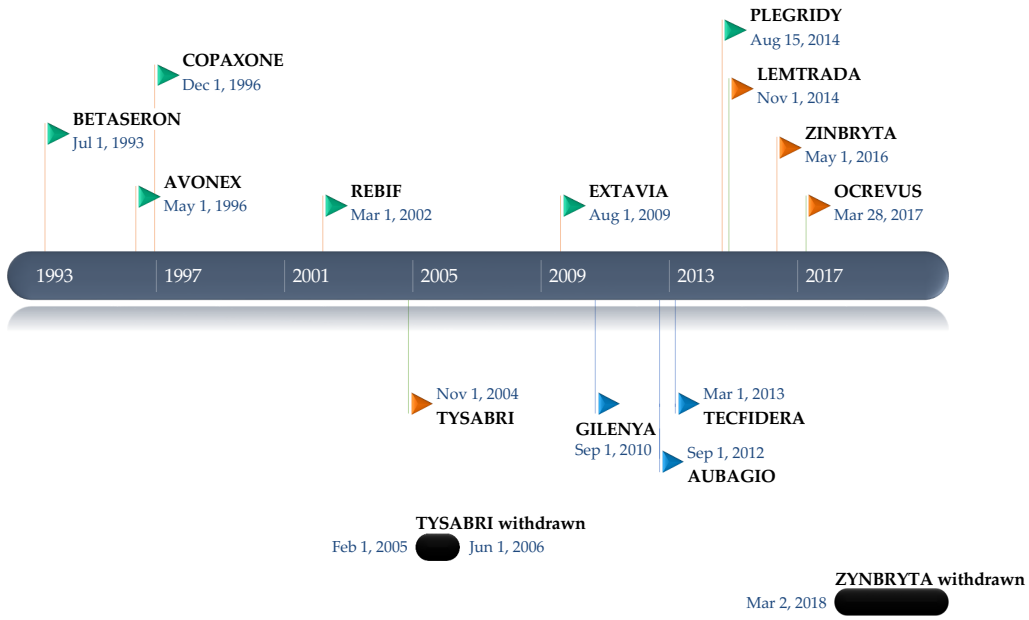


Figure 3.2: *Timeline of product approval in the MS market*

they can limit the occurrence of relapses.

DMTs for RRMS can be roughly divided in three groups (Figure 3.2). The first group of treatments consists of 6 injectables and includes the first three DMTs to be approved: interferon beta-1a (brand name Avonex), interferon beta-1b (Betaseron), and glatiramer acetate (Copaxone). The other three products in the class (Rebif, Extavia, and Plegridy) are also interferons. The second group of treatments consists of three oral drugs, approved between 2010 and 2013. Fingolimod (Gilenya) and dimethyl fumarate (Tecfidera) are generally though to be more effective than the injectable therapies, though they may carry an increased risk of severe side effects. The superiority of teriflunomide (Aubagio) over interferons has not been established. The third group consists of four monoclonal antibodies. These drugs are generally more effective, but also carry the risk of significant side effects. With the exception of ocrelizumab (Ocrevus), all drugs in this group carry a black box warning from the FDA, and two of them have been withdrawn for at least some time after approval due to safety concerns.

Data

Our analysis of this market relies on two main sources of data. The firs is a database of clinical trials and associated outcomes maintained by used one of the DMTs approved for RRMS, and were performed on a sample of patients with RRMS (we excluded trials on early-stage patients). The most common endpoint in MS clinical trials is the Annualized Relapse Rate (ARR), which measures the average number of relapses for a patient in a given year. This is also our measure of efficacy. Our data also reports the incidence of serious adverse events, which we use to evaluate the safety profile of each drugs. Table 3.1 reports the number and identification number of each trial we use in our analysis.

To evaluate drug quality in post-approval data we use a large claims database maintained by MarketScan, containing data on outpatient, inpatient, and prescription claims of employees (and dependents) of several large US firms. The data covers the period from 1995-2013, and we select a sample of patients who were diagnosed with MS on or after 2003. Figure 3.3 and Figure 3.4 show the evolution of market shares and prices for the five most popular products over the horizon of our analysis. Three of these products have been available since the late 1990s, while only two were introduced after the year 2000. Figure 3.3 shows that while market shares for older products are declining, they are still prescribed to the vast majority of the patient population. We also find that prices for these products are remarkably similar, and growing fast over the horizon we consider. This is not uncommon for prescription drugs, and subject to the caveat that what we observe in the data are list prices, which do not incorporate rebates and discounts.

3.4.2 Efficacy and Safety in Clinical Trials

In the first step of the analysis, we compare drug performance in clinical trials. Clinical trials are often the only source of information on new drugs when they first arrive on the market. In this section, we are interested in examining two things. First, how noisy estimates from clinical trials are, and second whether these estimates are reflected in the FDA assessment.

Figure 3.5 shows relative performance of DMTs for which we observe at least one clinical

Table 3.1: *Trial Characteristics*

Drug	# of Clinical Trials	Total Patient Enrollment	ct.gov Trial IDs
Aubagio	7	2391	NCT00134563 NCT00475865 NCT00489489 NCT00751881 NCT00811395 NCT00883337 NCT01252355
Avonex	2	501	NCT00605215 NCT00676715
Copaxone	1	350	NCT00451451
Gilenya	4	2542	NCT00289978 NCT00340834 NCT00355134 NCT00537082
Lemtrada	3	1024	NCT00050778 NCT00530348 NCT00548405
Nerventra	3	1188	NCT00349193 NCT00509145 NCT00605215
Ocrevus	1	110	NCT00676715
Plegridy	1	1012	NCT00906399
Rituxan	1	69	NCT00097188
Tecfidera	2	1530	NCT00420212 NCT00451451
Tysabri	2	1216	NCT00027300 NCT00030966
Zinbryta	1	919	NCT01064401

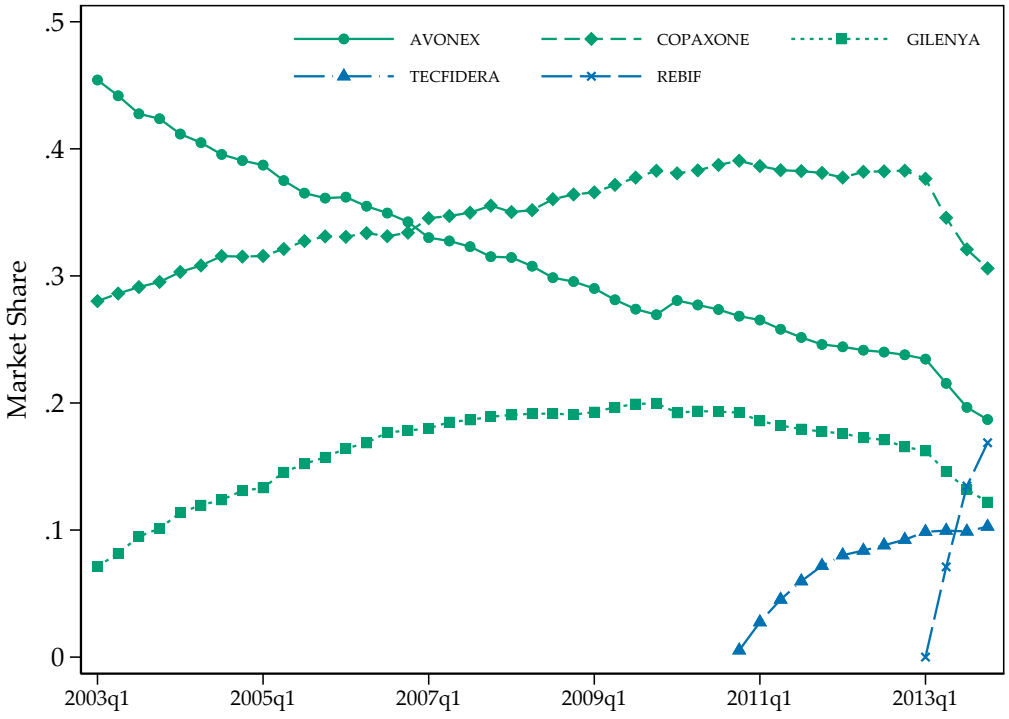


Figure 3.3: *Market shares of the top 5 products in the MS class, 2003-2013*

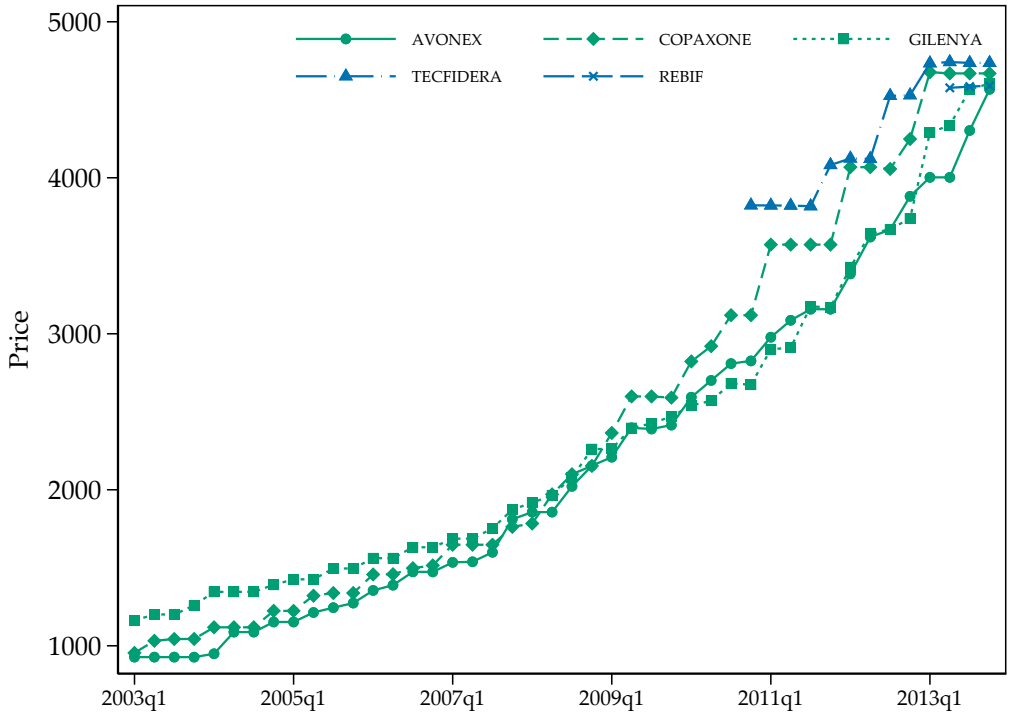
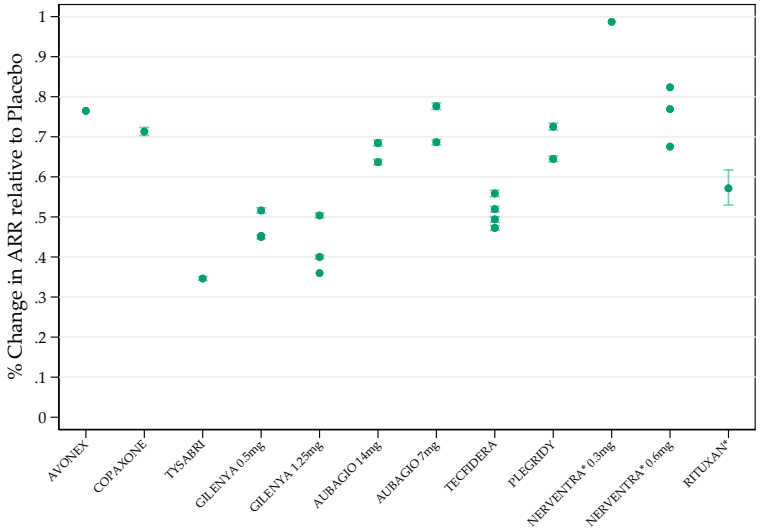
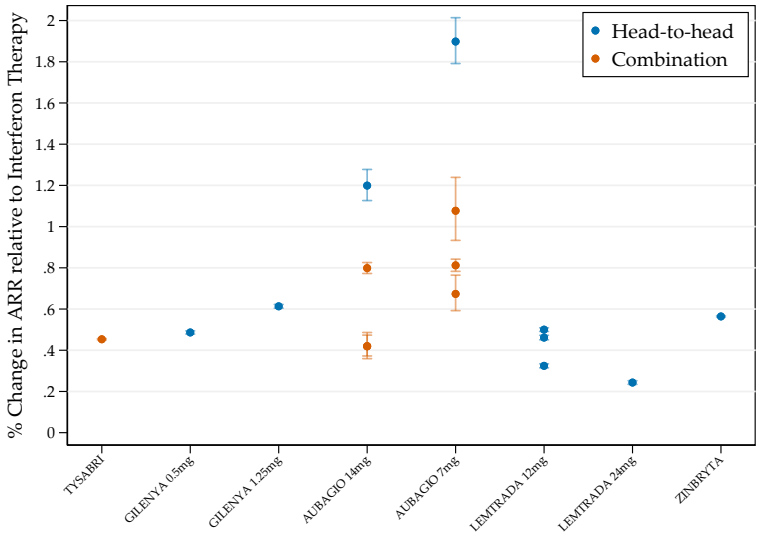


Figure 3.4: *Prices for the top 5 products in the MS class, 2003-2013*



(a) *Drug Efficacy relative to Placebo*



(b) *Drug Efficacy relative to Interferon-based Therapy*

Figure 3.5: *Drug Efficacy in Clinical Trials*

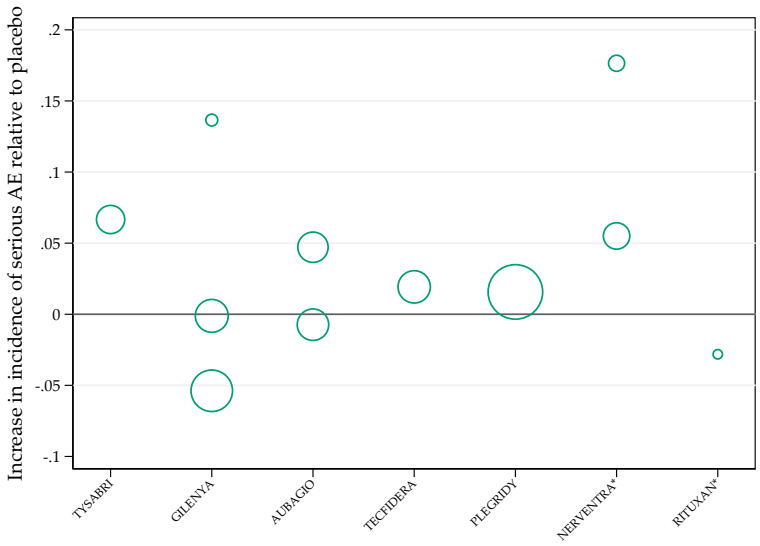
This figure plots the ratio between the Annualized Relapse Rate between each drug and a control group that received either a placebo (Sub-Figure 3.5a), or an interferon-based therapy (Sub-Figure 3.5b). Standard errors are calculated whenever possible, using the Fieller method (some trials did not report confidence intervals for their estimates). Combination refers to clinical trials where the treatment group was administered the drug together with interferons as a combination therapy, while Head-to-head refers to clinical trials in which the treatment group was administered the drug as the only therapy. * refers to therapies that have not yet received FDA approval (as of March 2018).

trial against placebo (panel 3.5a) and interferon-based therapy (panel 3.5b). Virtually all drugs significantly reduce the annual relapse rate when compared to a placebo. The point estimates for each drug are generally very precise, though we detect some variation across clinical trials. Point estimates range from 0.8 (a 20% reduction) to 0.35 (a 65% reduction), suggesting that there are substantial differences across products. The new generation therapies seem to perform better than the interferons in the comparison with placebo, and this impression is confirmed when we look at trials that pit these drugs against one another (panel 3.5b). With the exception of Aubagio, all new drugs significantly outperform interferons in clinical trials.

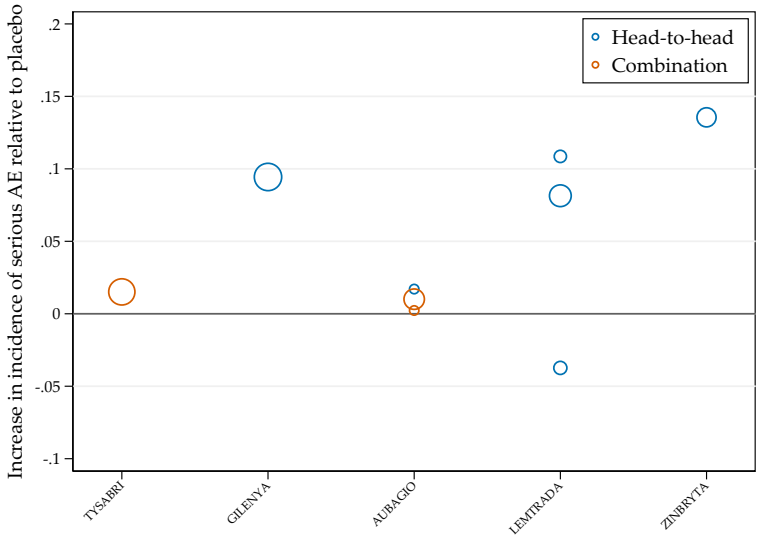
Figure 3.6 shows the relative incidence of serious adverse events.[14] We once again consider performance against placebo (panel 3.6a) and interferon therapy (panel 3.6b). Some of the new drugs, like Lemtrada and Nerventra appear less safe than the alternative. Patients on Tysabri experience a higher rate of serious adverse events than patients in the placebo control arm, but in the comparison with interferon therapy the two seem identical. The reverse holds for Gilenya, which appears safer when compared to a placebo than when compared to interferon-based therapy. Combined, these results suggest that the estimate are quite noisy. Indeed, conventional wisdom about clinical trials suggests that while efficacy can generally be precisely measured, the enrollment numbers of the typical clinical trial are far too small to accurately measure the safety profile of a drug (Rothwell, 2005; 2006).

An examination of FDA labels reveals that the assessment of adverse events is only partially influenced by the results of clinical trials. Sometimes label warnings are added only after the drug is approved. For example, Tysabri was temporarily withdrawn a few months after marketing started because a dangerous side effect was spotted during post-clinical monitoring. The drug was then re-introduced with a black box warning. Furthermore, some warnings are based on information not reflected in clinical trials. Aubagio's label lists two warnings: the first, for hepatoxicity, is based on the effects of a similar chemical, leflumonide,

[14]Data on adverse events is reported less accurately than data on efficacy. Confidence intervals are not included, and our data often reports aggregate results for multiple arms, which forces us to discard some observations in instances when collapsed arms received different DMTs.



(a) *Drug Safety relative to Placebo*



(b) *Drug Safety relative to Interferon-based Therapy*

Figure 3.6: *Drug Safety in Clinical Trials*

This figure plots the difference in the incidence of serious adverse events between each drug and a control group that received either a placebo (Sub-Figure 3.6a), or an interferon-based therapy (Sub-Figure 3.6b). Scatter plot size is proportional to the number of patients in the treatment group. Combination refers to clinical trials where the treatment group was administered the drug together with interferons as a combination therapy, while Head-to-head refers to clinical trials in which the treatment group was administered the drug as the only therapy.

which has been linked to severe liver failure; the second, for pregnancy effects, is based on animal testing. Our overview thus suggests that measuring the effect of a new drug from clinical trial data may lead to inaccuracies, especially in terms of the safety profile of drugs.

3.4.3 Efficacy and Safety in Post-Approval Population Data

Next, we move to the analysis of drug value in population data. For this analysis we use our database of patient claims and select a sample of patients with at least one prescription claim for an MS disease-modifying treatment (DMT). The analysis is performed at the patient-quarter level, and patients are assigned a single drug in each quarter.[15] Our main goal is to obtain a reduced-form measure of efficacy and safety from claims data.

We identify relapses using an algorithm first introduced by Ollendorf *et al.* (2002), and validated by Chastek *et al.* (2010), which defines a relapse as an inpatient stay associated to an MS diagnosis code, or a combination of an outpatient stay followed by a prescription for a corticosteroid drug within 7 days. We compare relapse rates between 2003 and 2013 for the five drugs with the highest market share as of 2013.[16]

Two trends emerge from the data. First, relapse rates are trending down over time (Figure 3.7). This could be a selection issue with the data (the MarketScan sample can change over time), or it could be due to the fact that the overall ability of physicians to prevent relapse rates is increasing over time. Another potential explanation could be a selection mechanism, whereby patients with a more severe condition may be switching to new therapies or stop taking DMTs altogether. Second, relapse rate are relatively homogeneous and converging across products.[17]

To identify side effects we look at inpatient stays that are not associated with an MS

[15] Patients are usually prescribed only one DMT at a time. In the few instances when a patient had prescriptions for more than one DMT, we assigned the DMT with the most prescriptions.

[16] Our measures of efficacy and safety for other drugs tend to be quite noisy, due to small sample sizes.

[17] We also detect a spike for all products around the time of product introduction. We suspect this is due to a selection effect: sickest patients, for whom older products may be ineffective, are likely to be the quickest to try the a new product.

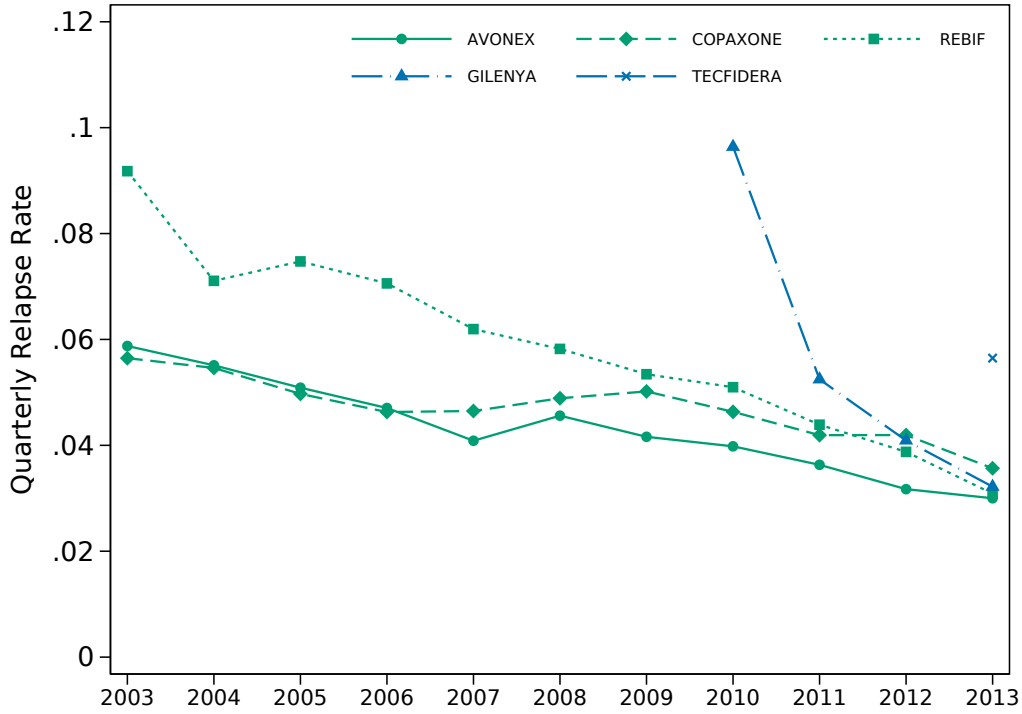


Figure 3.7: *Relapse rates by drug*

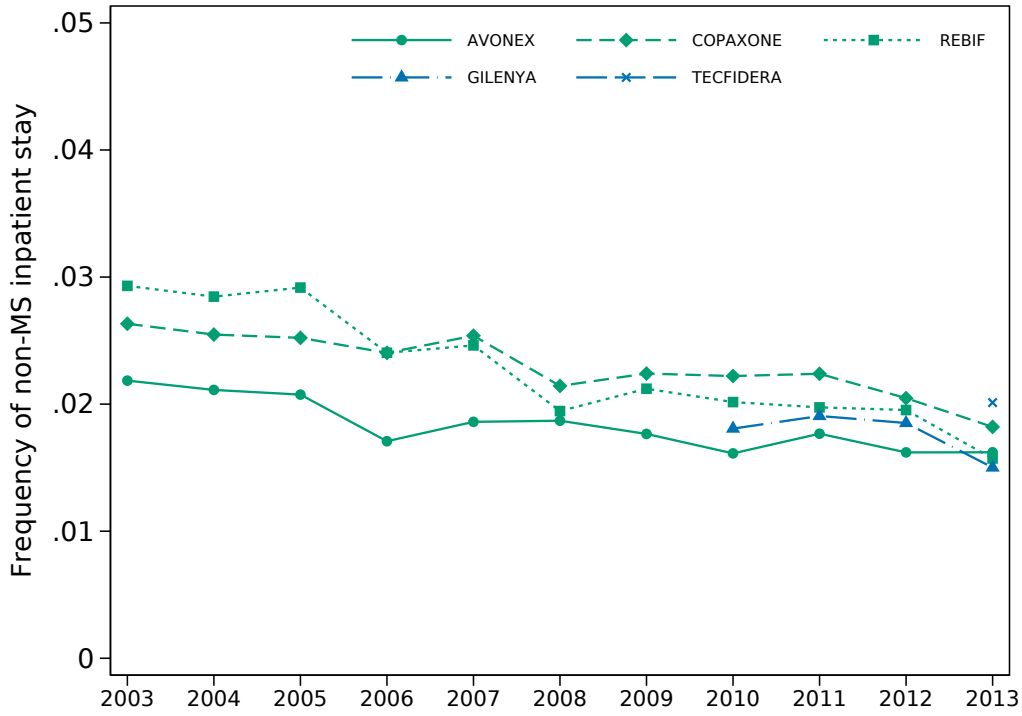


Figure 3.8: *Incidence of side effects by drug*

diagnosis code. This is obviously a relatively crude measure of side-effects: it will not pick up minor side effects that might nonetheless be important for patients, and it will not distinguish between different types of inpatient stays. However, due to sample size issues, it would be difficult to perform an analysis that is less aggregated than this. Similar to our results with relapse rates, we find that incidence of side effects is relatively similar across drugs (Figure 3.8).

In this simple comparison we do not correct for the naturally occurring selection that will happen in a non-experimental setting. Unlike in clinical trials, patients will not be randomly assigned to a drug, but will instead choose what they believe to be most appropriate therapy, together with the help of a physician. We try to address this concern by running an event study using the approval of Gilenya (in 2010) as a natural experiment. This identification strategy, first used in Feng (2018), is based on the fact that patients tend to display a certain degree of inertia. Hence, those who are diagnosed right before or right after the introduction

of a new drug will sometimes have different uptake rates even many years later.

We do this exercise using Gilenya because it is the only new drug with significant uptake that was introduced during our sample window.[18] Figure 3.9 shows the first stage results for a treated group of individuals who first started taking a DMT for MS in the 6 months following the launch of Gilenya, and a control group of individuals who started taking a DMT for the first time in the 6 months before Gilenya's launch. We split the control group in three sub-groups according to the DMT that was first prescribed. The split yields interesting results: patients in the control group who were first assigned to Rebif do not display signs of inertia, and the share of individuals who is on Gilenya is almost identical to the treated group a mere year after the launch of Gilenya. Conversely, patients on Avonex and Copaxone display a much lower propensity to switch: the fraction of individuals on Gilenya from these two groups is only about 5% three years after the launch of Gilenya, compared to 15% for the treated group.

When we test for differences in efficacy and safety in these two groups, our results seem to mimic the patterns observed in the pure observational data. Relapse rates for individuals in the treated arm appear similar to the relapse rates of individuals in the three subgroups of the control group (Figure 3.10).[19] The same holds for our measure of side effects, though the relationship is more noisy (Figure 3.11).[20]

The results from this analysis underscore the fact that capturing the value of a drug from population data is very complicated. Both our naive analysis and our event study yielded a similar, null result. This could be for a variety of factors. On the one hand, it is likely that we have not completely removed the effect of selection using our instrument. The fact that patients on Rebif immediately switched to Gilenya may be an indication that those

[18]Tecfidera also has significant uptake, but is only available in the last year of our sample, making it impossible to examine long run outcomes. Rebif has significant uptake, but is very similar to other interferons, and first-stage results (not included) show that the identification strategy is not very powerful. Tysabri and Aubagio do not achieve large enough market shares.

[19]The fact that relapse rates are very high at the beginning is due to the fact that most patients start a DMT in response to a relapse event.

[20]The spikes in both relapse events and side effects at the beginning of the sample are due to the fact that many patients start therapy with a DMT in response to an inpatient episode.

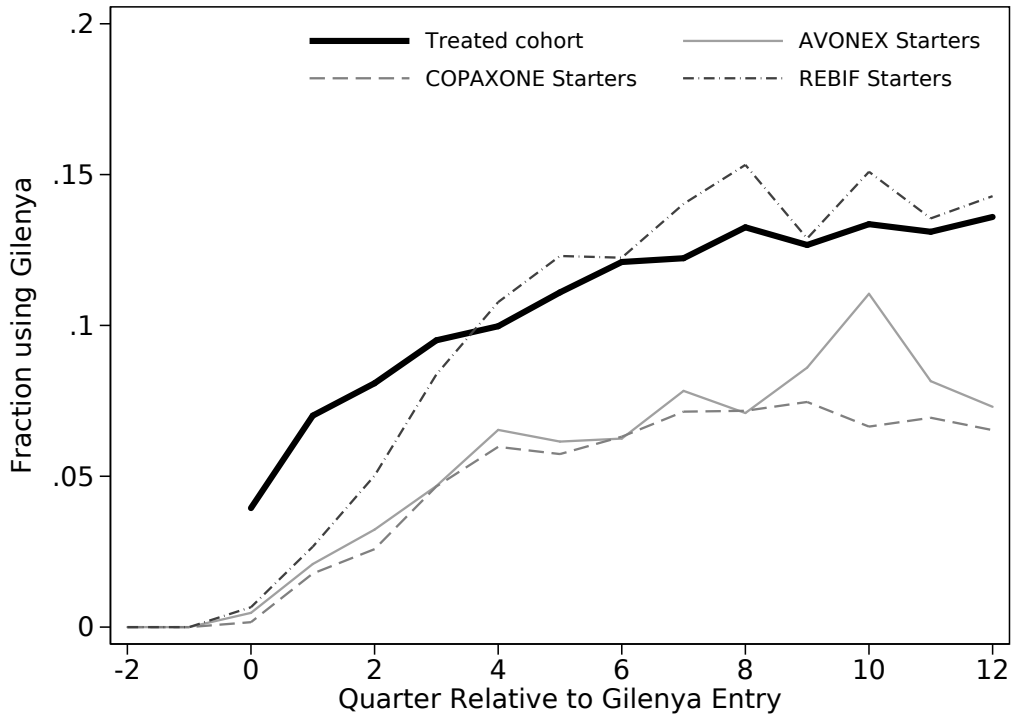


Figure 3.9: *Event study: First-Stage*

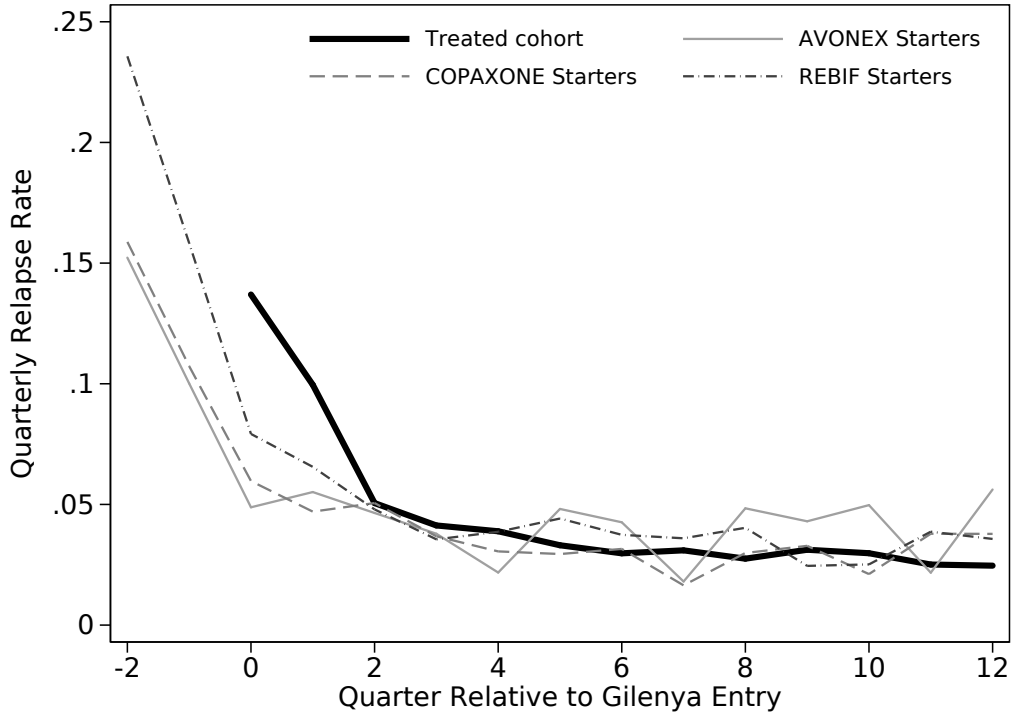


Figure 3.10: *Event study: relapse rates by group*

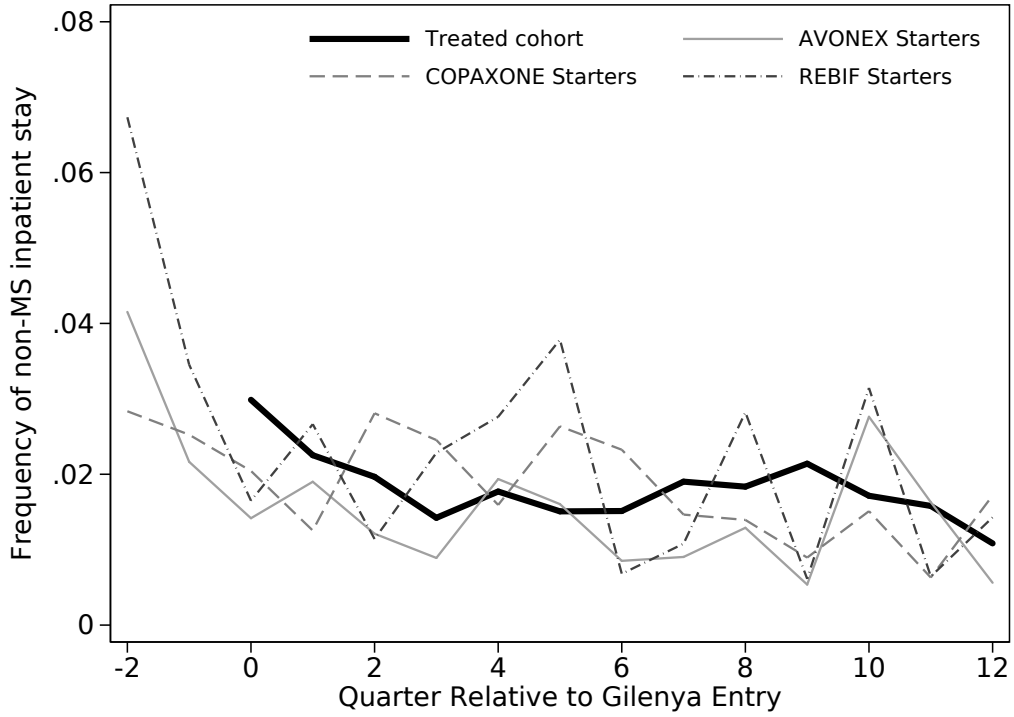


Figure 3.11: *Event study: incidence of side effects*

Table 3.2: *Switching patterns, 2013Q4*

	AUBAGIO	AVONEX	BETASERON	COPAXONE	EXTAVIA	GILENYA	REBIF	TECFIDERA	TYSABRI
AUBAGIO	93.60%	0.62%	0.31%	0.31%	0%	0.16%	0.62%	4.21%	0.16%
AVONEX	0.45%	93.80%	0.04%	0.45%	0.02%	0.43%	0.24%	4.53%	0.04%
BETASERON	0.31%	0.26%	91.04%	0.31%	0.10%	0.99%	0.26%	6.62%	0.10%
COPAXONE	0.66%	0.22%	0.06%	93.12%	0.01%	0.73%	0.32%	4.76%	0.12%
EXTAVIA	0.98%	0%	0.98%	2.94%	91.18%	0.98%	0%	2.94%	0%
GILENYA	0.26%	0.15%	0.04%	0.33%	0%	96.21%	0.04%	2.90%	0.07%
REBIF	0.65%	0.06%	0.14%	0.73%	0%	1.15%	91.03%	6.16%	0.08%
TECFIDERA	0.95%	0.64%	0.18%	1.22%	0.06%	0.70%	0.46%	95.69%	0.09%
TYSABRI	0.85%	0%	0%	0.21%	0%	1.06%	0.21%	5.71%	91.97%

patients were sicker to begin with (Figure3.7 seems to suggest that patients on Rebif suffer from slightly higher relapse rates relative to other interferons and Copaxone). Moreover, our results in the next section suggest that patients are more likely to switch drugs after suffering a relapse. Hence the patients that are not switching to Gilenya in our event study are more likely to be healthier. On the other hand, it is also possible that the performance of drugs in population data may differ relative to clinical trials. If this is the case, then it would suggest that using clinical trials to evaluate drug value may set up a biased benchmark.

3.4.4 A Revealed Preference Approach to Value

Finally, we estimate a demand system to infer a revealed-preference measure of value. For this analysis we build a panel of patients who were either diagnosed with MS or started treatment with an MS DMT between 2003 and 2013. The panel is balanced (no exit), but patients are included in the analysis only after they are diagnosed, or they start treatment. Hence, the total market size is increasing over the time period.

Raw switching patterns (Table 3.2) unsurprisingly reveal a strong degree of inertia. Inertia in individual choices is a well-documented fact in the choice of drug insurance (Ericson, 2014; Handel and Kolstad, 2015), but it has only recently been documented in the choice of

prescription drugs (Feng, 2018). Inertia in the choice of a DMT for MS is likely to be at least partially rational. Patients usually need a transition period during which the likelihood of experiencing side effects is greater. Another contributing factor is that treatment strategies for MS are based on escalating therapy after suboptimal responses (Comi *et al.*, 2017). In other words, patients and physician tend to stick with the current approach until it no longer seems to work.

We account for inertia in our demand estimates by including a control in the utility function of the agent. We parametrize the utility of patient i, for choosing drug j in period t as

$$u_{ijt} = \beta x_{ijt} + \alpha_{it} p_{jt} + \xi_{jt} + \varepsilon_{ijt}$$

where x_{ijt} includes a drug fixed effect and a series indicators for whether the drug chosen is the same as the drug chosen in the previous period (*incumbent* in Table 3.3), and whether the patient had an inpatient claim in the current period (*hospitalization*). In additional specifications we also separate hospital stays that had MS as the primary diagnosis code (*MS-primary*), had MS as either the primary or secondary diagnosis (*MS-hospitalization*, and *non-MS-hospitalization* for the reverse). We use an estimate of the patient cost-sharing for price. We treat ξ_{jt} as a random disturbance.[21]

We report results for a variety of specifications in Table 3.3. In column 1 we report results from a model with drug fixed effects and cost-sharing as the only controls. This model finds much higher coefficients for older drugs: Avonex, Copaxone, Rebif, and Betaseron have the highest coefficients. The only new drug with a similar coefficient is Tecfidera, which was launched towards the end of our sample periods, but managed to quickly achieve a large market share nonetheless. Even so, its coefficient is only the fourth largest.

Inserting an incumbent control has a large effects on values. Tecfidera has the largest coefficient, three times as big as the drug with the second-highest coefficient (Copaxone).

[21] In general, ξ_{jt} will be correlated with price. In this case, since patients usually pay a small fraction of the price or a copay, we disregard this possibility. Estimation using cost-sharing rates from other insurance plans as an instrument did not yield qualitatively different results (not shown).

Even so, we can see that these coefficients do not perfectly line up with the results suggested by clinical trials, according to which, for example, Gilenya is a much more effective drug than Avonex or Copaxone.

In columns (3) - (6) we explore whether we can identify active choice episodes in the data (i.e. periods in which agents re-evaluated their choice of therapy). We find that including a coefficient for hospitalization decreases the advantage of the incumbent drug. The effect is particularly strong if we isolate hospitalization episodes for which MS is the primary diagnosis. This finding is consistent with the idea that patients re-evaluate their choice of drug only when it stops working (i.e. when they experience a relapse). However, this only has a minor impact on the drug coefficients.

3.4.5 Discussion

The three approaches we used to measure drug quality yield different results. Using the first method, newer oral drugs like Gilenya and Tecfidera, and monoclonal antibodies like Tysabri and Lemtrada, appear to be the best options available, largely due to much higher efficacy, and, in most cases, a similar safety profile to other therapies. Using the second method, we could not pick up any meaningful difference in performance among the DMTs we observe in the data, though we were somewhat limited by sample size in our analysis. Using the third method, we find that choice patterns in the data indicate that older therapies are preferred, though the effect decreases when we control for inertia.

We perform a more careful comparison of the measures we obtained from clinical trials and from revealed preferences.[22] In Figure 3.12 we plot the drug FE obtained from our set of regressions against the average normalized annualized relapse rate in clinical trials.[23] When we do not account for inertia (Panel 3.12a), the drug quality implied by our estimates is actually positively correlated with ARR, which runs counter to common intuition. However,

[22]We leave out the measures obtained with the second methodology for now, as we believe that they may be affected by selection bias that we have not been able to account for yet.

[23]We normalized relapse rates with placebo at 1, and averaged results from multiple clinical trials to obtain this measure. Levels are therefore to be interpreted as percentage changes relative to placebo therapy.

Table 3.3: *Demand Estimates*

	(1)	(2)	(3)	(4)	(5)	(6)
cost-sharing	-0.0003***	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
AUBAGIO	0	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)	(.)
AVONEX	2.187***	0.388***	0.381***	0.383***	0.382***	0.381***
	(0.107)	(0.0905)	(0.0907)	(0.0906)	(0.0907)	(0.0907)
BETASERON	1.410***	-0.223*	-0.229*	-0.227*	-0.228*	-0.229*
	(0.108)	(0.0910)	(0.0912)	(0.0911)	(0.0912)	(0.0912)
COPAXONE	2.311***	0.602***	0.595***	0.598***	0.597***	0.595***
	(0.107)	(0.0904)	(0.0906)	(0.0905)	(0.0906)	(0.0906)
EXTAVIA	0.700***	-0.360**	-0.367**	-0.365**	-0.365**	-0.366**
	(0.154)	(0.118)	(0.118)	(0.118)	(0.118)	(0.118)
GILENYA	0.956***	0.143	0.137	0.140	0.139	0.138
	(0.113)	(0.0944)	(0.0947)	(0.0946)	(0.0946)	(0.0947)
REBIF	1.559***	-0.0302	-0.0373	-0.0345	-0.0361	-0.0372
	(0.107)	(0.0908)	(0.0910)	(0.0909)	(0.0910)	(0.0910)
TECFIDERA	1.469***	1.967***	1.975***	1.972***	1.974***	1.975***
	(0.114)	(0.0980)	(0.0982)	(0.0981)	(0.0982)	(0.0982)
TYSABRI	0.074	-0.867***	-0.873***	-0.872***	-0.872***	-0.873***
	(0.124)	(0.100)	(0.100)	(0.100)	(0.100)	(0.100)
incumbent		4.761***	4.781***	4.774***	4.778***	4.781***
		(0.0192)	(0.0195)	(0.0193)	(0.0194)	(0.0195)
incumbent × hosp.			-0.707***			
			(0.0881)			
incumbent × MS primary				-1.489***		
				(0.140)		
incumbent × MS-hosp.					-0.797***	-0.801***
					(0.101)	(0.101)
incumbent × non-MS hosp.						-0.439*
						(0.177)
N	714,983	714,983	714,983	714,983	714,983	714,983

Standard errors in parentheses

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

Table 3.4: *Regression of FE on efficacy and safety measures*

	FE from (1)			FE from (2)		
Efficacy	0.735		-2.472	0.445		-3.034
Safety		-6.938	-23.256		-32.518	-52.543

this may be explained by the fact that older drugs tend to benefit from inertia, and also tend to be less effective than newer therapies. When we control for inertia the relationship switches (Panel 3.12a), though the fixed effects for Gilenya and Tysabri still seem low relative to their efficacy as measured in clinical trials.

In Figure 3.13 we compare drug FE from our regression to the average normalized incidence of serious adverse events in clinical trials.[24] The results in the two specifications are similar, and show a negative relationship between revealed-preference quality and incidence of adverse events. In particular, the fact that both Gilenya and Tysabri have a higher incidence of serious adverse events may explain why their coefficients are low relative to their effectiveness.

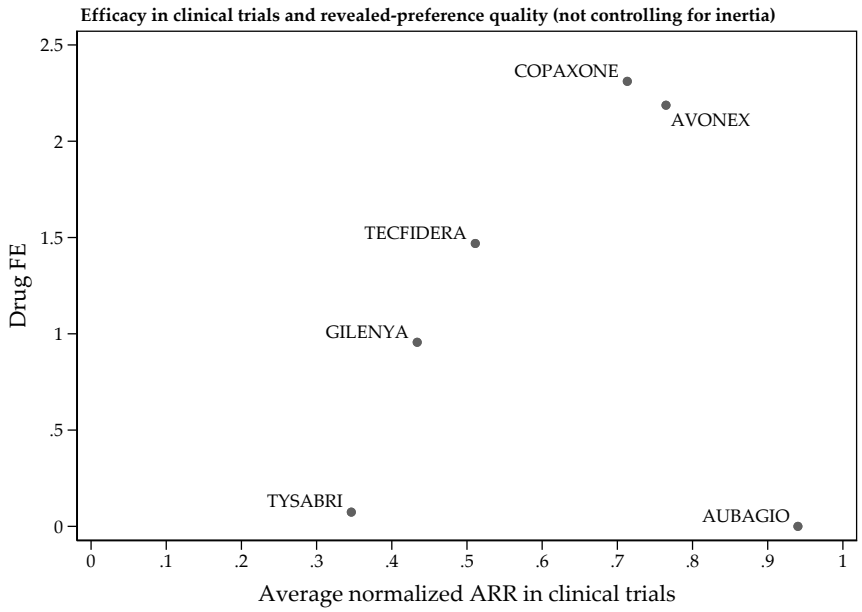
To check how fixed effect estimates correlate to both safety and efficacy, we run a set of simple linear regressions (Table 3.4).[25] The most notable observation from this simple exercise is that if we do not control for incidence of adverse events, the fixed effects from both specifications are positively correlated with relapse rates in clinical trials.

3.5 Conclusion

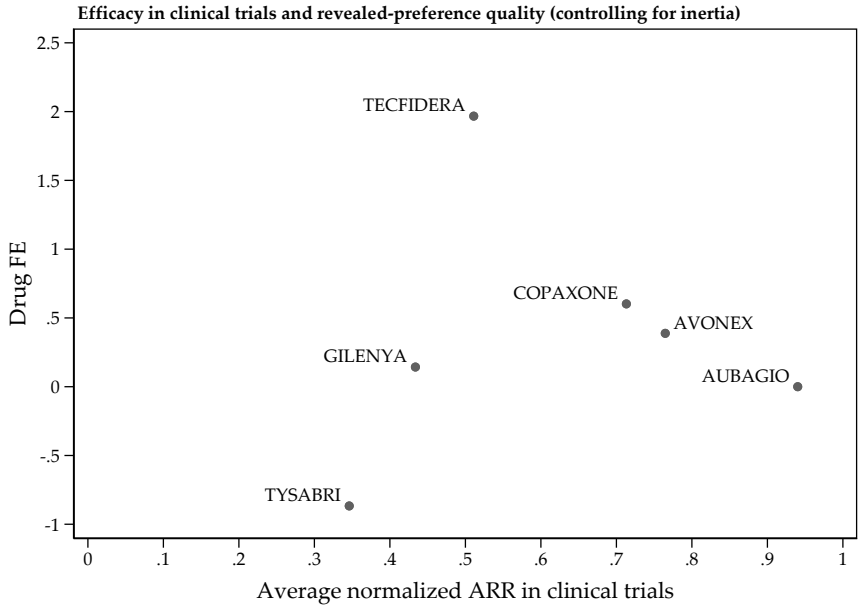
In this paper, we examine the theoretical and practical implications of a value-based pricing of pharmaceutical products. We rely on simple, but general economic models to demonstrate

[24]We normalized incidence of adverse events with placebo therapy at 0, and averaged results from multiple clinical trials to obtain this measure. Levels are therefore to be interpreted as percentage differences relative to placebo therapy. We did not have clean measures of safety for Copaxone and Avonex, so they do not appear in the graph.

[25]Since we only have a sample size of 6 for efficacy regressions and 4 for safety regressions we do not report standard errors.

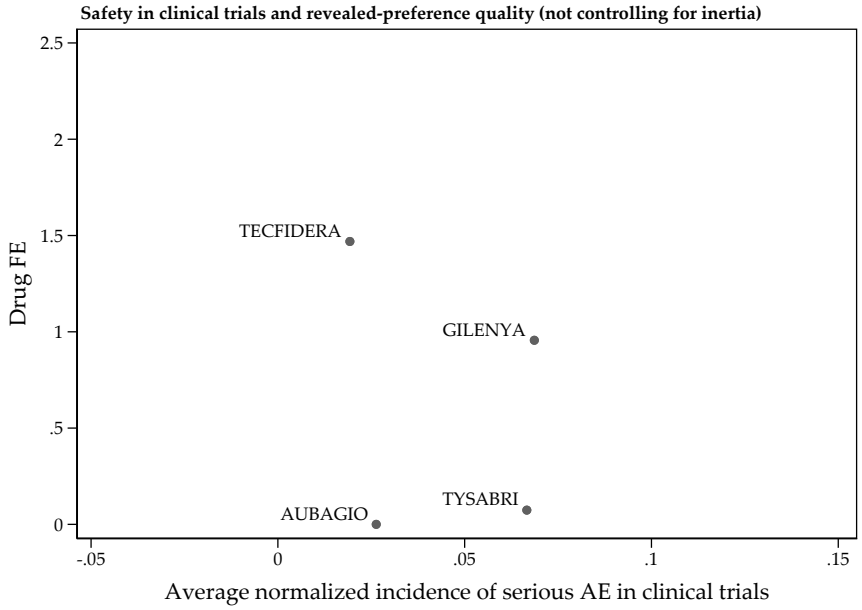


(a) *Drug FE estimates from column (1) of Table 3.3*

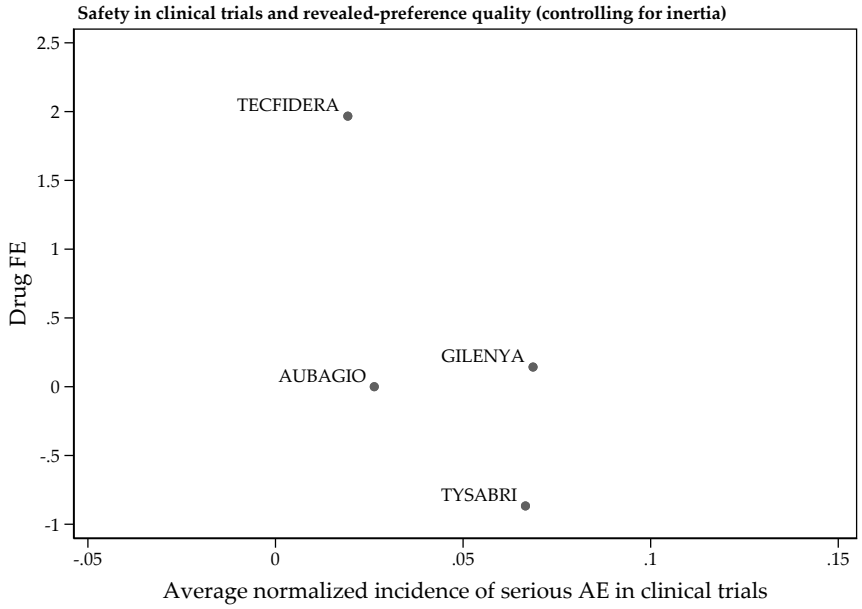


(b) *Drug FE estimates from column (2) of Table 3.3*

Figure 3.12: *Correlation between revealed preference and efficacy in clinical trials*



(a) *Drug FE estimates from column (1) of Table 3.3*



(b) *Drug FE estimates from column (2) of Table 3.3*

Figure 3.13: *Correlation between revealed preference and safety in clinical trials*

that value-based pricing of pharmaceutical does not lead to welfare maximization. We also show however, that the properties of value-based pricing make it a desirable and potentially easily implementable policy for a social planner that does not have any information on the cost of R&D, but still can measure drug value in some objective way. In the empirical part of the paper, we check whether drug value can be measured accurately. We test various methodologies and find that estimates are only loosely correlated.

Value-based pricing is intuitively appealing because it would ensure that patients do not overpay for medication, while also delivering the correct incentives to firms that engage in R&D. At the same time, the feasibility of implementing value-based pricing on a large scale has been questioned, both because of issues with the accuracy of value measurements, and because it may allow firms to capture a dominant fraction of consumer surplus.

This paper shows that both concerns are valid. From a theoretical point of view, we find that overall welfare would be greater if prices were below value. Furthermore, our empirical analysis suggests that measurements of drug value can be very noisy, which suggests that value-based pricing policies carry the risk of assigning the wrong price to many drugs.

There are other concerns with these policies that we have not addressed in the paper, but should be explored in future work. First, value-based pricing may affect the design of clinical trials. Firms already have an incentive to obtain a positive result, but their only objective is to clear a minimum threshold to win FDA approval. If the price of their product is tied to clinical trial performance, the incentive to obtain a large result will be magnified. Reporting of results and selection of patients will also be affected. Second, the value of drugs will fluctuate over time. For example, if a competitor loses exclusivity and a generic equivalent enters, the relative cost effectiveness of a branded product will decrease immensely. Whether and how this should be taken into account in value calculations is an open question. Finally, how to define value is something we only briefly touched upon, but that will have important repercussions. For example, if the definition of value does not include a premium for orphan diseases, the incentives to develop these drugs will be very small. What to include and what to not include in the definition is going to be a central issue in policy design.

References

ADAMSKI, J., GODMAN, B., OFIERSKA-SUJKOWSKA, G., OSIŃSKA, B., HERHOLZ, H., WENDYKOWSKA, K., LAIUS, O., JAN, S., SERMET, C., ZARA, C., KALABA, M., GUSTAFSSON, R., GARUOLIENÈ, K., HAYCOX, A., GARATTINI, S. and GUSTAFSSON, L. L. (2010). Risk sharing arrangements for pharmaceuticals: potential considerations and recommendations for European payers. *BMC Health Services Research*, **10**, 153–168.

AITKEN, M., BERNDT, E. R., CUTLER, D., KLEINROCK, M. and MAINI, L. (2016). Has The Era Of Slow Growth For Prescription Drug Spending Ended? *Health Affairs*, **35** (9), 1595–1603.

ARCIDIACONO, P., ELLICKSON, P. B., LANDRY, P. and RIDLEY, D. B. (2013). Pharmaceutical followers. *International Journal of Industrial Organization*, **31** (5), 538–553.

ARROW, K. J. (1963). Uncertainty and the Welfare Economics of Medical Care. *The American Economic Review*, **53** (5), 940–973.

— (1968). The Economics of Moral Hazard: Further Comment. *The American Economic Review*, **58** (3), 537–539.

AVORN, J. (2015). The $2.6 Billion Pill – Methodologic and Policy Considerations. *New England Journal of Medicine*, **372** (20), 1877–1879.

BACH, P. B. (2014). Indication-Specific Pricing for Cancer Drugs. *Journal of the American Medical Association*, **312** (16), 1629–1630.

— (2015). New Math on Drug Cost-Effectiveness. *The New England Journal of Medicine*, **373** (19), 1797–1799.

—, SALTZ, L. B. and WITTES, R. E. (2012). In cancer care, cost matters.

BAICKER, K., MULLAINATHAN, S. and SCHWARTZSTEIN, J. (2015). Behavioral Hazard in Health Insurance. *The Quarterly Journal of Economics*, **130** (4), 1623–1667.

BARNIEH, L., CLEMENT, F., HARRIS, A., BLOM, M., DONALDSON, C., KLARENBACH, S., HUSEREAU, D., LORENZETTI, D. and MANNS, B. (2014). A Systematic Review of Cost-Sharing Strategies Used within Publicly-Funded Drug Plans in Member Countries of the Organisation for Economic Co-Operation and Development. *PLoS ONE*, **9** (3), e90434.

BERNDT, E. R., COCKBURN, I. M., GRILICHES, Z., KEELER, T. E. and BAILY, M. N. (1996). Pharmaceutical Innovations and Market Dynamics: Tracking Effects on Price Indexes for Antidepressant Drugs. *Brookings Papers on Economic Activity. Microeconomics*, **1996**, 133.

BERRY, S. and REISS, P. (2007). Empirical Models of Entry and Market Structure. In *Handbook of Industrial Organization*, vol. 3, pp. 1845–1886.

BIRG, L. (2016). External Reference Pricing and the Choice of Country Baskets and Pricing Rules. *SSRN Electronic Journal*.

BORJA, G. L. (2014). Modeling Global Pricing and Launching of New Drugs.

BREKKE, K. R., CANTA, C. and STRAUME, O. R. (2016). Reference pricing with endogenous generic entry. *Journal of Health Economics*, **50**, 312–329.

—, HOLMAS, T. H. and STRAUME, O. R. (2015). Price Regulation and Parallel Imports of Pharmaceuticals. *Journal of Public Economics*, **129**, 92–105.

—, KÖNIGBAUER, I. and STRAUME, O. R. (2007). Reference pricing of pharmaceuticals. *Journal of Health Economics*, **26** (3), 613–642.

BRESNAHAN, T. F. and REISS, P. C. (1991). Entry and Competition in Concentrated Markets. *Journal of Political Economy*, **99** (5), 977–1009.

BRONNENBERG, B. J., DUBÉ, J.-P., GENTZKOW, M. and SHAPIRO, J. M. (2015). Do Pharmacists Buy Bayer? Informed Shoppers and the Brand Premium. *The Quarterly Journal of Economics*, **130** (4), 1669–1726.

BROWN, A. L., CAMERER, C. F. and LOVALLO, D. (2013). Estimating Structural Models of Equilibrium and Cognitive Hierarchy Thinking in the Field: The Case of Withheld Movie Critic Reviews. *Management Science*, **59** (3), 733–747.

BROWN, L. D., CAI, T. T. and DASGUPTA, A. (2001). Interval Estimation for a Binomial Proportion (with Discussion). *Statistical Science*, **16** (2), 101–133.

BUDISH, E., ROIN, B. N. and WILLIAMS, H. (2015). Do Firms Underinvest in Long-Term Research? Evidence from Cancer Clinical Trials. *American Economic Review*, **105** (7), 2044–2085.

CARDELL, N. S. (1997). Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity. *Econometric Theory*, **13** (02), 185.

CARONE, G., SCHWIERZ, C. and XAVIER, A. (2012). *Cost-containment policies in public pharmaceutical spending in the EU*. September.

CENTERS FOR MEDICARE & MEDICAID SERVICES (2016). Medicare Program; Part B Drug Payment Model; Proposed Rules. *Federal Register*, **81** (48), 13230–13261.

CHANDRA, A. and GARTHWAITE, C. (2017). The Economics of Indication-Based Drug Pricing. *New England Journal of Medicine*, **377** (2), 103–106.

CHASTEK, B. J., OLEEN-BURKEY, M. and LOPEZ-BRESNAHAN, M. V. (2010). Medical chart validation of an algorithm for identifying multiple sclerosis relapse in healthcare claims. *Journal of Medical Economics*, **13** (4), 618–625.

CHAUDHURI, S., GOLDBERG, P. K. and JIA, P. (2006). The Effects of Global Patent Protection in Estimating Pharmaceuticals : A Case Study of Quinolones in India. *American Economic Review*, **96** (5), 1477–1514.

CHERNEW, M. E., ROSEN, A. B. and FENDRICK, A. M. (2007). Value-Based Insurance Design. *Health Affairs*, **26** (2), w195–w203.

COCKBURN, I. M., LANJOUW, J. O. and SCHANKERMAN, M. (2016). Patents and the global diffusion of new drugs. *American Economic Review*, **106** (1), 136–164.

COMI, G., RADAELLI, M. and SØRENSEN, P. S. (2017). Evolving concepts in the treatment of relapsing multiple sclerosis. *The Lancet*, **389** (10076), 1347–56.

CONGRESSIONAL BUDGET OFFICE (2005). Prices for Brand-Name Drugs Under Selected Federal Programs. (June).

COSTINOT, A., DONALDSON, D., KYLE, M. and WILLIAMS, H. (2016). The More We Die, The More We Sell? A Simple Test of the Home-Market Effect.

COUNCIL OF EUROPEAN COMMUNITIES (1988). Council Directive of 21 December 1988 relating to the transparency of measures regulating the pricing of medicinal products for human use and their inclusion in the scope of national health insurance systems. *Official Journal of the European Communities*, **L** (40), 8–11.

CRAMER, J. A. (2004). A Systematic Review of Adherence With Medications for Diabetes. *Diabetes Care*, **27** (5), 1218–1224.

CRAWFORD, G. S. and SHUM, M. (2005). Uncertainty and learning in pharmaceutical demand. *Econometrica*, **73** (4), 1137–1173.

CRAWFORD, V. P. and MENG, J. (2011). New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income. *American Economic Review*, **101** (5), 1912–1932.

DAKIN, H., DEVLIN, N., FENG, Y., RICE, N., O'NEILL, P. and PARKIN, D. (2015). The Influence of Cost-Effectiveness and Other Factors on Nice Decisions. *Health Economics*, **24** (10), 1256–1271.

DANZON, P., TOWSE, A. and MESTRE-FERRANDIZ, J. (2015). Value-Based Differential Pricing: Efficient Prices for Drugs in a Global Context. *Health Economics*, **24** (3), 294–301.

DANZON, P. M. (2003). Prices And Availability Of Pharmaceuticals: Evidence From Nine Countries. *Health Affairs*, pp. 521–536.

— and EPSTEIN, A. J. (2012). Effects of regulation on drug launch and pricing in interdependent markets. *Advances in Health Economics & Health Services Research*, **23**, 35–71.

— and FURUKAWA, M. F. (2006). Prices And Availability Of Biopharmaceuticals: An International Comparison. *Health Affairs*, **25** (5), 1353–1362.

—, WANG, Y. R. and WANG, L. (2005). The impact of price regulation on the launch delay of new drugs?evidence from twenty-five major markets in the 1990s. *Health Economics*, **14** (3), 269–292.

DECAROLIS, F. (2015). Medicare Part D: Are Insurers Gaming the Low Income Subsidy Design? *American Economic Review*, **105** (4), 1547–1580.

DELOITTE (2013). *Impact of austerity on European pharmaceutical policy and pricing: Staying competitive in a challenging environment Contents*. Tech. rep.

DEVLIN, N. and PARKIN, D. (2004). Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Economics*, **13** (5), 437–452.

DICKSTEIN, M. and MORALES, E. (2018). What do Exporters Know?

DIMASI, J. A., GRABOWSKI, H. G. and HANSEN, R. W. (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics*, **47**, 20–33.

—, HANSEN, R. W. and GRABOWSKI, H. G. (2003). The price of innovation: New estimates of drug development costs. *Journal of Health Economics*, **22** (2), 151–185.

—, —, — and LASAGNA, L. (1991). Cost of innovation in the pharmaceutical industry. *Journal of Health Economics*, **10** (2), 107–142.

DORASZELSKI, U., LEWIS, G. and PAKES, A. (2018). Just Starting Out: Learning and Equilibrium in a New Market. *American Economic Review*, **108** (3), 565–615.

DRUMMOND, M. F. (2003). Will there ever be a European drug pricing and reimbursement agency? *The European Journal of Health Economics*, **4** (2), 67–69.

DUBOIS, P. and KYLE, M. (2017). The Effects of Pharmaceutical Innovation on Cancer Mortality Rates.

— and LASIO, L. (2017). Identifying Industry Margins with Unobserved Price Constraints: Structural Estimation on Pharmaceuticals.

DUGGAN, M. and SCOTT MORTON, F. M. (2006). The Distortionary Effects of Government Procurement: Evidence from Medicaid Prescription Drug Purchasing. *Quarterly Journal of Economics*, **121** (1), 1–30.

DUNN, A. (2012). Drug innovations and welfare measures computed from market demand: The case of anti-cholesterol drugs. *American Economic Journal: Applied Economics*, **4** (3), 167–189.

DUSO, T., HERR, A. and SUPPLIET, M. (2014). The Welfare Impact of Parallel Imports: a Structural Approack Applied to the German Market for Oral Anti-Diabetics. *Health Economics*, **23** (9), 1036–1057.

EIZENBERG, A. (2014). Upstream innovation and product variety in the U.S. home PC market. *Review of Economic Studies*, **81** (3), 1003–1045.

ELLISON, G. (2006). Bounded Rationality in Industrial Organization. In R. Blundell, W. K. Newey and T. Persson (eds.), *Advances in Economics and Econometrics*, vol. 2, Cambridge: Cambridge University Press, pp. 142–174.

ERICSON, K. M. 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.

EUROPEAN COMMISSION (2012). *Impact Assessment Accompanying the Document Proposal for a Directive of the European Parliament and of the Council relating to the transparency of measures regulating the price of medicinal products for human use and their inclusion in the scope of public*. Tech. rep.

EUROPEAN FEDERATION OF PHARMACEUTICAL INDUSTRIES AND ASSOCIATIONS (2014). *Principles for application of international reference pricing systems*. Tech. rep.

EUROPEAN FEDERATION OF PHARMACEUTICAL INDUSTRIES AND ASSOCIATIONS (2016). *The Pharmaceutical Industry in Figures*. Tech. rep.

EUROPEAN PARLIAMENT (2001). Directive 2001/83/EC of the European Parliament and of the Council. *Official Journal of the European Communities*, **L** (311), 67–128.

EUROPEAN PARLIAMENT (2004). Directive 2004/27/EC, of 31 March 2004 amending Directive 2001/83/EC on the Community code relating to medicinal products for human use. *Official Journal of the European Union*, **L** (726), 34–57.

FENG, J. (2018). History-Dependent Demand and Intermediaries: Explaining Prescription Drug Pricing Dynamics.

— and MAINI, L. (2016). Where Is the Good Stuff ? Drug Quality Measures and Economic Research. *NBER-IFS Network on the Value of Medical Research White Paper*.

GARRISON, L. P. and TOWSE, A. (2017). Value-based pricing and reimbursement in personalised healthcare: Introduction to the basic health economics. *Journal of Personalized Medicine*, **7** (3).

GOLDMAN, D. P., JENA, A. B., LAKDAWALLA, D. N., MALIN, J. L., MALKIN, J. D. and SUN, E. (2010). The value of specialty oncology drugs. *Health Services Research*, **45** (1), 115–132.

GRENNAN, M. (2013). Price Discrimination and Bargaining: Empirical Evidence from Medical Devices. *American Economic Review*, **103** (1), 145–177.

— and SWANSON, A. (2016). *Transparency and Negotiated Prices: The Value of Information in Hospital-Supplier Bargaining*. Tech. rep., National Bureau of Economic Research.

HANDEL, B. R. and KOLSTAD, J. T. (2015). Health insurance for "humans": Information frictions, plan choice, and consumer welfare. *American Economic Review*, **105** (8), 2449–2500.

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

HOLMES, T. J. (2011). The Diffusion of Wal-Mart and Economies of Density. *Econometrica*, **79** (1), 253–302.

HOUY, N. and JELOVAC, I. (2015). Drug Launch Timing and International Reference Pricing. *Health Economics*, **24** (8), 978–989.

HUGHES, D. A. (2011). Value-Based Pricing. *PharmacoEconomics*, **29** (9), 731–735.

HUSER, M. A., EVANS, T. S. and BERGER, V. (2005). Medication Adherence Trends With Statins. *Advances in Therapy*, **22** (2), 163–171.

ILLANES, G. (2016). Switching Costs in Pension Plan Choice.

JACOBSON, M., EARLE, C. C., PRICE, M. and NEWHOUSE, J. P. (2010). How medicare's payment cuts for cancer chemotherapy drugs changed patterns of treatment. *Health Affairs*, **29** (7), 1391–1399.

JAFFE, S. and SHEPARD, M. (2017). Price-Linked Subsidies and Health Insurance Markups.

JAYADEV, A. and STIGLITZ, J. (2009). Two Ideas To Increase Innovation And Reduce Pharmaceutical Costs And Prices. *Health Affairs*, **28** (1), w165–w168.

JENA, A. B. and PHILIPSON, T. J. (2008). Cost-effectiveness analysis and innovation. *Journal of Health Economics*, **27** (5), 1224–1236.

KALTENBOECK, A. and BACH, P. B. (2017). Outcomes-Based Drug Contracts Do Not Move Us Closer to Value. *Morning Consult*.

KANAVOS, P. and COSTA-FONT, J. (2005). Pharmaceutical parallel trade in Europe: stakeholder and competition effects. *Economic policy*, **20** (October), 751–798.

—, VANDOROS, S., IRWIN, R., NICOD, E. and CASSON, M. (2011). Differences in costs of and access to pharmaceutical products in the EU. *Policy Department A - Economic and Scientific Policy*.

KATZ, M. (2007). *Essays in Industrial Organization and Technological Change*. Ph.D. thesis.

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

— (2007). Pharmaceutical Price Controls and Entry Strategies. *Review of Economics and Statistics*, **89** (February), 88–99.

— (2011). Strategic Responses to Parallel Trade. *The B.E. Journal of Economic Analysis & Policy*, **11** (2).

— and QIAN, Y. (2013). Intellectual Property Rights and Access to Innovation: Evidence from TRIPS.

LAKDAWALLA, D. and SOOD, N. (2009). Innovation and the welfare effects of public drug insurance. *Journal of Public Economics*, **93** (3-4), 541–548.

LANTHIER, M., MILLER, K. L., NARDINELLI, C. and WOODCOCK, J. (2013). An improved approach to measuring drug innovation finds steady rates of first-in-class pharmaceuticals, 1987-2011. *Health Affairs*, **32** (8), 1433–1439.

LEE, L. J., LI, Q., REYNOLDS, M. W., PAWASKAR, M. D. and CORRIGAN, S. M. (2011). Comparison of utilization, cost, adherence, and hypoglycemia in patients with type 2 diabetes initiating rapid-acting insulin analog with prefilled pen versus vial/syringe. *Journal of Medical Economics*, **14** (1), 75–86.

LEOPOLD, C., VOGLER, S., MANTEL-TEEUWISSE, A., DE JONCHEERE, K., LEUFKENS, H. and LAING, R. (2012). Differences in external price referencing in Europe - A descriptive overview. *Health Policy*, **104** (1), 50–60.

LUCARELLI, C. and NICHOLSON, S. (2009). A quality-adjusted price index for colorectal cancer drugs.

LUTTMER, E. G. J. (1996). Asset Pricing in Economies with Frictions. *Econometrica*, **64** (6), 1439–1467.

MAINI, L. and PAMMOLLI, F. (2017). Reference Pricing as a Deterrent to Entry: Evidence from the European Pharmaceutical Market.

MANSKI, C. F. (2003). *Partial Identification of Probability Distributions*. Springer Series in Statistics, New York: Springer-Verlag.

MATTEUCCI, G. and REVERBERI, P. (2017). Drug innovation, price controls, and parallel trade. *International Journal of Health Economics and Management*, **17** (2), 159–179.

MCCABE, C., CLAXTON, K. and CULYER, A. J. (2008). The NICE cost-effectiveness threshold: what it is and what that means. *PharmacoEconomics*, **26** (9), 733–44.

MCGUIRE, A., RAIKOU, M. and KANAVOS, P. (2008). Pricing pharmaceuticals: Value based pricing in what sense? *Eurohealth*, **14** (2), 3–6.

MORALES, E., SHEU, G. and ZAHLER, A. (2017). Extended Gravity.

NAVARRIA, A., DRAGO, V., GOZZO, L., LONGO, L., MANSUETO, S., PIGNATARO, G. and DRAGO, F. (2015). Do the current performance-based schemes in Italy really work? "Success fee": A novel measure for cost-containment of drug expenditure. *Value in Health*, **18** (1), 131–136.

OECD HEALTH POLICY STUDIES (2008). *Pharmaceutical Pricing Policies in a Global Market*.

OLLENDORF, D. A., JILINSKAIA, E. and OLEEN-BURKEY, M. (2002). Clinical and economic impact of glatiramer acetate versus beta interferon therapy among patients with multiple sclerosis in a managed care population. *Journal of Managed Care Pharmacy*, **8** (6), 469–476.

PAKES, A. (2010). Alternative Models for Moment Inequalities. *Econometrica*, **78** (6), 1783–1822.

—, PORTER, J., HO, K. and ISHII, J. (2015). Moment Inequalities and Their Application. *Econometrica*, **83** (1), 315–334.

PANTELI, D., ARICKX, F., CLEEMPUT, I., DEDET, G., ECKHARDT, H., FOGARTY, E., GERKENS, S., HENSCHKE, C., HISLOP, J., JOMMI, C., KAITELIDOU, D., KESKIMÄKI, P. K., KRONEMAN, M., BASTIDA, J. L., BARROS, P. P., RAMSBERG, J., VUORENKOSKI, P., SPILLANE, S., VOGLER, S., VUORENKOSKI, L., WOUTERS, O., KILDEMOES, H. W. and BUSSE, R. (2016). Pharmaceutical regulation in 15 European countries. *Health Systems in Transition*, **18** (5), 1–118.

PAULY, M. V. (1968). The Economics of Moral Hazard: Comment. *The American Economic Review*, **58** (3), 531–537.

PAWASKAR, M. D., CAMACHO, F. T., ANDERSON, R. T., COBDEN, D., JOSHI, A. V. and BALKRISHNAN, R. (2007). Health Care Costs and Medication Adherence Associated with Initiation of Insulin Pen Therapy in Medicaid-Enrolled Patients with Type 2 Diabetes: A Retrospective Database Analysis. *Clinical Therapeutics*, **29** (6 PART 1), 1294–1305.

PHARMACEUTICAL INDUSTRY COMPETITIVENESS TASK FORCE (2006). *Competitiveness and Performance Indicators 2005*. Tech. rep.

REICH, M. R. (2000). The global drug gap. *Science*, **287** (March), 1979–1981.

ROTHWELL, P. M. (2005). External Validity of Randomised Controlled Trials: "To Whom Do the Results of This Trial Apply?". *Lancet*, **365** (9453), 82–93.

— (2006). Factors That Can Affect the External Validity of Randomised Controlled Trials. *PLOS Hub for Clinical Trials*, **1** (1), e9.

RUST, J. (1994). Structural estimation of markov decision processes. In *Handbook of Econometrics*, vol. IV, pp. 3081–3143.

SALLEE, J. M. (2014). Rational Inattention and Energy Efficiency. *The Journal of Law and Economics*, **57** (3), 781–820.

SCHNELL, M. and CURRIE, J. (2017). Addressing the opioid epidemic: Is there a role for physician education? *American Journal of Health Economics*.

SEIM, K. (2006). An empirical model of firm entry with endogenous product-type choices. *RAND Journal of Economics*, **37** (3), 619–640.

SHAPIRO, B. T. (2018). Promoting Wellness or Waste ? Evidence from Antidepressant Advertising.

SINKINSON, M. and STARC, A. (2018). Ask Your Doctor? Direct-to-Consumer Advertising of Pharmaceuticals. *The Review of Economic Studies*, (January), 1–46.

STABILE, M., THOMSON, S., ALLIN, S., BOYLE, S., BUSSE, R., CHEVREUL, K., MARCHILDON, G. and MOSSIALOS, E. (2013). Health care cost containment strategies used in four other high-income countries hold lessons for the United States. *Health Affairs*, **32** (4), 643–52.

STARGARDT, T. and SCHREYÖGG, J. (2006). Impact of cross-reference pricing on pharmaceutical prices: Manufacturers' pricing strategies and price regulation. *Applied Health Economics and Health Policy*, **5** (4), 235–247.

STERN, S. (1996). Market Definition and the Returns to Innovation: Substitution Patterns in Pharmaceutical Markets.

SUSSEX, J., TOWSE, A. and DEVLIN, N. (2013). Operationalizing value-based pricing of medicines: A taxonomy of approaches. *PharmacoEconomics*, **31** (1), 1–10.

THOMSON, S. and MOSSIALOS, E. (2010). Primary care and prescription drugs: coverage, cost-sharing, and financial protection in six European countries. *Issue brief (Commonwealth Fund)*, **82** (March), 1–14.

TOUMI, M., RÉMUZAT, C., VATAIRE, A.-L. and URBINATI, D. (2013). *External reference pricing of medicinal products : simulation- based considerations for cross- country coordination*. Tech. rep.

TOWSE, A., PISTOLLATO, M., MESTRE-FERRANDIZ, J., KHAN, Z., KAURA, S. and GARRISON, L. (2015). European Union Pharmaceutical Markets: A Case for Differential Pricing? *International Journal of the Economics of Business*, **22** (2), 263–275.

VOGLER, S., LEPUSCHÜTZ, L., SCHNEIDER, P. and STÜHLINGER, V. (2015). *Study on enhanced cross-country coordination in the area of pharmaceutical product pricing - Final Report*. Tech. Rep. Lot 2.

WILSDON, T., FIZ, E. and KIRKPATRICK, H. (2013). *The international impact of Swiss drug regulation*. Tech. rep.

Appendix A

Appendix to Chapter 1

A.1 Mathematical Appendix

A.1.1 Proof of Theorem 1

Theorem 1. *Assume that*

$$\int_{\mathcal{A}_i}\int_{\mathcal{A}_{-i}} r^2\left(a_i, a_{-i} z_i^o; \hat{\theta}^r\right)\cdot f_{\mathbf{s}_{-i}}\left(a_{-i}\middle|\,\theta^s\right) da_{-i}\, f_{\mathbf{s}_i}\left(a_i\middle|\,\theta^s\right) da_i < \infty$$

Then, for any $\varepsilon > 0$, there exists M' such that

$$\frac{1}{M}\sum_{k=1}^{M} r\left(a_{ik}^o, a_{-ik}^o, z_{ik}^o; \hat{\theta}^r\right) - \frac{1}{M}\sum_{k=1}^{M} R\left(\mathbf{s}_{ik}^\star, \mathbf{s}_{-ik} z_i^o; \theta_0^s\right) < \varepsilon$$

for all $M > M'$.

Proof. Let $\mathbf{R}_{ik}$ (boldface here denotes a random variable) be the measurement of the payoff obtained by agent i in market k when choosing the optimal strategy. The realization of this random variable depends on the realization of the actions a_{ik}.

First, note that by definition, the expected value of $\mathbf{R}_{ik}$ is

$$\mathbb{E}\left[\mathbf{R}_{ik}\right] = R\left(\mathbf{s}_{ik}^\star, \mathbf{s}_{-ik}, z_{ik}^o; \theta_0^s\right)$$

where $R(\cdot)$ is defined as in equation 1.5:

$$R\left(\mathbf{s}_{ik}^{\star}, \mathbf{s}_{-ik}, z_{ik}^{o}; \theta_0^s\right) = \int_{\mathcal{A}_i} \int_{\mathcal{A}_{-i}} r\left(a_{ik}, a_{-ik} z_{ik}^{o}; \hat{\theta}^r\right) \cdot f_{\mathbf{s}_{-ik}}\left(a_{-i} \middle| \theta_0^s\right) \mathrm{d}a_{-i} \, f_{\mathbf{s}_{ik}^{\star}}\left(a_i \middle| \theta_0^s\right) \mathrm{d}a_i$$

The measurement of the realized payoff of the agent – $r\left(a_{ik}, a_{-ik} z_{ik}^{o}; \hat{\theta}^r\right)$ – is a random draw from the distribution of $\mathbf{R}_{ik}$. Hence, the statement of the theorem is equivalent to saying that the series of random variables $\{\mathbf{R}_{ik}\}_{k=1}^{K}$ satisfies the strong law of large numbers.

A sufficient condition for this result is the convergence of the sequence $\omega_r = \sum_{k=1}^{r} \frac{\sigma_k^2}{k^2}$, where σ_k^2 is the variance of $\mathbf{R}_{ik}$ (Kolmogorov criterion). Notice that the statement of the theorem includes as an assumption that payoffs have bounded variance. Hence, the Kolmogorov criterion is satisfied, and the strong law of large numbers applies, which concludes the proof. ■

A.1.2 Proof of Theorem 2

Theorem 2. *Assume that*

$$\int_{\mathcal{A}_i} \int_{\mathcal{A}_{-i}} r^2\left(a_i, a_{-i} z_i^{o}; \hat{\theta}^r\right) \cdot f_{\mathbf{s}_{-i}}\left(a_{-i} \middle| \theta^s\right) da_{-i} \, f_{\mathbf{s}_i}\left(a_i \middle| \theta^s\right) da_i < \infty$$

Then, for any θ^s, and for any $\varepsilon > 0$, there exists M' such that

$$\frac{1}{M} \sum_{k=1}^{M} \left(\int_{\mathcal{A}_i} r\left(a_{ik}, a_{-ik}^{o}, z_{ik}^{o}; \hat{\theta}^r\right) \cdot f_{\mathbf{s}'_{ik}}\left(a_{ik} \middle| \theta^s\right) da_{ik} \right) - \frac{1}{M} \sum_{k=1}^{M} R\left(\mathbf{s}'_{ik}, \mathbf{s}_{-ik}, z_{ik}^{o}; \theta^s\right) < \varepsilon$$

for all $M > M'$.

Proof. I follow the proof of the previous theorem step by step. Denote $\mathbf{R}'_{ik}$ as the measurement of the payoff obtained by agent i in market k when playing strategy $\mathbf{s}'_{ik}$, and holding constant the strategy of other players. By definition, the expected value of $\mathbf{R}'_{ik}$ is

$$\begin{aligned} \mathbb{E}\left[\mathbf{R}'_{ik}\right] &= R\left(\mathbf{s}'_{ik}, \mathbf{s}_{-ik}, z_{ik}^{o}; \theta^s\right) \\ &= \int_{\mathcal{A}_i} \int_{\mathcal{A}_{-i}} r\left(a_{ik}, a_{-ik} z_{ik}^{o}; \hat{\theta}^r\right) \cdot f_{\mathbf{s}_{-ik}}\left(a_{-i} \middle| \theta_0^s\right) \mathrm{d}a_{-i} \, f_{\mathbf{s}_{ik}^{\star}}\left(a_i \middle| \theta^s\right) \mathrm{d}a_i \end{aligned}$$

Simulate a random draw from the distribution of $\mathbf{R}'_{ik}$ as

$$R'_{ik} = \int_{\mathcal{A}_i} \left(a_{ik}, a^o_{-ik} z^o_{ik}; \hat{\theta}^r\right) f_{\mathbf{s}'_{ik}} \left(a_{ik} \middle| \theta^s\right) \mathrm{d}a_{ik}$$

Under the assumptions of the theorem, the series of random variables $\left\{\mathbf{R}'_{ik}\right\}_{k=1}^K$ satisfies the strong law of large numbers, hence there exists M' such that

$$\frac{1}{M}\sum_{k=1}^{M}\left(\int_{\mathcal{A}_i} r\left(a_{ik}, a^o_{-ik}, z^o_{ik}; \hat{\theta}^r\right) \cdot f_{\mathbf{s}'_{ik}} \left(a_{ik} \middle| \theta^s\right) \mathrm{d}a_{ik}\right) - \frac{1}{M}\sum_{k=1}^{M} R\left(\mathbf{s}'_{ik}, \mathbf{s}_{-ik}, z^o_{ik}; \theta^s\right) < \varepsilon$$

for all $M > M'$. ■

A.1.3 Proof of Theorem 3

Theorem 3. *Let $R\left(\cdot; \theta^s\right)$ be defined as in equation 1.5 and $r\left(\cdot\right)$ be defined as in equation 1.4. The following two conditions hold:*

1. *If $\frac{\partial R(\cdot;\theta^s)}{\partial \theta^s} > 0$ for all θ^s, then*

$$\mathcal{M}\left(\left\{\mathbf{s}'_{ik}\right\}_{k=1}^K, \theta^s\right) = 0$$

   for all $\left\{\mathbf{s}'_{ik}\right\}_{k=1}^K$, $\theta^s \leq \theta^s_0$;

2. *if $\frac{\partial R(\cdot;\theta^s)}{\partial \theta^s} < 0$ for all θ^s, then*

$$\mathcal{M}\left(\left\{\mathbf{s}'_{ik}\right\}_{k=1}^K, \theta^s\right) = 0$$

   for all $\left\{\mathbf{s}'_{ik}\right\}_{k=1}^K$, $\theta^s \geq \theta^s_0$.

Proof. Consider the first statement. By assumption,

$$R\left(\mathbf{s}^\star_{ik}, \mathbf{s}_{-ik}, z^o_{ik}; \theta^s_0\right) > R\left(\mathbf{s}^\star_{ik}, \mathbf{s}_{-ik}, z^o_{ik}; \tilde{\theta}^s\right)$$

for all $\tilde{\theta}^s < \theta^s_0$. Moreover, by definition,

$$R\left(\mathbf{s}^\star_{ik}, \mathbf{s}_{-ik}, z^o_{ik}; \tilde{\theta}^s\right) \geq R\left(\mathbf{s}'_{ik}, \mathbf{s}_{-ik}, z^o_{ik}; \tilde{\theta}^s\right)$$

for all $\tilde{\theta}^s$, and for any $\mathbf{s}'_{ik}$. Hence, given any set of strategies $\left\{\mathbf{s}'_{ik}\right\}_{k=1}^K$, the following holds:

$$\frac{1}{K}\sum_{k=1}^{K} R\left(\mathbf{s}^\star_{ik}, \mathbf{s}_{-ik}, z^o_{ik}; \theta^s_0\right) - \frac{1}{K}\sum_{k=1}^{K} R\left(\mathbf{s}'_{ik}, \mathbf{s}_{-ik}, z^o_{ik}; \hat{\theta}^r\right) > 0$$

By Theorem 1

$$\frac{1}{K}\sum_{k=1}^{K} r\left(a_{ik}^{o}, a_{-ik}^{0}, z_{ik}^{o}; \hat{\theta}^{r}\right) - \frac{1}{K}\sum_{k=1}^{K} R\left(\mathbf{s}_{ik}', \mathbf{s}_{-ik}, z_{ik}^{o}; \hat{\theta}^{r}\right) > 0$$

which in turn implies that

$$\mathcal{M}\left(\left\{\mathbf{s}_{ik}'\right\}_{k=1}^{K}, \theta^{s}\right) = 0$$

The proof of the second statement is analogous. ■

A.1.4 Proof of Theorem 4

Theorem 4. *Let $R(\cdot;\theta^s)$ be defined as in equation 1.5 and $r(\cdot)$ be defined as in equation 1.4. Furthermore, let $F^{ub}\left(\mathbf{s}_{ik}, \mathbf{s}_{-ik}, z_{ik}^{o}; \theta^{s}\right)$ be a function such that*

$$F^{ub}\left(\mathbf{s}_{ik}, \mathbf{s}_{-ik}, z_{ik}^{o}; \theta^{s}\right) \geq R\left(\mathbf{s}_{ik}, \mathbf{s}_{-ik}, z_{ik}^{o}; \theta^{s}\right)$$

for all θ^s. Then, if

$$F^{ub}\left(\mathbf{s}_{ik}, \mathbf{s}_{-ik}, z_{ik}^{o}; \tilde{\theta}^{s}\right) < \frac{1}{K}\sum_{k=1}^{K} r\left(a_{ik}^{o}, a_{-ik}^{o}, z_{ik}^{o}; \hat{\theta}^{r}\right)$$

for some $\tilde{\theta}^s \in \Theta$, $\tilde{\theta}^s$ is rejected from the identified set.

Proof. Suppose that $\tilde{\theta}^s$ is the true value of θ^s, i.e. $\tilde{\theta}^s = \theta_0^s$. Then

$$\begin{aligned}\frac{1}{K}\sum_{k=1}^{K} r\left(a_{ik}^{o}, a_{-ik}^{o}, z_{ik}^{o}; \hat{\theta}^{r}\right) &= R\left(\mathbf{s}_{ik}, \mathbf{s}_{-ik}, z_{ik}^{o}; \theta_0^{s}\right) \\ &\leq F^{ub}\left(\mathbf{s}_{ik}, \mathbf{s}_{-ik}, z_{ik}^{o}; \theta^{s}\right) < \frac{1}{K}\sum_{k=1}^{K} r\left(a_{ik}^{o}, a_{-ik}^{o}, z_{ik}^{o}; \hat{\theta}^{r}\right)\end{aligned}$$

which is a contradiction.

■

Appendix B

Appendix to Chapter 2

B.1 Theoretical Derivations

In this Appendix we derive some results from the main body of the paper, and present a few complementary results that are not crucial to the results of the paper, but add robustness to the framework we use.

B.1.1 Logit Model

The utility of consumer ℓ, in country j, from consuming drug i (molecule m), belonging to therapeutic class κ, in year t is given by

$$u_{i(m,\kappa)\ell(j)t} = \delta_{ijt} + (\zeta_{m,\kappa} + (1-\sigma_\kappa)\,\varepsilon_{i\ell t})$$

where $\zeta_{m,\kappa}$, is common for all $i \in m$, and distributed according to the unique distribution such that if $\varepsilon_{i\ell t}$ is an extreme value random variable, then so is $\zeta_g + (1-\sigma)\,\varepsilon_{i\ell t}$ (Cardell, 1997). δ_{ijt} is parametrized as in Equation 2.2.

With this setup, one can show that the country j market share of *i within subset set m* is given by

$$s^m_{ijt} = \frac{\exp\left(\frac{\delta_{ijt}}{1-\sigma_\kappa}\right)}{D_m\left(\mathbf{X}_{mt}\right)} \tag{B.1}$$

where

$$D_m\left(\mathbf{X}_{mt}\right) = \sum_{k \in m} \exp\left(\frac{\delta_{kjt}}{1-\sigma_\kappa}\right)$$

and the market share of set *m within the overall market* is given by

$$s_{m/jt} = \frac{D_m\left(\mathbf{X}_{mt}\right)^{1-\sigma_\kappa}}{1 + \sum\limits_{h \in G_\kappa} D_h\left(\mathbf{X}_{ht}\right)^{(1-\sigma_\kappa)}} \tag{B.2}$$

where G_κ is the set of all molecules in class κ. Hence, the *overall market share of drug i* is

$$s_{ijt} = \frac{\exp\left(\frac{\delta_{ijt}}{1-\sigma_\kappa}\right) D_m\left(\mathbf{X}_{mt}\right)^{-\sigma_\kappa}}{1 + \sum\limits_{h \in G_\kappa} D_h\left(\mathbf{X}_{ht}\right)^{(1-\sigma_\kappa)}} \tag{B.3}$$

Derivation of the estimating equation Notice that the share of the outside option can be expressed as

$$s_{0jt} = \frac{1}{1 + \sum\limits_{h \in G_\kappa} D_h\left(\mathbf{X}_{ht}\right)^{(1-\sigma_\kappa)}} \tag{B.4}$$

Consider the log ratio of the market share of item i in group m to the outside good. According to the model, this can be expressed as

$$\ln\left(s_{ijt}\right) - \ln\left(s_{0jt}\right) \quad = \quad \left(\frac{\delta_{ijt}}{1-\sigma_\kappa}\right) - \sigma_\kappa \sigma \ln\left(D_m\left(\mathbf{X}_{mt}\right)\right)$$

Combining equations B.2 and B.4 we also obtain

$$\ln\left(D_m\left(\mathbf{X}_{mt}\right)\right) = \frac{\ln\left(s_{m/jt}\right) - \ln\left(s_{0jt}\right)}{1-\sigma_\kappa}$$

Hence we can write

$$\ln\left(s_{ijt}\right) - \ln\left(s_{0jt}\right) = \frac{\delta_{ijt}}{1-\sigma_\kappa} - \frac{\sigma_\kappa}{1-\sigma_\kappa \sigma}\left(\ln\left(s_{m/jt}\right) - \ln\left(s_{0jt}\right)\right)$$

which implies

$$\begin{aligned}
(1-\sigma_\kappa)\left(\ln\left(s_{ijt}\right) - \ln\left(s_{0jt}\right)\right) &= \delta_{ijt} - \sigma_\kappa\left(\ln\left(s_{m/jt}\right) - \ln\left(s_{0jt}\right)\right) \\
\implies (1-\sigma_\kappa)\ln\left(s_{ijt}\right) - \ln\left(s_{0jt}\right) &= \delta_{ijt} - \sigma_\kappa \ln\left(s_{m/jt}\right) \\
\implies \ln\left(s_{ijt}\right) - \ln\left(s_{0jt}\right) &= \delta_{ijt} + \sigma_\kappa \ln\left(\frac{s_{ijt}}{s_{m/jt}}\right)
\end{aligned} \tag{B.5}$$

B.1.2 Proof of Theorem 5

To prove the result of the theorem, we first prove the following Lemma:

Lemma 7. *Denote*

$$\lambda_{ijt} = \gamma_j \exp\left(\beta_Z Z_{ijt-1}\right)$$

and let $\lambda_{it} = \left\{\lambda_{ijt}\right\}_{j \in \mathcal{N}_i}$. *Then, there exists a set of weights* $\omega_{ijkt}\left(S_{t-1}, \{\lambda_{i\tau}\}_{\tau=1}^{t}\right)$ *such that for any drug* i*, country* j*, and year* t

$$p_{ijt}^{ref} = \sum_{\tau=1}^{t} \sum_{k \in R_{j\tau}} \omega_{ijkt}\left(S_{t-1}, \{\lambda_{i\tau}\}_{\tau=1}^{t}\right) p_{ik\tau-1}^{gov}\left(D_{ijt}\left(\xi_{jt}\right)\right)$$

Proof. This Lemma states that we can write p_{ijt}^{ref} as a linear function of government prices in all previous periods, using weights that depend only on the current entry sequences of all firms and structural parameters of the model other than the drug fixed effect. We use induction over t, where t indexes time starting with the year after the product was first approved (in the year of approval there cannot be any reference prices).

Consider $t = 1$. We want to show that for all j,

$$p_{ijt}^{\text{ref}} = \sum_{k \in R_{j1}} \omega_{ijk1}\left(S_0, \lambda_{i0}\right) p_{ik0}^{\text{gov}}\left(D_{ij1}\left(\xi_{j1}\right)\right) \tag{B.6}$$

The definition of the reference price is

$$p_{ij1}^{\text{ref}}\left(E_{i0}, D_{ij1}\left(\xi_{j1}\right)\right) = F_{j1}^{\text{ref}}\left(\left\{p_{ik0}\left(D_{ij1}\left(\xi_{j1}\right)\right)\right\}_{k \in \left(R_{j1} \cap E_{i0}\right)}\right)$$

Since reference pricing cannot be applied at time $t = 0$, the prices that can be referenced must be government prices:

$$p_{ik0}\left(D_{ij1}\left(\xi_{j1}\right)\right) = p_{ik0}^{\text{gov}}\left(D_{ij1}\left(\xi_{j1}\right)\right)$$

Then, if F_{j1}^{ref} is a linear function, like an average, equation B.6 is satisfied for constant weights. F_{j1}^{ref} can also be the minimum function, or the average of the three lowest prices, which are not linear.[1] These functions however, can be expressed as weighted averages, where the

[1] See Figure 2.1.

weights will depend on the relative ranking of the volume adjusted country government prices, which in turn will depend on λ_{ij0}.

In these cases weights can be constructed as follows. Recall that $E_{i0} = \{j : s_{j0} \neq 0\}$ is the set of countries where the product is available in period 0. This set can be obtained from information contained in S_0. We assume WLOG that E_{i0} is not empty (if it is, then there can be no possible reference and the case of $t = 1$ is identical to $t = 0$). Let n_k denote the rank of $k \in E_0$ in increasing order of λ_{i1}. In other words, if $n_k = 1$, then

$$\lambda_{ik1} = \min \{\lambda_{i\ell 1} : \ell \in (R_{j1} \cap E_0)\}$$

and, more generally,

$$\lambda_{ik1} = \min \{\lambda_{i\ell 1} : \ell \in (R_{j1} \cap E_0) \wedge n_\ell \geq n_k\}$$

Hence, the country that ranks first is the one with the lowest government price, the one that ranks second has the second-lowest price, etc. Finally, let $m_{ij1} = \min \{|E_0|, 3\}$, where the operator $|\cdot|$ indicates the cardinality of a set.

If F_{j1}^{ref} is the average of the three lowest prices, the weights can be written as

$$\omega_{ijk1}(S_0, \lambda_{i0}) = \begin{cases} 1 & \text{if } n_k = 1 \\ 0 & \text{otherwise} \end{cases}$$

If F_{j1}^{ref} is the average of the three lowest prices instead, construct the weights as

$$\omega_{ijk1}(S_0, \lambda_{i0}) = \begin{cases} \frac{1}{m_{ij0}} & \text{if } n_k \leq m_{ij0} \\ 0 & \text{otherwise} \end{cases}$$

These weights are written as a function of S_0 and λ_{i0} only, hence they satisfy the premise of the proposition.

To conclude the proof, suppose that the assertion of the proposition is true for $\tau \in \{1, \ldots, t-1\}$, and show that it must hold for t as well. This is easy to prove. We can walk through the same exact steps as we did for $t = 1$, but substituting $p_{ikt-1}(E_{it-1}, D_{ijt}(\xi_{jt}))$

for λ_{i1}. This will give us weights for p_{ijt}^{ref} as a linear function of the prices in the previous period. By construction, prices in the previous period are a weighted average of government prices and reference prices. The reference prices are themselves linear functions of adjusted government prices by the inductive assumption. Since the sum of linear functions is also linear, the proposition must hold for period t as well. ■

The Lemma gives us a way to write the reference price as a weighted average of government prices. Since government prices are a multiplicative function of θ_i, so are reference prices. Hence we can write

$$p_{ijt}^{\text{ref}}(\cdot) = \tilde{p}_{ijt}^{\text{ref}}(\cdot) \cdot \theta_i$$

where $\tilde{p}_{ijt}^{\text{ref}}$ is not a function of θ_i. Then, we can write

$$p_{ijt}(\cdot) = \begin{cases} \theta_i \gamma_j \cdot \exp\left(\beta_Z Z_{ijt} + \beta_D \ln\left(D_{ijt}\right)\right) & \text{if } p_{ijt}^{\text{ref}}(\cdot) \geq p_{ijt}^{\text{gov}}(\cdot) \\ \theta_i \gamma_j \cdot \exp\left(\beta_Z Z_{ijt} + \beta_D \ln\left(D_{ijt}\right)\right) + \mu_j \tilde{p}_{ijt}^{\text{ref}}(\cdot)\,\theta_i & \text{if } p_{ijt}^{\text{ref}}(\cdot) < p_{ijt}^{\text{gov}}(\cdot) \end{cases}$$

which shows that $p_{ijt}(\cdot)$ is a multiplicative function of θ_i. Hence, when we consider $\frac{p_{ijt}}{p_{ikt+1}}$, θ_i will appear in both the denominator and the numerator and we can eliminate it, giving us the result in equation 2.9.

■

B.1.3 Foundations for the Price-Setting Equation

In the main body of the paper we do not provide a micro foundation of the price setting equation from a utility- or revenue-based optimization model. The characteristics of such a model are less relevant for our paper. However, in this section we provide two possible sets of assumptions that could justify an estimation equation identical to the one we use.

Foundations as a static Nash Bargaining Model

We assume that the firm and the government play a Nash Bargaining game. The game is repeated every period, but for simplicity the two parties only split the static welfare gains

from the current period. In reality, prices impose dynamic constraints through reference pricing that both agents should take into account. To eliminate these dynamic considerations one could assume that the government is a myopic agent and that the firm's bargaining unit is only tasked with carrying out the negotiation, without concerns for the future ramifications of the agreed-upon price.[2]

The equilibrium price in a standard Nash Bargaining model is given by

$$p^{\star} = \arg\max_{p} \left[\Delta W_{ijt}\right]^{b_j} \times \left[\Delta \Pi_{ijt}\right]^{1-b_j}$$

where ΔW_{ijt} represents the change in the welfare of the government from having drug i available, $\Delta \Pi_{ijt}$ represents the incremental change in revenue, and b_j is the bargaining power of country j. Notice that the interpretation of ΔW_{ijt} is not necessarily welfare, but could more generally be described as the objective function of the government agent tasked with completing the negotiation.

Under our assumptions of static bargaining, $\Delta \Pi_{ijt}$ is simply the potential revenue in country j, and since demand is price inelastic, we can divide through by demand to recast the problem as a negotiation over the unit price of the product (instead of total revenue). We abstract away from marginal costs of production since for brand drugs they are a negligible fraction of prices. The simplified problem can be written as

$$p^{\star} = \arg\max_{p} \left[\Delta w_{ijt} - p\right]^{b_j} \times \left[p\right]^{1-b_j}$$

where the interpretation of Δw_{ijt} is the average change in the welfare function from obtaining an additional unit of drug i. The standard Nash bargaining solution, can then be written as

$$p^{\star} = \Delta w_{ijt} \left(1 - b_j\right)$$

[2]There are several reasons why short-term considerations could in fact play a major role for most government agencies. First, the main goal of pharmaceutical agencies is to keep spending within the limits of their budget, which is often specifically carved out for prescription drugs thus limiting the ability to generate trade-offs such as paying more for cost-effective drugs that would save money in other areas of health care, such as inpatient care. Turnover of government officials might also contribute to the failure of adopting long-term strategies. On the pharmaceutical company side, most firms have a separate bargaining unit for each country. Informal conversations with industry insiders seem to suggest that these unit operate in relative independence from one another.

This price denotes the equilibrium in the absence of reference pricing, and represents the *government price* p_{ijt}^{gov}.

To account for the impact of reference pricing we propose that the government can negotiate more effectively by eliciting a signal about what prices are charged abroad. We incorporate this possibility in the model by assuming that the signal (i.e. the reference price) affects the bargaining weight assigned to the government. The reference price p_{ijt}^{ref} is calculated as described in the main body of the paper.[3] Given p_{ijt}^{ref}, we write the bargaining weight of the government as

$$B_{ij}\left(p_{ijt}^{\text{ref}}\right) = b_j + \left(1 - b_j\right)\mu_j \left(1 - \frac{p_{ijt}^{\text{ref}}}{p_{ijt}^{\text{gov}}}\right) \cdot \mathbb{I}_{\left\{p_{ijt}^{\text{ref}} < p_{ijt}^{\text{gov}}\right\}}$$

where $p_{ijt}^{\text{gov}} = \Delta w_{ijt}\left(1 - b_j\right)$ is a function of model parameters that reflects the price that the government would have obtained without using reference pricing. We define the bargaining weight of the firm as $1 - B_{ij}\left(p_{ijt}^{\text{ref}}\right)$.

The function $B_{ij}\left(\cdot\right)$ has several attractive properties. First, it reduces to the base case whenever $p_{ijt}^{\text{ref}} < p_{ijt}^{\text{gov}}$. This has the intuitive implication that observing a reference price that is higher than the country's own internal benchmark does not affect negotiations. Second, the bargaining weight is inversely proportional to the reference price, meaning that a lower reference price lets the government extract a greater discount. Third, as long as $\mu_j \in (0, 1)$ the bargaining weight is also lies on the unit interval, which insures an interior solution for the first-order condition.

The first-order condition of the Nash Bargaining problem with the specified bargaining weights is

$$[p]: \quad \left(\Delta w_{ijt} - p\right)^{-1+b+\mu_j - b_j\mu_j - \frac{\mu_j p_{ijt}^{\text{ref}}}{\Delta w_{ijt}}} \cdot \left(\left(1 - b_j\right)\left(1 - \mu_j\right)\Delta w_{ijt} + \mu_j p_{ijt}^{\text{ref}} - \Delta w_{ijt}\right) p^{-b_j\left(1-\mu_j\right)+\mu_j\left(1-\frac{p_{ijt}^{\text{ref}}}{\Delta w_{ijt}}\right)} = 0$$

[3] As a side note, the reference price does not necessarily need to be linked to other prices, but can be anything else that might affect negotiations.

and has three roots:

$$\begin{aligned} p_1^\star &= 0 \\ p_2^\star &= \Delta w_{ijt} \\ p_3^\star &= \left(1-\mu_j\right)\left(1-b_j\right)\Delta w_{ijt} + \mu_j p_{ijt}^{\text{ref}} \end{aligned}$$

Notice that $p_{ijt}^{\text{ref}} \leq \left(1-b_j\right)\Delta w_{ijt}$ whenever the reference price binds. Hence, $p_1^\star < p_3^\star < p_2^\star$.

The second-order condition is given by

$$\begin{aligned} SOC: \quad & p^{-1-b\left(1-\mu_j\right)-\mu_j\left(1-\frac{p_{ijt}^{\text{ref}}}{\Delta w_{ijt}}\right)} \cdot \left(\Delta w_{ijt}-p\right)^{-2+b_j+\mu_j-b_j\mu_j-\frac{\mu_j p_{ijt}^{\text{ref}}}{\Delta w_{ijt}}} \cdot \\ & \left(\Delta w_{ijt}\left(1-b_j\right)\left(1-\mu_j\right)+\mu_j p_{ijt}^{\text{ref}}\right)\left(-b_j\Delta w_{ijt}\left(1-\mu_j\right)+\mu_j\left(p_{ijt}^{\text{ref}}-\Delta w_{ijt}\right)\right) \end{aligned}$$

and, for $p \in \left(0, \Delta w_{ijt}\right)$, is proportional to

$$\begin{aligned} SOC &\propto \left(\Delta w_{ijt}\left(1-b_j\right)\left(1-\mu_j\right)+\mu_j p_{ijt}^{\text{ref}}\right)\left(-b_j\Delta w_{ijt}\left(1-\mu_j\right)+\mu_j\left(p_{ijt}^{\text{ref}}-\Delta w_{ijt}\right)\right) \\ &\propto \left(-b_j\Delta w_{ijt}\left(1-\mu_j\right)+\mu_j\left(p_{ijt}^{\text{ref}}-\Delta w_{ijt}\right)\right) < 0 \end{aligned}$$

Hence the objective function is maximized for $p = p_3^\star$. The two other roots of the first-order condition are also roots for the second-order condition, therefore they represent points of inflection.

The final solution to this bargaining problem is therefore made up of two equations:

$$p_{ijt} = \begin{cases} \left(1-b_j\right)\Delta w_{ijt} & \text{if } p_{ijt}^{\text{ref}} \geq \left(1-b_j\right)\Delta w_{ijt} \\ \left(1-\mu_j\right)\left(1-b_j\right)\Delta w_{ijt} + \mu_j p_{ijt}^{\text{ref}} & \text{if } p_{ijt}^{\text{ref}} < \left(1-b_j\right)\Delta w_{ijt} \end{cases}$$

This solution will have the same form of our estimating equation as long as $\left(1-b_j\right)\Delta w_{ijt}$ can be written as a function of the observables we have included in our parametric function for the government price.

B.1.4 Proof of Theorem 6

To prove the theorem we will rely on the strong law of large numbers applied to non-identical, independent random variables. For this reason it will be useful to define the payoff of firm l as a random variable. We will then show that the set of these random variables (one for each firm l) satisfies the Kolmogorov criterion, which in turn implies the strong law of large numbers.

Let $\tilde{\mathbf{V}}_{lt}\left(\mathcal{A}_{it}; S_{lt-1}, S_{-lt-1}\right)$ denote the payoff of the firm starting in period t, conditional on the value of state variable in period $t-1$ and on firm l following strategy $\mathcal{A}_{lt}$. Note that by definition, $\mathbb{E}\left[\tilde{\mathbf{V}}_{lt}\left(\mathcal{A}_{it}; S_{lt-1}, S_{-lt-1}\right)\right] = \tilde{V}_t\left(\mathcal{A}_{lt}; S_{lt-1}, S_{-lt-1}\right)$.

For the proof of the Theorem we need the following Lemma and Corollary.

Lemma 8. *Let $\Pi\left(S_l, S_{-l}\right)$ be defined as in equation 2.8. Then $\Pi\left(S_l, S_{-l}\right)$ is finite.*

Proof. We prove that $\Pi_\tau\left(S_l, S_{-l}\right)$ is finite by showing that period profits are bounded. The realization of period profits depends on ξ_{jt}. Define

$$\Pi_\tau\left(S_l, S_{-l}, \xi_{j\tau}\right) = \sum_{i \in \mathcal{I}_l} \sum_{j \in S_{i\tau}} p_{ij\tau}\left(S_{l\tau-1}, S_{-l\tau}, D_{ij\tau}\left(\xi_{j\tau}\right)\right) \cdot D_{ij\tau}\left(S_{-l\tau}, \xi_{j\tau}\right)$$

For any given product i and country j, we can write demand as

$$D_{ij\tau}\left(S_{-l\tau}, \xi_{j\tau}\right) = MS_{j\tau} \cdot \frac{\exp\left(\alpha_{ij} + \beta_i \text{age}_{i\tau} + \eta_i NF_{ij\tau} + \xi_{j\tau}\right)}{1 + \sum_{\ell \in E_{-l\tau}} \exp\left(\alpha_{\ell j} + \beta_\ell \text{age}_{\ell\tau} + \eta_\ell NF_{\ell j\tau} + \xi_{j\tau}\right)}$$

where $MS_{j\tau}$ is the market size in country j in period τ. Hence, $D_{ij\tau}\left(S_{-l\tau}, \xi_{jt}\right) \in \left(0, MS_{j\tau}\right)$.

Price is bounded above by the government price:

$$p_{ij\tau}\left(S_{l\tau-1}, S_{-l\tau}, D_{ij\tau}\left(\xi_{j\tau}\right)\right) \leq p_{ij\tau}^{\text{gov}}\left(S_{l\tau-1}, S_{-l\tau}, D_{ij\tau}\left(\xi_{j\tau}\right)\right)$$

Moreover, using the definition of government price in equation 2.4 we can rewrite

$$p_{ij\tau}^{\text{gov}}\left(S_{l\tau-1}, S_{-l\tau}, D_{ij\tau}\left(\xi_{j\tau}\right)\right) \cdot D_{ij\tau}\left(\xi_{j\tau}\right) = p_{ij\tau}^{\text{gov}}\left(S_{l\tau-1}, S_{-l\tau}, 1\right) \cdot \left(D_{ij\tau}\left(\xi_{j\tau}\right)\right)^{1+\beta_D}$$

Hence, the period payoff of a single drug in any given country is bounded above by

$$p_{ij\tau}^{\text{gov}}\left(S_{l\tau-1}, S_{-l\tau}, 1\right) \cdot \left(MS_{j\tau}\right)^{1+\beta_D}$$

and bounded below by 0. This implies that the period payoff in any given country and period is finite, and therefore $\Pi_\tau\left(S_l, S_{-l}\right)$ is also finite. ■

Corollary 9. $\tilde{\mathbf{V}}_{lt}\left(\mathcal{A}_{it}; S_{lt-1}, S_{-lt-1}\right)$ *has finite variance.*

Proof. This lemma follows directly from Lemma 8. $\tilde{\mathbf{V}}_{lt}\left(\mathcal{A}_{it}; S_{lt-1}, S_{-lt-1}\right)$ is defined as the discounted sum of the expected period payoffs. By Lemma 8 the expected period payoffs are finite. Let

$$\Pi^{UB} = \max_{(S_l, S_{-l})} \Pi_\tau\left(S_l, S_{-l}\right)$$

Then, the support of $\tilde{\mathbf{V}}_{lt}\left(\mathcal{A}_{it}; S_{lt-1}, S_{-lt-1}\right)$ is bounded above by $(T-t)\cdot\Pi^{UB} < \infty$.

Since $\tilde{\mathbf{V}}_{lt}\left(\mathcal{A}_{it}; S_{lt-1}, S_{-lt-1}\right)$ is also bounded below by 0, it must have finite variance. ■

At this point we are ready to prove Theorem *6*.

Proof of Theorem 6. For any given drug i, $V_t\left(S_l^o, S_{-l}^o\right)$ represents a draw from the distribution of $\tilde{\mathbf{V}}_{lt}\left(\mathcal{A}_{it}^\star; S_{lt-1}, S_{-lt-1}\right)$. By Corollary 9, $\text{Var}\left(\tilde{\mathbf{V}}_{lt}\left(\mathcal{A}_{it}^\star; S_{lt-1}, S_{-lt-1}\right)\right) < \infty$. Moreover, the random variables $\tilde{\mathbf{V}}_{lt}\left(\mathcal{A}_{it}^\star; S_{lt-1}, S_{-lt-1}\right)$ are independently distributed. Thus, our premise satisfies the Kolmogorov criterion, which implies that the strong law of large numbers applies to the sequence of random variables $\tilde{\mathbf{V}}_{lt}\left(\mathcal{A}_{it}^\star; S_{lt-1}, S_{-lt-1}\right)$, and the sample average of the realized payoffs will converge to the average of their expected values. Formally, for any $\varepsilon > 0$, we can find M' such that

$$\frac{1}{M}\sum_{i=1}^{M}\left(V_t\left(S_l^o, S_{-l}^o\right) - \tilde{V}_t\left(\mathcal{A}_{lt}^\star; S_{lt-1}, S_{-lt-1}\right)\right) < \varepsilon$$

for all $M > M'$. This concludes the proof.

■

B.2 Dataset Construction

In this section I describe the IMS MIDAS database, the variables it contains, its missing data, and how the missing data is imputed. I also provide a brief description of additional data sources.

B.2.1 Dataset Characteristics

Coverage

We have data for 27 European countries, giving us almost complete coverage of the European Economic Area.[4] The data usually covers sales through the retail channel (i.e. pharmacies) and the hospital channel for 48 quarters, starting in 2001 (either Q2 or Q3) and ending in 2013 (either Q1 or Q2). Several countries have partial data. The most common data issues concern missing years, missing hospital sales data, and having sales data in Euro instead of local currency (for countries where the Euro is not the local currency). The problem with having data in Euros instead of the local currency is that IMS converted the data using a fixed exchange rate (for the entire 12-year sample) instead of a quarter-specific exchange rate. This affects makes the sales figures unreliable in countries that do not have a fixed exchange rate with the Euro. Table B.1 describes our data coverage by country.

Since the problem that we study concerns spillover effects across countries, it is imperative that we try to include in our sample as many countries as possible. Ultimately, we end up discarding 2 of the 27 countries in our sample due to unresolvable data problems: Czech Republic and Slovakia. These two countries had non-Euro currencies with fluctuating exchange rates during our sample period, which means that their sales data is unreliable. We include them in the reduced form results that do not involve prices, but drop them from the structural analysis.

Twelve of the remaining 25 countries have some data issue: Denmark, Estonia, Greece,

[4] We are missing data for Cyprus, Iceland, Lichtenstein, and Malta. Croatia is also a current member of the EEA, but only entered in 2015, which is outside our sample.

Table B.1: *Data Availability by Country*

Country	Years	Hospital Sales Data	Currency
Austria	2001Q2-2013Q1	Available	Euro
Belgium	2001Q2-2013Q1	Available	Euro
Bulgaria	2001Q3-2013Q2	Available	Local
Czech Republic	2001Q2-2013Q1	Available	Euro[a]
Denmark	2001Q2-2013Q1	Not distinguished separately	Euro[b]
Estonia	2001Q3-2013Q2	Not available	Euro[c]
Finland	2001Q2-2013Q1	Available	Euro
France	2001Q2-2013Q1	Available	Euro
Germany	2001Q2-2013Q1	Available	Euro
Greece	2001Q2-2013Q1	Not available	Euro
Hungary	2001Q3-2013Q2	Available	Local
Ireland	2001Q2-2013Q1	Available starting 2006Q1	Euro
Italy	2001Q2-2013Q1	Available	Euro
Latvia	2001Q3-2013Q2	Not available	Local[d]
Lithuania	2001Q3-2013Q2	Available starting 2002Q3	Local[d]
Luxembourg	2001Q2-2013Q1	Not available	Euro
Netherlands	2007Q2-2013Q1	Not available	Euro
Norway	2001Q3-2013Q2	Available	Local
Poland	2001Q3-2013Q2	Available	Local
Portugal	2001Q2-2013Q1	Available starting 2010Q1	Euro
Romania	2001Q3-2013Q2	Available starting 2005Q1	Local
Slovakia	2001Q3-2013Q2	Available	Euro[e]
Slovenia	2001Q3-2013Q2	Not distinguished separately	Euro[f]
Spain	2001Q2-2013Q1	Available	Euro
Sweden	2004Q1-2013Q2	Available	Local
Switzerland	2001Q3-2013Q2	Available	Local
UK	2001Q3-2013Q2	Available	Local

[a] The Czech koruna appreciated from 34.021 Ck for 1€in 2002Q1, to 25.167 Ck for 1€in 2012Q4.
[b] The value of a Danish Krona oscillated between 0.13398 and 0.13468 Euros from 2001 to 2012.
[c] The Estonian kroon was pegged to the Euro (15.6466 krooni per 1€) prior to Euro adoption in 2011.
[d] Latvia and Lithuania adopted the Euro in 2014 and 2015 respectively.
[e] The Slovak koruna appreciated from 42.234 Sk for 1€in 2002Q1, to 30.35 Sk for 1€in 2008Q4.
[f] The value of the Slovenian tolar remained constant from 2004Q3 until the country adopted the Euro in 2007.

Ireland, Latvia, Lithuania, Luxembourg, Netherlands, Portugal, Romania, Slovenia, and Sweden. Some of these have are minor data issues that do not affect the analysis. Denmark and Slovenia report retail and hospital sales jointly, which limits our ability to separate retail and hospital products in some robustness checks, but leaves our main specification unaffected. Denmark and Estonia only have sales in Euros, but their currency was tightly pegged to Euro throughout the duration of our sample period, so the conversion leaves the sales figures unaffected. Slovenia also only reports sales in Euros, and for the first 3 years of our sample data its currency was not pegged to the Euro. However, the Slovenian Tolar fluctuated considerably less than the Czech and Slovak currencies, so we ultimately decided to keep the data in the sample (without additional corrections).

The remaining data issues concern the lack of hospital sales data in 9 countries, and missing years for Netherlands and Sweden. We address these problem by performing a series of imputation steps that allow us to keep all these countries in the sample. These steps are detailed in section B.2.3.

Variables

In its raw form, the data is stored as excel files, one for each country. The variables contained in each file are described in Table B.2. Each product is identified by a combination of active ingredient (molecule), marketing firm, therapeutic class (defined by the ATC4 classification), product name (both international and country-specific), form, strength, and package.[5] Besides revenue and quantity sales, the data provides a few additional characteristics: whether the product is a prescription or an over-the-counter medication, the launch date, the estimated patent expiration date, the product type (brand or generic), protection status, and licensing information.

Main Sample Identification

Our main sample of drugs includes products that satisfy the following requirements:

[5]For more details on the ATC classification, please see `http://www.ephmra.org/classification`.

Table B.2: *Variables in the IMS MIDAS Database*

Variable Name	Description	Variable Type
country	Country	
distributionchannel	Distribution Channel	Hospital/Retail
mlist	Active Ingredient	Product definition
crp	Marketing Firm	Product definition
atc4	ATC4	Product definition
internationalproduct	International Product Name	Product definition
prd	Local Product Name	Product definition
productform	Product Form	Product definition
productinternationalstrength	Strength	Product definition
internationpack	International Package Identifier	Product definition
localpack	Local Package Identifier	Product definition
rxorotc	Rx or OTC	Product characteristics
localproductlaunchdate	Local Product Launch date	Product characteristics
estpatentexpirydate	Estimated Patent Expiration Date	Product characteristics
protection	Protection Status	Product characteristics
producttype	Product Type	Product characteristics
licensinginformation	Licensing Information	Product characteristics
sales_mnf_qtr_*'qrt'*_*'yr'*_local_curr	Sales in local currency	Sales information
sales_mnf_qtr_*'qrt'*_*'yr'*_lceuro	Sales in Euro	Sales information
standard_units_qtr_*'qrt'*_*'yr'*	Quantity sold	Sales information

- First launched on or after January 1st, 1995
- Patent expiration occurred on or after January 1st, 2003 (giving at least one full year of potential observations)
- At least one launch in a country on or after January 1st, 2002
- The firm selling the product is classified as the originator firm (i.e. the first recorded launch date of the molecule is associated to that firm)
- At least 50% of the launches occurred prior to patent expiration (this condition eliminates a handful of products that did not receive EU-wide approval)
- The product satisfies at least one of these three conditions:
  - It received approval from the EMA
  - It received at least one approval using the Mutual Recognition Procedure
  - It is sold in at least 11 countries as of December 31st, 2012

This gives us a sample of 481 products.

B.2.2 Construction of Therapeutic Classes

Broadly speaking, our therapeutic classes are defined at the ATC3 level, which corresponds to a therapeutic-pharmacological subgroup within one of the 14 main systems (Table B.3). In addition, we aggregate classes that contain classes of products that share the same molecules (e.g. D7A&D7B | CORTICOSTEROID, TOPICAL, PLAIN & COMBO is a therapeutic class that combines both plain and combination corticosteroids), and classes of products that have broad applications (e.g. oncologics, which are separated in three large classes: cytotoxics, hormonals, and targeted therapies). We also separate classes of products that have similar pharmacological profile, but are used to combat different diseases (e.g. J5B, which includes non-HIV antivirals, is separated in J5B1, Viral Hepatitis; J5B3, Herpes Antivirals; J5B4, Flu Antivirals; and J5B5,9, Other Respiratory Antivirals). Finally, we make a few adjustments for

Table B.3: *ATC Main Systems*

Code	Contents
A	Alimentary tract and metabolism
B	Blood and blood forming organs
C	Cardiovascular system
D	Dermatologicals
G	Genito-urinary system and sex hormones
H	Systemic hormonal preparations, excluding sex hormones and insulins
J	Antiinfectives for systemic use
L	Antineoplastic and immunomodulating agents
M	Musculo-skeletal system
N	Nervous system
P	Antiparasitic products, insecticides and repellents
R	Respiratory system
S	Sensory organs
V	Various

This table describes the classification of the first digit of the ATC code (or ATC1). Each letter roughly corresponds to an organ system. For more details on the ATC classification system, see `http://www.ephmra.org/classification`

complex diseases with therapies that come from a broad spectrum of therapeutic areas (e.g. Multiple Sclerosis includes Beta interferons (the L3B2 ATC4 code), plus a handful of drugs from other classes, mainly monoclonal antibodies).

B.2.3 Imputation of Missing Data

There are three types of missing data instances that we need to account for. First, some products are missing some years of data, or have negative sales recorded in a small number of instances. Second, a handful of countries are partially or completely missing data for the hospital channel. Third, the Netherlands and Sweden have missing data for a few years in the initial part of the sample.

Imputation of Missing Years of Data for Individual Products

Occasionally, we see some years of missing data or negative sales for some drugs. We treat negative sales as missing when we do this imputation.

There are three instances in which this can happen. First, we may see sales recorded in non-consecutive years (e.g. 2003 and 2005), but not in years in between (e.g. 2004). Second, in some cases we observe a launch date, but no sales (e.g. a product is recorded as having being launched in 2002 but first sale is recorded in 2004). Finally, in other cases we see a product disappear.

In the first instance, we use simple linear interpolation to fill in the missing years. In the second instance, we adjust the launch date instead of the sales variable, and replace it with the date when sales are first recorded. In the last instance we apply no adjustment and simply assume that the product exited exogenously (i.e. the firm did not choose to withdraw the product). We manually checked all these cases, and they all fall in one of two categories. Either the product had been declining in sales for a few years, until it disappeared; or the product was subject to a forced withdrawal by the EMA (we note these instances in Table B.4).

Imputation of Hospital Missing data

Ireland, Lithuania, Portugal, and Romania are missing hospital sales for a few years at the beginning of the sample, but all have at least a few years of hospital sales available. Estonia, Greece, Latvia, Luxembourg, and the Netherlands are missing hospital sales entirely. We impute hospital sales using information contained in other years and other countries.

Wherever possible, we calculate drug-country specific share of sales through the hospital channels, and use it to project sales for missing years. This strategy is only available for countries that have partially missing hospital sales. For countries that are missing hospital sales entirely, we use the drug-specific share of hospital sales.

The imputation works in two steps. First, we calculate the share of hospital sales s_{ij}^h (or s_i^h, for the drug-specific share) using data from other years and countries. Then, hospital sales

Table B.4: *Withdrawn Products*

Molecule Name	Approval Date	Withdrawal Year
DROTRECOGIN ALFA (ACTIVATED)	22aug2002	2011
EPOETIN DELTA	01jan2007	2009
GLIMEPIRIDE#ROSIGLITAZONE	01nov2006	2010
LUMIRACOXIB	01jan2005	2007
METFORMIN#ROSIGLITAZONE	20oct2003	2010
RIMONABANT	19jun2006	2009
ROSIGLITAZONE	11jul2000	2010
SIBUTRAMINE	01apr2001	2010
SITAXENTAN	10aug200	2010
VALDECOXIB	27mar2003	2008

This table lists all molecule that received approval by the EMA and later had that approval rescinded for safety reasons. An exact withdrawal data is not available since recalls generally take place over at least a few months.

are equal to

$$S_{ijt}^{\text{Hospital}} = S_{ijt}^{\text{Retail}} \cdot \frac{s_{ij}^h}{1 - s_{ij}^h}$$

We do the same imputation for quantity sold and revenue sales.

For countries that are missing hospital sales in all years, we cannot impute sales of products that are sold exclusively through the hospital channel. This is a shortcoming of the data, but unfortunately, not one that we can address in any way.

Imputation of Missing Years of Data for Countries

The only information we have on the fully missing years for Sweden and the Netherlands is what products are sold during those years, which we can deduct using the launch date.[6] To impute quantity sold we use the demand primitives implied by our structural demand system. To impute prices, we simply use the first available price.

[6] If a product that was available in a missing year disappeared before it could be recorded in our data we will not be able to observe it. However, product withdrawals are rare.

B.2.4 Additional Data Sources

Additional data sources are reported in Table B.5. We use approval dates from the EMA and the Heads of Medicines Agencies to calculate launch delays. We use data from the Global Burden of Disease study to calculate market size for the demand system. We use exchange rates to transform sales reported in local currencies to Euros. Finally, we use data on GDP and population to run some simple reduced-form checks.

All these additional data sources are publicly available (see the links in Table B.5, which are accurate as of November 20th, 2017).

B.3 Institutional Details

In this section we present some evidence on the impact that the introduction of the EMA had on delays. We also include the reference pricing matrices that we constructed for each year.

B.3.1 The Effect of the European Medicines Agency on Delays

Prior to 1995, pharmaceutical companies seeking to sell prescription drugs in Europe needed to separately apply for marketing approval in each country. The EMA was founded in 1995 in an effort to reduce the administrative burden faced by companies.

Figure B.1 plots the number of new molecular entities approved each year since 1995. We distinguish between approvals obtained directly from the EMA, through the centralized procedure, approvals obtained through the Mutual Recognition Procedure (which was also installed after the EMA was introduced), and other approvals. We can see that within a few years virtually every drug is using one of the two main procedures introduced by the EMA. In recent years, drugs are also increasingly reliant on the centralized procedure, rather than the Mutual Recognition Procedure.

In the data, we observe a market decrease in the average launch delay before and after the founding of the EMA (Figure B.2). A variety of factors likely contribute to this decrease. First, the new centralized approval procedure should reduce the fixed cost of entry in each country.

Table B.5: *Additional Data Sources*

Data	Use	Source	Source Link
Drug Approval Date (EMA)	Calculate launch delays	European Medicines Agency	http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/landing/epar_search.jsp
Drug Approval Date (Mutual Recognition Procedure)	Calculate launch delays	Heads of Medicines Agencies	http://mri.cts-mrp.eu/Human/about
Incidence of Disease	Calculate market size for each market (therapeutic area and country)	Global Burden of Disease Study	http://ghdx.healthdata.org/gbd-results-tool
Exchange Rates	Adjust sales to Euros	European Central Bank	https://www.ecb.europa.eu/stats/policy_and_exchange_rates/euro_reference_exchange_rates/html/index.en.html
GDP and Population	Regression controls and market size construction	Eurostat	http://ec.europa.eu/eurostat/web/national-accounts/data/database

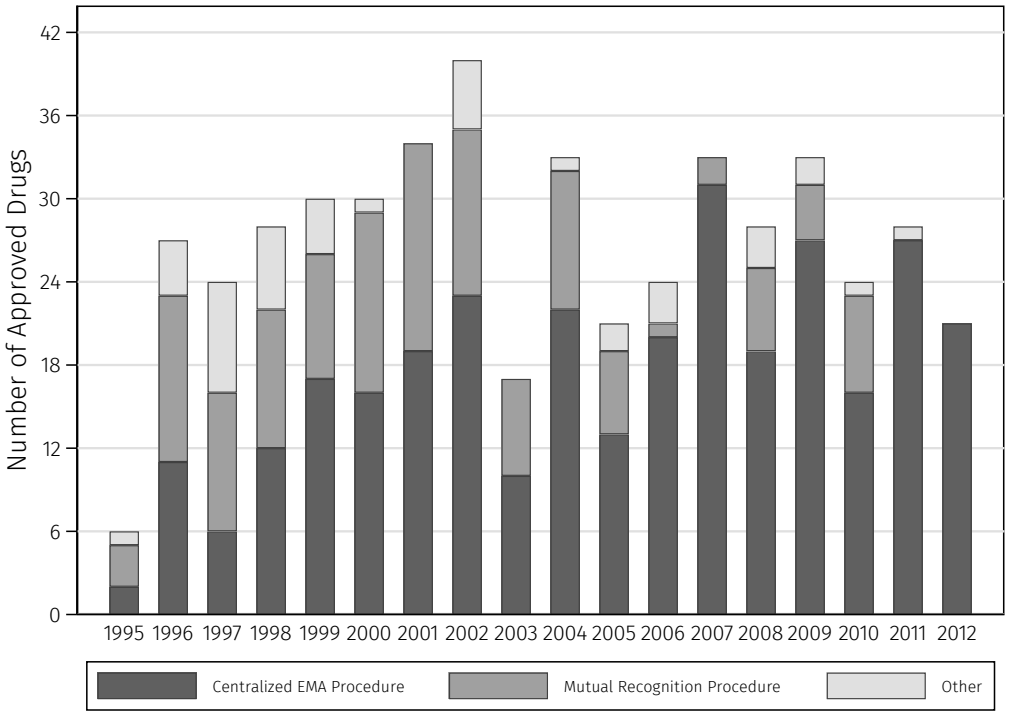


Figure B.1: *Number of Drug Approvals By Year*

This stacked bar chart shows the number of new molecular entities approved in Europe by year, starting in 1995 (the year the EMA was founded). Approvals are divided according to the procedure used by the firm (for more details on each procedure, please consult Section 2.1 of the paper).

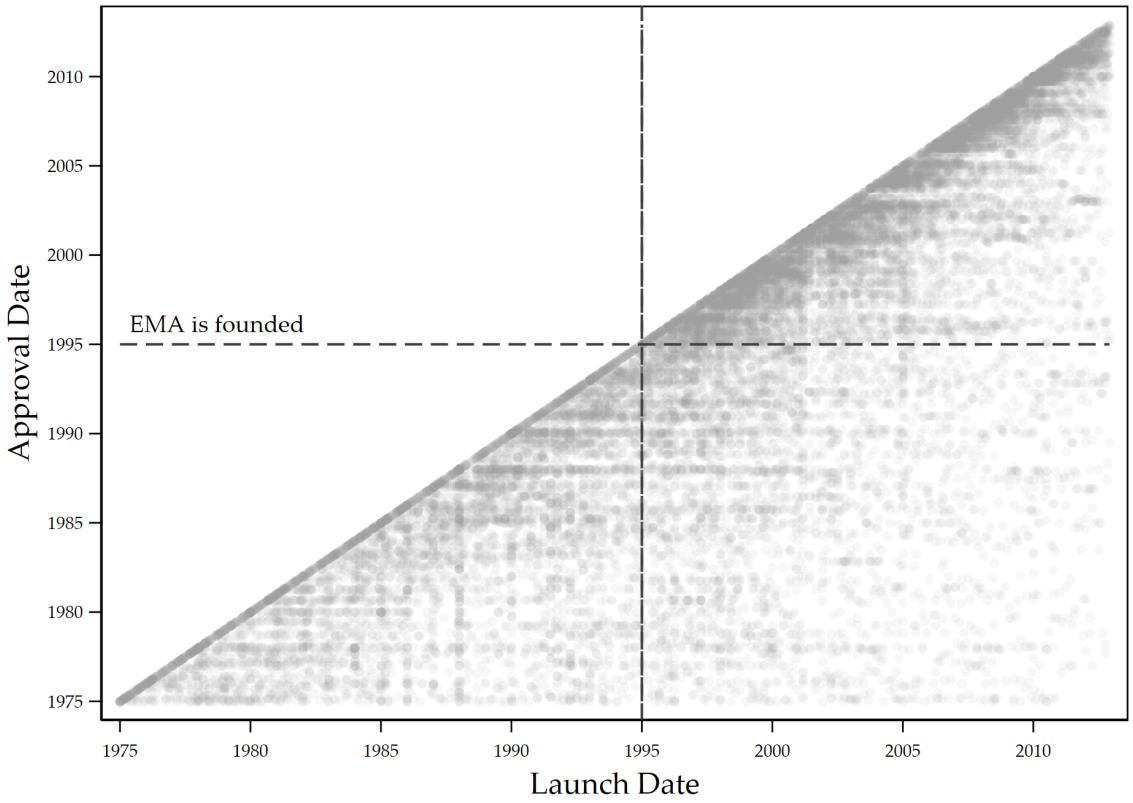


Figure B.2: *Delay patterns before and after the EMA*

This scatter-plot shows the distribution of the launch of new drugs. Each launch in each country is recorded separately, and is defined by two coordinates: the date in which the product was approved, and the date in which it was launched. If there were no delays, then all points should lie on the 45-degree line. It's very clear from the plot that drugs approved after 1995 experienced fewer delays (this is clear from the increased concentration of dots close to the 45-degree line), which suggests that the EMA had a large negative effect on delays. This result is what motivated our choice to focus only on the post-1995 sample.

This should be especially helpful for small low-income countries. Second, it should reduce the time necessary to receive approval. Third, it may reduce the protectionist tendency of member states to prevent entry from firms whose drugs compete with drugs produced by domestic manufacturers.

The clear discontinuity in delays between the pre-1995 sample, and the post-1995 sample is what motivated our choice to restrict the analysis to drugs that were first launched after January 1st, 1995.

Table B.6: *Changes to Reference Pricing Functions*

Year	Country	Implemented change
2005	AUSTRIA	Added new EU member states (CZ EE HU LT LV PL SL SK).
2008	AUSTRIA	Added new EU member states (BG RO).
2005	BELGIUM	Added new EU member states (CZ EE HU LT LV PL SL SK).
2008	BELGIUM	Added new EU member states (BG RO).
2011	BELGIUM	Changed formula to average.
2005	FINLAND	Added new EU member states (CZ EE HU LT LV PL SL SK).
2008	FINLAND	Added new EU member states (BG RO).
2010	GREECE	Removed DK, EE, and SE. Changed formula to average of 3 lowest prices.
2011	HUNGARY	Added BG, CH, DK, EE, LT, LU, LV, NL, NO, RO, SE, UK.
2005	ITALY	Added new EU member states (CZ EE HU LT LV PL SL SK).
2008	ITALY	Added new EU member states (BG RO).
2012	POLAND	Added AT, BG, EE, FI, LV, NO, RO, SL, SK.
2007	PORTUGAL	Added Greece.
2012	PORTUGAL	Changed basket to Italy, Spain, Slovenia.
2008	SPAIN	Added Slovenia (new EURO member).
2009	SPAIN	Added Cyprus and Malta (new EURO members).
2010	SPAIN	Added Slovakia (new EURO member).
2012	SPAIN	Added Estonia, Latvia, and Lithuania (new EURO members).

This table lists all the changes that were made to reference pricing functions, sorted by country and year. Most changes involved the addition of new EU/EURO Member States. Figure **??** also plots the full reference pricing matrices for all countries and years. Please refer to it for more details.

B.4 Additional Reduced-Form Evidence

In this section we present some additional reduced-form evidence on the correlation between market outcomes and delays. Part of this analysis motivated our choice of functional forms for the demand and price system.

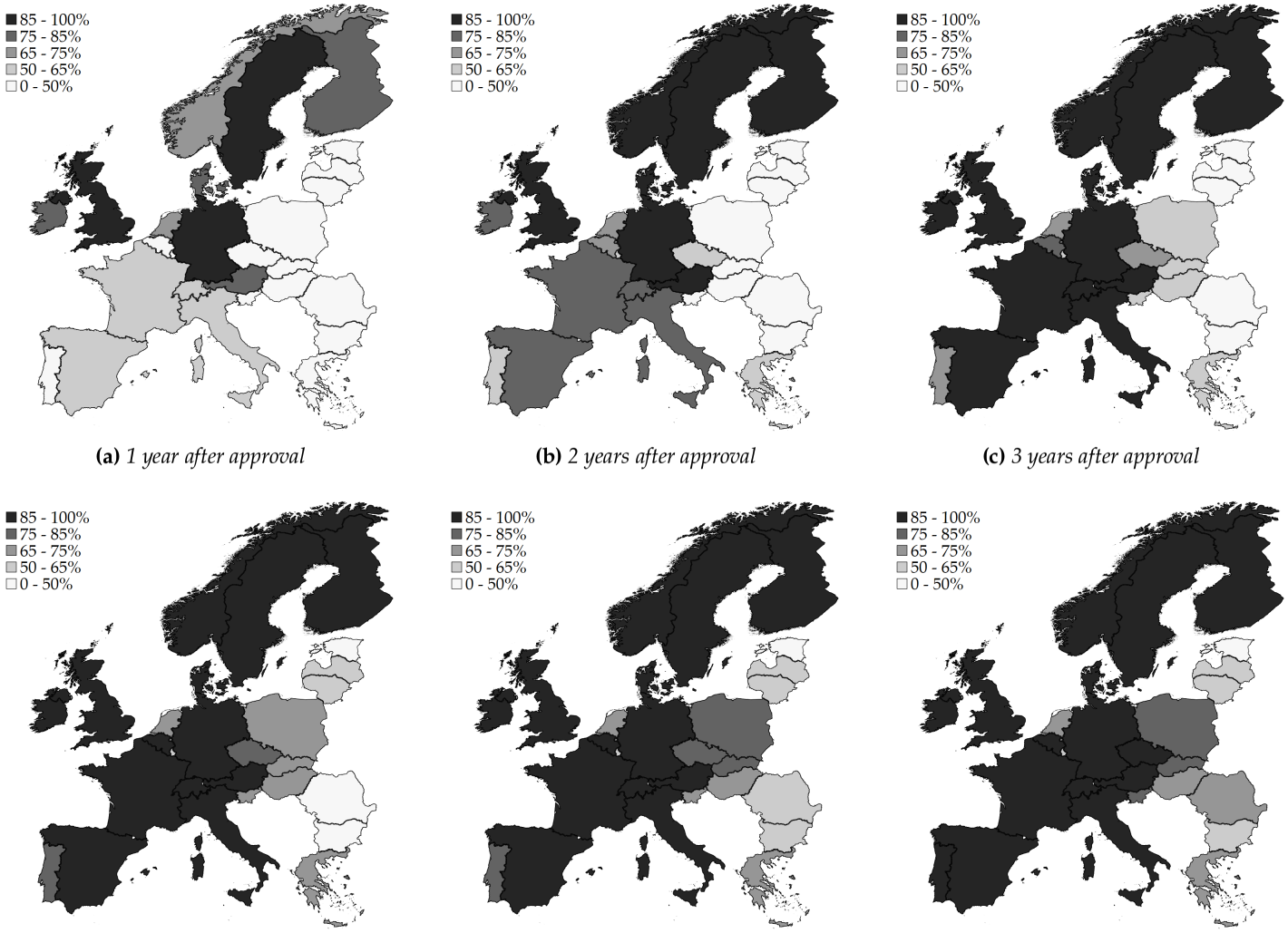
We also discuss some additional evidence of heterogeneous delays across drugs and countries by looking at delays for innovative products, and across therapeutic classes.

B.4.1 Delays Across Countries

Figure B.3 contains a series of maps that display the fraction of drugs launched in each country within the first 6 years of the marketing approval date. We select a balanced subsample of 142 drugs that received approval through the centralized procedure on or before December 31st, 2006 (out of the 481 in our main sample) to ensure that we can observe the first 6 years of their life-cycle in our data.

Several interesting patterns emerge. First, Eastern Europe is lagging behind the rest of the continent. 6 years after the original approval of a product, Bulgaria, Estonia, Latvia, and Lithuania are the nations with the lowest diffusion rates, followed by Romania and Hungary.[7] Second, even though the remaining European countries achieve similar diffusion rates by year 4, entry patterns differ considerably even between countries with similar market size and income level. For example, the UK has market size and prices similar to France and Italy, but almost all drugs are available in the UK within 12 months, whereas France and Italy only achieve a comparable diffusion rate after 3 years. The same comparison can be made between Austria and Belgium, or Sweden and Switzerland.

[7]The Netherlands also appear to have very low diffusion rates. Rather than reflecting a real empirical phenomenon, this result is caused by missing data. Our coverage for the Netherlands starts in 2007 and includes only retail sales (and not sales through the hospital channel), which means that many products never appear in the data because they are only available in hospitals or they had already exited the market in 2007. Unfortunately, the sample selected to draw this figure draw heavily on this earlier cohort of drugs, exacerbating the problem. In reality, the Netherlands have one of the highest (and fastest) diffusion rates, comparable to those of Germany or Sweden.



(a) *1 year after approval* **(b)** *2 years after approval* **(c)** *3 years after approval*

(d) *4 years after approval* **(e)** *5 years after approval* **(f)** *6 years after approval*

Figure B.3: *Diffusion of EMA-approved drugs in European Countries*

B.4.2 Correlation between Quantity, Prices, and Delays

In the reduced-form section of the paper we show that price levels and market size appear to be inversely correlated with delays across countries. In this section we present some additional evidence of the correlation between delays and market outcomes (quantity sold and price). We look for evidence that products with relatively higher price (or quantity) are launched earlier.

We test this hypothesis by first isolating within-country variation in demand and price. We do this by regressing quantity sold and price on drug and country fixed-effects:

$$\begin{aligned} \ln\left(q_{ijt}\right) &= \theta_i^q + \gamma_j^q + \varepsilon_{ijt}^q \\ \ln\left(p_{ijt}\right) &= \theta_i^p + \gamma_j^p + \varepsilon_{ijt}^p \end{aligned}$$

In these regressions i indexes products, j indexes country, and t indexes periods. The residuals ε_{ijt}^q and ε_{ijt}^p reflect both a fully stochastic component (reflecting uncertainty that the firm cannot observe in advance), as well as information that is not reflected in the drug and country fixed effect, but that the firm sees in advance. For example, if Italy had a disproportionately large number of diabetes patients, we would expect ε_{iITAt}^q to be large for insulin and other diabetes medication.

To check whether within-country variation in market outcomes is correlated with firm behavior, we check whether the residual is correlated with delays and order of entry, by running the following linear regressions:

$$\begin{aligned} y_{ij} &= \gamma_j^y + \beta_q \varepsilon_{ijt^0}^q + \varepsilon_{ij}^y \\ y_{ij} &= \gamma_j^y + \beta_p \varepsilon_{ijt^0}^p + \varepsilon_{ij}^y \end{aligned}$$

As independent variables y_{ijt} we use both the entry delay in months, and the order of launch (i.e. the rank of the country j in the launch sequence of drug i). As dependent variable we

Table B.7: *Correlation of Within-Country Variation and Firm Behavior*

	Launch Order		Delay	
ε^q_{ijt}	-0.435^{***}		-1.546^{***}	
	[-0.487,-0.383]		[-1.759,-1.333]	
ε^p_{ijt}		0.068		0.068
		[-0.116,0.252]		[-0.686,0.821]
Country FE	Y	Y	Y	Y
N	8,894	8,894	8,894	8,894
R^2	0.50	0.48	0.23	0.22

This table show the results of a regression of launch order (or delay) on country fixed effects plus a residual from a regression of log quantity (or price) on country and drug fixed effects (see Section B.4.2 for more details on the specification).

use the value of the residual in the launch year t^0.[8]

We report our results in Table B.7. We find that $\varepsilon^q_{ijt^0}$ is negatively correlated with both months of delay as well as order of launch, meaning that within each country, products with low demand tend to enter later.

Surprisingly, we do not find the same result for $\varepsilon^p_{ijt^0}$. There are two possible reasons to explain this phenomenon. First, it could be that $\varepsilon^p_{ijt^0}$ only contains information that the firm does not observe in advance. In this case, we would not expect the firm's actions to reflect anything that is unknown about price prior to entry. Second, it is possible that there is not enough within-country, cross-drug variation in price. In the data, we note that a price regression on drug and country fixed effects achieves a coefficient of variation of 0.98, meaning that these two variable soak up a lot of variation. The equivalent regression on sales has a much lower coefficient (around 0.77), which suggests that there may be more variation in quantity that the firm can act upon.

These regressions also serve to motivate our choice of demand and price functions. In our demand functions we include drug-country fixed effects, since our regressions suggest that using a separate set of fixed effects may fail to capture important variation that is known to

[8]When we do not observe sales in the launch year (because launch occurred before 2002), we use the residual from the first available year.

the firm. In our price functions however, we include drug and country fixed effects separately. The regression results in this section suggest that this is enough to capture the variation in price that is relevant for firms actions.

B.4.3 Delays for Innovative Drugs

We also look at whether innovative drugs experience fewer delays. To classify drugs according to their innovative status we rely on work by Lanthier *et al.* (2013), who classify drugs approved by the FDA in three categories: first-in-class, advance-in-class, and addition-to-class. First-in-class drugs are the first to be approved within their respective class. Advance-in-class are not the first to be approved within the class, but received a priority review designation, which usually means that the drug has the potential to represent a major advance in treatment. All other drugs are considered to be addition-to-class drugs.

We apply this classification to our drugs and calculate average delays for all matched drugs by country.[9] Figure B.4 shows our results. We find some evidence that innovative drugs tend to experience fewer delays, although the difference is only a couple of months. Interestingly, the effect is much smaller in Eastern European countries.

B.4.4 Delays across therapeutic classes

To check whether there are significant differences in delays across therapeutic classes, we run a regression of delays on country fixed effects and class fixed effects. Since most classes only have a handful of products, we aggregate them at the ATC1 level, which indicates the main system for which drugs are used (see Table B.3).

We plot the coefficient from this regression in Figure B.5. We find virtually no evidence of meaningful differences across therapeutic classes. With a single exception, all the coefficients from the regression indicate that differences across these classes are within 6 months. While some of the coefficients indicate that the difference in delays is significant, this difference is

[9] We are able to match 285 out of the 481 drugs in our sample. The difference is due to the fact that the FDA and the EMA sometimes approve different drugs, the different time-span our paper and Lanthier's paper cover, and the fact that our merge on molecule string may be imperfect.

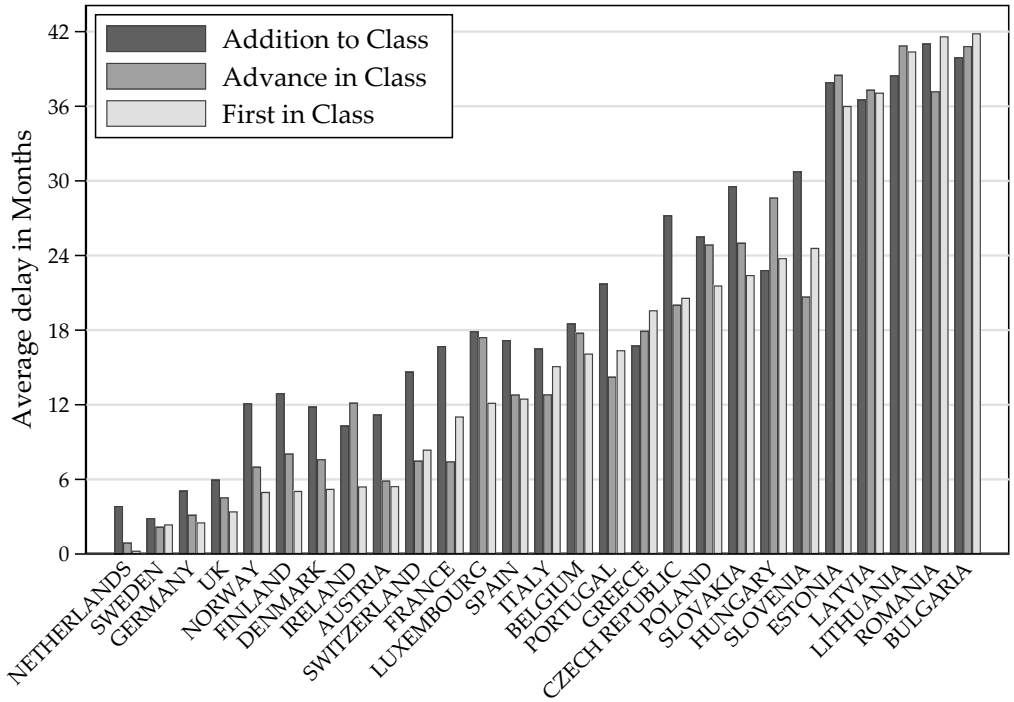


Figure B.4: *Delays by Innovative Status*

This chart plots the average delays of drugs separated according to their innovative status. First-in-class drugs (i.e. drugs who are the first to treat a certain disease or to exploit a specific treatment path) and drugs that represent and advance-in-class tend to experience fewer delays, but only in Western Europe (see Section B.4.3 for details on the classification).

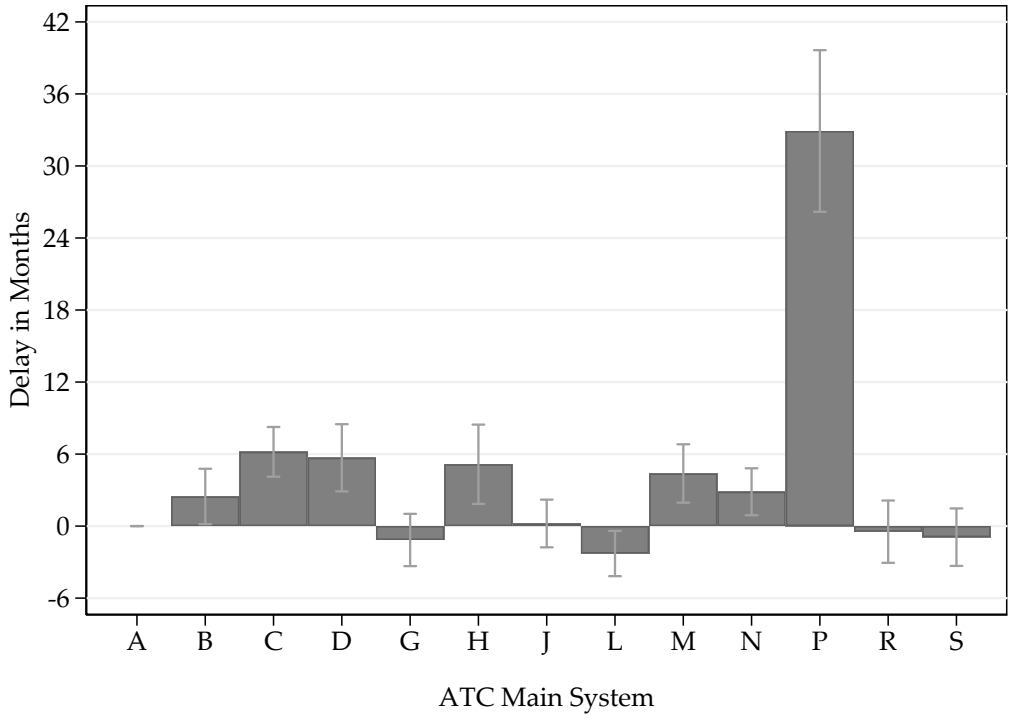


Figure B.5: *Delays by ATC Main System*

This chart plots the average delays of drugs separated according to their ATC1 classification (see Table B.3 for a description of what each letter stands for). We do not find large differences across therapeutic classes, with the exception of Antiparasitic drugs (P). The only products from our sample that belong to this class are antimalarial treatments.

very small. The exception is ATC code 1, which indicates antiparasitic products. Specifically, all products in this group belong to the therapeutic class of Antimalarial medications. Given that malaria is virtually non-existent in Europe, this result is unsurprising.[10]

B.5 Moment Inequality Extensions

In this section we describe extensions of the moment inequality approach used in the paper, which, subject to additional assumptions, can be used to improve the bounds obtained in the estimation.

B.5.1 Moment Inequalities with Entry and Approval Data

To derive a lower bound using entry data we calculate the overall probability of a delay across all drugs in the sample. The assumption we use in the estimation is that the probability of a delay is country-specific, but homogeneous across all drugs. Under this assumption, the overall probability of a delay in any given subsample of drugs should be higher than the probability of an idiosyncratic delay (as long as the subsample was not selected based on entry data). This suggests that we could obtain a tighter upper bound by simply looking at moment inequalities based on subsamples of the data instead of only looking at the full sample.

Formally, let G be a partition of our set of drugs. For each set $g \in G$, let $\left(1 - \bar{\psi}_{gj}\right)$ be the probability that product $i \in g$ will enter in country j in a given year. Since by assumption ψ_j is constant across drugs, it follows that $\bar{\psi}_{gj} \geq \psi_j$ for all g, and therefore

$$\psi_j \leq \psi_{Gj}^{\min} = \min_{g \in G} \left\{ \bar{\psi}_{gj} \right\}$$

Figure B.6 plots the values and confidence intervals for ψ_{gEU10} for the partition of the sample generated by the therapeutic class variable. Since we may be concerned about false

[10] Malaria was officially eradicated in Europe in 2016 (http://www.euro.who.int/en/media-centre/sections/press-releases/2016/04/from-over-90-000-cases-to-zero-in-two-decades-the-european-region-is-malaria-free).

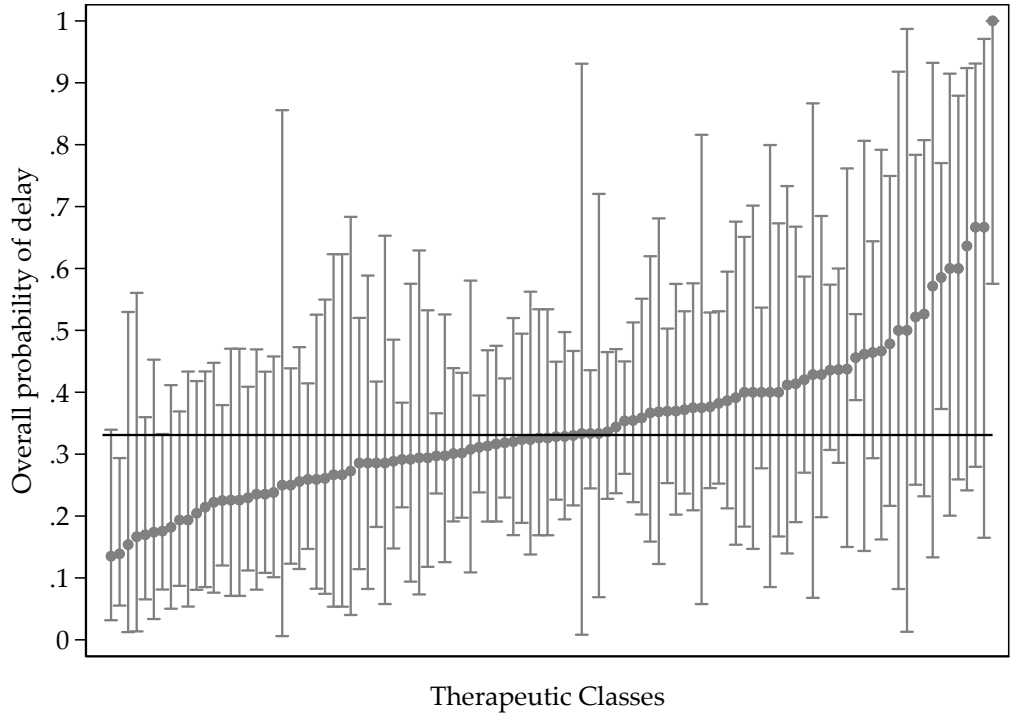


Figure B.6: *Probability of Delay in Eastern Europe by Therapeutic Class*

This graph plots the overall probability of delay in Eastern European countries for drugs separated according to their therapeutic class. Vertical bars represent the 99% confidence interval, while the black horizontal line is the estimate for ψ_{EU10} obtained from the overall sample.

positives (the partition includes around 100 sets), we use 99% confidence intervals around the point estimates. We find that in three instances the parameter estimate for $\bar{\psi}_{gj}$ is lower and significantly different from the estimate obtained from the full sample. The three classes are erectile dysfunction, cytotoxics (i.e. drugs for chemotherapy), and nasal corticosteroids without anti-infectives. The point estimates for the first two classes fall within the confidence interval for ψ_{EU10} that we calculated in the paper (curiously, the point estimate for erectile dysfunction drugs almost exactly coincides with our estimate for the lower bound). The latter category only contains one drug, fluticasone furoate (Avamys, by Glaxosmithkline), which entered with no delay in all Eastern European countries.

The main downside of this approach is that it relies heavily on the assumption that ψ_{EU10} is constant across drugs. This may not be true if, for example, the government has accelerated

procedures for drugs that are considered important. This may be the case with oncologics (though it is unclear why the same distinction would apply to the other two categories), and suggests that further work on heterogeneity across drugs may be warranted.

B.5.2 Moment Inequalities with Expected Profits Data

In the empirical implementation of the revenue-based moment inequalities we were able to generate a lower bound on the probability of an idiosyncratic delay, but not an upper bound. This is a result of certain features of the model and the data. In particular, we find that the expected profits of the firm are decreasing in the probability of an idiosyncratic delay. This implies that the inequalities will generate a one-directional bound on the parameter (for details, see Section 2.6 of the paper).

In this section we discuss an extension of the moment inequalities which can be helpful in generating an upper bound in these situations. The idea of this extension is to prove that for a given value of the idiosyncratic delay parameter no strategy will ever yield expected profits as high as what the firm obtained in the data.

The basic idea is to use the expected profit of the firm when facing a simplified environment that is tweaked to be more favorable to the firm (relative to the true environment). For example, one obvious candidate for $F^{ub}(\cdot)$ in the empirical application we consider is the expected profits of the firm when ERP is eliminated. In this scenario, the firm always earns more money, since the externality of ERP is eliminated. Moreover, with this structure, the model has a clear, unique solution: applying for entry everywhere right away.

This function satisfies two main criteria: it is easy to compute, and it is always an upper bound for $V(\mathcal{A}^{\star}(\psi); \psi, \cdot)$. Unfortunately this function also does not deliver a meaningful bound when tested empirically: it was unable to reject any value of the parameters that had not already been rejected using our original moment inequalities. Nonetheless, this approach could provide a promising avenue for future research.