Patent challenge and generic entry

Xin Zhang* Hanlin Zhao†

December 9, 2025

Please click here for the latest version.

Abstract

Pharmaceutical innovation depends on strong primary patents that allow originators to recoup R&D costs. However, drug companies often engage in evergreening that prolongs patent protection by filing follow-on patents with little therapeutic gain. We study a policy lever that works with market forces to screen out weak follow-on patents: the Hatch-Waxman Act, which incentivizes challenges to evergreening patents by granting the first successful challenger a period of marketing exclusivity. We investigate how the length of first-filer exclusivity shapes generic firms' incentives to initiate challenges, which can curb the extra monopoly protection created by evergreening while preserving incentives for genuine discovery and protecting consumer welfare through earlier generic entry. Using a two-stage structural model that endogenizes challenge and entry decisions, we estimate the fixed costs of generic entry with moment inequalities. We find that the current 180-day exclusivity raises challenge rates by about 4 percentage points. Extending exclusivity primarily activates challenges in markets that would otherwise go unchallenged: a two-year exclusivity increases the challenge rate to 15.38%. Effective exclusivity is highly heterogeneous across therapeutic classes: reaching a 20% challenge rate requires roughly two years for antimicrobials but less than one year for genitourinary drugs.

Keywords: Patent, generic drugs, entry, exclusivity, pharmaceutical industry

*Toulouse School of Economics. Email: xin.zhang@tse-fr.eu
†Toulouse School of Economics. Email: hanlin.zhao@tse-fr.eu

1 Introduction

Patent protection plays a central role in incentivizing innovation by granting inventors exclusive rights for a fixed period, enabling them to recover research and development (R&D) investments. In the pharmaceutical industry, drug development is characterized by high R&D costs and lengthy, expensive clinical trials, making innovation particularly risky and resource-intensive. As a result, patent protection is critical for firms to recoup their investments. Beyond primary patents covering novel active ingredients, pharmaceutical companies increasingly rely on secondary patents, which protect auxiliary features such as formulations (e.g., dosage forms, routes of administration) and manufacturing processes.

Secondary patents are frequently criticized as a form of "evergreening," a strategy that prolongs market exclusivity without offering significant therapeutic improvements. This practice is widespread in the pharmaceutical industry. In 2009, the ratio of primary to secondary patents was around 1:7 (European Competition Commission (2009)). Additionally, 78% of drugs associated with new patents were existing drugs, and 70% of drugs had more than one secondary patent (Feldman (2018)). Moreover, secondary patents defer generic entry, restrict competition, sustain high prices, and hurt consumers. Gupta (2023) finds that the presence of multiple patents delays generic entry by over three years per drug and imposes an additional $25.6 billion in costs on consumers.

Addressing the consumer welfare losses caused by excessive secondary patenting is a challenging task. One potential response is to reform patent law either by narrowing the scope or duration of secondary patents or by enhancing the scrutiny of their validity. Legal reform, however, is often slow, politically contentious, and vulnerable to industry lobbying. Furthermore, international agreements such as the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS)[1] also constrain unilateral policy changes. A more feasible and targeted alternative is to strengthen incentives for generic firms to challenge weak or invalid patents in court.

Generic entry typically occurs after patent expiration, but it can happen earlier through Paragraph IV (PIV) challenges when submitting the Abbreviated New Drug Application (ANDA) to the United States Food and Drug Administration (FDA). To further encourage early generic entry, the Hatch-Waxman Act of 1984 grants the first PIV challenger a 180-day period of marketing exclusivity, during which no other ANDAs for the same drug can

[1]The TRIPS Agreement sets minimum standards for the protection and enforcement of intellectual property rights among WTO member states. See in `https://www.wto.org/english/docs_e/legal_e/27-trips_01_e.htm`

be approved. This exclusivity provides a temporary period during which only the brand-name drug can share the market with this generic challenger. It offers significant financial incentives to offset the risks and costs associated with patent litigation. This regulatory framework, combining litigation incentives with temporary exclusivity, is intended to facilitate timely generic entry and enhance competition, all without fundamentally altering the legal structure of pharmaceutical patenting. It plays a critical role in balancing two competing goals: promoting pharmaceutical innovation and ensuring affordable access to medicines.

Although the 180-day exclusivity period is intended to spur earlier generic entry, many branded products, especially in certain classes such as respiratory, face few PIV challenges[2]. This paper studies how exclusivity shapes generic firms' decisions to initiate PIV challenges. On the benefit side, a successful first-filer can earn sizable rents during exclusivity (e.g., up to $60 million[3]), which we frame as the exclusivity-rent effect. On the cost and risk side, initializing a PIV challenge entails substantial fixed costs (roughly $5–$10 million[4]), litigation uncertainty, and the prospect that rents materialize only if the challenger both prevails and secures first-filer status. Consequently, expected profits may be insufficient to induce entry. Moreover, stronger exclusivity incentives can attract additional challengers, intensifying competition for first-filer status and dissipating expected rents, a business-stealing effect that can further deter entry. The net effect of exclusivity is therefore theoretically ambiguous ex ante. We build structural models to quantify those forces and exploit policy levers that operate on both margins to encourage patent challenges. One approach is to extend the exclusivity period for drugs with few challenges, thereby broadening the high-profit window[5]. Another approach is to reduce fixed costs by streamlining administrative procedures and improving FDA–industry communication. To summarize, we analyze how exclusivity duration and fixed-cost reductions jointly determine the incidence of PIV challenges.

We study the U.S. prescription drug market. Our dataset covers 9,137 drug applications and 14,437 NDCs (National Drug Codes) from 2003 to 2022. We document two patterns from the data. First, average PIV challenge rates are low and vary markedly across therapeutic classes. For example, central nervous system, cardiovascular, and antimicrobial

[2] https://www.healthaffairs.org/doi/abs/10.1377/hlthaff.2022.00873

[3] https://www.sciencedaily.com/releases/2009/10/091015141507.htm

[4] https://www.sciencedaily.com/releases/2009/10/091015141507.htm

[5] In light of the relatively low PIV penetration and challenge rates in the U.S., some researchers have advocated for extending exclusivity for certain drugs with limited patent challenges (source).

drugs together account for nearly half of all ANDA submissions, yet their PIV challenge rates fall below the overall average. By contrast, classes such as blood products, antiparasitics, and antidotes attract far fewer applications but exhibit substantially higher challenge rates. Second, generics associated with PIV challenges enter the market, on average, seven years earlier than those without challenges, suggesting that patent challenges substantially accelerate the entry of generics.

To investigate the role of exclusivity and the scope for policies that encourage PIV entry, we construct a structural model to back out the fixed costs of challenge and simulate counterfactual policies. The model has two stages. In the first stage, each generic firm selects one of three strategies: (1) not developing and not entering the market; (2) investing in PIV generic development; or (3) investing in non-PIV generic development. All generic firms incur reverse engineering costs to develop bioequivalent products. PIV challengers incur additional legal and technical costs to design around or invalidate the patent and face litigation uncertainty. Firms that invest without targeting a PIV challenge become regular generic entrants and enter the market only after patent expiration. Firms that pursue a PIV challenge initiate litigation against the brand-name manufacturer. The outcome of the challenge is uncertain and depends on the strength of the underlying patent. Generic firms form rational expectations over litigation outcomes, future market structure, and competitors' pricing behavior. They make entry decisions by comparing the expected net present value of profits to fixed costs. Each firm selects the strategy that yields the highest expected surplus.

In the second stage, firms compete in prices to maximize their product-level profits. On the demand side, we employ a nested logit model, which allows for different substitution patterns between branded drugs and generics, and captures consumers' higher willingness to pay for branded products and inertia in switching to later generic entrants. Using the demand estimates, we recover marginal costs and simulate firm-level profits under different market structures. Firms' entry decisions are modeled by comparing these predicted profits against fixed costs, which are specified as a linear function of observables and an unobservable cost shock. [6]

We identify the parameters of the fixed cost function using inequality restrictions derived from Nash equilibrium conditions. Observed entry choices reveal profit orderings

[6]Branded-drug firms may compensate generic firms to delay market entry, a practice known as pay-for-delay or reverse payments. We do not have access to relevant data on this issue and therefore abstract from it in this paper. Drake and McGuire (2025) provides an analysis of reverse payments.

across strategies, generating moment inequalities that we exploit in estimation. Because these restrictions need not deliver point identification, inference explicitly accommodates partial identification.

Our estimates suggest that the average profit from 180-day exclusivity is $5.7 million, while the average fixed costs of patent challenge are $6.04 million, approximately $2.1 million higher than the fixed costs of regular (non-PIV) generics. This helps explain low challenge rates. Moreover, we evaluate the importance of exclusivity and alternative policies with larger exclusivity rent and smaller fixed costs of challenge. A 180-day exclusivity substantially increases the PIV challenge rate to 14.37% compared with 10.27% in the absence of exclusivity. In addition, longer exclusivity periods provoke patent challenges noticeably in previously unchallenged groups, with a 2-year exclusivity leading to a challenge rate of 15.38%. Reducing fixed costs can produce a similar effect: a 10% cost reduction is equivalent to 1.5-year exclusivity. Additionally, the required exclusivity periods for a significant increase in challenge rate vary across therapeutic classes. For instance, 2-year exclusivity leads to a 20% challenge rate for antimicrobials, whereas blood products or genitourinary drugs would require less than one year. Taken together, our findings show that finely tuned first-filer exclusivity, a market-based incentive embedded in the existing regulatory framework, can curb evergreening, accelerate generic entry, and enhance consumer welfare without weakening core rewards for true pharmaceutical innovation.

Related Literature This paper speaks to the economics of innovation and intellectual property. Pharmaceutical R&D is costly and lengthy, with per-molecule estimates ranging from $1 billion to $3 billion (Schlander et al. (2021)). Patents play a central role by securing returns for innovators, but they also confer temporary monopoly power that delays generic competition. Beyond initial patents, branded manufacturers often accumulate additional patents to extend effective exclusivity, a strategy known as "evergreening", which allows them to sustain higher profits. Empirical studies document the prevalence of such patent accumulation, and show that it delays generic entry and harms consumers (Feldman (2018) and Gupta (2023)).

Working on patent protection design in this setting, we address an important policy question: how can policymakers incentivize earlier entry of low-cost generics while preserving rewards for true pharmaceutical innovation? One approach is to expand financial incentives to drug companies. Policies designed to extend profitable windows for drugs, such as priority review and transferable exclusivity extensions, illustrate this strat-

egy. For instance,Ridley et al. (2006) propose Priority Review Vouchers for neglected diseases, which shorten FDA review times by roughly one year and accelerate access to drugs. Dubois et al. (2022b) analyze transferable patent-extension vouchers that grant additional exclusivity and study implications for R&D incentives and social welfare. While these studies primarily focus on the entry of branded drugs, our focus is on generic entry. Patent challenges enable successful generic firms to enter the market before patent expiration. Moreover, under the Hatch-Waxman Act, the first successful PIV filer receives 180-day marketing exclusivity, a mechanism similar to a targeted voucher for challenging weak patents. A growing empirical literature shows that patent challenges and exclusivity incentives disproportionately target the highest-sales drugs, lower-quality patents, and later-expiring patents (Grabowski and Kyle, 2007; Hemphill and Sampat, 2011; Panattoni, 2011; Hemphill and Sampat, 2012; Grabowski et al., 2017). Branstetter et al. (2016) uses a Nested Logit model to quantify the welfare gains of around $78 billion from PIV-facilitated generic entry by comparing the scenarios with and without PIV-facilitated generic drugs, which provides an upper bound on the potential welfare gain from patent challenges. We build on this line of work by explicitly modeling firms' entry and challenge decisions, investigating how the length of exclusivity influences generic challenge behavior.

Methodologically, this paper is related to the literature on endogenous entry in differentiated-product markets (Bresnahan and Reiss, 1991; Mazzeo, 2002; Seim, 2006; Wollmann, 2018; Alam and Conti, 2024; Starc and Wollmann, 2025). This framework enables us to estimate fixed costs of entry and perform counterfactual policy analysis. Among these papers, Starc and Wollmann (2025) is most closely related and study entry in the U.S. pharmaceutical industry within the context of collusion. We extend their model and derive new moment inequalities to quantify the additional entry costs associated with PIV challenges and analyze a different question: how the length of exclusivity shapes generics' entry decisions.

Finally, this paper relates to the literature on limiting prescription drug prices. Håkonsen et al. (2009) assesses different price control policies and Dubois et al. (2022a) evaluates the effects of an international reference pricing policy on U.S. drug prices. Rather than imposing price ceilings, we propose to vary the length of the first-filer exclusivity to induce patent challenges and earlier generic entry, which can lower prices through competitive forces without direct price regulation.

The rest of the paper is organized as follows. Section 2 presents the data and empirical evidence. Section 3 describes the structural model and estimation. Section 5 conducts counterfactual policy analyses, and section 6 concludes.

2 Data and empirical evidence

2.1 Industry background

The U.S. prescription drug market is large and growing. In 2023, its value was approximately $602 billion, an increase of over 30% relative to six years earlier. Per capita drug expenditure in the U.S. is about twice that of leading European countries. Concerns about high prices for branded therapies have, in turn, motivated policies that accelerate generic entry.

To enter the pharmaceutical market, manufacturers must submit an application to the FDA. Brand-name drugs file a New Drug Application (NDA), while generic drugs are submitted through an Abbreviated New Drug Application (ANDA). The FDA grants approval if the application demonstrates safety and efficacy, and there are no unresolved legal issues. Once approved, the drug is marketed under one or more National Drug Codes (NDCs), which identify specific products by their ingredients, dosage form, strength, and packaging. A single application may correspond to multiple NDCs.

The regulatory framework governing generic entry is the Hatch-Waxman Act. Generic manufacturers seeking approval file an ANDA to demonstrate pharmaceutical equivalence and bioequivalence of their products to the reference listed drug (RLD)[7]. For each patent listed for the RLD by the FDA, the ANDA must include one of four certifications:

1. Paragraph I Certification: No relevant patent exists or has been filed.
2. Paragraph II Certification: The patent has already expired.
3. Paragraph III Certification: The generic will not enter the market until after patent expiration.
4. Paragraph IV Certification: The patent is invalid, unenforceable, or will not be infringed by the generic product.

Although generic entry typically occurs after patent expiration, it can happen earlier through Paragraph IV (PIV) challenges when submitting the ANDA to the FDA. This legal pathway requires generic manufacturers to invest not only in reverse engineering and development to show that the patents are invalid or not infringed, but also in litigation to defend their claims. For example, Cephalon's branded drug Provigil was protected by a

[7] See the FDA Glossary of Terms for details.

primary patent on the active ingredient, modafinil, which expired in 2010, as well as three secondary patents that extended protection in the U.S. until 2024. Several generic manufacturers, including Teva, Mylan, Ranbaxy, Barr, and Apotex, challenged these secondary patents. The secondary patent covering particle size[8] was invalidated in 2011, while the other two formulation-related patents were circumvented by generic firms through alternative drug designs. As a result, generic versions of Provigil entered the U.S. market in 2012.

A Paragraph IV certification (PIV) triggers a sequence of events. The generic applicant must notify the brand-name manufacturer and the patent holder, who then have 45 days to initiate an infringement lawsuit. If litigation is initiated within that window, the FDA is barred from granting final approval for up to 30 months, unless the case is resolved earlier or the generic challenger prevails. If no suit is filed within 45 days, the FDA may approve the ANDA immediately.

Hatch–Waxman also creates a specific entry incentive: the first applicant to file a complete ANDA with an approved PIV certification is eligible for 180 days of marketing exclusivity. During this exclusivity period, additional generic competitors referencing the same RLD are excluded from the market. Olson and Wendling (2018) documents that the initial PIV entry leads to substantial drug price reductions, with further decreases as subsequent generic products enter. These institutional features, specifically the 180-day first-filer exclusivity, create policy-driven returns to early challenge and structure the strategic entry problem that we analyze in the remainder of the paper.

2.2 Data

We combine several datasets for our analysis. The primary source is the Medicaid State Drug Utilization Data (SDUD), published by the Centers for Medicare & Medicaid Services. This dataset provides quarterly records of the number of prescriptions and total reimbursements at the National Drug Code (NDC) level across all U.S. states. Each NDC uniquely identifies a product by active ingredient, dosage form, strength, and package size. We aggregate NDCs by ingredient, dosage form, and strength to align product definitions across sources. We measure quantities by the number of prescriptions and define prices as the average reimbursement per prescription (total reimbursement divided by prescriptions).

[8]Particle size refers to the size of the solid particles of the drug substance before they are compressed into a tablet or suspended in a solution. It affects dissolution rate, absorption, bioavailability, and stability of the drug.

As Medicaid primarily serves low-income beneficiaries, SDUD-based quantities likely understate national utilization and thus firms' total sales and profits. Unfortunately, we do not have access to comprehensive national-level sales data (e.g., data from IQVIA). To address this limitation, we follow the approach in Starc and Wollmann (2025), who combine SDUD with national sales data from IQVIA and find close alignment after rescaling. We apply their scaling factor to adjust our quantities to the national level.

We augment the SDUD with FDA sources. First, the FDA's Orange Book[9] provides detailed information on each drug's active ingredients, dosage form, strength, manufacturer, and patent listings. Additionally, we scraped data on PIV certification status and ANDA submission dates from FDA approval letters[10].

Finally, we extract the therapeutic classes at the NDC level using public mapping code from Kury and Bodenreider[11]. Our final sample covers all drugs sold between 2003 and 2022, which includes 14,437 NDCs and 9,137 application numbers, of which 1,197 have PIV certification.[12]

2.3 Descriptive evidence

In this section, we present two descriptive facts from the data that frame our analysis of PIV challenges.

Heterogeneity in challenge rates across therapeutic classes. Figure 1 presents PIV challenge rates by 27 therapeutic classes, defined using the Veterans Affairs (VA) classification system[13]. We define the challenge rate as the share of generic applications in a class with a PIV certification. The average challenge rate is low, around 15%, and varies markedly across therapeutic classes. The central nervous system, cardiovascular, and antimicrobial classes together represent approximately half of all ANDAs, yet each exhibits a below-average challenge rate. This pattern suggests that, even though these classes are attractive to generic manufacturers, firms in these classes frequently choose to wait for patent

[9] `https://www.fda.gov/drugs/development-approval-process-drugs/drug-approvals-and-databases`.

[10] `https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm`.

[11] Fabricio Kury and Olivier Bodenreider proposed R codes to map NDC and therapeutic class (Anatomical Therapeutic Chemical (ATC) or Veterans Administration (VA) classes): `https://github.com/fabkury/ndc_map/blob/master/ndc_map.R`.

[12] Throughout this paper, we distinguish between applications (NDA or ANDA), defined at the ingredient-dosage form level, and products (NDC), aggregated at the ingredient-dosage form-strength level.

[13] See `https://www.ihs.gov/RPMS/PackageDocs/PSN/psn318u2.pdf`.

expiration rather than initiate patent challenges, which incur the up-front costs and risks of litigation. In contrast, blood products, antidotes, and antiparasitics have fewer ANDA filings but display substantially higher PIV challenge rates.

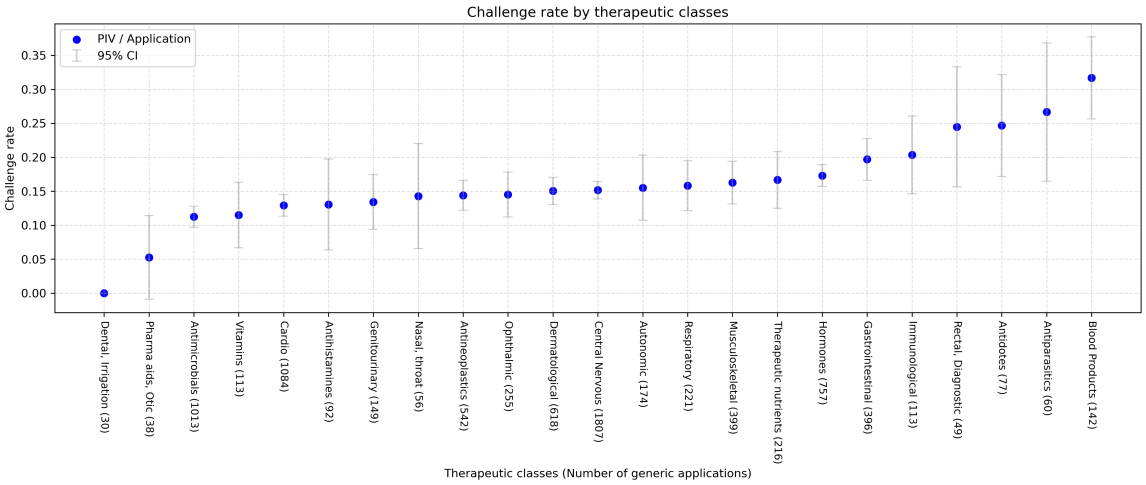


Figure 1: Challenge rate across therapeutic classes

PIV challengers enter about seven years earlier. Figure 2 compares the timing of market entry between PIV generic drugs and regular (non-PIV) generic drugs. We measure delay as the number of years between the branded drug's initial marketing date and the generic drug's market entry. The distribution of delays differs substantially: PIV generics are concentrated at much shorter delays, whereas regular generics exhibit a wider and later distribution. Specifically, on average, PIV generic drugs enter the market about seven years earlier than non-challenging generics, consistent with patent challenges accelerating the pathway to entry.

In summary, these findings emphasize the significant impact of PIV challenges in facilitating early generic competition, though the overall challenge rate is modest on average and highly uneven across therapeutic classes. These facts motivate our structural analysis of how exclusivity incentives and fixed costs shape challenge decisions, thereby simulating alternative policies to encourage PIV challenge.

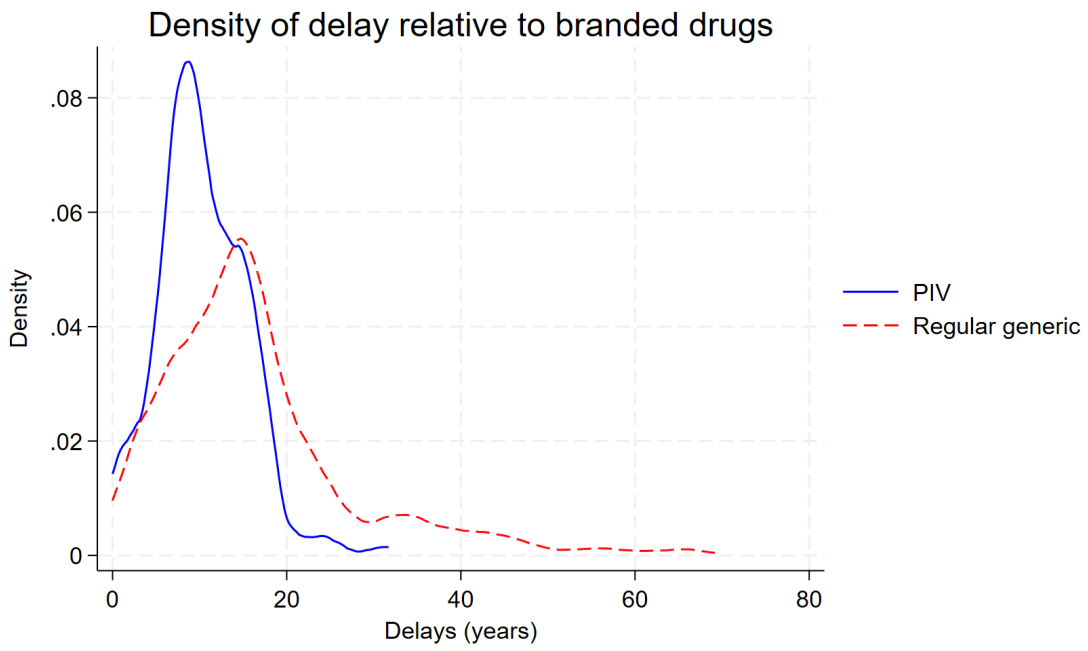


Figure 2: Density of delay in start marketing date to the branded drugs

3 Model

In this section, we present a two-stage model of generic entry and price competition under the Hatch-Waxman Act. The model captures (1) generic firms' entry choices, including whether to initiate a PIV challenge, and (2) subsequent Bertrand-Nash pricing competition in differentiated product markets. We solve by backward induction: Stage II characterizes pricing and profits for any realized market structure. Stage I maps these profits into entry payoffs and equilibrium choices.

3.1 Stage II: Price competition

3.1.1 Demand

On the demand side, we define a market mt by the combination of an active ingredient m and a half-year period t.[14] A product d is defined at the ingredient-dosage form-strength-firm level (i.e., the first nine digits of the NDC).[15]

[14]This half-year time definition aims to match with the 180-day (i.e., one half-year) exclusivity window. We obtain similar results using annual periods.

[15]Throughout, we index firms implicitly via products d.

We use a Nested Logit model to estimate demand. In each market mt, a buyer i selects a nest $g \in \{\text{Brand}, \text{Generic}\}$, either a branded drug or a generic drug.[16] Conditional on the nest, the buyer i chooses the drug d offering the highest indirect utility within the nest g or the outside option of not purchasing. The indirect utility of buyer i for drug d in market mt is

$$\begin{aligned} v_{idmt} = \lambda_m + \lambda_t + \alpha \ln p_{dt} + \beta_1 \mathbf{1}\{\text{Brand}\}_d + \beta_2 Npack_{dt} + \beta_3 \mathbf{1}\{\text{FirstG}\}_d \\ + \xi_{dmt} + \zeta_{igmt} + (1-\sigma)\epsilon_{idmt} \end{aligned}$$

with outside utility $v_{i0mt} = \epsilon_{i0mt}$. λ_m and λ_t represent the ingredient and time fixed effects, respectively. $\ln p_{dt}$ is the logarithm of the price[17]. To capture buyers' specific interests in branded drugs, we include a dummy variable $\mathbf{1}\{Brand\}_d$ indicating if the drug d is branded. $Npack_{dt}$ measures the number of available package sizes for drug d, capturing buyers' interests for products of more flexible packaging. As the first generic entrants secure early market access and establish solid relationships with wholesalers, we include a dummy variable $\mathbf{1}\{FirstG\}_d$ indicating whether the drug d is the first generic entrant, to capture its comparative advantage over later entrants. ξ_{dmt} is the unobserved product-specific shock, such as unobserved quality. The unobservables $\xi_{dmt}, \zeta_{igmt}, \epsilon_{idmt}$ are independently and identically distributed. The idiosyncratic shock ϵ_{idmt} follows an i.i.d. Type-I extreme value distribution. ζ_{igmt} is a nest-specific shock and distributed such that $\zeta_{igmt} + (1-\sigma)\epsilon_{idmt}$ also follows the Type-I extreme value distribution. $\sigma \in (0,1)$ is the nesting parameter, which governs the degree of correlation in unobserved utility within a group.

Let $\delta_{dmt} = \lambda_m + \lambda_t + \alpha \ln p_{dt} + \beta_1 \mathbf{1}\{\text{Brand}\}_d + \beta_2 Npack_{dt} + \beta_3 \mathbf{1}\{\text{FirstG}\}_d + \xi_{dmt}$ denote the mean utility. Define the within-nest inclusive value considering all products d in group g and ingredient m, i.e., $d \in \mathcal{G}_{gmt}$

$$D_{gmt} = \sum_{d \in \mathcal{G}_{gmt}} \exp\left(\frac{\delta_{dmt}}{1-\sigma}\right).$$

[16] We assume buyers in our framework are intermediaries, such as wholesalers, group purchasing organizations, and large retail chains, rather than end consumers or patients. This distinction matches the structure of the Medicaid data, in which transaction prices are based on reimbursements to these intermediaries, rather than out-of-pocket costs or list prices paid by the patients.

[17] As our sample includes drugs across different ingredients and time, we use the logarithm of the price to accommodate wide price dispersion. See, e.g., Dubois et al. (2022a) and Atal et al. (2022).

The market share of product d in ingredient m at period t is

$$s_{dmt} = \bar{s}_{d|g,mt} \cdot \bar{s}_{gmt} = \frac{\exp\left(\frac{\delta_{dmt}}{1-\sigma}\right)}{D_{gmt}^{\sigma}\left[\sum_g D_{gmt}^{1-\sigma}\right]}$$

where $\bar{s}_{d|g,mt}$ is the selection probability of drug d conditional on group g in ingredient m at period t

$$\bar{s}_{d|g,mt} = \frac{\exp\left(\frac{\delta_{dmt}}{1-\sigma}\right)}{D_{gmt}},$$

and $\bar{s}_{gmt}$ is the unconditional probability of the group g in ingredient m at period t

$$\bar{s}_{gmt} = \frac{D_{gmt}^{1-\sigma}}{\sum_g D_{gmt}^{1-\sigma}}.$$

3.1.2 Supply

On the supply side, firms set prices p_{dt} to maximize the profits of individual drugs. Let M_d denote the potential market size for drug d of ingredient m.[18]. Product d's period profit is

$$\pi_{dt} = (p_{dt} - mc_{dt})\, s_{dt} M_d$$

where mc_{dt} is the marginal cost. We parameterize marginal costs as

$$\ln(mc_{dt}) = \gamma_m + \gamma_t + \omega_{dt} \tag{1}$$

with ingredient and time cost fixed effects (γ_m, γ_t) and i.i.d. shocks ω_{dt}.

3.2 Stage I: Entry

In the first stage, a generic firm has a one-time opportunity[19] to make the entry decision for an ANDA in ingredient-dosage form j[20]. Each ANDA applies to all strengths of the drug.

[18]We follow Starc and Wollmann (2025) and proxy the market size as 1.5 times the maximum observed quantity over time for drugs of ingredient m.

[19]We assume once the firms complete the research, they will directly apply to the FDA to file an ANDA. They do not strategically choose the entry timing. Moreover, we focus on whether an ANDA challenges patents, rather than the timing of the challenge, as we believe that encouraging more challenges is of first-order importance, given the low average challenge rate in the industry.

[20]Drug companies make entry decision in ingredient-dosage form j, but the price competition is in ingredient m to account for the substitution across dosage forms. For example, ingredient Paracetamol has several

Firms must choose one of three options for ingredient-dosage form j: invest in PIV generic development, invest in non-PIV generic development, or not invest and not enter.

Each option incurs distinct fixed costs. Investing in non-PIV generic development entails fixed costs to replicate the branded drug and prove the ability to produce on a large scale, which are considered reverse-engineering costs. Those regular generics enter the market after the patent has expired. Investment in PIV generic development involves additional costs to challenge existing patents beyond the reverse-engineering cost. For example, if an ANDA challenges a branded drug's formulation patent, the firm must conduct further experiments to develop an alternative formulation and demonstrate its efficacy and safety. Additionally, the firm incurs further application and litigation fees. If the generic firm files a PIV ANDA, it enters before the patent expiration and may obtain a 180-day exclusivity period upon winning the lawsuit with a certain probability. If not, the ANDA becomes a regular generic.

After filing an ANDA, there is an uncertain duration of delay, denoted by D, during which the FDA reviews the application, and the case is settled upon challenge. The review process is simpler for regular generics, so we assume different distributions $F_{D,\text{reg}}$ for non-PIV filings and $F_{D,\text{PIV}}$ for PIV filings (reflecting more complex review and litigation). Drug exit typically results from supply disruptions outside the firm's control. Therefore, the model assumes that generic firms only consider entry decisions and do not plan for exit ex ante.

Firms are assumed to know the distributions of demand shocks $\mathcal{F}_\xi$, marginal cost shocks $\mathcal{F}_\omega$, and delay durations $\mathcal{F}_{D,reg}$ and $\mathcal{F}_{D,PIV}$. They form rational expectations regarding future realizations of these random variables. Moreover, firms know the fixed effects on utility (λ_m and λ_t) and on marginal costs(γ_m and γ_t). The set of drugs for an ANDA of firm f in ingredient-dosage form j is $d \in \mathcal{D}_{fj}$, specified at the ingredient-dosage form-strength level. To characterize the market structure for an ingredient-dosage form, it is necessary to determine the number of applications (ANDAs and NDAs) and the number of products (NDCs) for each application. We assume that firms can perfectly anticipate, for each ingredient-dosage form, the maximum number of branded drug applications (NDA) $N_{B,j}$, first PIV successful applicants $N_{F,j}$, subsequent successful PIV applicants $N_{S,j}$, regular ANDAs with the same ingredient-dosage form $N_{R,j}$, and ANDAs with the same ingredient

dosage forms, including tablets and suspension.

but different dosage forms $N_{R,-j}$[21][22].

Regular (non-PIV) generics. The value function of a regular (non-PIV) generic ANDA fj is a sum of its products d: [23]

$$V_{fj}^{R}(N_{B,j}, N_{F,j}, N_{S,j}, N_{R,j}, N_{R,-j}, t_{allowed,j}, t_{n_{R,j}=N_{R,j}}) = \sum_{d \in \mathcal{D}_{fj}} \Bigg\{ \sum_{t=t_{allowed,j}}^{t_{n_{R,j}=N_{R,j}}} \delta^t \times F_{D,reg}(t) \times$$

$$\sum_{n_{R,j}=0}^{N_{R,j}-1} \sum_{n_{R,-j}=0}^{N_{R,-j}} \underbrace{\rho(N_{R,j}-1, n_{R,j}, t, F_{D,reg})}_{\text{Prob of } n_{R,j} \text{ out of } N_{R,j}-1 \text{ in the market}} \times \underbrace{\rho(N_{R,-j}, n_{R,-j}, t, F_{D,reg})}_{\text{Prob of } n_{R,-j} \text{ out of } N_{R,-j} \text{ in the market}}$$

$$\times \int_{\xi} \int_{\omega} \pi_{fdt}(N_{B,j} + N_{F,j} + N_{S,j} + n_{R,j} + n_{R,-j} + 1) dF_{\xi} dF_{\omega} \times \mathbf{1}\{t \geq t_{allowed}\}$$

$$+ \sum_{t=t_{n_{R,j}=N_{R,j}}}^{T} \delta^t \int_{\xi} \int_{\omega} \pi_{fdt}(N_{B,j} + N_{F,j} + N_{S,j} + N_{R,j} + N_{R,-j}) dF_{\xi} dF_{\omega} \times \mathbf{1}\{t \geq t_{mature}\} \Bigg\}. \tag{2}$$

where δ is the discount factor. The value function of a regular generic ANDA starts from the time that it is allowed to enter ($t_{allowed}$). Entry can occur at patent expiration or after the 180-day exclusivity period if a PIV ANDA invalidates the branded drug's patents. The value function consists of two parts. In the first part, the market already contains branded drugs and PIV generics, so only regular generics are now entering. Among these regular generics, those with the same ingredient and dosage form as ANDA fj are denoted by $n_{R,j}$, while those with the same ingredient but a different dosage form are denoted by $n_{R,-j}$. The incumbents in the market include branded drug NDAs ($N_{B,j}$), as well as first PIV applicants ($N_{F,j}$), who benefit from exclusivity, and second PIV applicants ($N_{S,j}$), who do not benefit from exclusivity but can enter after it and before patent expiration. The number of new regular generics entrants, n, can vary from 0 to the maximum number N in each

[21] The drugs with the same ingredient but different dosage form matter since we allow substitution cross dosage forms in the demand model.

[22] In practice, brand-name firms may launch authorized generics (AGs) i.e., generic versions of their own drugs marketed without the brand name. AGs can enter before patent expiration and are sometimes used strategically by brand-name firms to deter generic entry. However, AG entry is relatively uncommon in our sample, so we model it as an exogenous event rather than as part of the brand's strategic behavior. Alam and Conti (2024) provide a detailed analysis of AG entry.

[23] In this expression, we abuse the notation and only illustrate the number of applications, but the real profits are for drugs (NDCs). We use the average number of NDCs across ingredient-dosage forms to compute the number of NDCs for each application.

period t. We include regular generics having the same ingredient and dosage form j, that is $n_{R,j} \in [0, N_{R,j}]$, and those having the same ingredient but different dosage forms, that is $n_{R,-j} \in [0, N_{R,-j}]$, since both sets of regular generic entrants compete in the competition stage. The flow profit is derived from the competition stage

$$\pi_{dt}(\cdot) = (p_{dt}(\cdot) - mc_{dt})s_{dt}(\cdot)M$$

For each possible market structure, composed by $(N_{B,j}, n_F, n_S, n_{R,j}, n_{R,-j})$ in a period t (before maturity), we compute the flow profits as in Starc and Wollmann (2025). Firms, knowing only the maximum number of applications, sum those flow profits weighted by the probability $\rho(\cdot)$ for each possibility combination to obtain the expected profit for a period. The probability $\rho(\cdot)$ follows a binomial distribution which is given by

$$\rho(a, b, t, F_D) = \frac{a!}{(a-b)!b!} F_D(t)^b [1 - F_D(t)]^{a-b} \tag{3}$$

Firms obtain expected profits of a potential market composition by integrating over the distributions of demand shock ξ and marginal cost shock ω. In the second part of the value function, no more drugs enter, and the market is mature. Firms obtain the same flow profit of ANDA fj afterwards[24].

PIV generics. The value function of a PIV generic ANDA fj is

$$\begin{aligned} &V_{fj}^{PIV}(N_{B,j}, N_{F,j}, N_{S,j}, N_{R,j}, N_{R,-j}, t_{end180,j}, t_{endpatent,j}, t_{n_{R,j}=N_{R,j}}) \\ &\quad = \sum_{d \in \mathcal{D}_{fj}} \sum_{t=0}^{T} \delta^t F_{D,PIV}(t) \times \left[P_W P_F V_{dt}^F + P_W (1 - P_F) V_{dt}^S + (1 - P_W) V_{dt}^{Fail} \right] \end{aligned} \tag{4}$$

[24]The value functions depend on the ingredient-dosage form-specific time thresholds, such as the timing that regular generics are allowed to enter ($t_{allowed,j}$). Firms do not know these time thresholds because of the uncertainty in the FDA review process. Therefore, their value functions need to take expectations over timing uncertainty, considering the distribution of delay F_D and all entry status scenarios across time. In each period t, we simplify that three potential scenarios arise as the market evolves before maturity: (1) during exclusivity, (2) subsequent PIVs enter following exclusivity, and (3) regular generics enter. The probability of each scenario is detailed in Appendix B

where P_W denotes the probability of a successful PIV challenge (i.e, winning the lawsuit[25], and P_F represents the probability of being the first challenger. These variables are central to linking possible outcomes in the value function.

The value function consists of three components, each representing a different outcome for the challenger. Specifically, V_{dt}^F is the value for an applicant who is both the first to challenge and succeeds. V_{dt}^S reflects the value for a successful challenger who is not the first. Finally, V_{dt}^{Fail} captures the value when the challenge is unsuccessful.

The value function of being a first successful applicant V_{dt}^F equals

$$
\begin{aligned}
V_{dt}^F(N_{B,j}, N_{F,j}, N_{S,j}, N_{R,j}, N_{R,-j}, t_{end180,j}, t_{endpatent,j}, t_{n_{R,j}=N_{R,j}}) = & \sum_{n_F=0}^{N_{F,j}-1} \sum_{n_{R,-j}=0}^{N_{R,-j}} \rho(N_{R,-j}, n_{R,-j}, t, F_{D,Reg}) \\
& \times \rho(N_{F,j}-1, n_F, t, F_{D,PIV}) \times \int_\xi \int_\omega \pi_{fdt}(N_{B,j}+n_F+1+n_{R,-j})\, dF_\xi\, dF_\omega \times \mathbf{1}\{t \le t_{\text{end180,j}}\} \\
& + \sum_{n_S=0}^{N_{S,j}} \sum_{n_{R,-j}=0}^{N_{R,-j}} \rho(N_{R,-j}, n_{R,-j}, t, F_{D,Reg}) \times \rho(N_{S,j}, n_S, t, F_{D,PIV}) \\
& \quad \times \int_\xi \int_\omega \pi_{fdt}(N_{B,j}+N_{F,j}+n_S+n_{R,-j})\, dF_\xi\, dF_\omega \times \mathbf{1}\{t_{\text{end180,j}} < t \le t_{\text{endpatent,j}}\} \\
& + \sum_{n_{R,j}=0}^{N_{R,j}} \sum_{n_{R,-j}=0}^{N_{R,-j}} \rho(N_{R,-j}, n_{R,-j}, t, F_{D,Reg}) \times \rho(N_{R,j}, n_{R,j}, t, F_{D,PIV}) \\
& \quad \times \int_\xi \int_\omega \pi_{fdt}(N_{B,j}+N_{F,j}+N_{S,j}+n_{R,j}+n_{R,-j})\, dF_\xi\, dF_\omega \times \mathbf{1}\{t_{\text{endpatent,j}} < t \le t_{n_{R,j}=N_{R,j}}\} \\
& + \int_\xi \int_\omega \pi_{fdt}(N_{B,j}+N_{F,j}+N_{S,j}+N_{R,j}+N_{R,-j})\, dF_\xi\, dF_\omega \times \mathbf{1}\{t > t_{n_{R,j}=N_{R,j}}\}.
\end{aligned}
$$

which includes four parts: 1) the PIV exclusivity period; 2) when PIV exclusivity ends but the existing secondary patent has not yet expired; 3) after the secondary patent has expired; and 4) the mature market stage.

[25]Here, we define a successful PIV challenge when we observe the PIV challenger's market entry before patent expiration. In addition to winning litigation, generics can achieve such entry if the innovator does not respond, if the case is settled out of court, etc.

The value function of being a subsequent successful applicant V_{dt}^{S} equals:

$$
\begin{aligned}
&V_{dt}^{S}(N_{B,j}, N_{F,j}, N_{S,j}, N_{R,j}, N_{R,-j}, t_{end180,j}, t_{endpatent,j}, t_{n_{R,j}=N_{R,j}}) = \\
&\quad \sum_{n_S=0}^{N_{S,j}-1} \sum_{n_{R,-j}=0}^{N_{R,-j}} \rho(N_{R,-j}, n_{R,-j}, t, F_{D,Reg}) \times \rho(N_{S,j}-1, n_S, t, F_{D,PIV}) \\
&\qquad \times \int_{\xi}\int_{\omega} \pi_{fdt}(N_{B,j}+N_{F,j}+n_S+1+n_{R,-j})\, dF_{\xi}\, dF_{\omega} \times \mathbf{1}\{t_{\text{end180,j}} < t \leq t_{\text{endpatent,j}}\} \\
&\quad + \sum_{n_{R,j}=0}^{N_{R,j}} \sum_{n_{R,-j}=0}^{N_{R,-j}} \rho(N_{R,-j}, n_{R,-j}, t, F_{D,Reg}) \times \rho(N_{R,j}, n_{R,j}, t, F_{D,PIV}) \\
&\qquad \times \int_{\xi}\int_{\omega} \pi_{fdt}(N_{B,j}+N_{F,j}+N_{S,j}+n_{R,j}+n_{R,-j})\, dF_{\xi}\, dF_{\omega} \times \mathbf{1}\{t_{\text{endpatent,j}} < t \leq t_{n_{R,j}=N_{R,j}}\} \\
&\quad + \int_{\xi}\int_{\omega} \pi_{fdt}(N_{B,j}+N_{F,j}+N_{S,j}+N_{R,j}+N_{R,-j})\, dF_{\xi}\, dF_{\omega} \times \mathbf{1}\{t > t_{n_{R,j}=N_{R,j}}\}.
\end{aligned}
$$

which includes three parts: 1) PIV exclusivity ends while the secondary patents remain active; 2) the secondary patents have expired; 3) the market has matured.

The value function of losing the lawsuit V_{fj}^{Fail} is the same as the value function of regular generic V_{fj}^{R},

$$
\begin{aligned}
&V_{fj}^{Fail}(N_{B,j}, N_{F,j}, N_{S,j}, N_{R,j}, N_{R,-j}, t_{allowed,j}, t_{n_{R,j}=N_{R,j}}) \\
&= V_{fj}^{R}(N_{B,j}, N_{F,j}, N_{S,j}, N_{R,j}, N_{R,-j}, t_{allowed,j}, t_{n_{R,j}=N_{R,j}}).
\end{aligned}
$$

Fixed costs. Similar to Starc and Wollmann (2025), we parameterize the fixed costs that firm f pays to introduce ANDA fj at ingredient-dosage form j are:

$$\theta_{fj} = \theta_0 + \theta_1 st_{fj} + \theta_2 ir_{fj} + \theta_3 PIV_{fj} + \eta_j$$

where st_{fj} is the number of strengths applied by firm f when submitting ANDA fj. ir_{fj} indicates whether the ANDA fj uses an irregular dosage form. PIV_{fj} is an indicator that equals 1 if ANDA fj claims patent challenge. This PIV_{fj} indicator captures additional fixed costs associated with the PIV challenge, such as lawsuit fees and technology development required to bypass or invalidate secondary patents. η_j represents an ingredient-dosage form-specific fixed cost shock that is unobservable to econometricians but known to firms at the time of entry. This shock is assumed to be symmetrically distributed and independent of the ingredient-dosage form characteristics (i.e., the number of strengths,

the delivery method, and the PIV status).

When making the entry decision for an ingredient-dosage form, each firm f forms a rational expectation of the fixed cost θ_{fj} based on its information set $\mathcal{I}_{fj}$:

$$\theta_{fj} = \mathbb{E}[\theta_{fj}|\mathcal{I}_{fj}] + \nu_{fj},$$

where ν_{fj} denotes each firm's expectations error, which, by construction, has a zero conditional mean $\mathbb{E}[\nu_{fj}|\mathcal{I}_{fj}] = 0$.

We assume an ingredient-dosage form-specific disturbance, η_j, that is common to all firms within ingredient-dosage form (i.e., market) j. The selection issue arises if markets with a high number of generic entrants have systematically different unobserved cost disturbances than those with few entrants.

While other specifications exist, they are less suitable for our context. One could, for instance, assume a firm-specific disturbance, η_f. In this case, the selection issue arises if firms that enter many markets have different cost draws than firms that enter few. However, it is more plausible that cost heterogeneity is driven by product characteristics rather than firm-level attributes, as pharmaceutical firms typically specialize in a limited range of drugs. Alternatively, one could assume a firm-market specific disturbance, η_{fj}. The selection issue arises when entrants into market j have more favorable cost draws than non-entrants (i.e., $\eta_{fj} < \eta_{f'j}$). Addressing this type of selection, particularly with a zero-mean assumption, is empirically challenging as it requires defining the set of all potential entrants for every market or the set of all potential markets for each firm.

Given these considerations, the η_j specification is the most appropriate. It aligns with industry realities and shifts the analytical focus from the identities of entrants to the number of entrants per market, which provides a more tractable approach to addressing the selection problem with moment inequalities that we detail in the next section.

The firm f compares the value function and expected fixed costs in ingredient-dosage form j. Generic firm f makes a decision for ANDA fj by choosing the action with the highest value from:

$$\{V_{fj}^{PIV} - \mathbf{E}[\theta_{fj}^{PIV}|\mathcal{I}_{fj}], V_{fj}^{R} - \mathbf{E}[\theta_{fj}^{R}|\mathcal{I}_{fj}], 0\} \tag{5}$$

We assume entry decisions form a Nash equilibrium. All firms make simultaneous decisions at $t = 0$ and commit to their future actions. Thus, we simplify the dynamic game into a static one, ensuring that each player chooses the best response.

4 Estimation

In this section, we present the estimation methods and estimators for our structural model.

4.1 Stage II: Demand

We utilize the market share inversion by Berry (1994), and the expression is as follows.

$$\begin{aligned}\ln s_{dt} - \ln s_{0t} = \lambda_m + \lambda_t + \alpha \ln p_{dt} + \beta_1 \mathbf{1}\{\text{Brand}\}_d + \beta_2 Npack_{dt} \\ + \beta_3 \mathbf{1}\{\text{FirstG}\}_d + \sigma \ln s_{d|g,t} + \xi_{dt}\end{aligned} \tag{6}$$

where $s_{d|g,t}$ is the share conditional on group g.

Drug prices $\ln p_{dt}$ and market shares $s_{d|gt}$ are endogenous. It is because $\ln p_{dt}$ correlates with the unobserved shock ξ_{dt}, and $s_{d|g,t}$ reflects an equilibrium outcome. Firms know the shock ξ_{dt} when setting prices, which leads to biased ordinary least squares (OLS) estimators for α and σ. We use the following instrumental variables: the number of competitors, the number of branded drugs, the number of products offered by a firm in period t, and indicator variables for therapeutic classes.

The estimation results are presented in Table 1. We find that $\hat{\sigma} = 0.406$, indicating a medium correlation among drugs within a group. The negative estimate, $\hat{\alpha} = -1.103$, indicates that buyers are sensitive to higher prices. Buyers have a higher willingness to pay for branded drugs, explaining the higher observed prices for these products. Drugs offered in a greater number of package sizes achieve higher market shares. And the first generic entrant enjoys a first-mover advantage by securing marketing channels and establishing connections with wholesalers. Buyers also show inertia in their purchasing behavior toward the first generic drug.

4.2 Stage II: Supply

From the FOC of profit maximization, we can recover the marginal cost as:

$$mc_{dt} = p_{dt} - (\frac{\partial s_{dt}}{\partial p_{dt}})^{-1} s_{dt} \tag{7}$$

where we recover $\frac{\partial s_{dt}}{\partial p_{dt}}$ from the demand estimation. After recovering $\hat{mc}_{dt}$, we can estimate the fixed effetcs $\hat{\gamma}_m$ and $\hat{\gamma}_t$, and the distributions of ω, $\hat{F}_\omega$ from equation 1.

Table 1: Demand estimates

	Estimate	SE
α	-1.103***	0.170
σ	0.406***	0.0859
Brand	1.012***	0.217
Npack	0.431***	0.0902
First Generic	0.286***	0.0559
FE ingredient	Y	Y
FE time	Y	Y
Observations	61 398	61 398

Standard errors are clustered at the drug-level.
* p<0.1, ** p<0.05, *** p<0.01

Using the demand estimates and the marginal costs, we can compute the expected profits for a hypothetical market structure.

4.3 Stage I: Successful challenge probability

The expected value function for PIV generics relies on two probabilities: the probability of being the first PIV challenger, denoted as P_F, and the probability of prevailing in patent litigation, denoted as P_W. These probabilities generate three outcomes for a PIV ANDA filer: first challenger and winner, subsequent challenger and winner, and unsuccessful litigant. The probability P_F is estimated using the sample mean of first-challenger status in the dataset, by assuming that firms behave according to the equilibrium path. The probability P_W is estimated with a Logit model, which relates the probability of a successful challenge to observable patent and firm characteristics for each ANDA.

We argue that a challenge's success depends on both the number and the robustness of patents listed by the innovator. While a larger number of patents may signal greater protection, it may also reflect the presence of weak or easily challengeable patents. Additionally, we believe branded drug firms defend more aggressively when their patents have a longer time left, as the stakes are higher. Moreover, we also allow the chance of a successful challenge to vary by therapeutic class.

To quantify these factors, we first utilize patent data from the FDA's Orange Book, and then supplement this with information scraped from the FDA drug database, including submission dates of ANDAs, their targeted innovator products, and associated patents. Using

these combined sources, we construct key variables, including the number of listed patents and the remaining patent duration at the time of PIV ANDA submission.

Having constructed these variables, we next estimate the probability of a successful PIV challenge using the following Logit specification.

$$P_W = P(Y = 1 \mid X) = \frac{e^{X\beta}}{1 + e^{X\beta}},$$

where $Y = 1$ indicates a successful challenge. The vector X includes the number of listed patents, the remaining patent length, and controls for submission year and therapeutic class.

4.4 Stage I: Fixed costs

We derive moment inequalities from the necessary conditions of a simultaneous-move Nash equilibrium to bound fixed costs. Specifically, if we observe that a PIV applicant f applies ANDA fj, then:

$$V_{fj}^{PIV}(N_{B,j}, N_{PIV,j}, N_{R,j}) \geq \theta_{fj}^{PIV}, \tag{8}$$

and if not, then[26]

$$V_{fj}^{PIV}(N_{B,j}, N_{PIV,j} + 1, N_{R,j}) < \theta_{fj}^{PIV}. \tag{9}$$

Similarly, if we observe that a regular applicant f applies ANDA fj, then

$$V_{fj}^{R}(N_{B,j}, N_{PIV,j}, N_{R,j}) \geq \theta_{fj}^{R}, \tag{10}$$

and if not, then

$$V_{fj}^{R}(N_{B,j}, N_{PIV,j}, N_{R,j} + 1) < \theta_{fj}^{R}. \tag{11}$$

In addition, a PIV applicant has no incentive to deviate and file a regular application:

$$V_{fj}^{PIV}(N_{B,j}, N_{PIV,j}, N_{R,j}) - \theta_{fj}^{PIV} \geq V_{fj}^{R}(N_{B,j}, N_{PIV,j} - 1, N_{R,j} + 1) - \theta_{fj}^{R}, \tag{12}$$

and conversely, a regular applicant has no incentive to deviate and file a PIV application:

$$V_{fj}^{R}(N_{B,j}, N_{PIV,j}, N_{R,j}) - \theta_{fj}^{R} \geq V_{fj}^{PIV}(N_{B,j}, N_{PIV,j} + 1, N_{R,j} - 1) - \theta_{fj}^{PIV}. \tag{13}$$

[26] When calculating the counterfactual value functions $V_{fj}^{PIV}(N_{B,j}, N_{PIV,j}+1, N_{R,j})$, we take $N_{PIV,j}+1$ as $N_{F,j} + 1$ since most PIV applicants in our sample are first filers. For $V_{fj}^{R}(N_{B,j}, N_{PIV,j}, N_{R,j} + 1)$, we take $N_{R,j} + 1$ as $N_{df} + 1$.

Inequalities (8)–(13) form the basis for estimation. However, they are not directly implementable due to the presence of unobserved fixed cost shocks η_j.[27]

To illustrate the selection problem, assume for simplicity that expectation errors are absent and that $\theta_0 = \theta_2 = \theta_3 = 0$, so that $\theta_{fj} = \theta_1 st_{fj} + \eta_j$. Inequalities (8)–(11) then imply:

$$\begin{aligned} \frac{V_{fj}^{PIV}(N_{B,j}, N_{PIV,j}+1, N_{R,j}) - \eta_j}{st_{fj}} < \theta_1 \leq \frac{V_{fj'}^{PIV}(N_{B,j}, N_{PIV,j}, N_{R,j}) - \eta_{j'}}{st_{fj'}} \\ \frac{V_{fj}^{R}(N_{B,j}, N_{PIV,j}, N_{R,j}+1) - \eta_j}{st_{fj}} < \theta_1 \leq \frac{V_{fj'}^{R}(N_{B,j}, N_{PIV,j}, N_{R,j}) - \eta_{j'}}{st_{fj'}} \end{aligned} \tag{14}$$

where j represents an ingredient-dosage form that has few generic entries, while j' has many generic entries. Selection bias arises when the unobserved components η_j are incorrectly assumed to be zero across all ingredient-dosage forms $\eta_j = \eta_{j'} = 0$. In general, when an ingredient-dosage form j' has many generic entries (conditional on observables st_{fj}, ir_{fj} and PIV_{fj}), this suggests a relatively favorable cost draw, i.e., $\eta_{j'}$ is likely small or even negative (from the left side of the distribution). Replacing this with zero would over-sample positive η_j and result in a downward bias in the upper bound. Conversely, for ingredient-dosage forms that have few entries, η_j is more likely to be large. Setting this to zero over-samples negative η_j, leading to an upward bias in the lower bound.

To address this, we construct feasible moment inequalities based on (8)–(13), following Starc and Wollmann (2025), and exploiting the assumptions that $\mathbb{E}[\eta_j] = 0$ and that the distribution of η is symmetric. The idea is that although the conditional expectation of η varies with observed entry in different ingredient-dosage forms, its unconditional expectation is nonetheless mean zero.

The crucial step is determining which individual inequalities should be averaged. One remark is that averaging within the same applicant type (PIV or regular) is generally unhelpful, because the average disturbance conditional on entry remains unsolved. Instead, we use entry and non-entry inequalities derived from observed and counterfactual value functions. To fix the idea, let us write down the fundamental inequalities behind (8)–(13) based on the actual and counterfactual value functions. To simplify the notation, we follow the setup in the illustrative example and further assume $\theta_1 = 0$. We conduct the analysis within each market j and for each generic type $k \in \{PIV, REG\}$. For a generic entrant

[27] While the expectational error ν_{fj} is also unobserved, it does not pose estimation difficulties. We omit it in this section and elaborate in Appendix C.

f_E that enters j and forms the actual mature market structure $(N_{B,j}, N_{G,j})$, we have

$$\underbrace{V^k_{f_E,j}(N_{B,j}, N_{G,j})}_{\text{actual entry}} \geq \eta_j, \tag{15a}$$

$$\underbrace{V^k_{f_E,j}(N_{B,j}, N_{G,j}+1)}_{\text{counterfactual entry}} < \eta_j, \tag{15b}$$

where the inequality (15a) rationalizes that the observed mature market structure, in which $N_{G,j}$ includes the firm f_E under study, is sufficiently profitable to justify f_E's entry. In contrast, inequality (15b) reflects a counterfactual scenario where we assume the market cannot support the entry of one additional generic firm of the same type as f_E, i.e., the transition $N_{G,j} \to N_{G,j}+1$ would not be profitable given the market-specific entry cost η_j.

By contrast, for a generic non-entrant f_{NE} such that the actual mature market structure $(N_{B,j}, N_{G,j})$ does not contain that firm f_{NE}, we have

$$\underbrace{V^k_{f_{NE},j}(N_{B,j}, N_{G,j}+1)}_{\substack{\text{counterfactual for} \\ \text{actual non-entry}}} < \eta_j, \tag{16a}$$

$$\underbrace{V^k_{f_{NE},j}(N_{B,j}, N_{G,j})}_{\text{counterfactual entry}} \geq \eta_j, \tag{16b}$$

where the inequality (16a) rationalizes why the firm f_{NE} under study does not enter: the actual mature market structure $(N_{B,j}, N_{G,j})$ is not sufficiently profitable to support its entry (i.e., the counterfactual with one more generic, $N_{G,j} \to N_{G,j}+1$, is not viable). Inequality (16b) considers the counterfactual in which the mature market had one fewer generic, $(N_{B,j}, N_{G,j}-1)$. In this case, it is assumed that entry by f_{NE} would have been profitable i.e., the transition $N_{G,j}-1 \to N_{G,j}$ would occur, making f_{NE} part of the observed market structure $(N_{B,j}, N_{G,j})$.[28] [29]

To average the disturbances η_j, we must combine inequalities of the same sign. Otherwise, averaging would involve subtracting disturbances across observations. There are two

[28] If we modeled firm-market-specific shocks η_{fj}, the right-hand side of the counterfactual inequalities (15b) and (16a) would contain η_{fj} for a hypothetical firm f. To ensure that the averaging inequalities (involving these hypothetical firms) still yield a mean of zero, we would need additional assumptions on the distribution of these disturbances (e.g., setting $\eta_{fj} \geq \eta_{f_E,j}$ in (15b) and $\eta_{fj} \leq \eta_{f_{NE},j}$ in (16a)).

[29] Our focus is on the fundamental counterfactuals. For example, we do not consider cases where the entrant f_E would still find it profitable to enter even if the market contained only $N_{G,j}-2$ generics (i.e., forming $N_{G,j}-1$ upon entry), as this is a natural implication of inequality (15a).

possible approaches: one combines inequalities (15b) and (16a), and the other combines (15a) and (16b). We adopt the first approach. The reason is that not all markets j have experienced at least one generic entry, making it impossible to construct inequality (16b) for markets where only branded drugs are observed. In such cases, there is no meaningful counterfactual in which one fewer generic exists, as the current number of generics is already zero.

An important implication of inequality (14) is that non-entering firms help identify the lower bound of the fixed cost parameter, while entrants inform the upper bound (under the plausible (and testable) assumption that all covariates are non-negative and all parameters are positive). Consequently, combining inequalities (15b) and (16a), which describe actual and counterfactual scenarios where the mature market cannot support an additional entrant, provides a natural basis for identifying the lower bound.

When constructing inequalities (15b) and (16a) from the type-specific (PIV or regular) conditions (8)–(11), we distinguish three types of markets. First, for markets j that have experienced both PIV and regular generic entries, we treat each entrant as a firm f_E and compute the counterfactual value $V^k_{f_E,j}(N_{B,j}, N_{G,j}+1)$, which enters inequality (15b). Second, for markets with only one type of generic entry (either PIV or regular), we again compute counterfactual values $V^k_{f_E,j}(N_{B,j}, N_{G,j}+1)$ for observed entrants. In addition, we consider a hypothetical firm of the *missing* type and compute the counterfactual value $V^k_{f_{NE},j}(N_{B,j}, N_{G,j}+1)$ to reflect why that type did not enter, yielding inequality (16a). Third, for markets with no generic entry (i.e., only branded drugs), we construct counterfactual values $V^k_{f_{NE},j}(N_{B,j}, N_{G,j}+1)$ for both hypothetical PIV and regular applicants, which again enter through (16a). In summary, for each firm $d(f)$, whether observed to enter or not, we compute the counterfactual value function $V^k_{d(f)j}(N_{B,j}, N_{PIV,j}+\mathbf{1}\{k=PIV\}, N_{R,j}+\mathbf{1}\{k=R\})$. Non-entering firms contribute to moment (17) via inequality (16a), while entering firms do so through inequality (15b), which we interpret as showing that the market could not profitably support an additional entrant of the same type under current conditions.

These actual and counterfactual value functions allow us to construct three sets of moment conditions. The first two sets rely on (8)–(11), while the third uses the deviation inequalities (12)–(13). Since we as econometricians do not observe the fixed cost shock η_j, we use $\mathbb{E}[\theta_{fj}|\mathcal{I}_{fj}]$. Below, we present the key ideas and moment conditions, with full derivations provided in Appendix C.

Moment for Lower Bound. As dicussed earlier, we collect inequalities (15b) and (16a) to estimate the lower bound. Because not all firms (PIV or regular) enter every ingredient-dosage form, for each j we can find at least one firm f that did not enter. Thus, for each j, we can always construct one valid instance of inequalities (15b) and (16a) (using (9) for a PIV entrant and (11) for a regular entrant). Averaging these across j pools exactly one η_j per j, yielding a set of unselected cost shocks:

$$\frac{1}{J}\sum_{j}\frac{1}{\mu_j}\sum_{d\in D_j}\frac{1}{2}\sum_{k\in\{PIV,R\}}\Big[V^{k}_{d(f)j}\big(N_{B,j},N_{PIV,j}+\mathbf{1}\{k=PIV\},N_{R,j}+\mathbf{1}\{k=R\}\big) - \theta_0-\theta_1 st_{d(f)j}-\theta_2 ir_{d(f)j}-\theta_3 PIV_{d(f)j}\Big]\times w_{d(f)j}<0, \tag{17}$$

where $w_{d(f)j}$ are non-negative weight functions including a constant, indicators for standard (and non-standard) dosage forms, indicators for whether the drug is available in one, two, or more than two strengths, and indicators for whether the drug involves a PIV applicant. J is the number of unique ingredient-dosage form combinations, and μ_j is the number of drugs associated with each j.

Moment for Upper Bound. Constructing upper bounds is more complicated because not every ingredient-dosage form j has experienced at least one entrant (PIV or regular). This prevents us from collecting inequalities (15a) and (16b) and applying (8) and (10) to every j, making it impossible to pool a set of unselected shocks with mean zero. To address this, we adopt a different approach that exploits the symmetry of the distribution of fixed cost shocks η_j.

The key idea is to compare the average of moments for ingredient-dosage forms with at least one entry against those with no entry. Under the assumption that the distribution of η_j is symmetric around zero, which implies that the difference in the two sample moments

cancels out the unobserved shocks. The resulting moment condition is:

$$\frac{1}{J_E}\sum_{j\in E}\frac{1}{\mu_j}\sum_{d\in D_j}\sum_{k\in\{PIV,R\}}\Bigg[\frac{\mathbf{1}\{N_k\geq 1\}\times V^k_{d(f)j}(N_{B,j},N_{PIV,j},N_{R,j})}{\mathbf{1}\{N_{PIV,j}\geq 1\}+\mathbf{1}\{N_{R,j}\geq 1\}+\mathbf{1}\{N_{PIV,j}=0,N_{R,j}=0\}} - \theta_0-\theta_1 st_{d(f)j}-\theta_2 ir_{d(f)j}-\theta_3 PIV_{d(f)j}\Bigg]\times w_{d(f)j}$$

$$-\frac{1}{J-J_E}\sum_{j\in NE}\frac{1}{\mu_j}\sum_{d\in D_j}\sum_{k\in\{PIV,R\}}\Bigg[\frac{1}{2}V^k_{d(f)j}(N_{B,j},N_{PIV,j}+\mathbf{1}\{k=PIV\},N_{R,j}+\mathbf{1}\{k=PIV\}) - \theta_0-\theta_1 st_{d(f)j}-\theta_2 ir_{d(f)j}-\theta_3 PIV_{d(f)j}\Bigg]\times w_{d(f)j}>0, \tag{18}$$

where E is the set of ingredient-dosage forms with at least one entrant, and $J_E = |E|$. For the non-entry group NE, we calculate the average of the PIV and regular applicants' counterfactual value functions in each j, weighted by a positive-valued function $w_{d(f)j}$, following the approach in Starc and Wollmann (2025):

$$V_j^+ = \sum_{d\in D_j}\frac{1}{2}\sum_{k\in\{PIV,R\}}V^k_{d(f)j}(N_{B,j},N_{PIV,j}+\mathbf{1}\{k=PIV\},N_{R,j}+\mathbf{1}\{k=PIV\})\times w_{d(f)j}.$$

To define the set NE, we order j by their V_j^+ values and select the $J-J_E$ ingredient-dosage forms with the smallest values.

Moment for entry costs related to PIV applications (parameter θ_3). We use inequalities (12) and (13) to construct moments that help identify our main parameter of interest, θ_3, which captures the incremental fixed cost associated with being a PIV applicant. Since the fixed cost shock η_j is assumed to be independent of PIV status, these inequalities are not subject to the selection problem. The assumption of symmetric distribution of η implies

that the unobserved disturbances cancel out in the moment condition:

$$\frac{1}{J_E}\sum_{j\in E}\frac{1}{\mu_j}\sum_{d\in D_j}\frac{1}{2}\sum_{k\in\{PIV,R\}}\Bigg[V^k_{d(f)j}(N_{B,j},N_{PIV,j},N_{R,j}) - V^{-k}_{d(f)j}\big(N_{B,j},N_{PIV,j}+(-1)^{\mathbf{1}\{k=PIV\}},N_{R,j}+(-1)^{\mathbf{1}\{k=R\}}\big) + (-1)^{\mathbf{1}\{k=PIV\}}\cdot\theta_3\Bigg]\times w_{d(f)j}\geq 0, \tag{19}$$

where V^{-k} denotes the counterfactual value if the firm had chosen the alternative application type, and $w_{d(f)j}$ is a non-negative weight function as defined previously. In practice, we divide this moment condition (19) into two moments by separately grouping observations of PIV and regular applicants, which respectively identify the upper and lower bounds of θ_3.

Inference. To construct confidence regions, we follow the moment selection and test inversion approach of Andrews and Soares (2010). Specifically, we evaluate a four-dimensional grid of candidate parameter vectors $(\theta_0,\theta_1,\theta_2,\theta_3)'$ and invert the test statistic to obtain a 95% confidence region for the identified set Θ.

Table 2: Fixed Cost Estimates: 95% Confidence Interval

Parameter	Estimates
Constant (θ_0)	[1e-6, 1.996]
Number of strengths (θ_1)	[0.879, 1.653]
$\mathbf{1}\{$irregular delivery$\}$ (θ_2)	[1e-6, 3.684]
$\mathbf{1}\{$PIV challenge$\}$ (θ_3)	[1.333, 2.896]
Observations	817 groups (3014 obs.)
Moments	18
Minimum fixed cost of entry	2.264
Average fixed cost of entry	4.086
Maximum fixed cost of entry	16.188

Notes: Fixed cost estimates are reported in millions of dollars. The bounds reported are 95% confidence intervals. To calculate the minimum, mean, and maximum fixed costs across ingredient–dosage form combinations, we set θ_0, θ_1, θ_2, and θ_3 equal to their respective midpoints.

Table 2 presents the 95% confidence sets for the fixed costs. The coefficient on the number of strengths, θ_1, ranges from \$0.9 million to \$1.7 million, suggesting that expanding a

product's strength portfolio meaningfully raises development costs. The coefficient on the irregular delivery indicator, θ_2, reaches up to \$3.7 million, suggesting substantial additional costs for products with irregular delivery. The coefficient on the PIV challenge indicator, θ_3, lies between \$1.3 million and \$2.9 million, reflecting that firms incur higher costs to pursue a PIV challenge compared with regular generics for a given ingredient–dosage form. These extra costs include expenses for bypassing patents, administrative procedures, and often most importantly, litigation.

The magnitudes align well with prior evidence. According to Parasrampuria et al. (2021), the average fixed cost for regular generics is approximately \$2.5 million, with litigation costs around \$1–\$3 million, consistent with our estimates. In our sample, total fixed costs range from \$2.3 million to \$16.2 million, with a mean of \$4.1 million. For comparison, Starc and Wollmann (2025) estimate fixed costs between \$1.4 million and \$13.9 million, and Gottlieb (2016) finds that a generic application typically costs between \$5 million and \$15 million. Overall, our estimates fall well within the range documented by the literature.

5 Counterfactuals

In this section, we explore the role of exclusivity by back-of-the-envelope analysis and simulate the effects of varying exclusivity durations. Moreover, we assess alternative policies to encourage PIV challenges.

5.1 Benefit from patent challenge

We first examine the costs and benefits of the exclusivity period. The first applicant to file an ANDA can benefit from earlier market entry and a 180-day marketing exclusivity period. This exclusivity provides an extended selling window and limited competition, enabling the initial generic entrant to set higher prices and achieve greater profit margins. In our sample, the average profit from 180-day exclusivity is estimated at around \$5.7 million, and it could be up to \$93.2 million. Early entry also confers a first-mover advantage that can persist after the exclusivity period. The average fixed cost for PIV generics is estimated at \$6 million, which is equivalent to around 6 years of regular generic profit of the same ingredient-dosage form. In comparison, the average fixed cost for regular (non-PIV) generics is estimated at \$3.7 million, which is equivalent to around 4 years of regular generic profit. Comparing

these values and fixed costs indicates that patent challenges are highly attractive, although their appeal depends on the probability of a successful challenge, which may significantly reduce the expected value.

5.2 Alternative policies

In this section, we simulate challenge rates under alternative policies, focusing on two key dimensions that capture the incentives for filing patent challenges: exclusivity length and fixed costs. To examine the role of exclusivity, we simulate a scenario without PIV exclusivity (reflecting the current European regulatory framework) and three alternative regimes that provide exclusivity periods of 1 year, 1.5 years, and 2 years. These are compared to the current benchmark of 0.5-year (180-day) exclusivity. To study the effect of fixed costs, we simulate arbitrary reductions in fixed costs of 2%, 4%, 6%, 8%, and 10%.

Figure 3 shows the relative expected values of PIV generics under different exclusivity regimes. Taking the current 0.5-year exclusivity as the benchmark, we compute expected values in two scenarios: (i) hypothetically fixing the market structure at the 0.5-year regime (blue bars) and hence shutting down drugs' entry and exit, and (ii) allowing drugs' entry and exit (green bars). In the absence of exclusivity, PIV generics lose profits from the most lucrative period, and the value of the challenge declines: the exclusivity rent effect is 10.5% less than the 0.5-year exclusivity benchmark. With longer exclusivity, the profitable window extends, and the expected values of challenge increase substantially when entry is restricted. Once entry is allowed, however, the picture changes: without exclusivity, fewer generics challenge, and the market is less competitive: the business stealing effect increases the value of PIV generics by 24.5 percentage points. In equilibrium, only the most profitable drugs find it worthwhile to challenge, which raises the average value of PIV generics by 14% compared with the 0.5-year exclusivity benchmark. For longer exclusivities, higher profits attract more PIV entrants, intensifying competition, therefore, the business-stealing effects are negative. The resulting business-stealing effect offsets the exclusivity rent effect, resulting in lower expected values (red bars) compared to the fixed-market case (blue bars). At shorter exclusivity lengths, the exclusivity rent effect dominates, whereas at longer durations, competition increasingly erodes profits. In terms of consumer welfare, Figure 4 shows that a 0.5-year exclusivity benefits consumers by 35.3 percentage points compared to the no-exclusivity regime. Longer exclusivity initially increases consumer welfare because it attracts more entrants and intensifies competition.

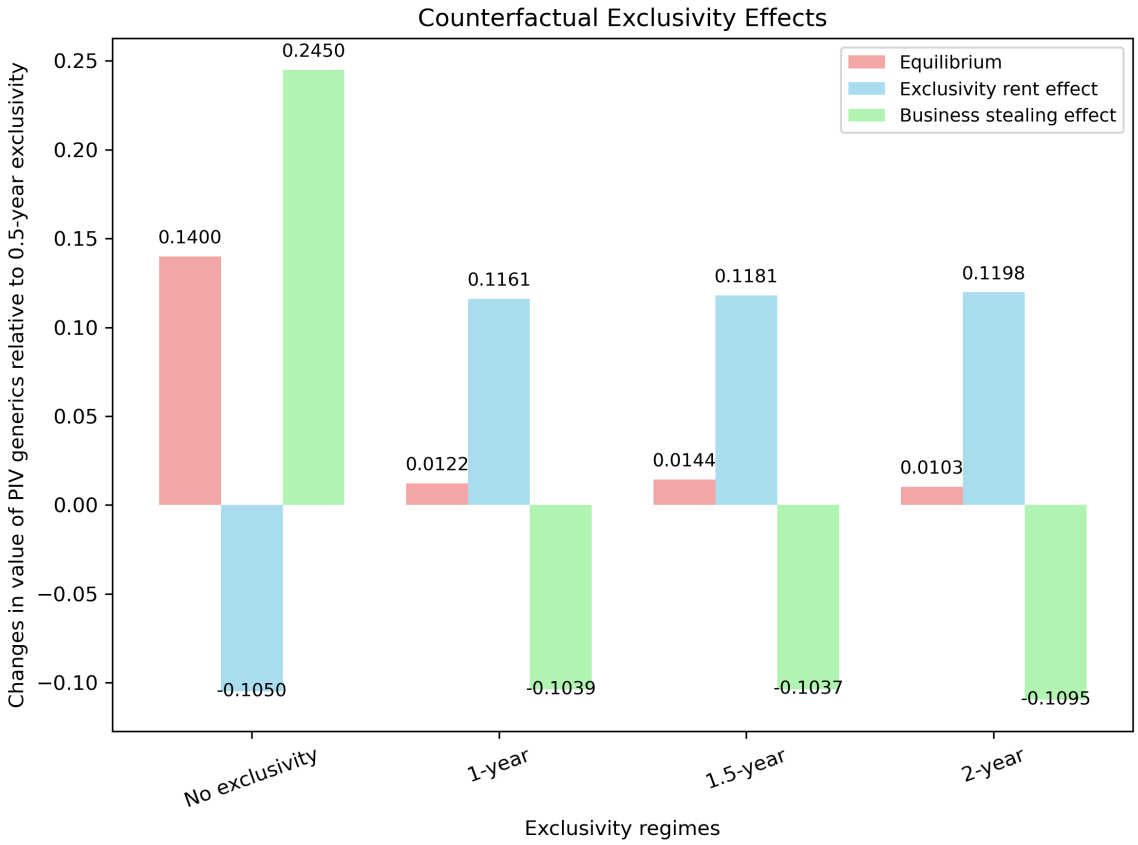


Figure 3: Expected value under different exclusivity regimes

Note: This figure shows the equilibrium value and its decomposition into exclusivity rent and business stealing under various exclusivity regimes relative to 0.5-year exclusivity. The equilibrium, exclusivity rent effect, and business stealing effect are average values of PIV generics.

However, with 2-year exclusivity, the consumer surplus decreases: the business-stealing effect stabilizes, and relatively high prices during the extended exclusivity period reduce consumer surplus.

Figure 5 presents the corresponding challenge rates under different exclusivity regimes. The observed challenge rate in the data is 16.01%, and our model closely predicts the challenge rate at 14.37%. Firms gain significantly from the 0.5-year exclusivity period: the 0.5-year exclusivity raises the challenge rate by approximately 4 percentage points compared to no exclusivity, whereas longer exclusivity has only a modest effect. However, for ingredient–dosage forms that were previously unchallenged, longer exclusivity strongly stimulates entry: a 2-year exclusivity yields a challenge rate of 15.38%. This suggests that firms do not have sufficient incentives to challenge in unchallenged markets under the

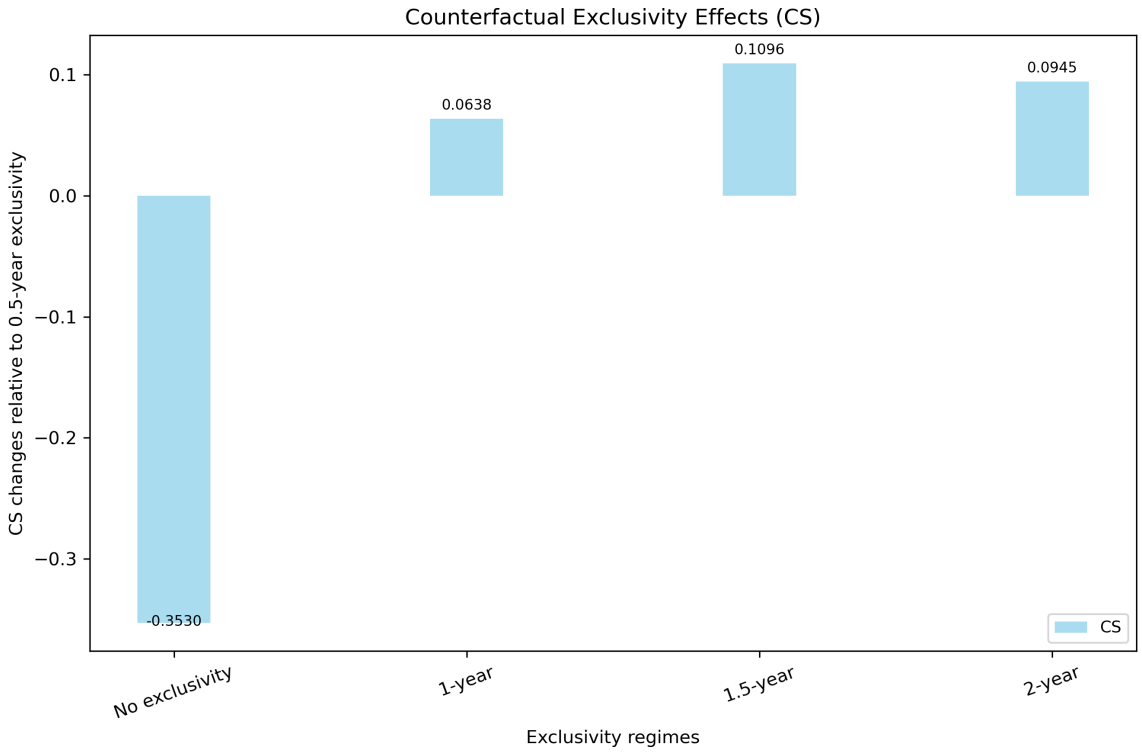


Figure 4: Expected value under different exclusivity regimes

Note: This figure shows the consumer surplus under various exclusivity regimes relative to 0.5-year exclusivity.

current 0.5-year exclusivity window. Extending exclusivity provides sufficient financial incentives to make challenging profitable, thereby encouraging earlier entry.

Figure 6 presents results for varying fixed-cost regimes, with reductions of 2%, 4%, 6%, 8%, and 10%. Lower fixed costs are associated with increased PIV challenge rates, but the effects are limited in magnitude. For example, a 10% reduction in fixed costs results in a mean PIV challenge rate of 15.06%, which remains below the rate simulated under 1.5-year exclusivity. The impact of fixed-cost changes is more pronounced in previously unchallenged markets. These groups, with no prior PIV entry, experience PIV entry at a rate of 10.58% following a 10% reduction in fixed costs. In comparison, varying the exclusivity period is both more effective and more feasible than altering fixed costs.

We further examine the exclusivity length required across therapeutic classes to achieve certain challenge rates. Figure 7 reports challenge rates in previously unchallenged markets by therapeutic class. Specifically, we categorize these products by class and calculate the average challenge rate under various exclusivity regimes. A 1-year exclusivity period is

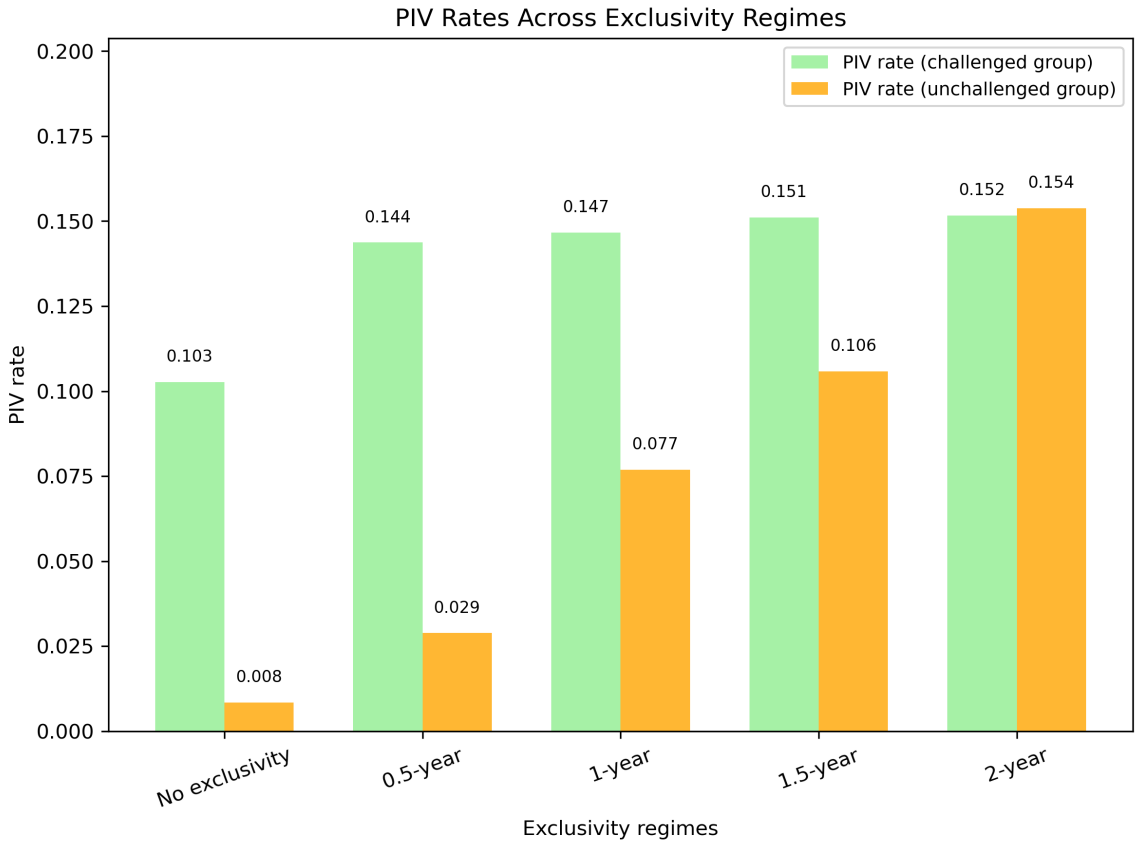


Figure 5: Challenge rate under different exclusivity regimes

sufficient to induce challenges for blood products and cardiovascular drugs. Genitourinary and hormone drugs require 1.5 years of exclusivity, while gastrointestinal and dermatological drugs respond only after 2 years. For antimicrobials, a 2-year exclusivity increases the challenge rate to 20%. The findings indicate that it may be preferable to grant varying exclusivity lengths depending on the therapeutic class.

6 Conclusion

Patent accumulation delays generic entry and sustains high drug prices. To mitigate this effect, the 180-day exclusivity provision was introduced, incentivizing generic firms to challenge weak secondary patents and enter the market earlier. Our data indicate that Paragraph IV challenges lead to generic entry approximately seven years earlier than non-challenge entry, though challenge rates remain low in general and vary across therapeutic classes.

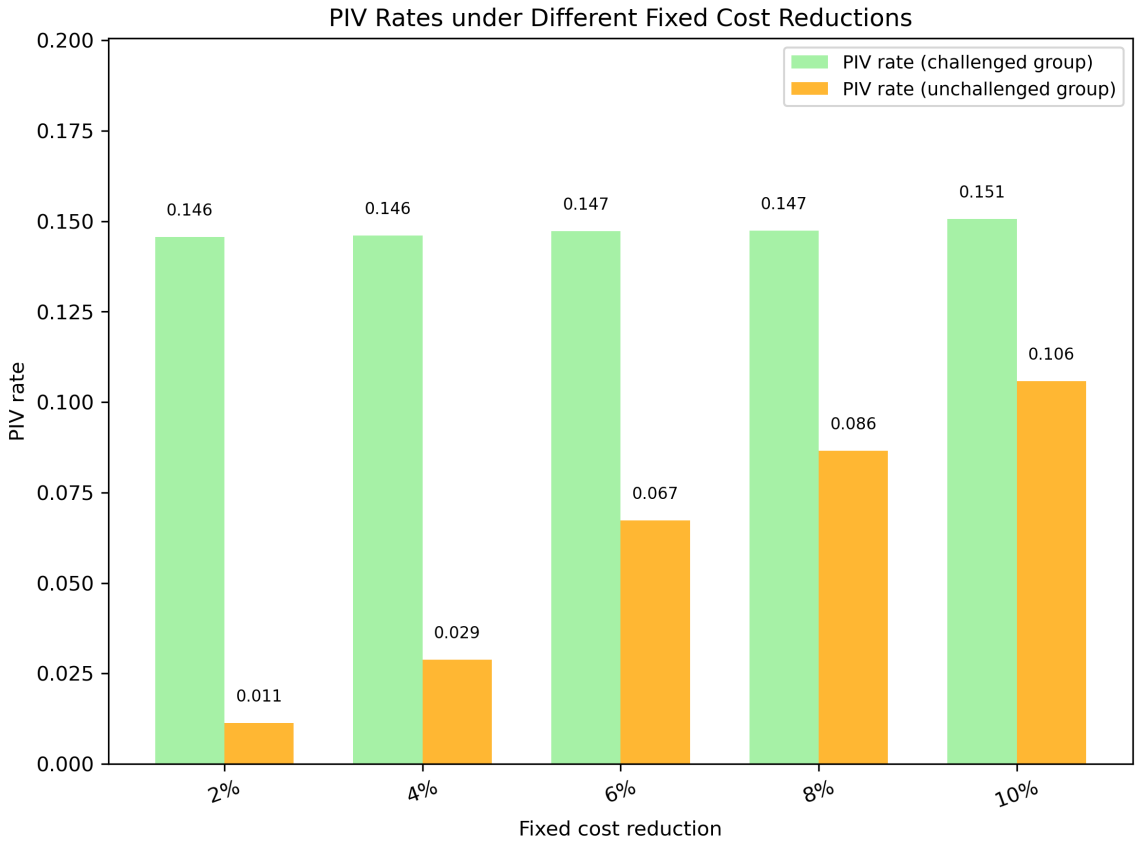


Figure 6: Challenge rate with different fixed costs

To explore alternative policies to encourage PIV entry, we build a structural model in which generic firms choose whether to enter a market, the type of entry (PIV or regular), and prices. Using moment inequalities, we estimate that the fixed costs of filing an ANDA range from $2.264 million to $16.188 million, with PIV entry costing on average $2.115 million more than regular (non-PIV) entry. We then simulate alternative policies under varying exclusivity regimes and fixed-cost structures. Our results show that 180-day exclusivity substantially increases patent challenges. Extending exclusivity beyond 180 days further encourages PIV entry in previously unchallenged markets. Reducing fixed costs has a similar effect, though extending exclusivity proves more effective and practical. Moreover, we find that a single rule does not fit all: heterogeneous exclusivity lengths should be applied across therapeutic classes.

This study suggests several promising avenues for future research. First, the timing of patent challenges merits further investigation. Firms face a trade-off. Filing early increases

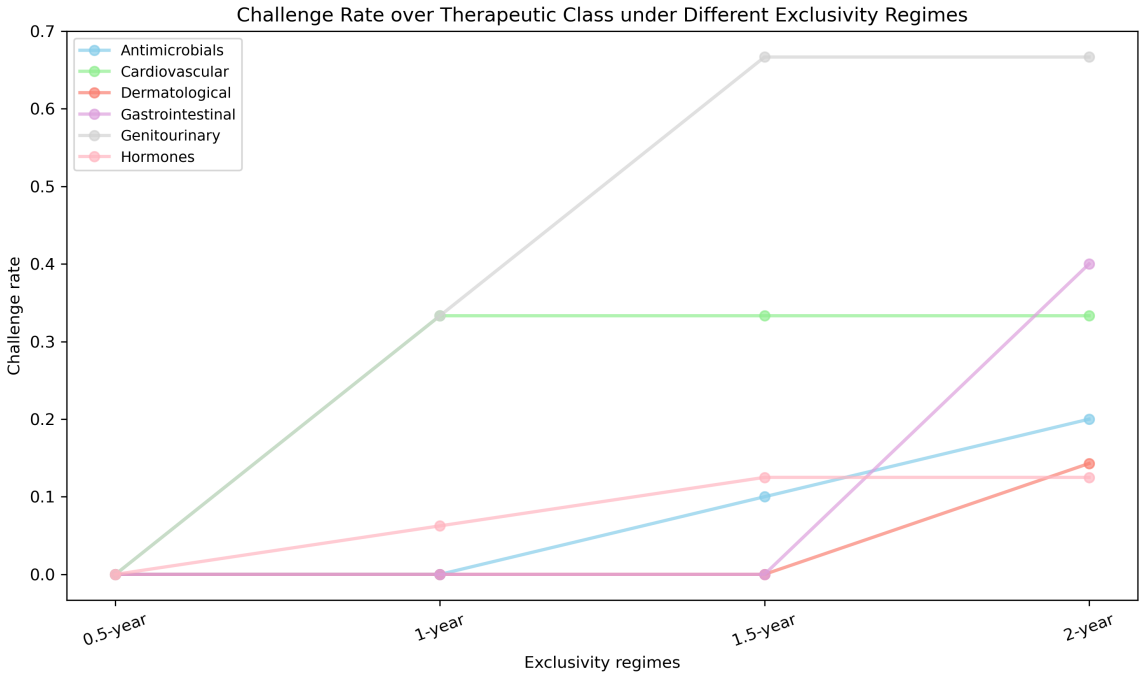


Figure 7: Challenge rate across therapeutic class under different exclusivity regimes

the chance of being the first challenger and securing exclusivity. However, delaying until closer to patent expiration may deter other generic entrants and yield a less competitive post-entry market. Additionally, the challenge costs may be lower due to a less aggressive innovator. Second, this analysis does not consider the influence of authorized generics or pay-for-delay settlements, both of which can substantially alter entry dynamics and competitive outcomes.

References

Alam, R. and Conti, R. (2024). Entry and pricing with fighting brands: Evidence from generics and authorized generics. Technical report, Working paper.

Andrews, D. W. K. and Soares, G. (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica*, 78(1):119–157.

Atal, J. P., Cuesta, J. I., and Sæthre, M. (2022). Quality regulation and competition: Evidence from pharmaceutical markets. Technical report, National Bureau of Economic Research.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 25(2):242–262.

Branstetter, L., Chatterjee, C., and Higgins, M. J. (2016). Regulation and welfare: evidence from paragraph iv generic entry in the pharmaceutical industry. *The RAND Journal of Economics*, 47(4):857–890.

Bresnahan, T. F. and Reiss, P. C. (1991). Entry and competition in concentrated markets. *Journal of political economy*, 99(5):977–1009.

Drake, K. M. and McGuire, T. G. (2025). Using stock price movements to estimate the harm from collusive drug patent litigation settlements. *Journal of Health Economics*, page 103054.

Dubois, P., Gandhi, A., and Vasserman, S. (2022a). Bargaining and international reference pricing in the pharmaceutical industry. *NBER Working Paper No. w30053*.

Dubois, P., Moisson, P.-H., and Tirole, J. (2022b). The Economics of Transferable Patent Extensions. TSE Working Papers 22-1377, Toulouse School of Economics (TSE).

European Competition Commission, D. (2009). Pharmaceutical sector inquiry. *Final Report*, 28.

Feldman, R. (2018). May your drug price be evergreen. *Journal of Law and the Biosciences*, 5(3):590–647.

Gottlieb, S. (2016). How obama's fda keeps generic drugs off the market. *Missouri Medicine*, 113(6):444.

Grabowski, H., Brain, C., Taub, A., and Guha, R. (2017). Pharmaceutical patent challenges: Company strategies and litigation outcomes. *American Journal of Health Economics*, 3(1):33–59.

Grabowski, H. G. and Kyle, M. (2007). Generic competition and market exclusivity periods in pharmaceuticals. *Managerial and Decision Economics*, 28(4/5):491–502.

Gupta, C. (2023). One product, many patents: Imperfect intellectual property rights in the pharmaceutical industry. *Many Patents: Imperfect Intellectual Property Rights in the Pharmaceutical Industry (November 11, 2023)*.

Hemphill, C. and Sampat, B. (2012). Evergreening, patent challenges, and effective market life in pharmaceuticals. *Journal of Health Economics*, 31(2):327–339.

Hemphill, C. S. and Sampat, B. N. (2011). When do generics challenge drug patents? *Journal of Empirical Legal Studies*. Columbia Law and Economics Working Paper No. 379, 5th Annual Conference on Empirical Legal Studies Paper.

Håkonsen, H., Horn, A. M., and Toverud, E.-L. (2009). Price control as a strategy for pharmaceutical cost containment—what has been achieved in norway in the period 1994–2004? *Health Policy*, 90(2):277–285.

Mazzeo, M. J. (2002). Product choice and oligopoly market structure. *RAND Journal of Economics*, pages 221–242.

Olson, L. M. and Wendling, B. W. (2018). Estimating the causal effect of entry on generic drug prices using hatchâ€“waxman exclusivity. *Review of Industrial Organization*, 53(1):139–172.

Panattoni, L. E. (2011). The effect of Paragraph IV decisions and generic entry before patent expiration on brand pharmaceutical firms. *Journal of Health Economics*, 30(1):126–145.

Parasrampuria, S., SERTKAYA, A., LORD, A., and BERGER, C. (2021). Cost of generic drug development and approval final.

Ridley, D. B., Grabowski, H. G., and Moe, J. L. (2006). Developing drugs for developing countries. *Health Affairs*, 25(2):313–324.

Schlander, M., Hernandez-Villafuerte, K., Cheng, C. Y., Mestre-Ferrandiz, J., and Baumann, M. (2021). How much does it cost to research and develop a new drug? a systematic review and assessment. *PharmacoEconomics*, 39(11):1243–1269.

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

Starc, A. and Wollmann, T. G. (2025). Does entry remedy collusion? evidence from the generic prescription drug cartel. *American Economic Review*, 115(5):1400–1438.

Wollmann, T. G. (2018). Trucks without bailouts: Equilibrium product characteristics for commercial vehicles. *American Economic Review*, 108(6):1364–1406.

A Tables

A.1 A table on notations

Table 3: Notation definitions

Notation	Definition
FDA	Food and Drug Administration
ANDA	Abbreviated New Drug Application for generic drugs
NDA	New Drug Application for branded drugs
PIV	Paragraph IV, which indicates patent challenge
NDC	National Drug Code, the identifier for a drug
VA class	Veterans Affairs class, a type of therapeutic class
m	Ingredient: a market in price competition
g	Nesting group in the demand model
d	A product, defined at the ingredient-dosage form-strength-firm level
j	Ingredient-dosage form: entry decision made in this level
fj	An ANDA of firm f in ingredient-dosage form j
ξ	Demand shock
ω	Marginal cost shock
η	Sunk cost shock
D	duration of delay: time difference between application and launch
$F_{D,Reg}$	Distribution of delay for regular generics
$F_{D,PIV}$	Distribution of delay for PIV generics
$N_{B,j}$	Max number of NDAs for an ingredient-dosage form j
$N_{F,j}$	Max number of first PIV applicants for an ingredient-dosage form j
$N_{S,j}$	Max number of subsequent PIV applicants for an ingredient-dosage form j
$N_{R,j}$	Max number of regular ANDAs for an ingredient-dosage form j
$N_{R,-j}$	Max number of regular ANDAs for the same ingredient as j but other dosage forms
n_F	Number of first PIV applicants
n_S	Number of subsequent PIV applicants
$n_{R,j}$	Number of regular ANDAS for an ingredient-dosage form j
$n_{R,-j}$	Number of regular ANDAs for the same ingredient as j but other dosage forms
V_{fj}^R	Value function of a regular ANDA of firm f in ingredient-dosage form j
V_{fj}^{PIV}	Value function of a PIV ANDA of firm f in ingredient-dosage form j
V_{dt}^F	Value function of a PIV generic drug d being first applicant at period t
V_{dt}^S	Value function of a PIV generic drug d being subsequent applicant at period t
V_{fj}^{Fail}	Value function of a PIV ANDA fj failing challenge
δ	Discount factor
θ_{fj}	Fixed costs of an ANDA fj
P_W	Probability of successful PIV challenge, i.e., winning the lawsuit
P_F	Probability of being the first PIV applicant
$\rho(a,b,t)$	Probability that b out of a players are in the market at period t

A.2 Challenge successful probability

Table 4: Probability of success (Logit)

	Estimate	Std. Error
#Patent	0.0332**	0.0154
Remaining length	-0.1422***	0.0507
Year FE	Y	Y
VA class FE	Y	Y

* p<0.1, ** p<0.05, *** p<0.01

A.3 Counterfactuals

Panel A: Values and PIV rates under different exclusivity regimes					
	No exclusivity	0.5-year exclusivity	1-year exclusivity	1.5-year exclusivity	2-year exclusivity
Value rate_no entry	0.8950	1.0000	1.0014	1.0029	1.0041
Value rate_with entry	1.1293	1.0000	1.0355	1.0069	1.0085
PIV rate_observed	/	0.1601	/	/	/
PIV rate_simulated	0.1027	0.1437	0.1466	0.1511	0.1517
PIV entry	0.0084	0.0288	0.0769	0.1058	0.1538
Panel B: Values and PIV rates under different Fixed cost regimes					
	2% off	4% off	6% off	8% off	10% off
Mean FC_PIV	5.7694	5.6516	5.5339	5.4161	5.2984
Mean FC_Reg	3.6971	3.6217	3.5462	3.4708	3.3953
PIV rate_simulated	0.1456	0.1460	0.1473	0.1474	0.1506
PIV entry	0.0113	0.0288	0.0673	0.0865	0.1058

Table 5: Summary statistic for different exclusivity regimes and fixed cost levels

Panel A notes: Table reports average values across exclusivity regimes. Value rates are relative to the 0.5-year exclusivity benchmark.
Panel B notes: Values are sample means. FC_PIV and FC_Reg are simulated Fixed costs, value refers to the simulated value with 0.5-year exclusivity, and PIV rate is the proportion of PIV cases to the generic applications.
We simulate PIV entry in groups that do not have generics in the data under different regimes and report the proportion of groups with PIV entry in the 'PIV entry' row.

Therapeutic class	0.5-year exclusivity	1-year exclusivity	1.5-year exclusivity	2-year exclusivity
Antimicrobials	0	0	0.1	0.2
Blood Products	0	0.5	0.5	0.5
Cardiovascular	0	0.3333	0.3333	0.3333
Dermatological	0	0	0	0.1429
Gastrointestinal	0	0	0	0.4
Genitourinary	0	0.3333	0.6667	0.6667
Hormones	0	0.0625	0.125	0.125

Table 6: Summary statistics for exclusivity regimes by therapeutic class

Note: The rates are challenge rates in previously unchallenged groups under different exclusivity regimes.

B Probabilities of entry scenarios at period t

We have 33 periods (16.5 years) in the data, and all markets become mature after 33 periods. Accordingly, we assume that the probability of maturity equals one after 33 periods. In terms of the other three entry scenarios, during exclusivity, entry by subsequent PIVs, and entry by regular generics, we compute the corresponding dynamic entry probabilities as follows.

Table 7: Probability distribution over time

Delay	$P(t$: during exclusivity$)$	$P(t$: subsequent PIVs entering$)$	$P(t$: regular generics entering$)$
0	$1-(1-F_D(0))^{\chi_{PIV}}$	0	0
1	$(1-F_D(0))^{\chi_{PIV}} \times (1-(1-F_D(1))^{\chi_{PIV}})$	$(1-(1-F_D(0))^{\chi_{PIV}}) \times P$(infringe)	$(1-(1-F_D(0))^{\chi_{PIV}}) \times P$(invalid)
τ	$\Pi_{t=0}^{\tau-1}(1-F_D(t))^{\chi_{PIV}} \times (1-(1-F_D(\tau))^{\chi_{PIV}})$	$\sum_{t=1}^{\tau-1} P$(t:during exclusivity$) \times P$(infringe)	$\sum_{t=1}^{\tau-1} P$(t:during exclusivity$) \times P$(invalid)

Notes: P(invalid) is an estimate from the data and P(infringe) $= 1 - P$(invalid).

C Derivation of moment inequalities

Throughout the paper, we assume that agents form rational expectations based on their information sets, so their subjective expectations $\mathcal{E}$ coincide with the empirical expectations $\mathbb{E}$. We further assume that our empirical estimates $\hat{V}$, $\hat{\pi}$, $\hat{F}_\xi$, $\hat{F}_\omega$, and $\hat{F}_D$ are measured without error relative to the beliefs used by agents at the decision-making stage, i.e., V, π, F_ξ, F_ω, and F_D.

Proposition 1. *Moment inequalities* (17) *produce consistent lower bounds for the parameters of interest* $\Theta = (\theta_0, \theta_1, \theta_2, \theta_3)'$.

Proof. Consider the individual inequalities (9) and (11). For $k \in \{PIV, R\}$:

$$V^k_{d(f)j}(N_{B,j}, N_{PIV,j} + \mathbf{1}\{k = PIV\}, N_{R,j} + \mathbf{1}\{k = R\}) - \mathbb{E}[\theta_{d(f)j} \mid \mathcal{I}_{d(f)j}] < 0,$$

where

$$\mathbb{E}[\theta_{d(f)j}|\mathcal{I}_{d(f)j}] = \theta_{d(f)j} - \nu_{d(f)j} = \theta_0 + \theta_1 st_{d(f)j} + \theta_2 ir_{d(f)j} + \theta_3 PIV_{d(f)j} + \eta_j - \nu_{d(f)j},$$

and $\nu_{d(f)j} \equiv \theta_{d(f)j} - \mathbb{E}[\theta_{d(f)j} \mid \mathcal{I}_{d(f)j}]$ denotes the firm's expectation error, which has mean zero $\mathbb{E}[\nu_{d(f)j} \mid \mathcal{I}_{d(f)j}] = 0$ by construction.

Substituting into the inequality, we get:

$$\begin{aligned} &V^k_{d(f)j}(N_{B,j}, N_{PIV,j} + \mathbf{1}\{k = PIV\}, N_{R,j} + \mathbf{1}\{k = R\}) \\ &- \theta_0 - \theta_1 st_{d(f)j} - \theta_2 ir_{d(f)j} - \theta_3 PIV_{d(f)j} - \eta_j + \nu_{d(f)j} < 0. \end{aligned}$$

Taking averages over all $d(f)j$ pairs, isolating the fixed cost disturbances and expectations errors on one side and applying the law of large numbers, we have :

$$\begin{aligned} &\frac{1}{J}\sum_j \frac{1}{\mu_j} \sum_{d \in D_j} \frac{1}{2} \sum_{k \in \{PIV,R\}} \Big[V^k_{d(f)j}(N_{B,j}, N_{PIV,j} + \mathbf{1}\{k = PIV\}, N_{R,j} + \mathbf{1}\{k = R\}) \\ &\qquad - \theta_0 - \theta_1 st_{d(f)j} - \theta_2 ir_{d(f)j} - \theta_3 PIV_{d(f)j}\Big] \\ &< \frac{1}{J}\sum_j \eta_j - \frac{1}{J}\sum_j \frac{1}{\mu_j} \sum_{d \in D_j} \frac{1}{2} \sum_{k \in \{PIV,R\}} \nu_{d(f)j} \xrightarrow{P} \mathbb{E}[\eta] - \mathbb{E}[\nu] = 0, \end{aligned} \tag{20}$$

where the last step follows from the law of total expectation: $\mathbb{E}[\eta] = \mathbb{E}[\mathbb{E}[\eta \mid \mathcal{I}]] = 0$ and $\mathbb{E}[\nu] = \mathbb{E}[\mathbb{E}[\nu \mid \mathcal{I}]] = 0$.

Since the value functions V and the fixed cost covariates st, ir, and PIV are all known to the firm at the decision-making stage, they are included in the firm's information set $\mathcal{I}_{d(f)j}$. Therefore, the population moment inequality (20) can be written as:

$$\begin{aligned} &\mathbb{E}\Big[V^k_{d(f)j}(N_{B,j}, N_{PIV,j} + \mathbf{1}\{k = PIV\}, N_{R,j} + \mathbf{1}\{k = R\}) \\ &\quad - \theta_0 - \theta_1 st_{d(f)j} - \theta_2 ir_{d(f)j} - \theta_3 PIV_{d(f)j} \,\Big|\, \mathcal{I}_{d(f)j}\Big] < 0, \end{aligned}$$

which is then translated into unconditional moments with weight $w_{d(f)j}$, which are non-negative functions of information variables such as st, ir and PIV that introduce addi-

tional variation to identify the parameters of interest Θ. We obtain unconditional moment inequalities that consistently estimate lower bounds on Θ:

$$\frac{1}{J}\sum_j \frac{1}{\mu_j}\sum_{d\in D_j}\frac{1}{2}\sum_{k\in\{PIV,R\}}\Big[V^k_{d(f)j}(N_{B,j}, N_{PIV,j}+\mathbf{1}\{k=PIV\}, N_{R,j}+\mathbf{1}\{k=PIV\}) \\ -\theta_0-\theta_1 st_{d(f)j}-\theta_2 ir_{d(f)j}-\theta_3 PIV_{d(f)j}\Big]\times w_{d(f)j}<0.$$

□

Proposition 2. *Moment inequalities* (18) *produce consistent upper bounds for the parameters of interest* $\Theta=(\theta_0,\theta_1,\theta_2,\theta_3)'$.

Proof. Consider the inequalities implied by (8)–(11). For $k\in\{PIV,R\}$, we have:

$$V^k_{d(f)j}(N_{B,j}, N_{PIV,j}+\mathbf{1}\{k=PIV\}, N_{R,j}+\mathbf{1}\{k=R\})-\mathbb{E}\Big[\theta_{d(f)j}\Big|\mathcal{I}_{d(f)j}, j \text{ has no entry}\Big]<0, \\ V^k_{d(f)j}(N_{B,j}, N_{PIV,j}, N_{R,j})-\mathbb{E}\Big[\theta_{d(f)j}\Big|\mathcal{I}_{d(f)j}, j \text{ has entry}\Big]\geq 0.$$

where

$$\mathbb{E}[\theta_{d(f)j}|\mathcal{I}_{d(f)j}]=\theta_{d(f)j}-\nu_{d(f)j}=\theta_0+\theta_1 st_{d(f)j}+\theta_2 ir_{d(f)j}+\theta_3 PIV_{d(f)j}+\eta_j-\nu_{d(f)j},$$

and $\nu_{d(f)j}\equiv\theta_{d(f)j}-\mathbb{E}[\theta_{d(f)j}\mid\mathcal{I}_{d(f)j}]$ denotes the firm's expectation error, which has mean zero $\mathbb{E}[\nu_{d(f)j}\mid\mathcal{I}_{d(f)j}]=0$ by construction.

Substituting into the inequality, we get:

$$V^k_{d(f)j}(N_{B,j}, N_{PIV,j}+\mathbf{1}\{k=PIV\}, N_{R,j}+\mathbf{1}\{k=R\}) \\ -\theta_0-\theta_1 st_{d(f)j}-\theta_2 ir_{d(f)j}-\theta_3 PIV_{d(f)j}-\eta_j+\nu_{d(f)j}<0$$

for ingredient-dosage forms without any entry, and

$$V^k_{d(f)j}(N_{B,j}, N_{PIV,j}, N_{R,j}) \\ -\theta_0-\theta_1 st_{d(f)j}-\theta_2 ir_{d(f)j}-\theta_3 PIV_{d(f)j}-\eta_j+\nu_{d(f)j}\geq 0$$

for ingredient-dosage forms having experienced entries.

Taking sample averages over ingredient-dosage forms with and without entry and sub-

tracting the latter from the former yields:

$$\frac{1}{J_E}\sum_{j\in E}\frac{1}{\mu_j}\sum_{d\in D_j}\sum_{k\in\{PIV,R\}}\Bigg[\frac{\mathbf{1}\{N_k\geq 1\}\times V^k_{d(f)j}(N_{B,j},N_{PIV,j},N_{R,j})}{\mathbf{1}\{N_{PIV,j}\geq 1\}+\mathbf{1}\{N_{R,j}\geq 1\}+\mathbf{1}\{N_{PIV,j}=0,N_{R,j}=0\}} - \theta_0-\theta_1 st_{d(f)j}-\theta_2 ir_{d(f)j}-\theta_3 PIV_{d(f)j}-\eta_j+\nu_{d(f)j}\Bigg]$$
$$-\frac{1}{J-J_E}\sum_{j\in NE}\frac{1}{\mu_j}\sum_{d\in D_j}\sum_{k\in\{PIV,R\}}\Bigg[\frac{V^k_{d(f)j}(N_{B,j},N_{PIV,j}+\mathbf{1}\{k=PIV\},N_{R,j}+\mathbf{1}\{k=PIV\})}{2} - \theta_0-\theta_1 st_{d(f)j}-\theta_2 ir_{d(f)j}-\theta_3 PIV_{d(f)j}-\eta_j+\nu_{d(f)j}\Bigg]>0.$$

Isolating the fixed cost disturbances and expectations errors on one side, we have

$$\frac{1}{J_E}\sum_{j\in E}\frac{1}{\mu_j}\sum_{d\in D_j}\sum_{k\in\{PIV,R\}}\Bigg[\frac{\mathbf{1}\{N_k\geq 1\}\times V^k_{d(f)j}(N_{B,j},N_{PIV,j},N_{R,j})}{\mathbf{1}\{N_{PIV,j}\geq 1\}+\mathbf{1}\{N_{R,j}\geq 1\}+\mathbf{1}\{N_{PIV,j}=0,N_{R,j}=0\}} - \theta_0-\theta_1 st_{d(f)j}-\theta_2 ir_{d(f)j}-\theta_3 PIV_{d(f)j}\Bigg]$$
$$-\frac{1}{J-J_E}\sum_{j\in NE}\frac{1}{\mu_j}\sum_{d\in D_j}\sum_{k\in\{PIV,R\}}\Bigg[\frac{V^k_{d(f)j}(N_{B,j},N_{PIV,j}+\mathbf{1}\{k=PIV\},N_{R,j}+\mathbf{1}\{k=PIV\})}{2} - \theta_0-\theta_1 st_{d(f)j}-\theta_2 ir_{d(f)j}-\theta_3 PIV_{d(f)j}\Bigg]$$
$$>\frac{1}{J_E}\sum_{j\in E}\frac{1}{\mu_j}\sum_{d\in D_j}\sum_{k\in\{PIV,R\}}\Big(\eta_j-\nu_{d(f)j}\Big)-\frac{1}{J-J_E}\sum_{j\in NE}\frac{1}{\mu_j}\sum_{d\in D_j}\sum_{k\in\{PIV,R\}}\Big(\eta_j-\nu_{d(f)j}\Big).$$

Now suppose the entry share converges, i.e., $J_E/J \xrightarrow{P} q\in(0,1)$. Applying the law of large numbers, we have

$$\frac{1}{J_E}\sum_{j\in E}\frac{1}{\mu_j}\sum_{d\in D_j}\sum_{k\in\{PIV,R\}}\Big(\eta_j-\nu_{d(f)j}\Big)-\frac{1}{J-J_E}\sum_{j\in NE}\frac{1}{\mu_j}\sum_{d\in D_j}\sum_{k\in\{PIV,R\}}\Big(\eta_j-\nu_{d(f)j}\Big)$$
$$\xrightarrow{P}\quad \mathbb{E}[\eta|\eta<F^{-1}(q)]-\mathbb{E}[\eta|\eta>F^{-1}(1-q)]-\Big(\mathbb{E}[\nu]-\mathbb{E}[\nu]\Big)=0,$$

where $F^{-1}(q)$ is the asymptotic entry threshold, $\mathbb{E}[\eta|\eta<F^{-1}(q)]$ and $\mathbb{E}[\eta|\eta>F^{-1}(1-q)]$ are equidistant from zero following the symmetry of the distribution F of η, which leads

to $\mathbb{E}[\eta|\eta < F^{-1}(q)] - \mathbb{E}[\eta|\eta > F^{-1}(1-q)] = 0$. Besides, ν is independent of η and hence $\mathbb{E}[\nu|\eta < F^{-1}(q)] = \mathbb{E}[\nu] = \mathbb{E}[\nu|\eta > F^{-1}(1-q)]$. We apply the law of total expectations to obtain the unconditional mean $\mathbb{E}[\nu] = \mathbb{E}\big[\mathbb{E}[\nu|\mathcal{I}]\big] = 0$. Finally, we obtain

$$\begin{aligned}
&\frac{1}{J_E}\sum_{j\in E}\frac{1}{\mu_j}\sum_{d\in D_j}\sum_{k\in\{PIV,R\}}\Bigg[\frac{\mathbf{1}\{N_k \geq 1\}\times V^k_{d(f)j}(N_{B,j}, N_{PIV,j}, N_{R,j})}{\mathbf{1}\{N_{PIV,j}\geq 1\}+\mathbf{1}\{N_{R,j}\geq 1\}+\mathbf{1}\{N_{PIV,j}=0, N_{R,j}=0\}} \\
&\qquad\qquad - \theta_0 - \theta_1 st_{d(f)j} - \theta_2 ir_{d(f)j} - \theta_3 PIV_{d(f)j}\Bigg] \\
-&\frac{1}{J-J_E}\sum_{j\in NE}\frac{1}{\mu_j}\sum_{d\in D_j}\sum_{k\in\{PIV,R\}}\Bigg[\frac{V^k_{d(f)j}(N_{B,j}, N_{PIV,j}+\mathbf{1}\{k=PIV\}, N_{R,j}+\mathbf{1}\{k=PIV\})}{2} \\
&\qquad\qquad - \theta_0 - \theta_1 st_{d(f)j} - \theta_2 ir_{d(f)j} - \theta_3 PIV_{d(f)j}\Bigg] > 0.
\end{aligned} \tag{21}$$

Since the value functions V and the fixed cost covariates st, ir, and PIV are all known to the firm at the decision-making stage, they are included in the firm's information set $\mathcal{I}_{d(f)j}$. Therefore, the population moment inequality (21) can be written as:

$$\begin{aligned}
&\mathbb{E}\Big[V^k_{d(f)j}(N_{B,j}, N_{PIV,j}+\mathbf{1}\{k=PIV\}, N_{R,j}+\mathbf{1}\{k=R\}) \\
&\quad - \theta_0 - \theta_1 st_{d(f)j} - \theta_2 ir_{d(f)j} - \theta_3 PIV_{d(f)j}\Big|\mathcal{I}_{d(f)j}, j \text{ has entry}\Big] \\
-&\mathbb{E}\Big[V^k_{d(f)j}(N_{B,j}, N_{PIV,j}+\mathbf{1}\{k=PIV\}, N_{R,j}+\mathbf{1}\{k=R\}) \\
&\quad - \theta_0 - \theta_1 st_{d(f)j} - \theta_2 ir_{d(f)j} - \theta_3 PIV_{d(f)j}\Big|\mathcal{I}_{d(f)j}, j \text{ has no entry}\Big] > 0,
\end{aligned}$$

which is then translated into unconditional moments with weight $w_{d(f)j}$, which are non-negative functions of information variables such as st, ir and PIV that introduce additional variation to identify the parameters of interest Θ. We obtain unconditional moment

inequalities that consistently estimate upper bounds on Θ:

$$\frac{1}{J_E}\sum_{j\in E}\frac{1}{\mu_j}\sum_{d\in D_j}\sum_{k\in\{PIV,R\}}\left[\frac{\mathbf{1}\{N_k\geq 1\}\times V^k_{d(f)j}(N_{B,j},N_{PIV,j},N_{R,j})}{\mathbf{1}\{N_{PIV,j}\geq 1\}+\mathbf{1}\{N_{R,j}\geq 1\}+\mathbf{1}\{N_{PIV,j}=0,N_{R,j}=0\}}\right.$$
$$\left.-\theta_0-\theta_1 st_{d(f)j}-\theta_2 ir_{d(f)j}-\theta_3 PIV_{d(f)j}\right]\times w_{d(f)j}$$
$$-\frac{1}{J-J_E}\sum_{j\in NE}\frac{1}{\mu_j}\sum_{d\in D_j}\sum_{k\in\{PIV,R\}}\left[\frac{V^k_{d(f)j}(N_{B,j},N_{PIV,j}+\mathbf{1}\{k=PIV\},N_{R,j}+\mathbf{1}\{k=PIV\})}{2}\right.$$
$$\left.-\theta_0-\theta_1 st_{d(f)j}-\theta_2 ir_{d(f)j}-\theta_3 PIV_{d(f)j}\right]\times w_{d(f)j}>0.$$

□

Proposition 3. *Moment inequalities* (19) *produce consistent bounds for the parameter θ_3 associated with the fixed costs generated by PIV challenges.*

Proof. Consider the individual inequalities (12) and (13). For firms entering as PIV applicants:

$$V^{PIV}_{d(f)j}(N_{B,j},N_{PIV,j},N_{R,j})-V^{R}_{d(f)j}(N_{B,j},N_{PIV,j}-1,N_{R,j}+1)-\theta_3\geq 0,$$

and for firms entering as regular applicants:

$$V^{R}_{d(f)j}(N_{B,j},N_{PIV,j},N_{R,j})-V^{PIV}_{d(f)j}(N_{B,j},N_{PIV,j}+1,N_{R,j}-1)+\theta_3\geq 0,$$

where we exploit the independence of the disturbance η and the applicant status PIV and the fact that

$$\mathbb{E}[\theta_{d(f)j}|\mathcal{I}_{d(f)j}]=\theta_{d(f)j}-\nu_{d(f)j}=\theta_0+\theta_1 st_{d(f)j}+\theta_2 ir_{d(f)j}+\theta_3 PIV_{d(f)j}+\eta_j-\nu_{d(f)j},$$

and $\nu_{d(f)j}\equiv\theta_{d(f)j}-\mathbb{E}[\theta_{d(f)j}\mid\mathcal{I}_{d(f)j}]$ denotes the firm's expectation error, which has mean zero $\mathbb{E}[\nu_{d(f)j}\mid\mathcal{I}_{d(f)j}]=0$ by construction.

Taking the average of all $d(f)j$ individual inequalities (with j having experienced an

entry), we have

$$\frac{1}{J_E}\sum_{j\in E}\frac{1}{\mu_j}\sum_{d\in D_j}\frac{1}{2}\sum_{k\in\{PIV,R\}}\Bigg[V^k_{d(f)j}\Big(N_{B,j},N_{PIV,j},N_{R,j}\Big) - V^{-k}_{d(f)j}(N_{B,j},N_{PIV,j}+(-1)^{\mathbf{1}\{k=PIV\}},N_{R,j}+(-1)^{\mathbf{1}\{k=R\}}\Big)+(-1)^{\mathbf{1}\{k=PIV\}}\theta_3\Bigg]\geq 0. \tag{22}$$

Since the value functions V and the fixed cost covariates st, ir, and PIV are all known to the firm at the decision-making stage, they are included in the firm's information set $\mathcal{I}_{d(f)j}$. Therefore, the population moment inequality (22) can be written as:

$$\mathbb{E}\Bigg[V^k_{d(f)j}\Big(N_{B,j},N_{PIV,j},N_{R,j}\Big)-V^{-k}_{d(f)j}(N_{B,j},N_{PIV,j}+(-1)^{\mathbf{1}\{k=PIV\}},N_{R,j}+(-1)^{\mathbf{1}\{k=R\}}\Big) + (-1)^{\mathbf{1}\{k=PIV\}}\theta_3\Bigg|\mathcal{I}_{d(f)j}\Bigg]\geq 0,$$

which is then translated into unconditional moments with weight $w_{d(f)j}$, which are non-negative functions of information variables such as st, ir and PIV that introduce additional variation and consistently estimate θ_3:

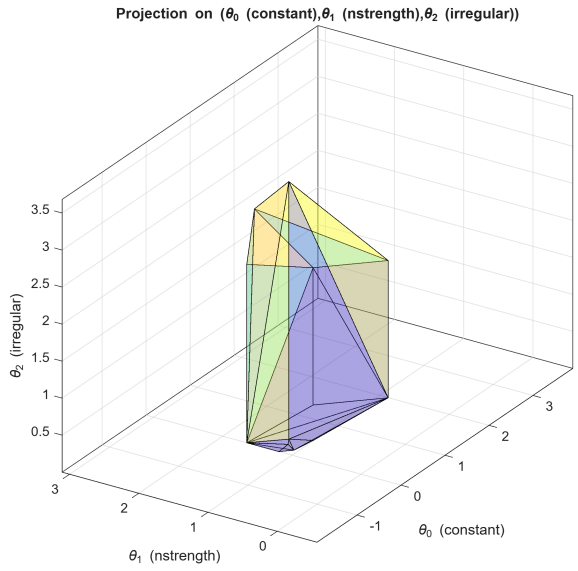
$$\frac{1}{J_E}\sum_{j\in E}\sum_{j}\frac{1}{\mu_j}\sum_{d\in D_j}\frac{1}{2}\sum_{k\in\{PIV,R\}}\Bigg[V^k_{d(f)j}\Big(N_{B,j},N_{PIV,j},N_{R,j}\Big) - V^{-k}_{d(f)j}(N_{B,j},N_{PIV,j}+(-1)^{\mathbf{1}\{k=PIV\}},N_{R,j}+(-1)^{\mathbf{1}\{k=R\}}\Big)+(-1)^{\mathbf{1}\{k=PIV\}}\theta_3\Bigg]\times w_{d(f)j}\geq 0.$$

□

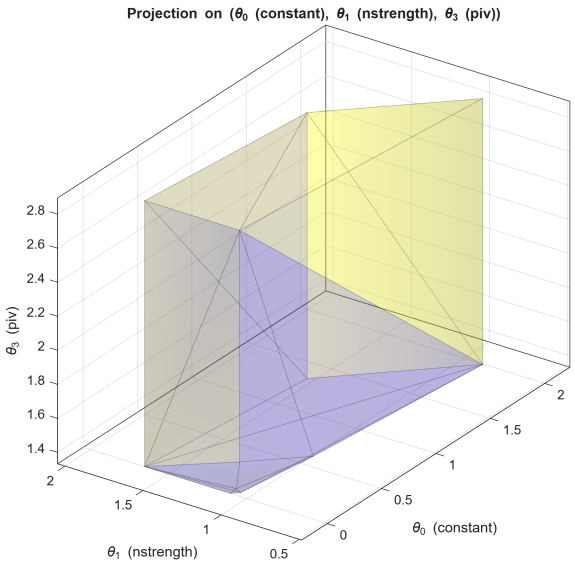
D Confidence intervals of fixed costs

In this section, we illustrate the shapes of the estimated identified set. We have four parameters $(\theta_0,\theta_1,\theta_2,\theta_3)$, and we display three at a time.

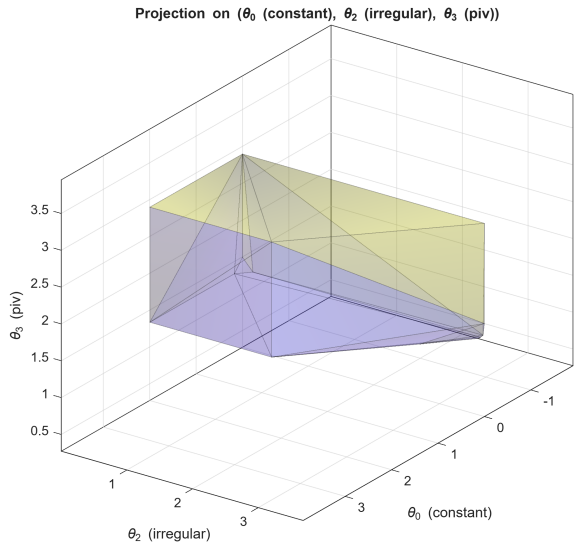
Figure 8: 3D polyhedron projections.



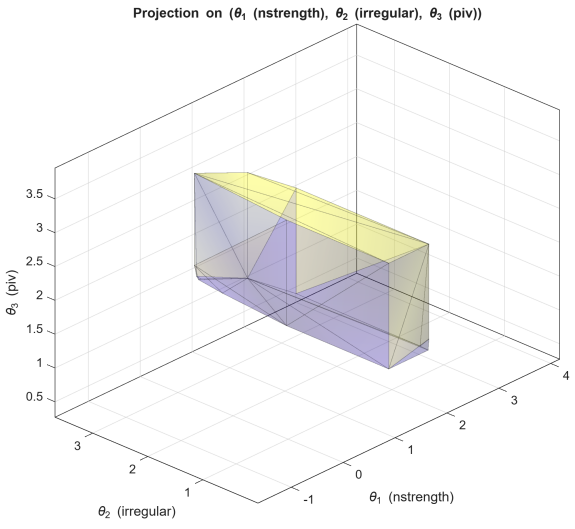
(a) Projection on $(\theta_0, \theta_1, \theta_2)$



(b) Projection on $(\theta_0, \theta_1, \theta_3)$



(c) Projection on $(\theta_0, \theta_2, \theta_3)$



(d) Projection on $(\theta_1, \theta_2, \theta_3)$