

Resilience of Global Supply Chains and Generic Drug Shortages

Anaïs Galdin[†]

Princeton University

JOB MARKET PAPER

November 3, 2023

[Click for Latest Version](#)

Since 2009, shortages have occurred with disproportionately high frequency in offshored generic injectable pharmaceutical markets. This paper investigates the factors behind these market failures, building a unique dataset that links drug products sold in the U.S. to upstream manufacturing facility locations. I offer new empirical evidence that links decreased pharmaceutical supply chain resilience to recent offshoring. Drugs produced in South-based plants exhibit 54% more frequent and 130 days-longer shortage spells. I introduce a structural model of spatial location choices and global drug procurement that endogenously generates shortages and underscores the incentive mechanisms behind supply disruption. The model (i) captures the industry trade-off between supply reliability and cost-cutting in off-patent drug markets and (ii) allows quantifying the amount of excess demand across markets. I use the estimated model to analyze the social-efficiency of enforcing failures-to-supply clauses, allowing for short-term price variations in response of shortages, or spatially redistributing plants. While heavily emphasized in recent policy debates, switching back to a re-shored production system does not achieve the first best equilibrium. Instead, refining procurement contracts to align pricing with quality assurance can substantially mitigate market failures (specifically, the market's limited ability to adequately capture and reward production reliability), enhancing Welfare by up to 37%. Frictions minimally affecting product availability in a "closed" economy might lead offshored markets to unravel if they lack pricing mechanisms to capture changes in product characteristics.

Keywords: Global Value Chains, Offshoring, Pharmaceuticals, Competition

JEL Codes: I11, F14, L11, L23, L44.

[†]Department of Economics, Princeton University, agaldin@princeton.edu

I am grateful to Eduardo Morales, Steve Redding and Janet Currie for their continual guidance and support on this project. I also thank Treb Allen, Felipe Barbieri, Andrew Bernard, Juanma Castro-Vincenzi, Davin Chor, Levi Crews, Jacob Dorn, Teresa Fort, Cecile Gaubert, Matthew Grant, Gene Grossman, Kate Ho, Douglas Irwin, Jakub Kastl, Benny Kleinman, Isabela Manelici, Isabelle Mejean, Joan Monras, Ezra Oberfield, Luigi Pascali, Nina Pavcnik, Meredith Startz, Swapnika Rachapalli, Richard Rogerson, Esteban Rossi-Hansberg, Jesse Silbert, Christopher Snyder, Bob Staiger and Lucas Zavala for helpful discussions and comments. I am indebted to Erin Fox for sharing the drug shortage data, as well as Andreas Schick and Qiyu Liu from the FDA for their helpful discussions about shortages. This research benefited from financial support from the International Economics Section (IES) at Princeton University, whose support is gratefully acknowledged. Part of this research was carried as a Ph.D. Fellow in International Economics at Dartmouth College. All errors are my own.

"If China and India decided to simply stop selling raw material to the U.S. – we would not be able to make about 80% of the drugs that we need"

Erin Fox, Senior Director of Drug Information at the University of Utah.

"The White House is talking about more regulation, generic manufacturers are calling for direct subsidies.(...) We need to create a market for high quality manufacturing, so you can allow generic manufacturers to make certain claims about the reliability and the quality of the manufacturing. And then for generic manufacturers that can make those claims, you pay them for that. You pay them for the fact that they have reliable manufacturing that might be domestic, that might be more modern, so it's going to be more reliable and less prone to shortages."

Scott Gottlieb, former FDA commissioner

1. Introduction

One should not expect shortages to occur in competitive markets, yet they occur with high frequency in generic injectable pharmaceutical markets. Early indicators of the vulnerabilities of generic drugs' global supply chains appeared around 2009 in the U.S., with a pronounced surge in the scarcity of injectable oncology drugs. By 2011-2012, the Food and Drug Administration (the FDA) documented a tenfold rise in such shortages from early 2000s levels. More strikingly, shortages are highly persistent, both in the U.S. and in Europe, such that longer ongoing shortages drive the growth in disruption rates (Figure 1). Shortages of sterile injectable drugs are conditions of excess demand that can occur due to short-run rigidities in both drug manufacturers' production capacities and in end-consumers' demand (hospital patients). Low-price elasticities of demand and supply for medically necessary prescription drugs means that product disruptions, whatever their causes, translate into shortages as higher drug prices cannot clear the market on the short-term.¹

The persistence of shortages led the American Medical Association in 2018 to urge the U.S. federal government to treat the dearth of generic drugs as a national-security issue.² Yet, the root causes of pharmaceutical shortages remain elusive. The wave of supply issues went hand in hand with a strong offshoring movement of manufacturing facilities to the South, and the multiplication of evidence against manufacturing-quality issues in production. While foreign plants only represented 15% of the U.S. generic pharmaceutical market in the early 2000s, 80% of active pharmaceutical ingredients (API) are now produced in South-East Asia. Why are shortage rates persistently high, and do we then see differences in the offshoring structure of injectable drugs affected by shortages? What are the fundamental drivers of differential in disruption probabilities across plants? Is the current equilibrium the first best? If not, could the market eventually be shifted to an equilibrium that endogenously provides higher supply reliability (and at what costs)?

This paper proposes answers to these questions with four contributions. First, I provide a comprehensive mapping of the global supply chains of injectable drugs, solving the fundamental traceability issues of the market. These novel data allow me in turn to offer novel empirical evidence on the causal

¹See Section 2.2 for a detailed definition of shortages. A drug is medically necessary if it "is used to treat or prevent a serious disease, and there is no other available source that is an adequate substitute." (Fox et al., 2009)

²Source: AHSP (2018), "Drug Shortages as a Matter of National Security". At the end of the second quarter of 2023, the number of medications in low supply reached a new 10-year peak, with 309 ongoing drug shortages (ASHSP)

impact of offshoring on shortages. Second, I build a structural partial-equilibrium model of global supply and demand for pharmaceuticals that endogenizes the occurrence of shortages in equilibrium. The model illustrates the market's trade-off between supply reliability and cost-cutting. Third, I estimate this model on my new dataset to quantify excess demand across drug markets and the geographical distribution of supply risk. Fourth, I use the estimated model to evaluate policies aiming to reduce the equilibrium occurrence of shortages, and quantify the implications of these policies for welfare, product prices and consumers' access to drugs.

The first part of the paper examines the connection between the global re-allocation of drug production plants to the South and shortages. To my knowledge, this paper is the first to assess whether increases in shortage frequencies and amplitudes in the North are disproportionately driven by disruptions in manufacturing *facilities* that are foreign. I achieve these goals by constructing a comprehensive panel dataset that maps generic sterile injectable drug products to their labelers and plants, and to their history of shortages, over the period 2002-2019. This dataset solves traceability issues in the globalized U.S. injectable drugs market and constitutes a major empirical contribution of this paper. The lack of a reliable registry of manufacturing plants and products has notably prevented estimating the true market share of foreign-made products and assessing potential links with supply disruptions.³ To link each production facility to drug products, I combine confidential FDA data on the geographical location of foreign and domestic manufacturing facilities with detailed digitized U.S. import customs records. I exploit probabilistic record linkages, combining text mining techniques and geocoding of plants and shippers to match manufacturers' names and addresses across datasets.

Thanks to this rich dataset, I document new reduced-form findings about offshoring and an increased likelihood of shortages. Drugs produced in the South experience a disproportionately higher frequency of disruptions and longer-lasting shortage spells. To strengthen the evidence for a causal interpretation, I use a dynamic difference-in-differences design to estimate the partial equilibrium impact of the reorganization of a drug's manufacturing structure from domestic to offshore production on shortages. The specification relies on a Coarsened Exact Matching (CEM) algorithm to match similar offshored (treatment) and non-offshored (control) sterile injectable drugs and control for selection on observables. Offshored sterile injectable drugs are on average 13% more likely to experience shortages after offshoring compared to similar drugs whose production remained U.S.-based.

To further overcome selection issues, I propose a shift-share instrumental variable that controls for the endogeneity of the offshoring decision. I isolate a plausibly exogenous source of variation in offshoring rates of individual drug products in order to explain the re-allocation of production plants from the U.S. to Asia, which takes the form of a "Predicted Diseases Burden" ([Acemoglu and Linn, 2004](#); [Costinot et al., 2019](#)). The instrument builds on the observation that demographic groups that are more likely affected by specific diseases are also more likely to consume pharmaceutical treatments that target those diseases (the "share"), and thus more likely to become net producers of those drugs as their markets grow (the "shifter") - a "Home-market Effect" ([Krugman, 1980](#)). I find that after a drug

³Hospitals, wholesalers and even the FDA, are not always able to find out which manufacturer makes which drugs or where factories are located, because manufacturing plants' locations and outsourcing relationships are considered trade secrets (Department of Homeland Security 2018). This information is the property of the company that holds the sales license.

is fully offshored to Asia, shortages are 54% more likely and 61% longer (130 more days on average).

In the second part of the paper, I propose a structural model of manufacturer's location and production choices in global procurement markets that endogenizes shortage occurrence. The model generates insights into the incentives of supply disruptions and rationalizes the offshoring correlations observed in the data. It permits (i) to analyze the manufacturers' response to the moral hazard incentives created by the offshoring option, and (ii) to quantify buyers and sellers' behavior in terms of economic parameters, such as the weight they place on future disruption outcomes in their decision-making.

In order to capture what drives the current market choices of buyers and drug manufacturers, and what may motivate manufacturers to let supply periodically fail, the model matches key characteristics of the current market. It features international procurement contracts between U.S. buyers (hospital conglomerates) who seek to procure a homogeneous drug product from a heterogeneous set of suppliers (pharmaceutical manufacturers) whose investments in production facilities are imperfectly observable. The buyer uses competitive bidding to attribute suppliers' market shares through years-long procurement contracts, whose characteristics create some short-term market rigidities (notably, prices and capacities are determined before disruptions are realized). The model features moral hazard in production (regulatory inspections with imperfect observation of production quality), a stochastic production process, as well as buyers' price sensitivity and potential taste for reliability in supply.

Under the baseline of the model, the occurrence of disruptions is determined by exogenous random supply shocks, but the probability of a subsequent stock-out event can be targeted by the firm (through the choice of building buffers or choosing less risky production locations). While choosing to invest in higher levels of production capacities or building plants in more regulated markets decreases the probability that a disruption turns into conditions of excess demand, it comes at a higher cost for suppliers (e.g.: cost of facility upgrades or quality checks), and eventually private buyers. The model thus endogenously generates market equilibria with supply shortages, defined as conditions of excess demand at the product level. The model is general enough that it can be easily brought to the data and expanded to other industries affected by supply disruptions, enabling researchers and policymakers to better quantify the complex trade-offs between offshoring, cost savings, and supply chain resilience.

I estimate the structural model in a multi-steps method of moments procedure to match the shortage and global production patterns documented in the data. The estimated model allows me to go beyond the qualitative evidence on offshoring and quantify the market-wide trade-off between resilience and cost-cutting. In line with the model's prediction, I find that, the shortage history of drug manufacturers does not significantly affect their future market shares. Private buyers display high price sensitivity in choosing suppliers for the drug products they source, but low tastes for past reliability of these suppliers. Higher valuation of a supplier's shortage history while drafting procurement contracts is nevertheless correlated with a lower probability of future shortages, lower shares of demand unmet, and higher equilibrium prices. The average U.S.-based manufacturing plant tends to build less buffer than their foreign, southern counterparts. Indeed, higher fixed and marginal cost levels make surplus more costly for U.S.-based firms on average, favoring just-in-time production (a well-known characteristic of the U.S. generic drugs market). Eventually, I estimate much lower average capacity yields (fraction of supply surviving a disruption) for plants offshored to the South, suggesting that the

average amplitude of shocks in foreign, southern plants is much larger than for the U.S.

Could the market be shifted to an equilibrium that endogenously generates fewer shortages, and at what cost? I use the estimated model to study the welfare implications of three counterfactual policies. The first counterfactual rests on the premise that the offshore production structure creates frictions that may not be alleviated. It then focuses on providing tax breaks to re-shore production to U.S. territories, implementing the suggested 10.5% reshoring subsidies on firms' revenues from sales. On the contrary, the second and third counterfactual experiments posit that the current system could be optimally leveraging regional comparative production advantages—allocating the production of innovative drug products to the North and generic drugs to the South. The second policy considers introducing short-run *ex post* price variation in response to disruptions. The third policy considers changing procurement contracts to allow penalties to be imposed in response to drug shortages.

Each considered counterfactual deals with one market failure. Subsidizing reshoring addresses regulatory deficiencies outside of the U.S. borders. Pricing capacity targets short-run price rigidities in procurement contracts. Enforcing penalties for failure to supply addresses the inability of the market to capture and reward reliability in production.

The counterfactual experiments show that the net welfare gains of reducing supply shortfall levels may be large and are not offset by the associated increase in drug prices. All three scenarios lead to higher Social Welfare, thus corroborating that the current level of shortages may not be achieving the first-best equilibrium. While under all scenarios, reaching higher welfare levels comes at the cost of higher unit prices and an increase in market concentration, this does not result in crowding out effects on the patients' side. On the contrary, enforcing penalties results in a 37% increase in consumption access to drug products and a 30% increase in Welfare, while allowing the market prices to react to excess demand results in a 3% increase in consumption access and a 5% increase in Welfare.

While heavily emphasized in the recent policy debates, switching back to a re-shored system achieves the lowest increase in welfare (around 1-2%) and the highest increase in price (around 30% at the suggested 10% tax break level). Its effectiveness in reducing shortage levels is moreover the most contrasted among all three scenarios: while the average probability of shortages decreases by 7.5%, the expected share of excess demand increases by 10%. This is partly driven by higher production costs in the U.S., which makes building surplus more expensive. By implementing changes in procurement contracts that help in aligning price mechanisms with quality assurance, counterfactual experiments that directly tackle the underlying roots of the market failures (specifically, the market's limited ability to adequately reward production reliability) allow reaching substantially higher welfare levels. The highest increase in welfare (30%) is achieved by implementing a system of penalty for failures-to-supply. This effectively provides higher margins for reliable suppliers, potentially allowing them to invest back in facilities maintenance and high-quality production processes.

Related Literature This paper advances three connected areas within *the International Trade literature*: the literature on global value chains and offshoring dynamics, studies of firms boundaries, foreign outsourcing and product life-cycle, and the emerging research on supply chain resilience.

I first contribute to the theoretical trade literature modeling **Global Value Chains (GVC)** ([Antras](#)

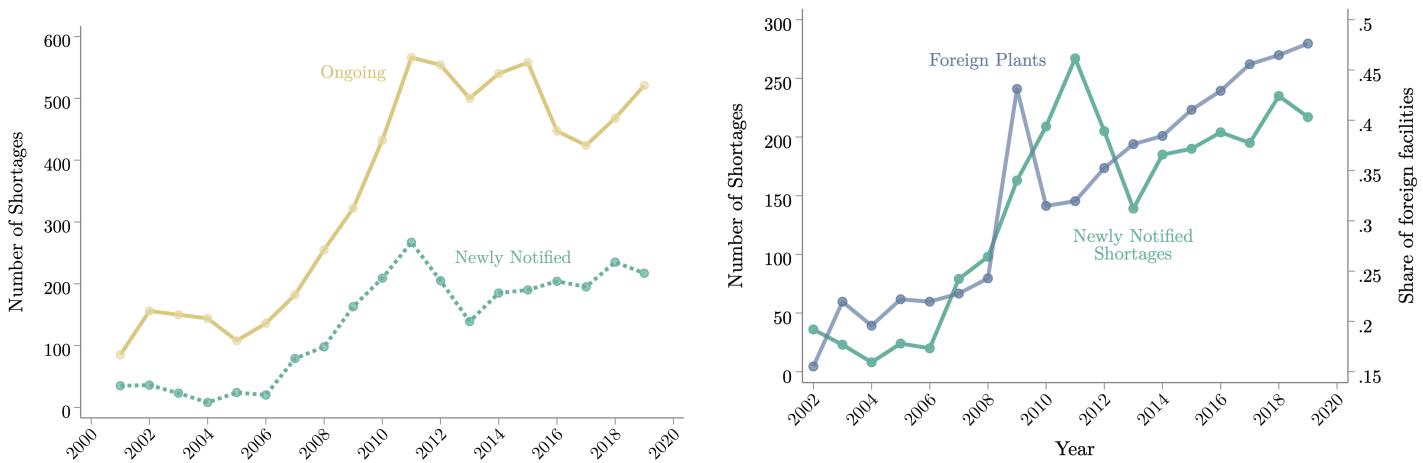


Figure 1: Yearly U.S. Drug Shortages and Offshoring

Notes: The left figure displays the yearly number of U.S. drug shortages as reported to the UUDIS. The green, dashed line reports the number of new drugs being notified as short in a given year. The yellow, solid line reports ongoing shortage events, thus capturing the length of disruptions in time. On the right graph, I superpose the time series of newly notified shortages with a plot of the share of foreign-based manufacturing facilities for generic drugs marketed in the United States. Plants location data are from the FDA's Center for Drug Evaluation and Research (CDER). Yearly records of foreign plants authorized to market generic drug products in the U.S. do not come with records of the drug products manufactured in the plants, so that the witnessed increase in the share of foreign plants over the 2000-2019 period does not necessarily correspond to the production of the drugs going short.

and Chor, 2013; Antras and Chor, 2018; Alfaro et al., 2019; Costinot, Vogel, and Wang, 2012; Costinot, Vogel, and Wang, 2013) and offshoring mechanisms (Grossman and Rossi-Hansberg, 2008; Grossman and Rossi-Hansberg, 2012; Antràs and Staiger, 2012; Baldwin and Venables, 2013; Baldwin and Robert-Nicoud, 2014). While prior theoretical works studying the impact of falling offshoring costs on factor prices have emphasized gains-from-trade in general equilibrium, I underscore that the complexity, length, and lack of transparency of global supply chains may also make it increasingly complicated to capture product characteristics in some commodity-like markets. I develop a tractable structural framework that can be brought to the data to quantify the manufacturers trade-off between investment in supply chain's reliability and cost-cutting. This partial-equilibrium model can be extended to a variety of settings studying firms' production and location choices.

Understanding what happens to products reaching the **end of the innovation life-cycle** is one of the central objects of the international trade literature (Chaudhuri, Goldberg, and Jia, 2006; Bilir, 2014). Mapping connections between patent expiration and the re-organization of production remains nevertheless challenging due to the intricacies of tracing products in long and complex supply chains. My paper adds to this literature by first, providing new empirical evidence that commodity-like products at the end of the life cycles are the most likely to be offshored and outsourced, and second, by solving supply chains transparency issues in the injectable drug market. My work in this regard particularly relates to the recent literature studying foreign outsourcing (Bernard et al., 2020; Antràs et al., 2023; Fort, 2023), which uses new survey data or microdata on multinational firms to fill empirical gaps about “factoryless goods producers”.

My paper shares the closest connection with Bernard et al. (2020), which highlights that trade lib-

eralization provided opportunities for innovative firms to offshore production of low-quality varieties, focusing production efforts on the development of high-quality goods and, hence exploiting comparative advantages. The boom in drugs offshoring was in part made possible by several waves of patents expiration. Often referred to as “patent cliffs”, these waves happened in the United States since the late 80s, thus allowing manufacturers in Southern countries to produce drugs with close to zero R&D costs⁴. The most significant of these patent cliffs coincided with the 2011-2012 peak of shortages, with patents expiring for multiple blockbuster drugs, including Lipitor (Atorvastatin) and Plavix (Clopidogrel)⁵.

I present evidence underscoring that the increased international division of production must go in hand with an adaptation of procurement systems to the new market structure, and stress that supply chain resilience should arise not solely from the supply side but also from the buyers’ side. This is necessary to ensure that pricing mechanisms effectively capture and incentivize quality, thereby preventing markets from unraveling. If U.S. buyers seek to procure goods meeting U.S. manufacturing standards at lower manufacturing costs, they must provide not only regulatory but also financial incentives for South manufacturers to meet these thresholds and differentiate their products to those they sell on their local markets. There cannot be product differentiation without financial incentives that recognize variations in product characteristics.

This paper centrally contributes to the literature on **supply chain resilience** and disruptions (Grossman, Helpman, and Lhuillier, 2023; Elliott, Golub, and Leduc, 2022; Khanna, Morales, and Pandalai-Nayar, 2022; Pichler et al., 2023) by expanding our understanding of the consequences of offshoring in critical industries, highlighting the importance of balancing cost savings with supply chain resilience. While Grossman, Helpman, and Lhuillier (2023) and Elliott, Golub, and Leduc (2022) provide theoretical foundations to understand factors leading to supply chains resilience, I rely on a new and detailed data mapping of products to suppliers to offer tools that allow for analysis of manufacturers’ incentives to opt for low resilience levels. I build on this dataset to provide new empirical evidence on the (lack) of resilience of global supply, to underline characteristics of markets are the most likely to get affected, and to explicit the mechanisms leading to changes in market resilience.

A large strand of the supply chains resilience literature investigates the propagation of supply shocks within firms networks, leveraging large exogenous shocks. Most of this literature leverages natural disasters to study the role of firm-level linkages in propagating disruptions. Boehm, Flaaen, and Pandalai-Nayar (2019), Carvalho et al. (2020) and Barrot and Sauvagnat (2016) study the propagation of negative shocks across regions after the Tohoku earthquake in Japan in 2011, while Castro-Vincenzi (2022) studying the car industry. Studying Indian firms networks, Khanna, Morales, and Pandalai-Nayar (2022) suggest that the most resilient supply chains are those with larger suppliers, more differentiated inputs, and a low number of alternative suppliers. Building on this

⁴Typically, pharmaceutical patents last for 20 years from the date of filing, although there can be several patents associated with a single drug (covering different uses, formulations, etc.). Generic entry is made possible after these patents expire, which is associated with a significant decrease in prices.

⁵Two other big patent cliffs marked the U.S. over the last 40 years: in the late 1980s and early 1990s, many drugs discovered in the 1970s (when modern drug discovery really took off) began to go off-patent. This included for instance the patent expiration in 1992 of captopril, a groundbreaking heart failure and hypertension drug. The late 1990s were marked again by patent expirations of several significant drugs, such as Prozac (Fluoxetine)

insight, fragmentation of supply chains, lack of products differentiation and high degrees of market contestability in generic pharmaceutical markets make them especially vulnerable to experience a lack of resilience.

This paper also contributes to three distinct strands of *the Health Economics literature*: the study of generic pharmaceutical entry, the analysis of factors causing drug shortages, and the Industrial Organization research on parallel trade and quality regulations in the pharmaceutical sector.

This paper directly connects to the reduced-form health literature examining **participation of generics in pharmaceutical markets**. A large body of work analyzes the entry of generics post patent expiration in the U.S. ([Scott Morton, 1999b](#); [Scott Morton, 2000](#)). [Frank and Salkever \(1997\)](#) pioneered the exploration into the relationship between generic entries and pharmaceutical price adjustments. The effect of exogenous firms entry on generic prices is further documented by [Conti, Berndt, and Murphy \(2017\)](#) and [Conti and Berndt \(2019\)](#), who find evidence of a decreasing number of suppliers over the period 2004-2016, attributable both to more exit and less entry. [Frank, Hicks, and Berndt \(2019\)](#) estimate that, between 2007 and 2016, generic prescription drugs sold at retail pharmacies experienced a eighty percent drop in the consumer price index.

The focus on recent offshoring trends in off-patent pharmaceuticals was established by [Conti, Berndt, and Murphy \(2018\)](#) and [Kaygisiz et al. \(2019\)](#), who discuss the geography of pharmaceuticals supplied to the United States and the regulatory shifts implemented under the Generic Drug User Fee Amendments (GDUFA) to process an influx of applications from overseas facilities. However, these studies do not possess information about the specific pharmaceutical products made in each plant.

The economic **literature studying drug shortages** is sparse. The lack of comprehensive data that ties manufacturing companies and sites to products has impeded research efforts to analyze the impact of changes in the geographical structure of production on shortages. On the health policy side, former FDA's commissioner Woodcock was among the firsts to describe how the lack of reward for manufacturing quality can reinforce price competition and lead to a market in which quality problems are severe and common enough to drive shortages ([Woodcock and Wosinska, 2013](#)). There is however a dearth of quantitative studies that examine the drug shortage problem. [Yurukoglu, Liebman, and Ridley \(2017\)](#) argue that the Medicare Modernization Act (MMA) of 2005 sharply lowered reimbursement rates for provider-administered sterile injectables drugs, which had in turn an adverse impact on profit margins for generic sterile injectable drugs.⁶ In a descriptive study, [Stomberg \(2016\)](#) further underlines the responsibility of drug manufacturers' capacity constraints and the FDA regulatory scrutiny.

I solve both limitations encountered in these literatures by: (i) linking manufacturing firms, sites and products and (ii) building a structural model of supply and demands in global pharmaceutical

⁶Although adjustments in provider reimbursements may have contributed to reduce product margins, drug prices in the U.S. are not subjects to direct regulation. Instead, they emerge through negotiations between private buyers (Group Purchasing Organizations) and suppliers. The Medicare reimbursement framework bases its rates on the average negotiated prices observed over the previous two quarters, at the product level. Industry reports moreover underscore that intense price competition was already a defining trait of the generic sterile injectable market well before the 2005 MMA's implementation. Further nuancing the MMA's implications, drug shortages have been observed to escalate in a concurrent manner in Europe, where health care systems are very different from the U.S. and where the MMA did not take place (see Section B.3).

markets to assess the impact of counterfactual policies on shortage outcomes.

Structural work related to pharmaceutical markets stems from the Industrial Organization literature studying **regulatory policies, parallel trade, and drug quality**. Part of this literature focuses on the effects of the expiration of drug patents on imports and trade, as [Chaudhuri, Goldberg, and Jia \(2006\)](#), who find that price controls may be welfare-improving when used to counterbalance the welfare losses from other regulatory measures, or on the effects of parallel imports on patented drugs ([Dubois and Saethre, 2018; Grossman and Lai, 2008](#)).

Another portion of this literature investigates the equilibrium implications of price regulations in developed countries, focusing on price dispersions, price discrimination, and cross-border effects ([Dubois, Gandhi, and Vasserman, 2022; Dubois and Lasio, 2018; Dubois, Gandhi, and Vasserman, 2022](#)). In a working paper, [Ganapati and McKibbin \(2019\)](#) study two elements of price dispersions of pharmaceutical prices across countries: non-tariff trade barriers and buyers' bargaining power. On their part, [Dubois, Gandhi, and Vasserman \(2022\)](#) estimate a structural model of demand and supply for pharmaceuticals to assess the effects of a hypothetical U.S. reference pricing policy that would cap prices in U.S. markets by those offered in Canada. All these papers rely on proprietary data from IQVIA, which link drug sales to labeler firms (patent holders), but do not actually provide information about the actual identity of manufacturing locations. My detailed mapping of supply chains allows analysis of cross-country trade of pharmaceuticals beyond drug labelers.

My paper is most closely related to [Atal, Cuesta, and Sæthre \(2022\)](#), who analyze the equilibrium effects of quality regulation in the generic pharmaceutical market. They underscore the trade-off of increasing quality regulations in Chile, which results in an increase in market concentration and prices.⁷ I emphasize that, although allowing the market to select for production quality comes at the cost of higher prices and market concentration, this does not result in crowding-out effects on the patients side and yields both lower disruptions in supply and higher welfare levels.

Outline The paper proceeds as follows. Section 2 describes institutional features of the market and the data used to map injectable drugs supply chains. Section 3 provides new empirical evidence on the links between South production and shortage spells. Section 4 builds a model of generic drug shortages that endogenizes supply disruptions. Section 5 estimates the model. Section 6 evaluates counterfactual policies. The last section concludes. An [Appendix](#) collects additional results and proofs.

2. Market Structure: Tracing Injectable Drugs' Supply Chains

This section first summarizes the functioning of the U.S. Market for Generic Drugs and provides a formal definition of shortage, before describing the data collected and the process of mapping injectable drugs supply chains.⁸ It then leverages this new dataset to portrait the sterile injectable drug markets and underline distinctive features of the market that make it especially prone to shortages.

⁷Quality regulation increased demand for generic drugs by resolving asymmetric information, leading in equilibrium to higher exit rates of low-quality manufacturers and entry of high-quality drug products.

⁸Figure ?? in Appendix summarizes the main datasets used in this paper

2.1. Institutional Features of the U.S. Generic Drug Markets

Entering Generic Drug Markets. To manufacture a new innovator drug (“branded drug”), one needs a patent. These patents generally have a limited life-span (typically twenty years in the United States)⁹. On the day a patent expires, numerous generic pharmaceutical companies —referred to as generic sponsors, or “labelers”— stand ready to market it and submit an Abbreviated New Drug Application (ANDA) (see Figure A1 for a graphical representation of the market).¹⁰

The institution in charge of the regulatory oversight of this market is the Food and Drug Administration (the FDA). The FDA is responsible both for the delivering of marketing licenses (ANDA for generics) and for ex-post drug quality assurance through plant inspections.

The manufacturing, involving the transformation of active pharmaceutical ingredients (API) to finished drug form products (FDF), can be conducted by labelers’ own plants or outsourced to contract manufacturing organizations (CMO). This process can be either fully or partially offshored. Throughout this paper, “labelers” will denote firms acting as generic sponsors, while “manufacturers” will pertain to the manufacturing facilities.

The U.S. Procurement System. Group Purchasing Organizations (GPOs) serve as pivotal buyers, working as intermediaries between hospital pharmacies and drug manufacturers. GPOs directly engage with drug *labelers* to negotiate bulk product prices, aggregating demands from member hospitals to negotiate volume discounts, thus securing drug prices that individual hospitals might not independently obtain. A substantial 96-98% of hospitals engage in drug procurement via GPOs, with 80% of hospital purchases occurring through these entities. According to the FDA, “GPOs account for over \$100 billion of the drugs purchased in this country in a given year”.¹¹ As of 2018, the four largest Group Purchasing Organizations (GPOs) accounted for 90 percent of the medical supply market. Subsequent to purchases, payers, such as insurers and Medicare, reimburse hospitals using fixed payments, based on the weighted average of sale prices in the previous two quarters). Refer to Section A.2.1 for a detailed overview of the U.S. procurement process.

Institutionalization of Perfect Substitutability. The “Drug Price Competition and Patent Term Restoration Act of 1984,” also known as the Hatch-Waxman Amendments, established bioequivalence as the basis for approving generic copies of drug products.¹² Marketing a generic product thus only requires labelers to demonstrate bioequivalence to the brand.

By making bioequivalence the basis for approving generic copies of drug products, the Hatch-

⁹Most drugs however lose their exclusivity after 7 years, and the first approved generic drug receives 180 days of exclusivity. Source: FDA(<https://www.fda.gov/drugs/development-approval-process-drugs/>)

¹⁰An abbreviated new drug application (ANDA) contains data that are submitted to FDA for the review and potential approval of a generic drug product. The estimated cost of an ANDA is \$1-\$5 million (Berndt and Aitken, 2011). See Appendix A.1 on the regulatory costs of entering the U.S. generic market.

¹¹ Burns and Lee (2008) unveil that although 41% of surveyed U.S. hospital providers affiliate with multiple GPOs, they route most of their purchases through a primary GPO, reserving secondary affiliations for specialized contracts.

¹²These Amendments permit the FDA to approve applications for generic versions of brand-name drugs without repeating costly clinical trials to establish safety and efficacy. See Bronnenberg et al. (2015) for a discussion of brand vs. generics.

Waxman Act (1984) institutionalized that all generic versions of a same innovator drug are perfectly substitutables. Conditional on the attribution of an ANDA marketing license, buyers (GPOs and hospital pharmacies) deem all marketed generic versions of a drug to be of satisfactory quality and fully interchangeable. In turn, the lack of transparency of supply chains means that buyers themselves cannot directly observe and reward quality, compelling them to prioritize the lowest-priced drug. This invariably propels prices to be set almost "auction-style": GPOs use competitive bidding through online reverse procurement auctions to select the lowest-cost supplier for member hospitals. Current contracting practices thus contribute to a "race-to-the-bottom" in pricing.

Low-Profit Margin and Price Competition in Generic Markets. High degrees of contestability in the market further discipline prices down, as manufacturers constantly face the threat of being undercut and lose the next procurement contract. All these factors create financial uncertainty for suppliers, which makes low-margin generic markets especially likely to "unravel" due to adverse selection.¹³ Theory predicts an improvement in production quality due to competition *only* if price is greater than marginal cost: in that case, firms facing competition have an incentive to increase their quality in order to attract patients ([Moscelli, Gravelle, and Siciliani, 2021](#)). If price is below marginal cost, then increases in competition can actually lead to reductions in manufacturing quality.

The more generic manufacturers entered the market, the lower the price of the drug. Average monthly prices are lower post-loss of U.S. patent exclusivity (LOE) and generic entry than pre-LOE, for all drug formulations. Aggregate price declines appear to be larger among physician-administered infused/injected drugs (38-46.4%) than among orally formulated drugs (25-26% decline) ([Conti and Berndt, 2016](#)).¹⁴ The average price of a drug product with a single generic manufacturer is 39 percent lower than the branded product, while this price falls down to 95 percent below the branded price with six or more generic manufacturers ([Conrad and Lutter, 2019](#)).¹⁵ Injectables with lower prices have more vulnerable supply chains: sixty percent of pediatric oncology drug products that cost less than \$10 were in shortage in 2022.¹⁶

2.2. Drug Shortages: Definition and Key Features

Defining Shortages: Theory versus Empirics. A shortage is a condition of *excess demand* - in contrast with surplus that is defined as excess supply. In a perfectly competitive market, shortage spells should be transitory; whenever demand exceeds supply, prices should increase and provide incentives for existing and new suppliers to increase production until there is enough supply to meet demand

¹³When producing a drug has low margins, only firms with very low production costs tend to survive, which makes producing the drug even more low margins, in a "doom loop" so that there may be no price at which the market "works". The survival of very low-costs firms only is a characteristic that is correlated with lower quality in production.

¹⁴For injectable or otherwise physician-administered drugs, [Conti and Berndt \(2016\)](#) also find that when the number of manufacturers increases from one to two, average prices fall about 25-30%.

¹⁵Scott Gottlieb, former FDA's commissioner, at CBS News in May 2023: "50 percent of generic drugs right now in a generic manufacturer portfolio lose money. So a generic manufacturer loses money on half of their drugs that they market."

¹⁶Premier, a leading GPO, reported in 2022 that of the more than 400 drugs they had under contract that cost \$3 or less per vial, 42 percent were actively in shortage, while only six percent of drugs that cost more than \$10 per vial were short.

again (market equilibrium, see Appendix Figure A3). In practice however, markets are usually not perfectly competitive, and many frictions may lead shortage spells to last longer. In this respect, the market for prescription drugs and especially generic drugs, which account for most drug shortages, display shortages that persist after supply disruptions, despite some price increases.

Shortages of sterile injectable drugs can occur due to short-run rigidities in both drug manufacturers' production capacities and in end-consumers' demand (hospital patients). This means that product disruptions, whatever their causes, translate into shortages as higher drug prices cannot clear the market. Three reasons explain the low-price elasticities of demand and supply for medically necessary prescription drugs. First, because these drugs are by definition medically necessary, they have few substitutes and consumption generally cannot be shifted over time.¹⁷ Second, shortages are defined at the *patient's* level, and final consumers' demand for prescription drug products is largely unaffected by changes in price, as they purchase through health insurance contracts.¹⁸ Third, on the supply side, costly specialized equipment is required to produce prescription medications. Production processes are complex, especially for sterile injectable products, and firms are required to adhere to Current Good Manufacturing Processes (CGMPs). Costly excess inventories and high costs of switching assigned production lines limit the industry's capacity to increase capacity in response to an increase in price.

Prices of drugs in shortage may increase, though the FDA only found sustained price increases for 18% of drugs that went into shortage in 2013-2017. According to a 2019 survey released by the American Society of Health-System Pharmacists (ASHSP), almost 80 percent of hospitals said drug shortages resulted in increased spending on drugs to a moderate or large extent.

Contracting Practices Create Short-term Price and Capacity Rigidities. In practice, market prices are determined during the procurement process and results from negotiations between buyers (Group Purchasing Organizations or Hospitals) and suppliers (drug labelers). Though these prices are not directly subject to government regulations in the U.S., the standard procurement contract binds for one to two years, which means that negotiated prices must be honored during this period).¹⁹ Sellers are thus generally not allowed to sell their products at a higher price while the contract is valid.²⁰

Governmental price controls only impact drug prices through the insurance market, by determining reimbursement amounts for drugs prescribed through Medicare. These regulated reimbursement prices are however based on market conditions as they adjust based on the past two-quarters market prices, which have been negotiated between suppliers and buyers. While in practice we do observe

¹⁷A drug is considered to be medically necessary if it "is used to treat or prevent a serious disease or medical condition, and there is no other available source of that product or alternative drug that is judged by medical staff to be an adequate substitute." (Fox et al., 2009)

¹⁸Price sensitivity is generally low for the kind of drugs going short (inpatient, mostly oncology and pediatric treatments). Although hospital providers may face price increases during shortages, their demand for prescription medications is not very responsive to the price paid to manufacturers for the drug, especially for low-price medically essential generic drugs.

¹⁹This was confirmed by a discussion with the President of the Group Purchasing Organizations Association, Todd Ebert. He stressed that GPO generally tries to work with firms to solve shortages in order for them to honor their contracts at the negotiated price, considering sourcing from alternative suppliers as the last-recourse solutions.

²⁰At least through the contract they negotiated with their own GPO; drug manufacturers theoretically *can* and sometimes *do* negotiate sales directly with provider hospitals and pharmacies, or through other GPOs, at higher prices

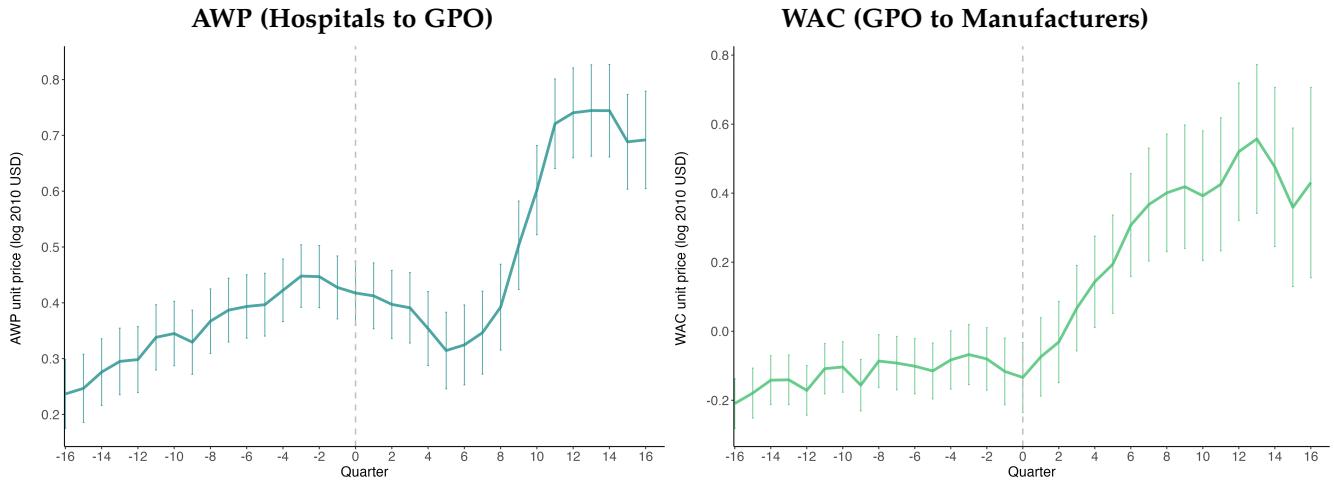


Figure 2: Price Variations Around the Start of Shortage Spells

Notes: Balanced panel of generic injectable drugs in shortage from 2002 to 2019 from the UUDIS. AWP and WAC prices are from IBM Micromedex ("Red Book"). All prices are adjusted for inflation using CPI (base year: 2010). Prices do adjust upwards post-shortage, even if the current structure of contracts between buyer GPOS and suppliers make the adjustment slow. For more discussions on pricing, see Appendix F.1.

significant price increases during and after shortages, the current procurement systems nevertheless imposes some short-term price rigidities and most prices adjustments take at least 3-4 quarters to occur (see Figure 2 and Appendix F.1 for more discussions on pricing.). This pricing feature that procurement contracts set some ex-ante per-unit level of prices that must be honored during the length of the contract validity is nevertheless common to many markets, in particular commodity-like markets.

The low price responsiveness of demand for injectable oncology drugs also has implications for inventories and capacity decision. If there is an excess supply of an oncology drug, there may be no market for it, even at a low price. The combination of limited ability to compel supply (through failure-to-supply clauses or contractual breach provisions) and low price elasticity means that manufacturers face an asymmetry of incentives: there is little cost (except reputational) of producing too little of one drug (rather than another), but a potentially high cost of producing too much of that drug.

Key Features of Shortages. Two features of U.S. (and European) shortages are critical to understand the problem. First, disruptions are supply-driven and due to "manufacturing-quality" issues. Second, disruptions mostly affect generic markets, and among them sterile injectables.

U.S. drug shortages mainly occur because of disruptions in the manufacturing process. 70% of the FDA's reported reasons for medication shortages were product-quality issues that caused a mandatory or voluntary recall or cessation of production. This may include chemical, bacterial contamination or equipment failures. Only 4% of all shortages in the UUDIS data are caused by a sudden increase in demand (see Figure A6). Discontinuations by some firms of older, medically necessary products are

also often triggered by capacity constraints and quality-related production line disruptions.²¹

In theory, supply disruptions do not need to result in product shortages; if there are enough excess capacities or redundancies that can be utilized as backups, or if frictions are small and the market quickly adjust production level and prices as a response to disruptions, then shortages could be avoided. Most of these “manufacturing-quality” issues are directly triggered by lack of incentives to produce less profitable drugs and invest in manufacturing quality and redundant capacity. Drug manufacturers navigate pivotal business decisions throughout a drug’s lifecycle, such as initiating new products, maintaining existing ones, or investing in manufacturing facility maintenance and enhancements. Management may opt not to launch a drug if anticipated profitability is low or uncertain. Further, diminished profitability of a marketed drug may prompt the firm to cease production or restrict manufacturing investments, potentially resulting in supply disruptions.²²

There is indeed little financial incentive to produce older generic, off-patent medications. We rarely see shortages of branded drugs, and the drug shortage problem is highly concentrated: *it almost entirely affects generic drugs*, which represent 90% of prescriptions in the U.S., *and in particular sterile injectable drugs*, including cancer drugs, anesthetics, and antibiotics. These drugs account for 80% of the shortages that occur despite only representing 29% of the entire generics market in volume. Sterile injectable drugs are mainly used to treat acute-care patients in hospitals and are generally administered by physicians. Two-third of sterile injectable drugs on the shortages list are classified in five disease areas: oncology, anti-infective, cardiovascular, central nervous system and pain management. Of the total market for generic sterile injectable drugs, fully 55% of these medicines were on the shortage list in 2011-2019 (see Figure A5).

2.3. *The Case of Sterile Injectable Drugs*

Technological Features of Injectable Drug Markets. Sterile injectable markets exhibits several technological features that make them prone to exacerbate shortage risks.

First, patients demand for this type of drug is very inelastic on the short and supply and demand must match in fine time intervals. Most of these drugs are cancer or pain management drugs used to treat acute care patients in hospitals, so that delaying treatment is costly for patients (health outcomes are conditioned by the consumption of the drug) and they are *de facto* “necessary goods”. Second,

²¹Most shortages caused by delays and capacity issues are actually driven by quality-related shutdowns. Note that companies often do not report the direct cause of a shortage, as this is not compulsory.

²²There are many quotes in FDA hearings reports about U.S. generic drug manufacturers announcing such product discontinuations: “*Right now our energy is focused primarily on the U.S. oral solids business...It's a unique situation. There are significant pricing declines. At least in the medium term, we don't see a shift to that situation, so we're assessing how best to optimize that, given that dynamic.*” (Comments made by Novartis (Sandoz) during the Q4, 2017 earnings call before announcing proposed sale of core U.S. generics business in Sept. 2018). Another quote from the generic manufacturer Teva in Q1 201: “*The overall situation on U.S. generics and pricing ...we had a consolidation on the buyers side and you've had a situation where suppliers may be accepting of lower prices because they used to have a healthy margin...and if you take a race to the bottom on price...the only way out of that negative spiral is of course to stop it ... About 80% of the product we will get out of and they will move to other suppliers and about 20% of the product we will see an increase in price*”.

because manufacturers are selling homogenous generic products (Hatch-Waxman Act, 1984), profit-margins can drop dramatically whenever there is more than one manufacturer of a molecule-form product. Third, storing sterile injectable pharmaceuticals is costly. They need to be kept sterile and can be sensitive to light and temperature. Sterile injectable products must be produced in highly specialized facilities with dedicated lines.²³ These injectable drugs are also harder to store, and most need to be refrigerated or are sensible to lights.²⁴ These high operational costs contribute to increase fixed production costs for the manufacturer. Markups not being high enough to cover fixed costs for this type of drugs may in turn explain the lack of manufacturer investments in facilities. Appendix Section (A.3.2) discusses injectable drug markets in more details.

The slim profit margins of generic sterile injectable drugs created low incentives for manufacturers to shell out factory updates in order to avoid supply disruptions.²⁵ In the case of sterile injectables, there is moreover very little margin for errors in the manufacturing process, contrary to drugs in tablets or capsules. Some of the drug shortages reported in the U.S. have been traced to facilities that have been in operation since the 1960s with almost no factory improvements (GAO, 2016). Moreover, re-purposing existing factories to manufacture antibiotics or cancer drugs can take years.

2.4. A Novel Dataset: Mapping Drugs Supply Chains

This project builds a new comprehensive panel dataset that maps generic sterile injectable drug products to their labelers and plants, and to their history of shortages. This dataset solves traceability issues in the globalized U.S. injectable drugs market and constitutes a major empirical contribution of this paper. The lack of a reliable registry of manufacturing plants and products has notably prevented estimating the true market share of foreign-made products and assessing potential links with supply disruptions. To my knowledge, this paper is the first to assess whether increases in shortage frequencies and amplitudes in the North are disproportionately driven by disruptions in manufacturing *facilities* that are foreign.

To solve traceability issues in the injectable drugs market, I build on three main blocks of data, the third one unifying the first two blocks. The data construction structure is illustrated in Figure 3.

²³Production must take place in a “clean room” environment to make them perfectly sterile

²⁴These factors may explain why supply disruptions are not easily corrected or why hospital pharmacists may not be easily able to stockpile these medicines in prevision of shortages.

²⁵The FDA Commissioner Gottlieb in July 2018 said that “the low profit margins, and the significant cost of manufacturing these complex drugs, has resulted in consolidation in the industry. The only way to produce these low-margin products profitably is to manufacture them at tremendous scale. This has resulted in fewer and fewer manufacturers for certain key products. (...) It has also resulted in an under-investment in manufacturing, especially of sterile injectable drugs. Unless drug makers are investing in their manufacturing facilities, it can create conditions that give rise to production stoppages”.

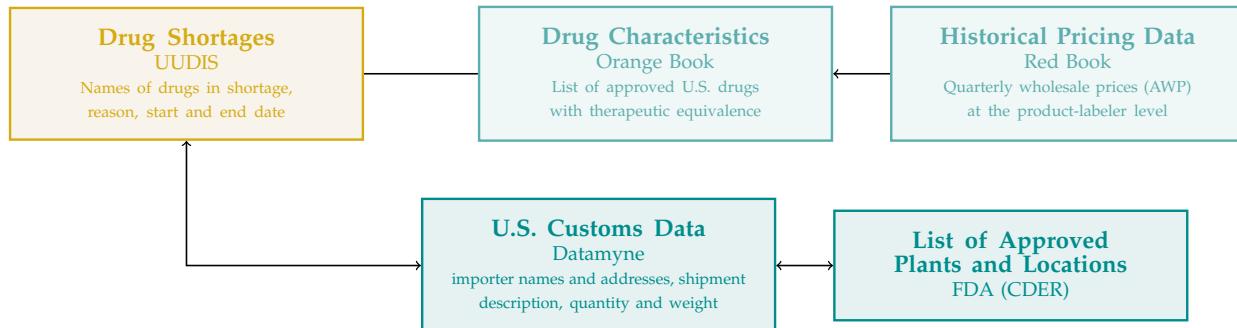


Figure 3: Overview of Data Structure

The first block build a drug-labeler level dataset that captures the characteristics of the sterile injectable drug markets. I first obtain the list of all injectable drugs marketed in the United States in each year (short and non-short) from the FDA's Approved Drug Products with Therapeutic Equivalence Evaluations (so called "Orange Book data"). The Orange Book contains information on a drug's generic/brand status, labeler names, drug ages and therapeutic category. I obtain historical prices data for all U.S. marketed drugs from the "Red Book", a privately-owned dataset proprietary of IBM Micromedex. The "RedBook" provides Average Wholesale Prices (AWP, the price charges by GPOs to hospitals) and Wholesle Acquisition Costs (WAC, the price paid by GPOs to manufacturers) at the NDC level.²⁶. Details on Pricing can be found in Appendix F.1. I complement the data with Medicare Part B's National Summary Files and pricing data to obtain quarterly sales quantities and Average Sale Prices (ASP).

Finally, shortages history were provided by the University of Utah Drug Information Service (UU迪斯), which compiles information on current and past shortages that were reported to the American Society of Health-System Pharmacists (ASHP) and the FDA. I processed two versions of the UUDIS datasets. The first version provides information on drug-level shortages (shortages defined at the molecule-form level). The second version provides information on labeler-drug level shortages and allows to determine which firms made default.²⁷ A main limitation of the data is that it does not quantify the missing product quantities when a shortage occurs.

The second block builds a plant-level dataset that captures the production structure of U.S.-marketed drugs. I first leverage confidential data obtained from the FDA's Center for Drug Evaluation and Research (CDER), which provides the geographical location of foreign and domestic drug manufacturing facilities over the period 2000 to 2022. The data contains the facility name and full address of each registrant, and a unique Facility Establishment Identifier (FEI or DUNS). I complement the data with records of the FDA's inspections and warning letters data, leveraging text mining algorithms to create data from text.

Note that there are two limitations with these data. First, registering to market a drug in the U.S. is

²⁶"NDC" stands for "National Drug Code", a unique numeric drug identifier at the labeler-product-dosage form level.

²⁷This information was stored as thousands of text Documents, which I digitized using automatic text processing and text mining techniques to extract data from text. More details and an overview of these documents can be found in Appendix F.2.

not tied to actually *producing* the drug.²⁸ Secondly, no information is available on the products actually manufactured at the site. As a consequence, these datasets do not allow the FDA to know exactly which facility makes which drug.²⁹

The third and last block unifies the two first blocks by linking products to manufacturing firms and labelers. Using detailed U.S. customs records from a proprietary source (Descartes Datamyne), I use text analysis techniques to identify injectable drug names in detailed descriptions of U.S. import shipments. For the purpose of matching drugs to plants, I then geo-locate addresses of each shippers' address from the customs through automated queries to the Google Maps API. I do the same with the CDER's list of manufacturing plant and use probabilistic records linkage techniques to match manufacturers to shippers based on names, addresses, longitudes and latitudes. These processes are described in details in Appendix F.5 and F.2.

The final data is a balanced panel of 1,220 sterile injectable drugs (defined at the molecule-form level) aggregated at the year-month-manufacturer level. After matching the U.S. Customs to the shortage data and the FDA list of manufacturing facility locations, I obtain the quarterly list of all shipments for which we find U.S. imports of sterile injectable drugs. I then match back the data to the shortage dataset, so that for each sterile injectable drug, we get a full list of manufacturers, quantities of products imported and indicators providing information about the status of shortages for all year-month dates between January 2004 and December 2019.³⁰

3. Reduced-Form Evidence: Offshoring and Shortages

I exploit this rich dataset to underline distinctive features of the drugs affected by shortages, and to provide new empirical evidence on i) offshoring of drug manufacturing facilities and ii) correlations between foreign-made products and higher probability of shortage spells.

While each of the three empirical findings described in Subsection 3.1 may not individually provide enough evidence to link recent spikes in drug shortages to offshoring of production facilities, they together form a constellation of evidence that strongly suggests both events are linked.

These correlations, however, are based on fixed effect panel regressions that may be subject to endogeneity and selection issues and do not establish causal relationships. In order to address selection of drug products and manufacturing firms into offshoring and the endogeneity of the offshoring decisions, I exploit in Subsection 3.2 an instrumental strategy building on the concept of Home Market Effect (Krugman, 1980) to find exogenous variations in offshoring rates across locations and products.

²⁸The FDA found that among all approved generic drugs in 2019, only 39 percent were observed to be marketed.

²⁹This was confirmed through a discussion call with FDA economists from the Drug Shortage Team in 2020

³⁰I do not see hospitals (or GPOs) direct purchases. On the demand side, I only have data at the labeler-molecule dosage form product-quarter level for the entire U.S. market, through Medicare Part B and Red Book Prices. Prices I use are thus average prices over all U.S. hospitals for a given drug product

3.1. 3 Stylized Facts on Offshoring

In this section, I define my measure of *offshoring* before documenting three new stylized facts linking offshoring and drug shortages, which leverage my novel dataset linking injectable drugs to their manufacturing facility locations and to their shortage status.

Definition of Offshoring. I define offshoring as *an increase in the aggregate share of a given molecule-form drug m being produced outside of the U.S. borders*. I remain agnostic about the structure of ownership: in what follows, “offshoring” could be either that (i) a U.S. labeler (generic sponsor) originally producing through U.S.-based plants decide to move production abroad (same labeler “ownership”), or (ii) a drug originally produced in the U.S. by an U.S.-owned and based firm is produced in the following period outside of the U.S. by a foreign-based and foreign-owned firm (different labeler “ownership”). My analysis hence focuses on *actual* production locations to define whether a drug is fully or partially offshored. Manufacturing plants themselves can either be owned by the labeler company or outsourced to Contract manufacturing Organizations (CMO).

The first fact linking offshoring and shortages is based on the single matching of the drug shortages dataset with drug shipments from U.S. custom records, thus capturing the process that linked drug products to their foreign manufacturers. The two following facts exploit the full matching made between the FDA’s list of yearly registered manufacturing locations and their drug products identified through customs, which captures the list of plants *actually* registered to market a drug in a given year.

Fact 1: Foreign imports drop during shortages. In Figure 4a, I study quarterly variations in the average number of foreign manufacturers of a drug around the occurrence of a shortage. The start of a U.S. shortage event corresponds to a sharp drop in the number of foreign drug manufacturers found in the U.S. Customs.³¹ This plot serves two validation purposes: first, it validates the matching of drug products to foreign plants made through the U.S. imports customs, as imports patterns match disruption patterns; second, foreign-based manufacturers are going short. Indeed, a manufacturer affected by supply-driven disruptions should stop or dramatically decrease exports to the U.S. and would mechanically not appear in the U.S. customs records during the disruption.

The following two facts are based on my comprehensive matched dataset of drugs to their registered plants³² Compared to the previous fact, this matching ensures that the presence of firms in the data is *not* affected by the occurrence of shortages, which means that, for each year, I capture the exact set of firms that were manufacturing injectable drug for the U.S. market.

Fact 2: Injectable drugs going short display higher shares of foreign-based manufacturers than sterile injectables that do not. Taking the full panel of sterile injectable drugs sold in the U.S. over the 2002-2019 period, I look at yearly variations in the share of foreign manufacturers for drugs that

³¹ Appendix figure B6 further shows that the 2011-2012 peak of U.S. shortages corresponds to a sharp drop in the number of foreign manufacturers found in the U.S. imports data.

³² It adds to the previous data of drugs products and shipments a second matching between the U.S. customs and the FDA’s dataset of manufacturing plant locations. This matching was made using probabilistic record linkages, both exploiting firms’ names and their geo-located addresses in the U.S. Customs and in the FDA data.

went into shortage *at least once* over the period and compare it to sterile injectable drugs that did not go short (see Figure 4b). Sterile injectable drugs experiencing shortages display a higher shares of foreign-based manufacturers compared to sterile injectables that never went into shortages over the 2002-2010 period, and this share also increases more over time (from 23% to 67% for drugs on the shortage lists, against 5% to 28% for non-short drugs). Imported drugs also face *longer* and *more frequent* shortage spells compared to non-imported drugs, as displayed in Figures B7) (3.8 shortages and 379 days short on average for imported drugs, against 2.1 shortages and 232 days short for non-imported drugs).

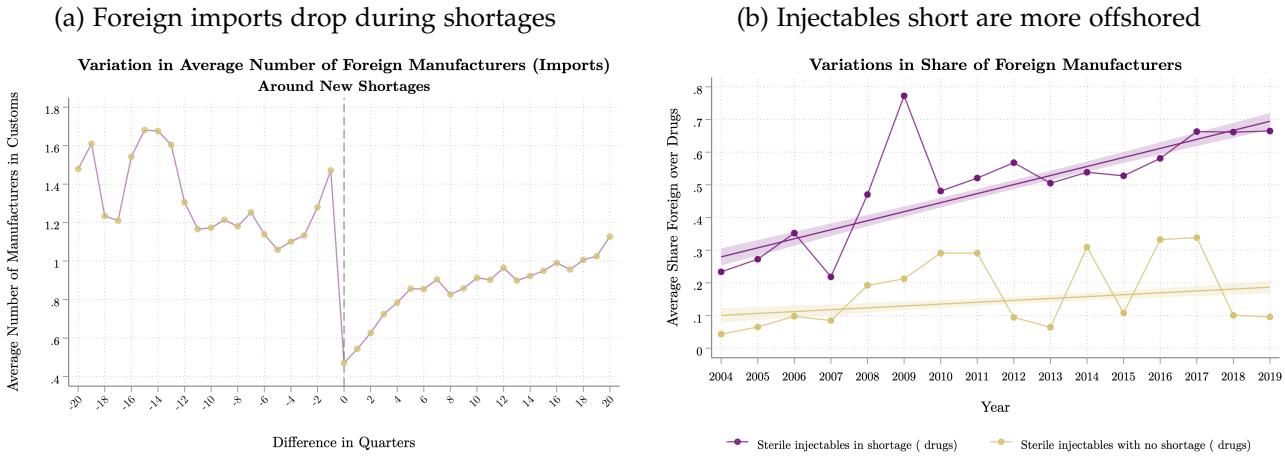


Figure 4: Correlations between Offshoring and Injectable Shortages

Notes: Figure 4a plots quarterly variations in the average number of foreign manufacturers for a drug around the occurrence of a shortage. For each drug, I first compute the average number of manufacturers found in the U.S. customs data by quarter (customs data are granular at the daily level). The average is then computed over all drugs at the quarterly level. A shortage event corresponds to a sharp drop in the number of foreign manufacturers found in the U.S. Customs at the exact time of the shortage. Figure 4b shows that injectable drugs going short display higher shares of foreign-based manufacturers than sterile injectables that do not. Leveraging a balanced panel of generic sterile injectable drugs, the share of foreign producers of drugs that go short go from 23% in 2004 to 67% in 2019. Generic injectable drugs that do not go short go from 4.3% of foreign manufacturers to 33% at their peak in 2017.

Fact 3: Sterile injectable drugs that went short over the last 15 years are the most heavily exported by Indian firms. In Figure 5, I use detailed descriptions of shipments from digitized Indian Customs Data to plot the evolution of Indian exports of sterile injectable drugs over years 2008-2019. I exploit similar text mining techniques already used to identify drugs in the U.S. custom manifests to look for text strings that match the list of sterile injectable drugs sold in the U.S. inside Indian imports and exports of pharmaceutical products.³³ Figure 5 plots the evolution of imported quantities and imported drug values over time. I then further split the sterile injectable drugs I capture in the Indian Customs Data by their shortage status in Figure 5 (a drug is labeled as "shortage" if it appears on the U.S. shortage list at some point).

Additional Results: Event Study and Dynamic DID In Appendix Section B.5, I describe an additional analysis whose goal is to mitigate concerns about endogeneity and selection. I develop a matched dynamic difference-in-differences (DID) design on a subset of data to estimate the partial equilibrium

³³I match 65% of sterile injectable drugs present in my data for the U.S. based on a matching of exact drug names. I did not use here such a comprehensive pre-cleaning of shipment descriptions and fuzzy-matching techniques as performed for the main U.S. customs data

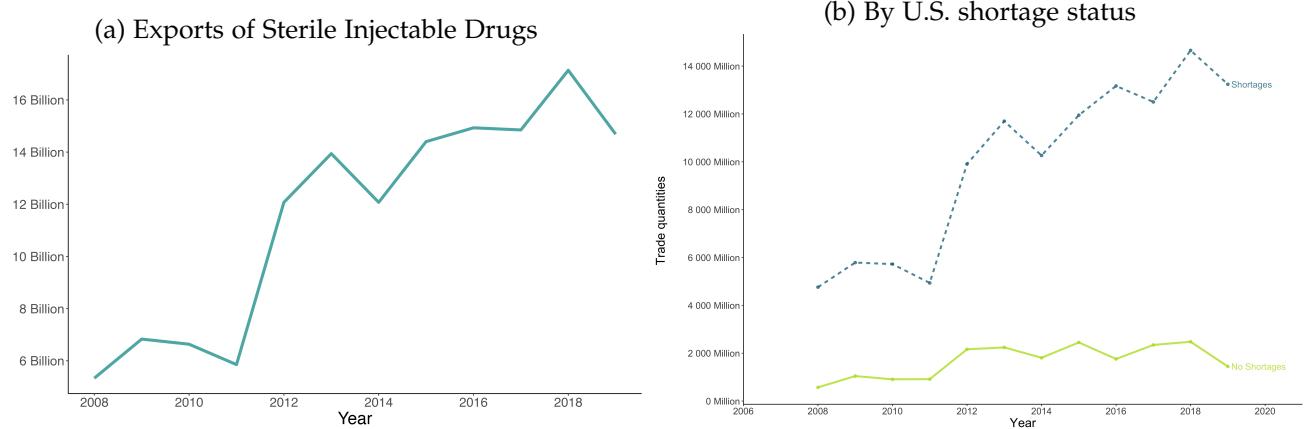


Figure 5: Global Exports of Sterile Injectable Drugs, India

Notes: This figure exploits detailed descriptions of shipments from digitized Indian Customs to plot the evolution of Indian exports of sterile injectable drugs over years 2008-2019. Panel (b) split the sterile injectable drugs I capture in the Indian Customs Data by their shortage status. A drug is labeled as in "shortage" if it appears on the U.S. shortage list at some point. I have approximately the same number of drugs in both groups (250 drugs categorized as "shortage" drugs and 260 drugs without U.S. shortages). Sterile injectable drugs that went into shortage at some point over the last 15 years are those that are the most heavily exported by Indian firms.

impact of shifting from domestic to offshored production on drug shortages. In order to address selection of drug products and manufacturing firms into offshoring and the endogeneity of the offshoring decisions, I match sterile injectable drugs whose manufacturing process was offshored with similar drugs that were never produced abroad. This matching, using a Coarsened Exact Matching (CEM) algorithm, controls for selection on observables, while the DID design address potential concerns about selection on unobservables.

Table C1 displays the regression results and Figures C1 its event-study version. All post-treatment coefficients are positive and strongly correlated with shortages. Offshored drugs are on average 12% more likely to experience shortages compared to similar drugs whose production remained U.S.-based. There is also a 6% higher probability of a new shortage being notified just after offshoring for these drugs, compared to their counterparts. Eventually, offshored drugs are strongly correlated with longer shortage events compared to non-offshored drugs - on average 125 days longer across specifications. These results are robust to alternative definitions of treatment groups (see Appendix Tables ?? and ??).

3.2. A Shift-Share Instrumental Variable Strategy for Offshoring

The decision of a manufacturing firm to offshore production is endogenous, and there is also selection of the type of drugs and firms that will decide to go abroad. In order to address these endogeneity concerns and strengthen the evidence for a causal interpretation, I develop in this section a shift-share instrumental variable for the offshoring of sterile injectable drugs. The instrument leverages the concept of Home Market Effect to isolate a plausibly exogenous source of variations in offshoring rates of individual drug products that provides variations at the drug product-location-year level.

Instrumental Variable Strategy. The instrument builds on the idea that offshoring of off-patent

medications may be driven by a shift in global demand for generic drugs. The last 20 years have witnessed a shift in the global distribution of consumption for medicine products, mostly driven by a boom in consumption for healthcare and generic drugs by “Pharmerging” giants like Brazil, Mexico, India, or China.³⁴ As the world’s consumption of drugs shifts, production re-allocates to get closer to this new booming market, and take advantage of economies of scales and cheaper production costs in the South. This is a classical “Home-market Effect” story à la Krugman (Krugman, 1980): countries with larger demand for some products at home tend to have larger sales of the same product abroad, because it lends a competitive advantage to local firms³⁵.

To capture the effect of an increase in demand for some category of drugs in South-Asian countries on the re-allocation of production, the instrument requires to isolate some plausibly exogenous source of demand variations at the drug product-year-geographical location level. I build on an idea that was first introduced by Acemoglu and Linn (2004) and exploited more recently by Costinot et al. (2019). This demand shifter builds on the observation that demographic groups who are more likely to die from a particular disease, because of exogenous characteristics of their population, are also more likely to demand pharmaceutical treatments that target those diseases.

Predicted Disease Burden. The predicted disease burden measure, detailed in equation (1), thus captures the average country-level disease burden that would be expected given a country’s demographic structure. This ratio varies at the country-disease-year level and leverages cross-country and cross-time variations in demographic structures and healthcare access. The empirical strategy thus exploits the fact that disease burdens vary over time and by demographic groups and that countries vary in their demographic composition (isolating the apparent extent to which drugs have a demographic bias in their relevance, as well as the extent to which countries differ in their demographic composition of their populations). Changes in demographic composition (the “shifter”) interacted with the propensity for a demographic group to use a drug (the “share”) provide an exogenous variation in market demand.

$$(PDB)_{it}^d = \sum_{a,g} \left[\text{InsuredPop}_{iagt} \times \left(\frac{\sum_{k \neq i} \text{disease burden}_{kagt}^d}{\sum_{k \neq i} \text{population}_{kagt}} \right) \right]. \quad (1)$$

The last ratio term measures the average disease burden per capita from disease d for gender g and age group a , calculated excluding the country of interest (that is, summing over all countries k except for country i). This ratio is then weighted by the share of population for that gender g and age group a , and summed across age and gender groups, for a given country i in year t . Because the rise in consumption of drug products in the South is in part driven by the rapid access of wealthier populations to health insurance, we weight these populations by the share of people insured by country in each year, using data from the World Health Organization. This provides additional variation at the country-year level.³⁶ Note that the fact that firms from country i are better at treating a given disease d

³⁴IQVIA estimates that the accelerating shift in global spending toward generics between 2010 and 2015, rising from 20% in 2005 to 39% of spending in 2015, was mostly driven by an increase of consumption from these four countries

³⁵Incidentally, the main categories of drugs in shortage in the United States are those that faced the greatest increase in consumption in South-Asia: generic drugs, drugs used in hospitals (contrary to OTC medications, which tend to be more expensive), and primarily oncology drugs and antibiotics.

³⁶Figure B8 additionally shows that the constructed predicted disease burden measures display significant variations over

may cause a lower burden for that disease in country i . Leaving out country i from the average disease burden per capita addresses this endogeneity issue.

To build this predicted disease burden measure, I use the World Health Organization (WHO)'s Global Burden of Disease (GBD). The data provide the country-year-disease measures of burden, broken down into six different demographic groups: three age groups (0–14, 15–59, and 60+) for each gender³⁷. I use as a primary measure of disease burden the number of lost disability-adjusted life years (DALYs, which combines data on the mortality and morbidity caused by each disease. I eventually use population data for each country in each of the six demographic groups for each year (2004–2019) from the WHO as well. To link drugs to diseases, I first create crosswalks that allows linking each therapeutic category in my dataset, based on the Ephmra classification, to the WHO's list of therapeutic categories (ATC). I then use the hand-coded mapping made by [Costinot et al. \(2019\)](#) between therapeutic categories and the (WHO)'s Global Burden of Disease dataset to map my drug products to disease burden.

IV Regression Results. Table 1 displays the results of a regression of shortage probability on offshoring, in which I instrument for offshoring of drug facilities with the constructed predicted disease burden measure for India and China. These regressions are based on a linear probability model (LPM) with a binary dependent variable, Shortage_{jt} , a dummy equals to one if drug j is in shortage at time t :

$$\text{Shortage}_{jt} = \alpha_j + \delta_t + \beta \text{Offshoring}_{jt} + \phi_p \log(p_{jt}^{\text{AWP}}) + \phi_M M_{jt} + \phi_P 1\{\text{PatentStatus}_{jt}\} + \nu_{jt}$$

where the first-stage is

$$\text{Offshoring}_{jt} = a_j + \Delta_t + \beta \text{PDB}_{jt}^{\text{Asia}} + \varphi_p \log(p_{jt}^{\text{AWP}}) + \varphi_M M_{jt} + \varphi_P 1\{\text{PatentStatus}_{jt}\} + \omega_{jt}$$

Offshoring_{jt} captures the share of drug j produced outside of the U.S. borders at t . $\ln(p_{jt}^{\text{AWP}})$ is the log-median AWP unit price (in 2010 dollars), M_{jt} measures the number of labelers for drug j at t (a measure of market concentration on the supplier side) and $1\{\text{PatentStatus}_{jt}\}$ is an indicator variable equal to one if the drug lost its patent status. All regressions include year and drug fixed effects.

All else equal, an increase in the share of foreign producers for a drug is strongly correlated with an increase in the probability that a shortage occurs and in the fraction of time during which a drug is in shortage, in a given year. For every 100% increase in the share of foreign manufacturers for a given drug, we see a 31 percentage point increase in the likelihood of a new drug shortage, all else being equal. This goes up to 56 percentage points for a 100% increase in the share of South-East Asian manufacturers. Similarly, a 100% increase in the share of foreign production of a drug is associated with a 36.5% percentage point increase in the length of the shortage events (61% for offshoring to Asia only). Similar regressions without controls (see table) or using alternative measures of the Predicted Disease Burden based on India and China only, provide coefficients of similar magnitude.

time and geographical regions.

³⁷Although there may be local variation in the collection of vital statistics that underpin these measures, the WHO ensures that these data are valid for cross-country and cross-disease comparisons.

Table 1: Probability of shortage on share foreign - Instrument: log(PDB)

Dependent Variable:	P(New Shortage=1)	Shortage Days		
<i>Second-Stage[†]</i>				
Facilities, Share Foreign	0.3129** [.1152]		0.3651** [.1447]	
Facilities, Share Asia		0.5415** [.2721]		0.6064*** [0.1873]
Log Median Unit Price	-0.00116 [0.00529]	0.00437 [0.00517]	-0.00617* [0.00357]	-0.00379 [0.00367]
Number Labelers	0.01184** [0.0059]	0.01689*** [0.0056]	0.00513 [0.00378]	0.00918** [0.00377]
Off patent dummy	0.01519 [0.03379]	-0.03058 [0.0425]	0.00447 [0.02206]	-0.0388 [0.02918]
<i>First-Stage</i>				
$\Delta \log(\text{PDB})$	0.3*** [0.109]	0.2254*** [0.07342]	0.3*** [0.109]	0.2254*** [0.07342]
Year Fixed Effects	✓	✓	✓	✓
Drug Fixed Effects	✓	✓	✓	✓
Std.Dev.	13.44	18.72	9.038	12.88
F-stat First-Stage [‡]	36.44	34.4	36.44	34.4
Beta Coefficient	0.4874 [0.3352]	0.8354** [0.3544]	0.9349** [0.3706]	1.512*** [.4009]
OLS Coefficient	0.00322 [0.01475]	0.06109***	-0.00389 [0.01013]	0.04124*** [0.01201]
Observations	16576	16576	16576	16576

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bootstrap SE in parenthesis. [†]Instrument: Predicted Disease Burden (log). F statistics are statistically significant following the "tF" test procedure given by Lee et al.(2020) for weak IVs. Table B7 displays the same regressions without control variables

4. A Model of Shortage in Global Procurement Markets

I propose in this section a structural model of manufacturers' location and production choices in global procurement markets that endogenizes shortage occurrence. The model serves several purposes. First, it generates insights into the incentive dynamics of supply disruptions and helps rationalize the offshoring correlations observed in the data. Second, it provides a framework to analyze the manufacturers' response to the moral hazard incentives created by the offshoring option, and allows to quantify buyers and sellers' behavior in terms of economic parameters, such as the weight buyers and sellers place on future disruption outcomes in their decision-making processes.

4.1. Model Summary and Timing

In order to capture what drives the current market choices of buyers and drug manufacturers, and what may motivate manufacturers to let supply periodically fail, the model matches key characteristics of

the injectable drug market. It features international procurement contracts between U.S. buyers (GPOs, see Appendix A.2.1 for institutional details) and a diversified group of suppliers (pharmaceutical manufacturers). These suppliers' investments in production plants (facilities maintenance and upgrades) are not fully observable by the regulator (the FDA) and are subject to periodic regulatory inspections to monitor production quality. Buyers leverage competitive bidding to allocate market shares to suppliers through year-long procurement contracts, establishing some short-term market inflexibility, as prices and capacities are predetermined before any disruptions materialize.

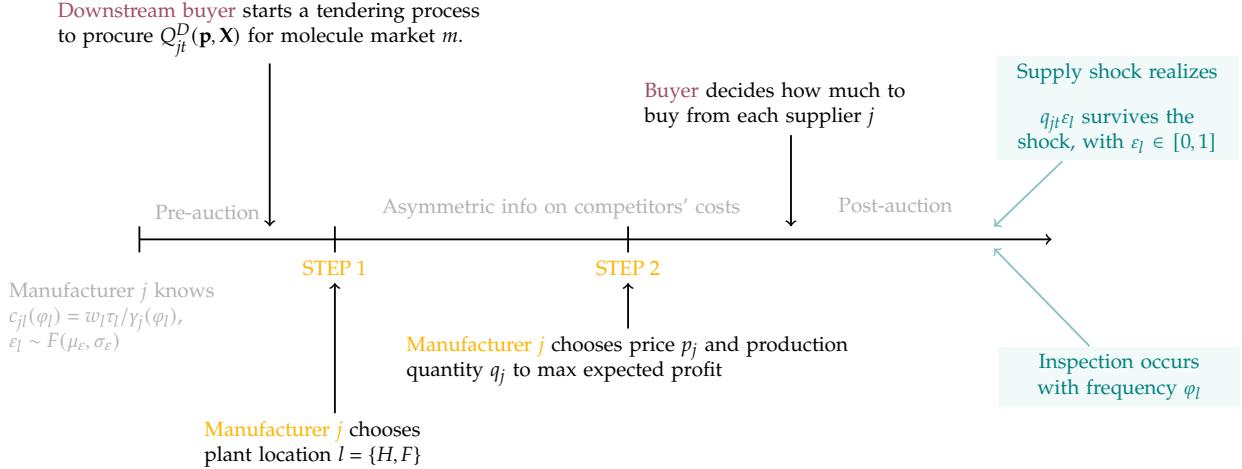


Figure 6: Model Timing

Notes: A downstream buyer (GPO) located in a single location (Home) starts a tendering process to procure quantity $Q_{mt}^D(\cdot)$ of a given molecule-form product m at time t . Molecule markets m are oligopolistic and contestable, and a set of J^{mt} upstream drug manufacturers enter each period in order to obtain procurement contracts. Each manufacturer j 's problem is decomposed into two steps. In a first step, each manufacturer makes *ex-ante* location choices for their unique plant. In a second step, they choose production quantities q_{jmt} and unit prices p_{jmt} to maximize their expected profit. While drugs are considered *homogeneous at the molecule level*, there is observable heterogeneity at the manufacturer level: buyers assign market shares based on manufacturers' proposed prices and observable characteristics. After contracts are assigned, a location-dependent exogenous random supply shock $\omega_{l(j)t}$ realizes and may reduce supply delivered relative to chosen production levels. Manufacturers have two available margins to insure against future expected supply shock: first, building capacity buffers; second, choosing production locations with low-disruption risks.

Featuring moral hazard and a stochastic production process, along with buyers' price sensitivity and a potential taste for supply reliability, the model underscores that while disruptions arise from exogenous random supply shocks, the likelihood of ensuing stock-out events can be strategically managed by firms through the creation of capacity buffer or by opting for less disruption-prone production locations. Opting for greater production capacities or establishing plants in more regulated markets minimizes the risk of disruptions escalating into conditions of excess demand, but incurs higher costs for suppliers (e.g.: cost of facility upgrades or quality checks), and eventually, buyers. Thus, the model endogenously generates market equilibria characterized by supply shortages - defined as conditions of excess demand at the product level.

This tractable model can be brought to a variety of data and expanded to other industries affected by supply disruptions, enabling researchers and policymakers to better quantify the complex trade-offs between offshoring, cost savings, and supply chain resilience. In particular, this model of manufacturer

locations and production choices in global procurement markets may accommodate many different demand systems - as long as one has data on prices, observed quantities sold, and indicators for manufacturer-level disruptions.

4.2. A Discrete Choice Logit Model of Buyers' Demand for Drugs

Setup. A sterile injectable drug market m is defined at the molecule-form level (e.g. injectable morphine). In each market m , there are $i = 1, \dots, M_{mt}$ patients at a given time t (a fiscal quarter). A downstream buyer (GPO) located in a single location (Home) starts a tendering process to procure quantity $Q_{mt}^D(\cdot)$ of a given molecule-form product m . Each buyer is assumed to be a monopsonist for its set of patients (Dafny, 2005).

Molecule markets m are oligopolistic, and a set of J^{mt} upstream drug manufacturers compute each period in a contestable market (Tirole, 1988; Chaudhuri, 1996) in order to obtain procurement contracts. We assume a manufacturer j enters a procurement auction if they are able to set regular capacities to at least Q_{jmt}^D . Buyers choose among $J_m + 1$ differentiated products where $j = 0, \dots, J_{mt}$ and good 0 denotes the outside good (no purchase), and contract quantities Q_{jmt}^D with each supplier such that $Q_{mt}^D = \sum_j^{J^{mt}} Q_{jmt}^D$. Buyers assign market shares based on players' proposed prices and observable characteristics. While drugs are considered *homogeneous at the molecule level*, there is observable heterogeneity at the manufacturer level: buyers assign market shares based on manufacturers' proposed prices and observable characteristics (brand versus generics, labeler size, or firm's reputation as measured by their average past number of stock-out events).

Logit Model. For a given molecule market m , patient i 's indirect utility from buying a drug product from supplier j in fiscal quarter t is given by a fixed-coefficient discrete choice logit model:

$$v_{ijmt} = X_{jmt}\beta - \alpha p_{jmt} + \lambda g_{jm} + \phi_{mt} + \epsilon_{ijmt} \quad \forall j \quad (2)$$

and $v_{i0mt} = \epsilon_{i0mt}$

where I normalize the utility v_{i0mt} for the outside good to zero. Drug characteristics are captured by (i) p_{jmt} , the endogenous manufacturer j 's price in market m at time t , (ii) g_{jm} , a binary variable indicating whether drug j is generic or brand and (iii) X_{jmt} , which captures an exogenous reliability score varying at the manufacturer-molecule form- fiscal quarter level. ϕ_{mt} is a molecule-quarter fixed effect controlling for unobservable factors that vary across molecule markets and over fiscal quarters (but are constant within a market-quarter), and ϵ_{ijmt} is an idiosyncratic error term.

Buyers disutility from higher prices is captured by the coefficient α , while taste for a supplier's "reliability" is captured by the coefficient β . I focus on purchases made in the hospital sector, and hospitals typically fully internalizing the prices of drugs that they purchase on behalf of patients.³⁸ I follow Dubois and Lasio (2018) by assuming that demand is static; hence, I am not allowing for consumer

³⁸Similarly to Dubois and Lasio (2018), I cannot observe data on the behavior of insurers, healthcare providers, and other intermediaries between patients and drug manufacturers, and hence I abstract away from modeling them and do not disentangle their role in aggregate revealed preferences.

learning. Although learning may be important due to the uncertainty over suppliers' reliability in this market over the considered period, the lack of transparency over supply chain links means that buyers actually have little ability to learn from previous contractual relationships with sellers.³⁹ Hence, buyers imperfectly observe a supplier's reliability and I assume that the reliability score X_{jmt} is an exogenous measure, which in practice, could be attributed by the FDA.⁴⁰ In practice, I build this reliability score measure based on a supplier's history of shortages, using as an index the fraction of time in the previous year during which the manufacturer was *not* short (so that higher scores correspond to higher reliability levels).

Following the logit assumption on the extreme-value type I distribution of the error term ϵ_{ijmt} , there is a direct mapping from indirect utility to model-predicted market shares for each manufacturer. The market share of manufacturer j in market m and quarter t is:

$$s_{jmt} = \frac{\exp \{X_{jmt}\beta - \alpha p_{jmt} + \lambda g_{jm} + \phi_{mt}\}}{1 + \sum_{k=1}^{jmt} \exp \{x_{kt}\beta - \alpha p_{kt} + \lambda g_{km} + \phi_{mt}\}} = \frac{\exp \{\delta_{jmt}\}}{\sum_{k=0}^{jmt} \exp \{\delta_{kt}\}} \quad (3)$$

Demanded quantities (quantities *contracted* as a result of the procurement auction process) can be obtained by multiplying logit market shares by market size M_{mt} :

$$Q_{jmt}^D(p_j, X_j, p_{-j}, X_{-j}) \equiv Q_{jmt}^D(\mathbf{p}, \mathbf{X}) = M_{mt} \cdot s_{jmt}$$

where $M_{mt} - \sum_{j=1}^{J_m} Q_{jmt} = Q_{0mt}$ is the size of the outside option and M_{jmt} is the number of prescriptions for manufacturer j in molecule-market m and M_{mt} is the total number of prescriptions for a given molecule-form drug. Each manufacturer j 's contracted quantity depends on prices and characteristics of all other manufacturing firms competing in the market.

4.3. Uncertainty and Shortages

The economy is subject to locations-dependent exogenous random supply shocks, which realize after contracts are assigned to firms (and thus, after manufacturers' pricing and supply decisions are made).⁴¹ The randomness in supply can give rise to firm-level shortages, where supply of firm j falls short of demand. A shock reduces capacity at facility j from $Q_{jmt}^D(\mathbf{p}, \mathbf{X})$ to $Q_{jmt}^D(\mathbf{p}, \mathbf{X})\omega_{l(j)t}$ with

³⁹GPOs themselves have no good vision of shortage history; first, because generic markets are highly contestable, such that firms enter and exit the market constantly. Second, because drug manufacturers generally sell one product variety through one GPO at a time, and may have previously sold this product through a competitor GPO - in which case, buyer i ignores the past relationship between this supplier and their competitors. Third, because drug supply chains are long, globalized, fragmented and labelers are not required to disclose the identity and location of their plants. As an example, Vizient, a leading GPO, started to contract in 2021 with the UUDIS in order to access information on shortages history and get warnings when shortages are about to start. The FDA recently advised GPOs to ask about the exact identity and location of their suppliers when drafting contracts, but there is no legislation enforcing this practice and the U.S. regulator still does not have the authority to intervene in procurement contracts.

⁴⁰Note that reputation does not matter *by itself* in the model; it only matters in that it may predict future disruptions. If the buyer believes that the yearly accreditation delivered by the FDA is sufficient measure of reliability, then additional signal from the sellers on product quality is unnecessary. See discussion in Section D.4.

⁴¹Following current market practices, contracted prices and quantities are assumed to be fixed during the length of the contract (one to two years). Prices are renegotiated over time when contracts expire or when the buyer meets new sellers.

$\omega_{l(j)t} \in [0, 1]$. $\omega_{l(j)t}$ is a random variable that captures the share of firm j 's supply that survives the exogenous shock and can be utilized to fulfill demand - a "capacity yield".⁴²

Definition 4.1 (Firm-level Shortages). *The probability that firm j goes short in a given fiscal quarter t and molecule market m is defined as the probability of positive excess demand for firm j :*

$$\begin{aligned} Pr(\text{shortage})_{jmt} &= Pr\left(\underbrace{q_{jmt}\omega_{l(j)t}}_{\text{Realized Supply}} < \underbrace{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}_{\text{Demand}}\right) \\ &= Pr\left(\omega_{l(j)t} < \frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}}\right) = F\left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}}\right) \end{aligned} \quad (4)$$

where $\omega_{l(j)t} \sim F(\mu_{\omega_l}, \sigma_\omega)$ is the realization of an exogenous random supply shock with cumulative distribution function $F(\cdot)$.

The realization of shortage spells in the economy affects the supply quantity effectively *observed* by the econometrician. In equilibrium, absent of shortages, the observed quantities sold are equal to demanded quantities. During shortage periods, the realized quantities sold fall below demanded quantity levels, so that observed market shares are no longer equal to the traditional logit shares. Defining an indicator equal to 1 during shortage periods and 0 otherwise, $\text{Short}_{jmt} \equiv \mathbb{1}\{q_{jmt}\omega_{l(j)t} < Q_{jmt}^D(\mathbf{p}, \mathbf{X})\}$, the observed quantities supplied by manufacturer j become

$$\begin{aligned} q_{jmt}^{\text{observed}} &= \min\{Q_{jmt}^D(\mathbf{p}, \mathbf{X}), q_{jmt}\omega_{l(j)t}\} \\ &= Q_{jmt}^D(\mathbf{p}, \mathbf{X}) - \text{Short}_{jmt} \left(\underbrace{Q_{jmt}^D(\mathbf{p}, \mathbf{X}) - q_{jmt}\omega_{l(j)t}}_{\text{Excess demand}} \right) \end{aligned} \quad (5)$$

where $Q_{jmt}^D(\mathbf{p}, \mathbf{X}) - q_{jmt}\omega_{l(j)t}$ is positive in situations of excess demand and negative for excess surplus.

4.4. Manufacturer's Profit Maximization Problem: Supply of Drugs

The choices of manufacturer j selling in an molecule oligopolistic and contestable market m at time t is decomposed into two sequential steps. In a first step, manufacturer j chooses the location of their unique plant. In a second step, all manufacturers competing in a molecule market simultaneously choose production quantities q_{jmt} and bid prices p_{jmt} to maximize expected profits.

The realization of shortages is directly endogenized into suppliers' choices. While manufacturer j has no control over the *realization* of exogenous shocks (natural disasters, trade shocks, or contaminations may occur independently on their facilities investments), they can choose the *probability* that a shortage occurs (i.e. that demand goes unfilled). In particular, manufacturers have two available margins to insure against shortage risks: first, they can build capacity buffers by choosing to produce above the initially contracted quantity levels; second, they can choose production locations with low-disruption risks (which is equivalent to choosing a less risky draw of $\omega_{l(j)t}$). Both assurance

⁴²An assumption widely used in the Operation Management Literature; see for instance Tang and Kouvelis (2011)

margins may however come at the price of higher marginal and fixed production costs). The pivotal manufacturer's trade-off between supply reliability and cost-cutting thus appears in both steps of the firm's problem: in step 2, the manufacturer trades-off price levels versus surplus quantities; in step 1, they trade-off disruption risk versus location-specific production costs.

The model is solved by backward induction; first computing the equilibrium prices and quantities conditional on a location, then choosing the location that brings the highest expected level of profits.

4.4.1. Step 1: Location Choice

Each manufacturer trades-off three location-specific factors: the distribution of supply shocks, marginal costs and fixed costs. In order to capture key features of the observed data, I focus on three groups of locations: "Home" (U.S.-based plants), "Foreign, North" (corresponding to OECD countries in the data) and "Foreign, South" (corresponding to South-East Asia, mostly India and China). In what follows, I denote these locations as $\{H, F_N, F_S\} \equiv \{H, F_N, F_S\}$.

Manufacturer j chooses the location l in which to produce in order to maximize *expected* profits net of fixed investment/entry costs:

$$l^* = \arg \max_{\{\pi_{jmt}^l\}} \mathbb{E} \left[\pi_{jmt}^l \left(\omega_{l(j)t}, c_{jmt}^l; q_{jmt}^*, p_{jmt}^* \right) - FC_{jmt}^l \right] \quad (\text{Eq. Location})$$

Choosing a location first boils down to choosing the distribution of the exogenous supply shock one wants to draw a $\omega_{l(j)t}$ (capacity yield) from. Choosing a location also boils down to choosing fixed and marginal cost levels. Fixed-costs are assumed to follow a location-specific log-normal distribution $\ln(FC_{jmt}^l) \sim N(\mu_{FC_l}, \sigma_{FC_l}^2)$ with firm-level logit shocks. Firm j 's marginal cost is parametrized as

$$c_{jmt}^l = \frac{w_{lt}\tau_{lt}}{\gamma_{l(j)t}(\varphi_l)} \quad (6)$$

, where w_{lt} is location-specific wage in the pharmaceutical manufacturing sector, $\gamma_{l(j)t}$ is a firm-specific idiosyncratic productivity shock, τ_{lt} is an iceberg trade cost between manufacturing location l and the destination market (the U.S.) and φ_l is inspection frequency in location l over the previous periods, with $\varphi_{F_S} < \varphi_{F_N} < \varphi_H$.

4.4.2. Step 2: Choice of Production and Prices

Given location l , manufacturer j operating in market m at a given time t , chooses (simultaneously with competitors) optimal output levels q_{jmt}^* and unit price p_{jmt}^* in order to maximize *ex-ante expected profit*, conditional on expectations over future supply shocks. Taking as given the production and pricing decisions of their competitors, firm j 's profit maximization problem can then be written as a weighted

average of expected profits in disrupted and undisrupted times:

$$\max_{q_{jmt}, p_{jmt}} \mathbb{E}_{\omega_l} \left[\pi_{jmt}^l(p_{jmt}, q_{jmt}; Q_{jmt}^D(\cdot), c_{jmt}^l, FC_{jmt}^l) \right] =$$

$$\underbrace{\left[1 - F \left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}} \right) \right]}_{\Pr(\text{no excess demand})} \underbrace{p_{jmt} Q_{jmt}^D(\mathbf{p}, \mathbf{X})}_{\mathbb{E}(\text{Revenue} \mid \text{no excess})} +$$

$$\underbrace{F \left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}} \right) p_{jmt} q_{jmt} \mathbb{E} \left[\omega_{l(j)t} \mid \omega_{l(j)t} < \frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}} \right]}_{\Pr(\text{excess demand})} - \underbrace{(c_{jmt}^l q_{jmt} + FC_{jmt}^l)}_{\text{Total Production Cost}}$$

(Eq. Profit)

where the probability of shortage $F \left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}} \right)$ can be directly written as a function the cumulative distribution function $F(\cdot)$ of the exogenous supply shock (see Definition 4.1).

This probability captures the fraction of time, during a given fiscal quarter t , during which firm j 's expects to supply below demanded quantities. In absence of shortages, firm j 's revenue simply corresponds to the product of unit prices p_{jmt} and demanded quantities $Q_{jmt}^D(\mathbf{p}, \mathbf{X})$. Under situations of excess demand, firm j 's expected revenue corresponds to the product of unit prices and the expected realized supply. This depends on the initially chosen production quantities q_{jmt} and on the conditional expectation over location-specific capacity yield $\omega_{l(j)t}$ when a shortage is realized (positive excess demand).

Production displays increasing return to scale, as the total production costs of firm j are equal to $c_{jmt}^l q_{jmt} + FC_{jmt}^l$ where FC_{jmt}^l represents location-specific draws of fixed costs. Importantly, note that whatever the scenario that realizes, the firm pays marginal costs c_{jmt}^l that correspond to the *actual* production choices made by the firm at the time contracts were negotiated. While the generic injectable industry is well-known for practicing just-in-time production,⁴³ I allow suppliers in the model to "overproduce" in order to insure against future disruptions. Consistently with the high storage costs of sterile injectable drug markets, I choose to abstract from storage and I do not explicitly model suppliers' choice of equilibrium inventory levels. Instead, I assume manufacturers insure through costly production buffers and free disposal of production surplus. This assumption helps for tractability while allowing to capture the key feature of interest: producing surplus is costly in these low-margin markets, and manufacturers may want to minimize losses by reducing "inventories."

A firm j 's conditional expectation over location-specific capacity yield $\omega_{l(j)t}$ when a shortage is realized can be re-written as:

$$\mathbb{E}[\omega_{l(j)t} \mid \text{shortage}_{jmt}] = \mathbb{E} \left[\omega_{l(j)t} \mid \omega_{l(j)t} < \frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}} \right] = \frac{1}{F \left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}} \right)} \int_0^{\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}}} e f(e) de$$

⁴³Drug manufacturers would allegedly not keep more than one to two weeks equivalent of stocks. Since the Covid-19 pandemic, the FDA has acted to support legislation that would enforce drug manufacturers to build at least three months of stocks - a buffer that many health policymakers still judge insufficient given the average lengths and depths of shortage events. As of October 2023, this legislation has not been enacted and remains at the state of "recommendations"

where $f(\cdot)$ denotes the probability density function of the exogenous supply shock $\omega_{l(j)t}$.

Using this formulation, I can rewrite firm j 's ex-ante expected profit as:

$$\mathbb{E}_{\omega_l} \pi_{jmt}(\cdot) = p_{jmt} \left(\underbrace{\left[1 - F\left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}}\right) \right] Q_{jmt}^D(\mathbf{p}, \mathbf{X})}_{\mathbb{E}[\text{sales|no shortage}]} + \underbrace{q_{jmt} \int_0^{\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}}} e f(e) de}_{\mathbb{E}[\text{sales|shortage}]} \right) - c_{jmt}^l q_{jmt} - FC_{jmt}^l \quad (7)$$

The structural form of the manufacturer's problem emphasizes that while the occurrence of disruptions is determined by exogenous random supply shocks, the probability of a subsequent stock-out event can be targeted by the firm (by building more buffers or choosing less risky production locations). While choosing to invest in higher levels of production capacities or building plants in more regulated markets decreases the probability that a disruption turns into conditions of excess demand, it comes at a higher cost for suppliers (e.g.: cost of facility upgrades or quality checks), and eventually buyers.

Equilibrium prices and supply quantities are obtained from the first-order conditions of the firm's problem (See Appendix D.1 for details on the model derivations). From these first-order conditions, I obtain the following relationships for equilibrium prices and quantities:

$$p_{jmt}^* = c_{jmt}^l \left[\int_0^{\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}}} e f(e) de \right]^{-1} \quad (\text{Eq. Price})$$

$$\underbrace{- (1 + \eta_{jjt})}_{\text{price elasticity}} \underbrace{\left[1 - F\left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}^*}\right) \right]}_{\Pr(\text{no excess demand})} \underbrace{\overbrace{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}^{\text{contracted quantities}}}_{\text{expected share of surviving supply in the event of a shortage}} \left[\underbrace{\int_0^{\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}^*}} e f(e) de}_{\mathbb{E}[\text{sales|shortage}]} \right]^{-1} = q_{jmt}^* \quad (\text{Eq. Supply})$$

where own-price elasticity $\eta_{jjt} \equiv 1 - \alpha p_{jmt}^* (1 - s_{jmt})$. s_{jmt} is the logit market share of firm j in drug market m .

I can rewrite this equilibrium function to get the classical price equals markup over marginal cost form, adjusted by shortage-weighted equilibrium supply as:

$$p_{jmt}^* = \underbrace{- \left(\frac{1}{1 + \eta_{jjt}^*} \right)}_{\text{markup over mc}} c_{jlt} \underbrace{\left(\frac{q_{jmt}^*}{Q_{jmt}^D(\mathbf{p}, \mathbf{X}) \left[1 - F\left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}^*}\right) \right]} \right)}_{\text{equilibrium supply adjusted by } 1/\Pr(\text{no shortage})} \quad (8)$$

where the markup here is variable and depends on firm size. Note that without the occurrence of exogenous random supply shocks in the economy, we would retrieve the traditional markup over marginal cost formula: the probability of shortage $F\left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}^*}\right)$ would be equal to zero, and the ratio of supplied to demanded quantities $\frac{q_{jmt}^*}{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}$ would be equal to one, thus cancelling the second part of the right-hand-side formula.

4.5. Social Welfare and the Regulator's Problem.

Social Welfare. I derived a measure of consumer welfare directly from the discrete choice logit model of demand ([andersonpalmathisse](#); [Capps, Dranove, and Satterthwaite, 2003](#)), using buyers' expected utility measure for each market m at time t , $EU_{mt}(\cdot)$:

$$\begin{aligned} EU_{mt}(\mathbf{p}_{mt}) &= M_{mt} \int \ln \left[1 + \sum_{jmt}^{J_m} \exp(v_{ijt}) \right] dF(\epsilon_{im}) \\ &= M_{mt} \int \ln \left[1 + \sum_{jmt}^{J_m} \exp(X_{jmt}\beta - \alpha p_{jmt} + \lambda g_{jm} + \phi_{mt}) \right] dF(\epsilon_{im}) \end{aligned}$$

Consumer welfare is given in the model by the sum of the expected utility produced by each drug available in market m , a measure of consumers' surplus. Based on a discrete choice model of demand for drug j that conditions on molecule markets and other observable (market size, number of producers, i.e., suppliers competition, price), we can thus predict ex-ante buyer's willingness-to-pay (WTP) for a given drug product (similar to [Capps, Dranove, and Satterthwaite \(2003\)](#)).

The Regulator's Problem. I define the regulator's payoff as the geometric mean of expected consumer welfare, measured using consumer surplus based on the logit expected utility measure, as in [Mermelstein et al. \(2020\)](#) and [Lim and Yurukoglu \(2018\)](#), and the manufacturer's value function (profit),⁴⁴ minus the cost of audit m_{lt} (plants inspections and certification):⁴⁵

$$u_R(\cdot) = \mathbb{E}[EU_{mt}(\cdot)]^\rho \mathbb{E}[\pi_{jmt}(\cdot)]^{1-\rho} - m_{lt}$$

where ρ is an exogenous weight parameter the regulator puts on consumer surplus (expected utility EU_{mt}) against manufacturers' profit $\pi_{jmt}(\cdot)$.

The regulator (the FDA) possess two margins of action. First, they choose the relative weight they put on consumer surplus versus drug manufacturer profits, ρ . Second, they choose their location-specific auditing intensity rate $\varphi_{lt} \in [0, 1]$, which comes at the monitoring cost of m_{lt} for a given location l . Higher weight on manufacturer profits can increase social welfare as it leads to stronger investment incentives for drug manufacturers, which in turn mitigates the supply disruption problem.

The regulator imperfectly observes suppliers' (drug manufacturers) manufacturing quality ([Taylor and Wiggins, 1997](#)). This requires inspecting the plant. I assume inspections reveal effort with certainty

⁴⁴The geometric mean specification is analogous to the Nash bargaining model in which players maximize the geometric mean of their utilities, while ensuring a non-negative value of the firm.

⁴⁵Monitoring costs are not negligible, especially for foreign-based facilities. For example, the FDA had estimated that the average cost of a foreign inspection was around \$62,500 for fiscal year 2009 and \$57,600 for fiscal year 2015, and the foreign inspection trips averaged three weeks in length. In contrast, the average estimated cost of a domestic inspection is only \$991. As a result, in FY19, the FDA planned 18,000 inspections, mostly targeted towards domestic plants. In FY 2021, the FDA's Office of Regulatory Affairs (ORA) spent a total of \$4,920,033 for domestic regulatory inspection against \$1,449,058 in FY 2020 on foreign inspection trips. The agency only inspected 6% of the 3000 overseas plants where drugs and their ingredients are produced in 2022, and only 3% of Indian manufacturers in 2022

and guarantees production quality.⁴⁶ However, inspection is costly, so that the regulator cannot inspect every period. The regulator's inspection frequency depends on the plant location l , and is defined as $\varphi_l \in [0, 1]$, with inspections of Foreign, South plants occurring less frequently than Foreign, North plants and less frequently than Home plants: $\varphi_{F_S} < \varphi_{F_N} < \varphi_H$. I set the cost of inspecting facilities located outside of the U.S. borders to be higher than the cost of inspecting at home, $m_{F_S} > m_{F_N} > m_H$ so that inspections outside of the U.S. happen with lower frequency. This captures administrative barriers for inspections abroad (necessary obtainment of visas for the FDA's inspectors, travel costs, regulatory compliance, etc.).

The regulator's problem thus reduces to two major choices: first choosing inspection frequency φ_l , which comes at audit cost m_l ; and second, choosing the potential weight to give to consumer's welfare versus firms' profits.⁴⁷

4.6. Predictions of the Model

I illustrate in this section the model's key dynamics by numerically simulating the model under a simple example. In Table 2, I consider two level of marginal costs; a "low" cost corresponding to a South production location, and a "high" cost, capturing production in the United States. Similarly, I consider two exogenous supply shock levels: a "good" shock corresponds to a high average capacity yield $\omega_{l(j)t}$ (70% of supply survives the disruption) and a "bad" shock corresponds to a low average capacity yield (70% of supply is destroyed by the shock).

For a given distribution of the shock (i.e. looking across columns for a given row in Table 2), we reach higher unit prices, higher shortage probabilities and higher shares of demand unmet by producing at Home than by offshoring to the South. The mechanisms are intuitive: lower marginal cost levels in the South means manufacturers may achieve higher markups than their U.S. counterparts, providing more incentives to build capacity buffers to assure against a shock. On the contrary, because production surplus is costlier for U.S.-based plants, they are more likely to practice "just-in-time" production. Similarly, for a given level of costs (i.e. looking at row variations for a given column in Table 2), worse shocks lead to higher levels of supply loss and hence higher market prices.

Table 2: Model Predictions: 4 scenarios

$c_{\text{low}} = 5$ ("Foreign, South")				$c_{\text{high}} = 10$ ("Home, North")			
p^*	Pr(Short)	Share Q^D unmet	$\mathbb{E}[\pi^*]$	p^*	Pr(Short)	Share Q^D unmet	$\mathbb{E}[\pi^*]$
"Good" $\mathbb{E}[\omega] = 0.7$	25.52	0.34	-0.07 (surplus)	1440.05	26.32	0.75	0.3
"Bad" $\mathbb{E}[\omega] = 0.3$	26.78	1.0	0.67	345.66	32.98	0.9	0.75

⁴⁶In practice, inspections are here to ensure that manufacturers follow the FDA's "Good Manufacturing Practices"

⁴⁷The regulator's weight ρ on consumer surplus is an exogenous measure that I do not observe directly. In counterfactuals, I will numerically solve the model under a variety of ρ parameter values to generate predictions of the model under different regulator's preference systems.

The main takeaways from this numerical example is that, in order for the model to match the observed patterns in the data (a higher frequency of shortages from South-East Asian plants), we need to observe *both* differences in marginal costs between the two locations *and* a different distribution of the supply shock. Under such setting, it could be optimal for drug manufacturers selling to the U.S. to operate from the South and let supply periodically fail in return for lower cost of production, enabling some markup. A central idea of the model is that the exogenous shock captures supply reliability in a given location. While U.S.-based firms may desire cutting on production costs further, they cannot afford to do so due to the stringent FDA enforcement of "Good Manufacturing Practices" (GMP) regulations within the U.S. borders; offshoring thus becomes a way for firms to target higher disruption probabilities than permissible in the U.S.⁴⁸ Essentially, the possibility of offshoring prompts firms to prefer foreign locations because the desired supply disruption level is unattainable domestically at prevailing price levels (or similarly, because firms cannot dis-invest enough in manufacturing process by reducing domestic production costs further (costs of facility maintenance, machines upgrades, or spare capacity holdings). The main purpose of the model estimation will then be to recover the distribution of the shock across locations.

5. Estimation of the Structural Model

In order to quantify the expected risk of disruption across locations and to capture the market trade-off between cost-cutting and resilience, I estimate the structural model presented in the previous section. I estimate twenty parameters separately, eighteen of which vary across production origin l , which I cluster into three groups of locations: "Home" (U.S.-based plants), "Foreign, North" (corresponding to OECD countries in the data) and "Foreign, South" (corresponding to South-East Asia, mostly India and China). I estimate these parameters following a four step procedure.

In the first step, I quantify the price-sensitivity of demand and buyers' taste for supply reliability by estimating the demand parameters from the discrete choice logit model. These parameters can be inferred directly from the data, using standard logit instruments to control for price endogeneity, separately from the rest of the system. In a second step, conditional on the first-stage demand parameters, I recover the distribution of exogenous supply shocks across locations using a Generalized Method of Moments (GMM) estimation procedure that matches observed shortage events with the model-predicted probability of shortage. In the third step, I make parametric assumptions about the form of marginal costs, trade costs and firms' productivity shocks (TFP) and use model inversions and a Simulated Method of Moments (SMM) approach to recover marginal costs, trade costs and the location-specific parameters of the distribution of the productivity shock. In the fourth and last step, I use the estimated parameters of demand, location-specific supply shocks and marginal costs to recover the location-specific distribution of fixed costs, using the decision rules from the location choice part of the model and a GMM approach.

⁴⁸Domestic facility inspections are more frequent in the U.S. than overseas (averaging every 2.5 years compared to once every 14 years outside the U.S.), and penalties may be enforced more rigorously. Consequently, equipment quality and facility upgrades typically meet higher standards domestically than in the South.

5.1. Sample Definition

Markets are defined at the molecule-form m and fiscal quarter level t , for the period January 2002 to December 2019. A unit of observation is a drug manufacturer (defined by a manufacturing plant), molecule-form product, fiscal quarter. A molecule-form market m is defined by a level 5 Anatomical Therapeutic Chemical (ATC-5) class. Consumer preferences for each drug within a homogeneous molecule market are defined according to a logit framework for differentiated products. Drug products j are defined at the FDA's National Drug Code (NDC) 5-4 level (which includes labeler in addition to molecule-dosage-form, not taking into account commercial package size). For the purpose of estimation, I compute for each drug product j the average unit price over package size. I use WAC prices (wholesale acquisition costs, the price paid by GPO to manufacturers to buy drug products).⁴⁹ Sale quantities in a quarter are from Medicare Part B. More details about the characteristics of drug markets, robustness to several definitions of markets, and summary statistics about the distribution of price, shortages and sales across markets and locations are reported in Appendix E.1.

Table 3: Parameters to Estimate

Model's step	Parameter description	Parameter	Estimation method
Demand estimation	Price parameter	α	Logit Model
Demand estimation	"Reputation" parameter	β	Logit Model
Demand estimation	Brand preference parameter	λ	Logit Model
Manufacturer's profit maximization	Distribution of supply shock	$[\mu_{\omega_H}, \sigma_{\omega_H}; \mu_{\omega_{F_S}}, \sigma_{\omega_{F_S}}; \mu_{\omega_{F_N}}, \sigma_{\omega_{F_N}}]$	GMM
Price/Marginal Cost Inversion	TFP distribution	$[\mu_{\gamma_H}, \sigma_{\gamma_H}; \mu_{\gamma_{F_S}}, \sigma_{\gamma_{F_S}}; \mu_{\gamma_{F_N}}, \sigma_{\gamma_{F_N}}]$	SMM
Manufacturer's location choice	Fixed Cost parameters	$[\mu_{FC_H}, \sigma_{FC_H}; \mu_{FC_{F_S}}, \sigma_{FC_{F_S}}; \mu_{FC_{F_N}}, \sigma_{FC_{F_N}}]$	GMM

5.2. First-Stage: Estimating Demand Parameters and Elasticities

The realization of exogenous supply shocks in the economy introduces a censoring problem as the *observed* market shares of firms operating in markets affected by shortages will be endogenously affected by stock-out events. As a consequence, one cannot simply estimate demand parameters using the traditional GMM logit procedure, which matches observed market shares to predicted logit shares. I describe in this section the estimation procedure I use to overcome challenges related to situations of excess demand in logit models.

Demand Specification. I can rewrite my fixed-coefficient logit specification (see Section 4.2) using demanded quantities to have:

$$\ln Q_{jmt}^D(\mathbf{p}, \mathbf{X}) - \ln Q_{0mt}^D(\mathbf{p}, \mathbf{X}) = \beta X_{jmt} - \alpha p_{jmt} + \lambda g_{jm} + \phi_{mt}$$

where m denotes molecule-form markets, t denotes fiscal quarter, g_{jm} are time-invariant molecule-product specific characteristics (brand/generic dummy, drug-labeler-specific characteristics), X_{jmt} is

⁴⁹Note that this does not include rebates, for which there are no data. However, rebates are generally marginal for generic markets, especially low-margin ones, compared to branded drug alternatives.

an exogenous reliability index measure (based on FDA's shortage records at time t), and ϕ_{mt} are molecule-quarter market fixed effects that control for any unobserved, market-specific time-variant factors. On the left-hand side, $Q_{jmt}^D(\mathbf{p}, \mathbf{X})$ are quantities sold of product j and $Q_{0mt}^D(\mathbf{p}, \mathbf{X})$ captures the outside option of not consuming. Market size is given by $M_{mt} = Q_{0mt}^D(\mathbf{p}, \mathbf{X}) + \sum_{j=1}^{J_m} Q_{jmt}^D(\mathbf{p}, \mathbf{X})$.

As neither the outside option $Q_{0mt}^D(\mathbf{p}, \mathbf{X})$ nor market sizes M_{mt} are directly observed, and there is no obvious definition of the *potential* molecule markets for sterile injectable drugs, I use the *difference across inside goods* to identify demand parameters. Under my fixed coefficient logit specification, the demand parameters α (price parameter), β (reliability parameter) and λ (brand preferences) are fixed across molecule markets⁵⁰ and I can use the functional form of the logit market share equations, *without* having to assume the actual size of the outside option or market size, and thus without using market shares. Differencing between two goods j and j' ,⁵¹

$$\ln Q_{jmt}^D(\mathbf{p}, \mathbf{X}) - \ln Q_{j'mt}^D(\mathbf{p}, \mathbf{X}) = \beta (X_{jmt} - X_{j'mt}) - \alpha (p_{jmt} - p_{j'mt}) + \lambda (g_{jm} - g_{j'm}) \quad (9)$$

where market-time fixed effects drop out and I can identify α, β, λ , by two-stage least squares with the usual instrumental variables to deal with price endogeneity. As quantities observed through quarterly sales are only equal to demanded quantities for firms not experiencing shortages in a given quarter t (there is no censoring issues for these firms, as $q_{jmt}^{\text{observed}} = Q_{jmt}^D(\mathbf{p}, \mathbf{X})$, see discussion below), I can estimate this equation on markets where at least two manufacturing firms are not short.⁵²

Conditions of Excess Demand and Censoring of Suppliers' Market Shares. The realization of exogenous random supply shocks introduce an endogeneity issue in observed market shares that goes beyond the usual limitations of using observed shares as an empirical counterpart of logit-predicted shares. Indeed, the observed drug quantities sold in the data are now directly affected by conditions of excess demand, which introduces a censoring issue:

$$q_{jt}^{\text{observed}} = \begin{cases} Q_{jt}^D(\mathbf{p}, \mathbf{X}) & \text{when } q_{jt}\omega_{l(j)t} \geq Q_{jj}^D(\mathbf{p}, \mathbf{X}) \quad [\text{no excess demand, realized supply above demand}] \\ q_{jt}\omega_{l(j)t} & \text{when } q_{jt}\omega_{l(j)t} < Q_{jt}^D(\mathbf{p}, \mathbf{X}) \quad [\text{excess demand}] \end{cases}$$

As a consequence, using observed firms' market shares to estimate the coefficients of the logit model

⁵⁰I am not using a random coefficients logit mode and I can thus recover the true demand parameters by two-stage least square as long as the number of observations is large enough. Fixed-coefficient logit is a reasonable assumption given all molecule-form markets are for U.S.-sold sterile injectable drug products

⁵¹In practice, I carry the differentiation with respect to one firm within each market, so that one firm is consistently used as the "outside option".

⁵²Note that without the unobserved logit heterogeneity term ξ_{jmt} , there is no selection issue here. See Appendix Note E.7 for additional discussions about estimating unobserved heterogeneity in situations of excess demand.

will under-estimate the shares of firm shorts. Within a given drug market m , observed shares are:

$$s_{jmt}^{\text{observed}} = \begin{cases} \frac{q_{jmt}\omega_{l(j)t}}{Q_{0t}^D(\cdot) + \sum_{k=1}^K q_{kt}\omega_{l(k)t} + \sum_{h=1}^H Q_{ht}^D(p, X)} = \frac{q_{jmt}\omega_{l(j)t}}{M_t} & (\text{i}) \quad \text{if } j \text{ short} \\ \frac{Q_{jmt}^D(p, X)}{Q_{0t}^D(\cdot) + \sum_{k=1}^K q_{kt}\omega_{l(k)t} + \sum_{h=1}^H Q_{ht}^D(p, X)} & (\text{ii}) \quad \text{if } j \text{ not short but } k \geq \text{firms short} \\ s_{jmt}^{\text{logit}} \equiv \frac{Q_{jmt}^D(p, X)}{M_t} & (\text{iii}) \quad \text{if no firms short (standard logit)} \end{cases}$$

where quantities for the outside option are

$$Q_{0t}^D(\cdot) = M_t - \underbrace{\sum_{k=1}^K q_{kt}\omega_{l(k)t}}_{\text{firm shorts}} - \underbrace{\sum_{h=1}^H Q_{ht}^D(p, X)}_{\text{firms not short}} = M_t - \sum_{jmt} q_{jmt}^{\text{observed}}$$

Conditional on correctly capturing the size of the outside option, situations (ii) and (iii) reduce to the traditional discrete choice logit market shares. Within a molecule-quarter market, I can use all firms not short in order to estimate a standard logit model, independently of the shortage status of other firms present in the market.⁵³ Note that under the scenario (i), the logit-predicted market shares will always be under-estimated, independently of whether we consider inner or outer shares.

Price Instruments. I use traditional cost-shifters to deal with the issue of endogeneity of price in the logit model, namely parameters of marginal cost: location-year specific wage in the pharmaceutical manufacturing sector and tariffs, and Ghandi-Houde (2019) instruments (price of products close in the characteristics space, i.e. price of drugs belonging to the same ATC4 therapeutic category but not same market (defined at the ATC5 or Molform level).⁵⁴ I do not restrict the instrument to be region specific (i.e. to belong to one of my three location groups, Home, Foreign South and Foreign North): location-specific wages and tariffs vary at the country and year levels, which allows to exploit more data variation.⁵⁵

Market size approximation and outside good. As is usual in logit demand models, one does need to take a stand on the potential market size in order to measure market shares. I use a nonlinear least squares calibration procedure similar to Dubois and Lasio (2018) and Dubois, Gandhi, and Vasserman (2022) in the IO of pharmaceutical markets literature to calibrate the potential quarterly market size

⁵³Note that if on the contrary we only consider the inner market shares (*without* the outside option), then in case (ii) the shares of firms not being short in markets where some firms are short will be over-estimated too.

⁵⁴Note that playing with different levels of therapeutic classes - i.e. using ATC3 or ATC2 level - do not seem to significantly impact the regression coefficients

⁵⁵Note that this does add computational complexity as I do not have any combinatorial problem: in my data, a manufacturing firm is defined as a plant and I do not assume pharmaceutical firms have to choose how many plants and locations they want to operate.

M_{mt} .⁵⁶ This method relies on a two-step procedure. In a first stage, I use the estimates of $\hat{\alpha}, \hat{\beta}, \hat{\lambda}$ recovered using the differencing technique described above. This allows me to compute logit-predicted market shares for all firms, independently of their shortage status. In a second stage, for each molecule market m , I estimate M_t that solves the following minimization problem:

$$\min_{M_{mt} \geq \sum_{j=1}^{J_m} Q_{jmt}^D(\mathbf{p}, \mathbf{X})} \sum_{t=1}^T \left([\hat{\alpha}(M_{mt}) - \hat{\alpha}]^2 + [\hat{\beta}(M_{mt}) - \hat{\beta}]^2 + [\hat{\lambda}(M_{mt}) - \hat{\lambda}]^2 \right)$$

where $\hat{\alpha}(M_{mt}), \hat{\beta}(M_{mt})$ and $\hat{\lambda}(M_{mt})$ are the two-stage least squares coefficient estimates of the following equation, using the same instruments as before:

$$\ln s_{jtm} - \ln \left(\frac{M_{mt} - \sum_{j=1}^{J_m} Q_{jt}^D(\mathbf{p}, \mathbf{X})}{M_{mt}} \right) = \beta X_{jmt} - \alpha p_{jmt} + \lambda g_{jm} + \phi_{mt}$$

This method estimates the market size as the solution to the maximization of the *fit* of the model that does not specify the outside option, but uses instead market shares relative to one inside good.⁵⁷ This method provides a good approximation of the market size, which avoids making more ad hoc assumptions. In a market where total sales grow over time, it is indeed particularly difficult to guess the potential market size, as one ignores whether it is stable over time or grows and at what rate.⁵⁸

Based on the estimated demand parameters α, β and λ , I can then estimate for each firm, short or not, demanded quantities in each molecule-quarter market, as $\hat{Q}_{jmt}^D(\mathbf{p}, \mathbf{X}) = s_{jmt} M_{mt}$. I can also recover each firm's own-price elasticity $\eta_{jjt} = -\alpha p_{jmt}(1 - s_{jmt})$ and cross-price elasticities with another firm k in the same market m : $\eta_{jkt} = \alpha p_{kmt}s_{kmt}$, where s_{jmt} is firm j 's logit market share.

5.3. Second-Stage: Estimating Supply Parameters

Conditional on the first-stage demand parameters, I estimate three sets of location-specific supply parameters. I proceed by decomposing this estimation into three consecutive sub-steps.

First, I recover parameters of the distribution of the exogenous supply shocks, which I assume follow a location-specific logit normal distribution $\omega_{l(j)t} \sim \text{LogitNorm}(\mu_{\omega_l}, \sigma_{\omega_l})$, by using a GMM procedure that matches observed shortage events with the model-predicted probability of shortage.

Second, I estimate marginal costs parameters, and in particular parameters of the distribution of a firm's total factor productivity, $\gamma_{l(j)t}$. Conditional on already estimated demand and supply shock

⁵⁶On average, [Dubois, Gandhi, and Vasserman \(2022\)](#) find that the estimated outside market share across ATC4 categories was 27.9% in Canada and 22.8% in the US. The original procedure is from [Huang and Rojas \(2014\)](#).

⁵⁷The generally accepted procedure is to assume a “market potential” that implicitly defines the size of the outside good; in practice, this means that an endogenous quantity is approximated by a reasonable guess thereby introducing the possibility of an additional source of error and, most importantly, bias.

⁵⁸As an alternative, I could use panel data techniques to produce unbiased structural estimates by treating the market potential as a fixed effect (known as a *correlated random effect* in the non-linear panel data literature). In particular, [Huang and Rojas \(2014\)](#) explores three possible solutions: a) controlling for the unobservable with market FEs, b) specifying the unobservable to be a linear function of average product characteristics, and c) a *demeaned* regression approach. They show that solution a) is feasible (and preferable) when the number of goods is large relative to the number of markets, whereas b) and c) are attractive when the number of markets is large. All three solutions are as effective in removing the bias.

parameters, I first use a model inversion to recover each firm's marginal cost. I then rely on a reduced-form regression of marginal costs on tariffs and distance to recover trade costs. I recover each firm's TFP by inverting the marginal cost function, and estimate the location-specific parameters of its distribution using a Simulated Method of Moments (SMM) approach.

The third and last part of the estimation consists in estimating the parameters of the distribution of location-specific fixed production costs. For that purpose, I make parametric assumptions about the form of the fixed cost distribution and use the model to simulate suppliers' location decisions in order to match data moments related to the share of firms located in each region.

5.3.1. Supply Step 1: Recovering the Distribution of Exogenous Supply Shocks

Based on first-stage estimates of demand parameters $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\lambda}$ and own and cross-price elasticities η_{jjt}, η_{jkt} , I start by estimating the parameters of the distribution of the exogenous supply shock $\omega_{l(j)t}$.

I recover six location-specific parameters, $\theta_\omega \equiv \{\mu_{\omega_H}, \mu_{\omega_{F_N}}, \mu_{\omega_{F_S}}, \sigma_{\omega_H}, \sigma_{\omega_{F_N}}, \sigma_{\omega_{F_S}}\}$, using a GMM estimation procedure that matches the model-predicted probability of shortage $F\left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}}\right)$ for firm j operating in drug market m with its empirical counterpart, the observed fraction of time short for each firm j in each quarter t , $\mathbb{E}\left[\mathbb{1}\{q_{jmt}\omega_{l(j)t} < Q_{jmt}^D(\mathbf{p}, \mathbf{X})\}\right] \equiv \rho_{jmt}$.

Moment Conditions. To identify the location-specific mean of the shock, μ_{ω_l} where $l = \{F_N, F_S, H\}$, I rely on moment conditions equating the *mean* probability of shortage across locations in the data and in the model:

$$\mathbb{E}\left[\left(\rho_{jmt} - F\left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}}; \mu_{\omega_l}, \sigma_{\omega_l}\right)\right) \times \mathbb{1}\{\text{location } l = \{F_N, F_S, H\}\}\right] = 0 \quad (\text{Moments Shock})$$

I use similar, higher-order moments (standard deviation and skewness of shortage frequency), to recover the variance of the shock across locations, σ_{ω_l} .

Nested Fixed Point Algorithm. Note that the theoretical probability of shortage $F\left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}}\right)$ depends both on the location-specific distribution of the exogenous supply shock ($F(\cdot)$ is the CDF of the distribution), on demanded quantities $Q_{jmt}^D(\mathbf{p}, \mathbf{X})$ (recovered for all firms from the demand estimation step) and on supplied quantities q_{jmt} . Though I never observe these quantities directly in the data,⁵⁹ I can use the model predictions for equilibrium produced quantities q_{jmt}^* (see Eq. Supply) to compute

⁵⁹Recall that due to the occurrence of supply shocks and firms' ability to build buffers to insure against shocks, I observe $q_{jt}^{observed} = \min\{Q_{jt}^D(\mathbf{p}, \mathbf{X}), q_{jt}\omega_{l(j)}\} = Q_{jt}^D(\mathbf{p}, \mathbf{X}) - \underbrace{\text{Short}_t(Q_{jt}^D(\mathbf{p}, \mathbf{X}) - q_{jt}\omega_{l(j)})}_{\text{Excess demand}}$ where $\text{Short}_t \equiv \mathbb{1}\{q_{jt}\omega_{l(j)} < Q_{jt}^D(\mathbf{p}, \mathbf{X})\}$

the theoretical probability that a firm experiences a stock-out event:

$$\frac{\overbrace{1 - F\left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}^*}\right)}^{\text{Pr(no excess demand)}} \overbrace{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}^{\text{contracted quantities}}}{\underbrace{(1 + \eta_{jlt})}_{\text{price elasticity}} \underbrace{\int_0^{\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}^*}} \epsilon f(\epsilon) d\epsilon}_{\text{expected share of surviving capacity in the event of a shortage}}} = q_{jmt}^* \quad (\text{Eq. Supply})$$

However, model-predicted supply quantities depend on the distribution of the shock, which in turns depend on supply quantities themselves. In order to estimate the parameters of the distribution of the shock, I thus incorporate a nested-fixed point algorithm within the GMM loop, using the equilibrium supply formula derived from the first-order condition of the firm's problem in (Eq. Supply) to iterate on the value of q_{jmt} and $\{\mu_{\omega_l}, \sigma_{\omega_l}\}$ until convergence of the GMM objective (see details in Appendix E.5). Upon convergence of the GMM algorithm, I recover each firm j 's production choice, conditional on the estimated shock parameters $\hat{q}_{jmt}(\hat{\mu}_{\omega_l}, \hat{\sigma}_{\omega_l})$.

This step therefore allows to estimate both the location-specific supply shock distributions, firms' equilibrium production choices q_{jmt} given the distribution of the shock, and firms' buffer, measured as the difference between firms' demanded quantities and supplied quantities:

$$\text{Buffer}_{jmt} \equiv \hat{q}_{jmt}(\hat{\mu}_{\omega_l}, \hat{\sigma}_{\omega_l}) - \hat{Q}_{jmt}^D(\mathbf{p}, \mathbf{X}) \quad (\text{Eq. Buffer})$$

Above all, it allows us to estimate the *expected* amount of excess demand within a market, measured as the expected difference between demand and *realized* supply, conditional thus on the estimated distribution of the supply shocks:

$$\hat{Q}_{jmt}^D(\mathbf{p}, \mathbf{X}) - \mathbb{E}[\omega_{l(j)t}] \hat{q}_{jmt}(\hat{\mu}_{\omega_l}, \hat{\sigma}_{\omega_l}) \quad (\text{Expected Excess})$$

5.3.2. Supply Step 2: Recovering Marginal Costs, Trade Costs and Firms' TFP

In this second sub-step, I recover firms' marginal costs, trade costs and firms' productivity shocks (TFP) sequentially, conditional on previously estimated demand and supply shock parameters.

Marginal Costs. I first recover the model-predicted marginal cost for each firm j operating in a given drug market m and fiscal quarter t by using a model inversion of the equilibrium price function. Given the estimated demand and supply shock parameters, I derived in previous steps the model-predicted demanded quantities, supplied quantities and probability of shortages. I then use observed data on unit prices to recover each firm's marginal cost by inverting the optimal pricing condition derived from the first-order conditions of the manufacturer's problem in Eq. Price:

$$\hat{c}_{l(j)t} = p_{jmt}^* \int_0^{\frac{\hat{Q}_{jmt}^D(\mathbf{p}, \mathbf{X})}{\hat{q}_{jmt}(\hat{\mu}_{\omega_l}, \hat{\sigma}_{\omega_l})}} e f(e) de$$

where $f(\cdot)$ is the probability density function of the exogenous supply shock $\omega_{l(j)t}$.

Trade Costs. I then rely on a reduced-form regression of marginal costs on tariffs and distances to recover trade costs. I first parametrize iceberg trade costs as a function of location-year specific tariffs in the pharmaceutical sector and haversine distance between a drug's production location l and unique destination (the U.S.):

$$\tau_{lt} = (1 + \text{tariff}_{lt})^{\phi^{\text{tariff}}} \text{dist}_l^{\phi^{\text{dist}}} \quad (10)$$

In order to identify the parameters ϕ^{tariff} and ϕ^{dist} , I then estimate the following log-linear equation:

$$\ln(c_{l(j)mt}) = \phi_0 + \phi^{\text{tariff}} \ln(1 + \text{tariff}_{lt}) + \phi^{\text{dist}} \ln \text{dist}_l + \delta_{m,l(j)} + \zeta_{l(j)t}^m$$

where $c_{l(j)mt}$ are firms marginal costs previously recovered from the price inversion. My dataset allows me to add a market-location fixed effect $\delta_{m,l(j)}$, which controls for unobserved determinant of marginal costs that could be market and origin-location specific. Through the inclusion of these fixed effects, I can control for unobserved production costs that could potentially bias the estimates. Since recovered $c_{l(j)mt}$ depend on the estimated shock distribution, demand parameters and observed average wholesale prices (AWP) prices, I interpret the error term $\zeta_{l(j)t}^m$ as a measurement error that is uncorrelated with the right-hand side variables.

Total Factor Productivity. I use these estimated firm-specific marginal costs and location-specific trade costs, along with the structural form of marginal costs $c_{jlt} = \frac{w_{lt}\tau_{lt}}{\gamma_{l(j)t}(\varphi_l)}$ to recover each firm's predicted TFP. Here, w_{lt} (location-specific average wage in the pharmaceutical manufacturing sector), τ_{lt} (iceberg-trade costs) and φ_l (average location-specific inspection frequency) are directly recovered from the data.⁶⁰ By simply inverting this marginal cost function, I can thus recover each firm j 's TFP as $\hat{\gamma}_{l(j)t} = \frac{w_{lt}\tau_{lt}}{\hat{c}_{jlt}(\varphi_l)}$.

In order to later compute firms' expected profits and marginal costs in counterfactual locations, I then need in a last step to estimate the location-specific distribution of firm's productivity. For this purpose, I assume that firms' productivity draws $\gamma_{l(j)t}$ are location dependent, and all plants in the same location receive random i.i.d draws from the same log normal distribution with mean μ_{γ_l} and variance σ_{γ_l} . I then use the recovered $\hat{\gamma}_{l(j)t}$ for each firm from the precedent step and a Simulated Method of Moments (SMM) approach to estimate the parameters of the distribution of the TFP shock such that the model-simulated TFP matches the recovered $\hat{\gamma}_{l(j)t}$ (see details in Appendix 2).

5.3.3. Step 3: Estimating the Distribution of Fixed Costs Across Locations

In this last step, I use the estimated parameters of demand, location-specific supply shocks and marginal costs to recover the location-specific distribution of fixed costs. To identify the parameters of the fixed cost distribution, I use the decisions rules from the model to simulate suppliers location decisions, and rely on a GMM procedure in order to match data moments capturing the share of firms located in each region.

Location Choices. Location choices are made before supply and pricing decisions are made, and before supply shock realizes. Drug manufacturer j chooses the plant location l that maximize their

⁶⁰Inspection frequencies are calibrated to follow the data to 1/2.5 for the United States (Home), 1/5 for the OECD (Foreign, North) and 1/14 for South-East Asia (Foreign, South).

ex-ante *expected* profit across locations (see Eq. Location). This profit function integrates several key components that vary at the location-level: marginal costs, supply shock distribution and fixed costs, the distribution of the first two having been estimated in the previous section. I parametrize fixed costs to follow a location-specific log-normal distribution $\ln(FC_{jlt}) \sim N(\mu_{FC_l}, \sigma_{FC_l}^2)$ with firm-level logit shocks, the last of which provides stochastic variations in profit. The logit shock is assumed to be i.i.d across firms and locations and is distributed Extreme Value Type I.

Let $\mathbb{E}\pi_{jmt}^l$ be the expected profit function of firm j operating in molecule market m at time t and producing from location l , before supply shocks are realized. Without loss of generality, we can re-express the firm's expected profit function from (Eq. Profit) in general term as

$$\mathbb{E}\pi_{jmt}^l = \mathbb{E}R_{jlt} - MC_{jlt} - FC_{jlt} + \epsilon_{jlt}$$

where $\mathbb{E}R_{jmt}$ is ex-ante expected revenue in location l (before shocks are realized), FC_{jlt} are fixed costs in location l , MC_{jlt} are marginal costs and ϵ_{jlt} is the logit shock of firm j in location l .

Estimation Approach. Given the logit assumption for the shock, the choice probability of firm j choosing location l is

$$P_{jlt} = \frac{\exp(\mathbb{E}\pi_{jmt}^l)}{\sum_k \exp(\mathbb{E}\pi_{jmt}^k)}$$

The share of firms selecting a specific location l can be thought of as an aggregate of these individual firm decisions. For a sufficiently large number of firms, the observed share S_{lt} of firms selecting location l converges to the expected value of the choice probability, i.e. $\hat{S}_{lt} \approx \mathbb{E}[P_{lt}]$.

Identification relies on two pivotal moments. As the mean μ_{FC_l} of the log-normal fixed cost distribution changes, it will shift the entire distribution of fixed costs. This will directly influence the share S_{lt} of firms selecting each location. To identify μ_{FC_l} , I thus equate the model-predicted share of firms choosing each location to their observed empirical share (see Equation 11).

The variance $\sigma_{FC_l}^2$ of the log-normal distribution determines how spread out the fixed costs are across locations. A higher variance means there's a wider range of fixed costs, implying greater heterogeneity among locations. To identify the variance, I leverage similar higher-order moments (differences between variance of the observed shares and the variance of the model's predicted shares, as well as skewness and kurtosis of the distribution of shares across locations since the log-normal distribution is skewed). Moment conditions used to identify the mean and variance of fixed cost distribution are then:

$$\begin{aligned} \mathbb{E}[S_{lt} - \mathbb{E}[P_{jlt}]] &= 0 \\ \mathbb{V}[S_{lt}] - \mathbb{V}[\mathbb{E}[P_{jlt}]] &= 0 \end{aligned} \tag{11}$$

These moment conditions are used in a GMM framework to jointly estimate μ_{FC_l} and $\sigma_{FC_l}^2$. Given I previously estimated all other model parameters varying at the location level, I can then simulate the expected share of firms choosing each location, and use the discrepancies between these simulated shares and the observed shares to identify the parameters of the fixed costs.

5.4. Estimation Results

Table 4 displays parameter estimates from the estimation of the logit demand model, while Appendix Table E4 reports the full range of estimated supply and demand coefficients.

The estimated model allows to go beyond the qualitative evidence on offshoring and quantify the market-wide trade-off between resilience and cost-cutting. In Table 4, I find that, consistently with anecdotal evidence, the shortage history of drug manufacturers does not significantly affect their future market shares. Buyers display high price sensitivity in choosing suppliers for the drug products they source, but low tastes for past reliability of these suppliers (Table 4). Higher valuation of a supplier's shortage history while drafting procurement contract is nevertheless correlated with lower probability of future shortages, lower shares of demand unmet and higher equilibrium prices.

Table 4: Demand Estimates (Beta Coefficients)

Dep. var: logit market shares	OLS			IV		
	(1)	(2)	(3)	(4)	(5)	(6)
Price diff	-0.012** [0.00481]	-0.01193** [0.00481]	-0.01188** [0.00481]	-0.6111** [0.255]	-0.6106** [0.2552]	-0.6101** [0.2551]
Reliability Score (0-1) diff		-0.00865 [0.00681]			-0.00418 [0.00744]	
Reliability Score, categorical diff			-0.01449** [0.0066]			-0.00964 [0.00725]
Brand Dummy		-0.0252*** [0.00096]	-0.0201*** [0.0013]		-0.01293*** [0.00035]	-0.01774 ** [0.0054]
Observations	69,816	69,816	69,816	66,931	66,931	66,931
F-stat	56.229	53.864	55.419	55.743	53.616	55.086
Drug Market Fixed Effects	✓	✓	✓	✓	✓	✓
Instruments :	✗	✗	✗	Location-specific wages & tariffs		

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All coefficients are beta coefficients (standardized) Standard errors in parenthesis are clustered at the market level. A market is defined at the molecule-form-year-quarter level. Balanced Panel of Sterile Injectable Drugs, from January 2002 to December 2019. All regressions include drug market fixed effects. The reliability measure "Reliability Score (0-1)" is computed as the fraction of time not short during the previous quarter, at the NDC 5-4 level. The alternative measure "Reliability Score, categorical" is defined as a three-levels categorical variable. Level 3 corresponds to no shortages in the previous quarter. Level 1 corresponds to less than 15% of time not short in the previous quarter (i.e. the firm is mostly short). Level 2 corresponds to anything in between.

Estimates of the supply parameters are reported in Appendix Table E4. The average U.S.-based manufacturing plant tends to build less buffer than their Foreign, South counterparts. Indeed, higher fixed and marginal costs levels make surplus more costly for U.S.-based firms on average, favoring just-in-time production (a well-known characteristics of the U.S. generic drugs market). I estimated much lower average capacity yield (fraction of supply surviving a disruption) for plants offshored to the South, suggesting that the average amplitude of shocks in Foreign, South plants is much larger than for the U.S. (Appendix Table E4).

6. Counterfactuals and policy applications

6.1. Objectives and Summary

Having quantified the baseline model, I study the welfare consequences of regulations aiming to correct market failures in generic drugs procurement and evaluate the impact of counterfactual policy designs. My analysis articulates around two pivotal policy avenues.

First, a policy assuming the frictions introduced by the current buyer-seller contracts cannot be significantly dampened, and thus focusing on switching back to the previous organizational structure, by re-shoring production to U.S. territories. Policies focused on shortening supply chains and bringing back production of essential manufacturing sectors home have been widely discussed since 2017, both in the U.S. and Europe [Washington Post, January 2023; The Economist Report, 2021].

Second, a set of two policies focused on solving market failures introduced by current procurement contracts inside the new globalized market. The first policy focuses on alleviating short-term price rigidities by authorizing the unit price of a product to increase in the amount of supply shortfall. The assumption is that authorizing short-term price variations after the occurrence of a shortage would allow the market to price capacity and help incentivizing production investments post-shock. The second policy focuses on alleviating barriers to contract on production reliability by enforcing failures-to-supply clauses, thus effectively moving the market closer to "Japanese-Type" contracts. The assumption is that such policy would prompt market players to internalize shortage costs on patients and would provide higher margins for reliable suppliers, allowing them to invest back in facilities maintenance and high-quality production processes.

Throughout these analyses, my outcomes of interest are the equilibrium probability of shortages and share of excess demand in the market, drug prices, consumer welfare, and firm profits.

6.2. Counterfactual 1: Subsidies to Re-Shoring

In this counterfactual scenario, I evaluate the impact of a policy designed to influence sourcing decisions by subsidizing re-shoring to U.S. territories, re-establishing the tax breaks that existed pre-2006 in the United States.

Context of the Proposed Reshoring Policy. From 1976 to 2006, the Internal Revenue Code exempted from taxation corporate income generated in U.S. territories. The tax break would have saved drug companies the 2022 equivalent of \$235 million per year, or 70 cents per American (Washington Post, January 2023), and its ending coincided with the start of offshoring of drug plants to the South.

Post-2020, U.S. and European governments started a push towards tax incentives that encourage reshoring (see details in Appendix F.1). In the U.S., key legislative proposals, such as the *Domestic Medical and Drug Manufacturing Tax Credit* would provide manufacturers with a tax credit of 10.5% of the net income from the sale of critical medical products produced domestically. If passed, this would ef-

fectively halve their corporate tax. Concurrently, President Joe Biden's 2021 *American Jobs Plan* pledges an investment of \$300 billion towards U.S. manufacturing, which includes a substantial allocation for reinforcing supply chains and ensuring pandemic readiness by reshoring Active Pharmaceutical Ingredients (APIs).⁶¹ In France, President Emmanuel Macron, publicly announced in June 2023 that France would reshore the production of 50 essential medicines, backed by significant public funding.

It is crucial to acknowledge that reshoring production of pharmaceuticals do not come free of frictions. The process of relocating a single product can span up to 2 years, inclusive of all logistical, regulatory, and administrative tasks. Establishing a new facility from the ground up can demand an investment upwards of \$2 billion and a timeline of 5 years. An added layer of complexity arises from the potential unforeseen consequences of reshoring. For instance, if companies find it more cost-effective to maintain larger buffers abroad, then compelling them to re-shore might inadvertently amplify supply disruptions.

Reshoring: Counterfactual Framework. Let s_{reshore} denote a subsidy payable to any firm that reshores production to U.S. territories. Assuming three locations, $l = \{H, F_N, F_S\}$, denoting Home, Foreign North, Foreign South respectively, supplier j 's location problem (Eq. Location) becomes:

$$l^* = \arg \max_{\{\pi_{jmt}^l\}} \left\{ \pi_{jmt}^H \left(c_{jt}^H, FC_{jmt}^H, \omega_{H(j)t}; q_{jmt}^*, p_{jmt}^* \right), \pi_{jmt}^{F_N} \left(c_{jt}^{F_N}, FC_{jmt}^{F_N}, \omega_{F_N(j)t}; q_{jmt}^*, p_{jmt}^* \right), \pi_{jmt}^{F_S} \left(c_{jt}^{F_S}, FC_{jmt}^{F_S}, \omega_{F_S(j)t}; q_{jmt}^*, p_{jmt}^* \right) \right\}.$$

where, for each drug market m

$$\mathbb{E}_{\omega_l} [\pi_{jlt}^l (\cdot)] = \overbrace{\left[1 - F \left(\frac{Q_{jt}^D(\mathbf{p}, \mathbf{X})}{q_{jt}} \right) \right] p_{jt} Q_{jt}^D(\mathbf{p}, \mathbf{X}) + F \left(\frac{Q_{jt}^D(\mathbf{p}, \mathbf{X})}{q_{jt}} \right) p_{jt} q_{jt} \mathbb{E} \left[\omega_{l(j)} \mathbb{1} \left\{ \omega_{l(j)} < \frac{Q_{jt}^D(\mathbf{p}, \mathbf{X})}{q_{jt}} \right\} \right]}^{\text{No excess demand}} - c_{jlt} (\varphi_l) q_{jt} - F_{jt}^c + s_{\text{reshore}} \mathbb{1}\{l = H\} \quad (12)$$

This equation modifies the baseline expected manufacturer's profit (Eq. Eq. Profit) by adding a subsidy to firms choosing to locate their plans at "Home" ($l = H$), $s_{\text{reshore}} \mathbb{1}\{l = H\}$, thus making the option of reshoring to the U.S. less costly for drug manufacturers. I define this subsidy as a percentage of the average expected firm's revenue within a market-year.

This subsidy adds an extra cost for the Regulator, which is proportional to the tax break amount and the number of firms choosing to re-shore:

$$u_R(.) = \mathbb{E}[EU(.)]^{\rho} \mathbb{E} [\pi_{jmt}(.)]^{1-\rho} - m_{lt} - \sum_{j=1}^{J_H} s_{\text{reshore}} \mathbb{1}\{l = H\}$$

Conterfactual Results. Figure 7 depicts the changes in equilibrium outcomes that follow the simulation of this counterfactual scenario under previously estimated parameter values. The y-axis displays changes in equilibrium outcomes (as prices, share of excess demand or consumer surplus), measured in percentage deviation from the current baseline measures. The x-axis displays changes in the rate of the reshoring tax break. I vary the subsidy between 0% (baseline) and 20% of the average expected firm's revenue within a market-year, the suggested U.S. tax break being around 10%.

⁶¹The White House's June 2021 100-day review of America's Supply Chains: "Building Resilient Supply Chains, Revitalizing American Manufacturing and Fostering Broad-Based Growth"

Subsidizing reshoring unequivocally results in an increase in the unit price of drug products. In panel 7a, the suggested 10% tax break translates into a 30% increase in prices over injectable markets. A 15 to 20% tax break could lead to up to a 50% increase in price across injectable markets.

Compared to the current situation, a reshoring tax break does achieve a reduction in the equilibrium probability that shortages occur: by 5% for a 10% tax break on firms' revenues, to 10% by increasing the tax break to 20% of revenues (Panel 7b). However, the economy-wide *expected* share of excess demand increases compared to the baseline levels: a 10% tax break on revenues increases the expected share of missing supply by 12% (Panel 7a). While the expected share of excess demand in the economy decreases slightly with higher levels of subsidies, a 20% tax rebate level for reshoring still corresponds to a 1% higher expected share of excess demand.

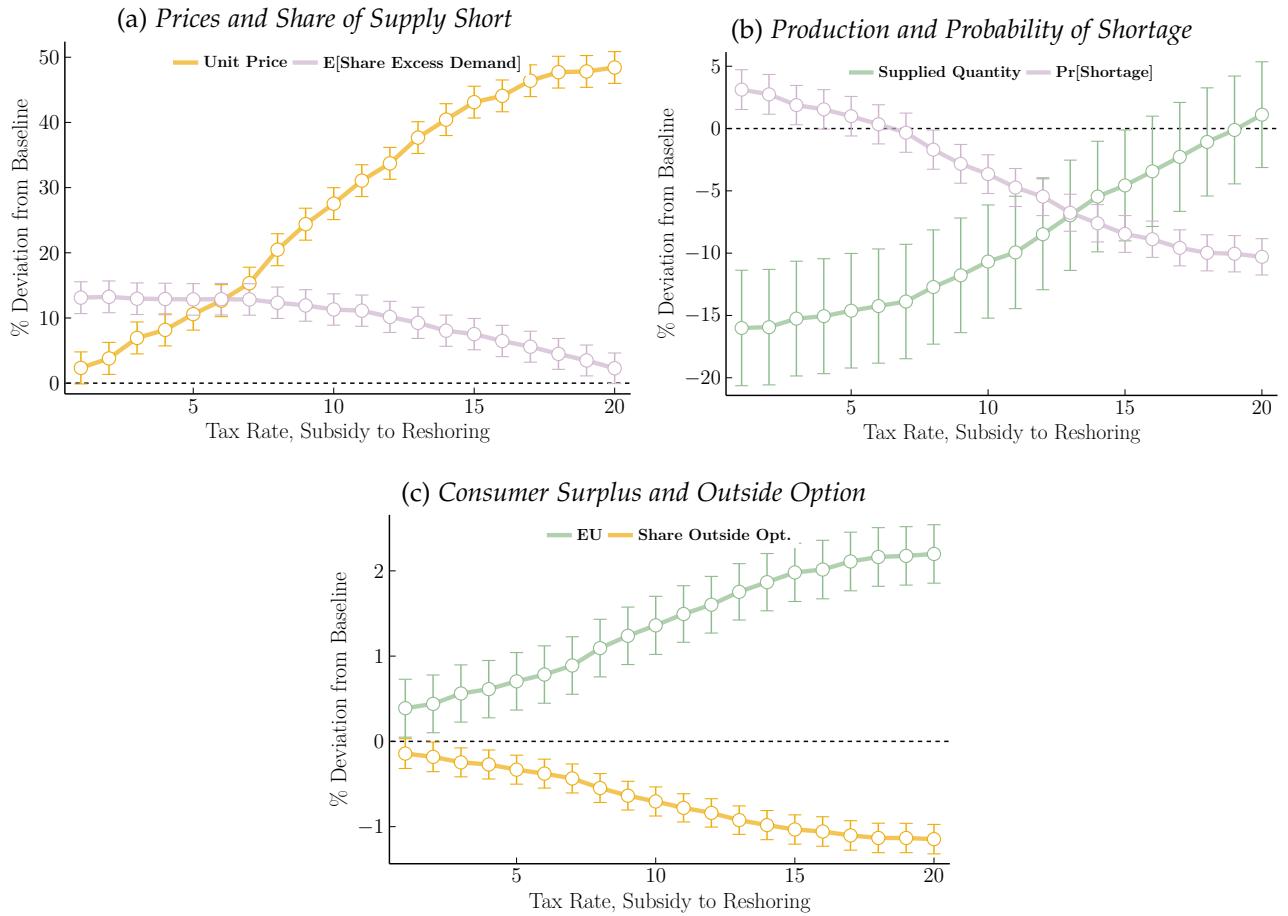


Figure 7: Changes in Equilibrium Outcomes as a Function of Reshoring Subsidy Levels

Notes: The y-axis displays changes in equilibrium outcomes (as prices, share of excess demand or consumer surplus), measured in percentage deviation from the current baseline measures. The x-axis displays changes in the rate of the reshoring tax break (the baseline is 0). The suggested U.S. tax break being around 10%. In panel 7a, both the average drug unit price and the expected share of excess demand are above baseline. Unit prices increase with the subsidy amounts while expected excess demand decreases. Panel 7b shows that both the average supplied quantity and the probability of shortages decrease compared to the baseline (around 10% lower supply and a 5% lower probability of shortage at the 10% tax break level). In Panel 7c, consumer surplus, measured using Expected Utility (EU) from the Logit, increases by 1-2% with a reshoring subsidy, while the share of patients choosing to not consume decreases slightly. More results can be found in Appendix F1.

While somewhat counter-intuitive, the mechanisms behind this result are simple: on Panel 7b, sub-

sidizing reshoring actually results in a *decrease* in average supplied quantities (around 12% at the 10% tax rebate), as U.S.-based manufacturing plants tend to build less buffer than their foreign counterpart. Indeed, higher fixed and marginal cost levels make surplus more costly on average (see estimated average costs from 5.4), favoring “just-in-time” production. Though disruptions become less likely on average on Panel 7b, small shocks are expected to result in immediate unfulfilled demand due to the absence of market-wide buffers.

At a 10% tax break level, we may expect up to a 50% relocation in the share of firms located in the U.S. compared to today (see Appendix Figure F1). The impact of such a policy on equilibrium shortage levels would thus depend on the actual *realization* of supply shocks, and may need to go in hands with policies incentivizing buffers or redundancies to effectively increase products availability.

Overall, reshoring results in a slight increase in consumer surplus, as measured by expected utility (1 to 2% compared to today), which mostly results from decreases in the share of patients choosing the outside option of not consuming (Panel 7).

6.3. Counterfactual 2: Pricing Market Capacity

I now relax the short-run price rigidities constraint introduced by procurement contracts and study how firms’ supply decisions and equilibrium excess demand measures are influenced by post-shock price variations. Specifically, I modify the baseline supply model (4.4.1) to permit unit prices to be indexed on supply conditions and increase proportionally with the magnitude of supply scarcity.

Price-Adjustment During Shortages: Counterfactual Framework. I assume that firm j' price variation at time t depends on *molecule-market excess demand*. Under this scenario, each supplier’s price p_{jmt} becomes a stochastic variable, whose value depends on the realization of supply.⁶²

I start by defining the market-level expected excess demand measure used in this counterfactual scenario before detailing the implementation of the new pricing scheme.

Definition 6.1 (Market-level Expected Excess Demand). At the market m level, the expected total amount of excess demand is defined as

$$\sum_{j=1}^{J_m} \max \left\{ Q_{jmt}^D(\mathbf{p}, \mathbf{X}) - q_{jmt} \mathbb{E} \left[\omega_{l(j)t} | \omega_{l(j)t} < \frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}} \right], 0 \right\}$$

where $Q_{jmt}^D(\mathbf{p}, \mathbf{X})$ is demanded quantities for firm j ’s product, q_{jmt} are quantities actually produced by firm j and $\mathbb{E} \left[\omega_{l(j)t} | \omega_{l(j)t} < \frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}} \right]$ measures the expected share of supply that survives the shock, conditional on a shortage occurring. This identity quantifies the expected missing supplied quantities to cover demand at the

⁶²The implementation of a spot market allowing to price for capacities was one of the solution implemented for the electricity market. While the delays necessary for capacity adjustment are greater for injectable drugs than for electricity generation, so that a "spot market" for drugs is not realizable, the underlying idea of pricing capacity is similar.

market level. The total expected share of excess demand is defined as:

$$\sum_{j=1}^{Jm} \max \left\{ \frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X}) - q_{jmt} \mathbb{E} \left[\omega_{l(j)t} | \omega_{l(j)t} < \frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}} \right]}{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}, 0 \right\} \equiv \sum_{j=1}^{Jm} \max \left\{ \delta_{jmt}^{\text{excess}}, 0 \right\} \quad (13)$$

The ensuing counterfactual pricing scheme relies on the premise that only supply that contributes to reliability is rewarded. If no firm is expected to be short and there is ample market capacity in period t , firm j 's price is fixed to previously contracted market price p_{jmt} (Equation 14, first case). The underlying assumption is that absent of excess demand in the economy, unused capacities bring no economic benefits (the model assume surplus disposal). On the contrary, under situation of excess demand (Definition 6.1), which arises whenever $k \geq 1$ firms are short in the market (which may include j itself), firm j 's price increases proportionally to the expected share of missing supply in the market:⁶³

$$\tilde{p}_{jmt} = \begin{cases} p_{jmt} & \text{if no firms short} \\ p_{jmt} + \gamma^{\text{excess}} \sum_{k=1}^{Jm} \max \left\{ \delta_{kmt}^{\text{excess}}, 0 \right\} p_{jmt} & \text{if } k \geq 1 \text{ firms short} \end{cases} \quad (14)$$

The parameter γ^{excess} in Equation 14 represents *price sensitivity* - the rate at which firm j 's price increases in proportion to the amount of excess demand. The maximum function ensures that price increases only occur when there is *positive* excess demand.

Based on previously estimated model parameters, I re-simulate the full model under this new pricing scheme, testing for different levels of price sensitivity. I allow price sensitivity to supply shortfall to vary between 0 (baseline scenario) and 2.0 (situation under which a disruption leading to 50% of missing quantities on the market results in a 100% increase in prices) and compare shortage outcomes, firms' outcomes and welfare measures under this scenario.

Counterfactual Results. Figure 8 displays the evolution of the model's equilibrium outcomes as a function of price sensitivity. The y-axis displays changes in equilibrium outcomes, measured in percentage deviation from the current baseline measures. The x-axis displays changes in the price sensitivity parameter γ^{excess} . In most sub-plots, I group together the U.S. and OECD locations, as they display similar trends in equilibrium outcomes, and consider the South-East Asian region separately, as counterfactual patterns often differ for drug suppliers located in this region.

Three main results arise from this counterfactual simulation.

First, compared to today's baseline, all three production regions (the U.S., OECD and South-East Asia) effectively witness a sharp *increase in the average probability that shortages occur* as price sensitivity to excess demand increases (Panel 8a). While unit prices of plants located in South-East Asia are slightly higher than baseline (up to 4%) for price increases that are less than one-to-one in the amount of supply shortfall, prices of U.S.-based plants always drop compared to today (Panel 8b). Higher price sensitivity always goes in hand with moderate decreases in average unit prices .

⁶³Depending on whether manufacturing firm j is itself short or not, the new price scheme under situation of excess demand enters firm j 's expected profit function through either the "non-short" or "short" part. See Appendix F.2 for details.

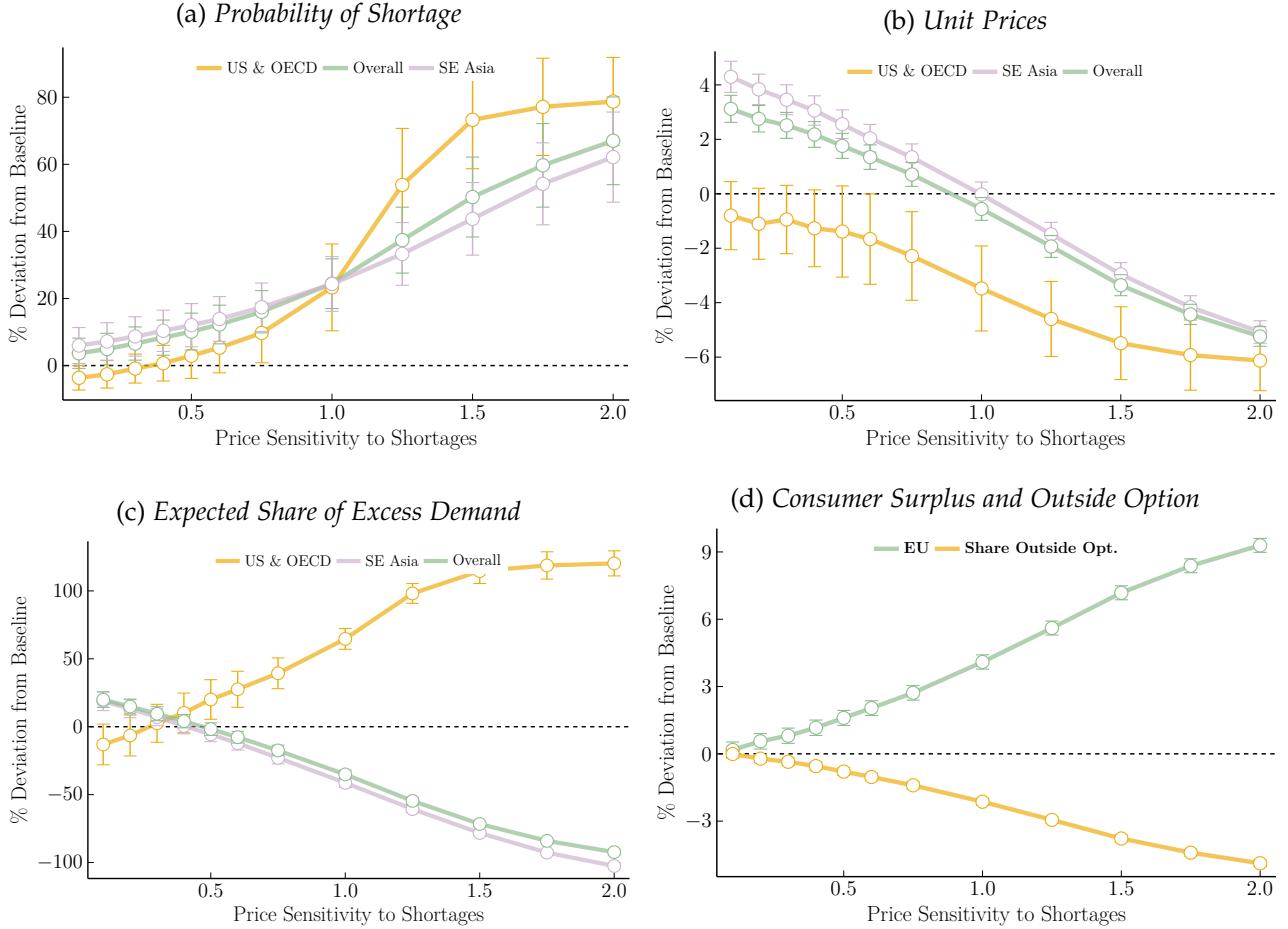


Figure 8: Changes in Equilibrium Outcomes as a Function of Price Sensitivity To Shortages

Notes: The y-axis displays changes in equilibrium outcomes (as prices, share of excess demand or consumer surplus), measured in percentage deviation from the current baseline measures. The x-axis displays changes in the price sensitivity parameter γ^{excess} from Equation 14 (the baseline is 0). Panel 8a shows that, the more sensitive to excess demand prices are, the more the probability of shortage increases, for all production locations. Panel 8b shows that the average contracted unit price decreases with price sensitivity. U.S. and OECD plants' prices are always below baseline, while South-East Asia locations are above baseline when prices vary less than one-to-one with excess demand. In Panel 8c, the expected share of excess demand decreases overall and for South-East Asia plants, but increases for U.S. and OECD-based plants. Panel 8d shows that the overall consumer surplus (Expected Utility) increases by 3 to 8%, while the share of patients choosing to not consume a drug decreases by 3% on average. More results can be found in Appendix F2.

The underlying mechanism is simple. This counterfactual scenario implicitly creates a capacity market in which firms who are able to supply during shortage periods are paid more, but the otherwise unit price remains low. When capacity is scarce the capacity price is high; when capacity is plentiful the capacity price is low. The negative correlation between equilibrium unit prices and price sensitivity to shortages can then be explained by larger and more frequent deviations in prices as a result of the overall higher probability of “disruptions” in the economy. As larger price sensitivity to excess demand means larger price increases post shortage, firms get incentives to let supply periodically fail to benefit from higher shortage-adjusted prices.

Second, the model predicts that price-sensitivity to shortages result in *large reductions in unfulfilled demand* from South-East Asian plants (between -40% and -70% for price sensitivity γ^{excess} between 1.0

and 1.5),⁶⁴ while the U.S./OECD-based locations experience on the contrary higher shares of excess demand (Panel 8c), and thus a worsening of shortage events. As more than 80% of the U.S. production of generic drugs is currently sourced from South-East Asia, this policy may prove effective in reducing the equilibrium shares of unmet demand, while keeping the current, globalized, organizational structure.

Perhaps counter-intuitively, even though pricing market capacity may result in higher probabilities of disruptions, it should also be resulting in less severe shortages. These dynamics take root in location differentials in both i) production costs and ii) average amplitude of shocks. When production losses are expected to be large, firms anticipating sharp price increases will insure against the shock by building more surplus,⁶⁵ and a lower share of these firms' demand will remain unfulfilled.

With large price sensitivity to shortages, production locations that were estimated to experience larger expected shocks (smaller average capacity yield ω_{lmt}), such as South-East Asia (see Section (5.4)) should now display less excess demand compared to the baseline (Panel 8c). On the contrary, when expected production losses are small, firms now have incentives to *restrict supply in order to take advantage of higher prices*. This corresponds to places with large estimated market-wide capacity yield $\omega_{l(j)t}$, as the U.S. and the OECD (Panel 8c).

Third, allowing price variations post-shock lead to higher offshoring levels, with a 10 to 20% increase in offshoring of production to South locations, and a sharp 50 to 70% decrease in the share of U.S.-based firms (see Appendix F2). Under the new counterfactual scenario, profit-margins of U.S.-based firms indeed become thinner (Figure F2) and producing at Home can only be sustained under higher level of disruptions. U.S.-based suppliers end up offshoring productions to places where they can increase profit margins, by trading lower production costs against the possibility of targeting higher disruptions levels than what can be achieved within the U.S. borders.

Conclusion. Overall, allowing prices to react to market conditions result in an overall increase in Welfare (Expected Utility increases by 2 to 8% compared to baseline levels, see Panel 8d) and to a 1 to 6% decrease in the share of patients choosing to not consume drug products (by 1 to 6%).

The overall effect of a policy allowing such short-term price variations is nevertheless ambiguous and depends on both price sensitivity levels and on the share of firms located in regions expecting large capacity yield $\omega_{l(j)t}$ post-shock. Firms will only choose larger supply levels to insure against higher disruption probabilities if the combination of current market prices and price-sensitivity make capacity investment profitable (see Panel 8c). Plants in more regulated, less risky locations, as the U.S. and OECD countries, should have more incentives than South plants to induce shortages in order to take advantages of ex-post price increases.

⁶⁴A γ^{excess} coefficient of 1.0 corresponds to a one-to-one price increase with supply shortfall. A coefficient of 1.5 means that if 10% of demand is unfulfilled post shock, prices will increase by 15%

⁶⁵Surplus is defined as the difference between realized supply (production choices adjusted for the shock) and demanded quantities: $\text{surplus}_{jmt} \equiv q_{jmt}\omega_{l(j)t} - Q_{jmt}^D(\cdot)$

6.4. Counterfactual 3. Changing Contracts: Enforcing Penalties

This third counterfactual scenario directly targets deficiencies in the prevailing procurement contracts. A salient feature of the contemporary procurement process is that sellers' reliability, as measured by their historical record of disruptions, does not markedly affect market shares (see Table 4 estimates). The question then arises: How can investments in production stability be incentivized?

Procurement Systems in Global Markets. The current procurement systems articulates around a system of auctions to select low-cost suppliers and relies on plants inspection to ensure adherence to U.S. "Good Manufacturing Practices" (see discussion from Section 2.1). Yet, in the face of the current globalized production landscape, these inspections have proven to be both deficient and expensive (see Sections 4.5 and A.6). An alternative procurement method, colloquially termed "Japanese-type contracts", operates on the premises that, when inspections become prohibitively costly, product quality is better assured by cultivating dependable relationships between buyers and suppliers. This approach hinges on extended-term contracts that offer suppliers assurances regarding future sales, coupled with the imposition of penalties if these contracts are breached.⁶⁶

Regarding generic injectable sales, regulatory enforcement of a penalty system for supply failures might prompt industry participants to internalize the substantial costs associated with excessively reducing production costs at the expense of supply reliability. Enforcing penalties may allow reliable suppliers to benefit from higher profit margins, thereby incentivizing investments in the production process. While contracts between GPOs and drug manufacturers do include clauses addressing failures to supply, these penalties are seldom implemented in real-world scenarios.⁶⁷

Failures-to-Supply Penalty: Framework. I modify the baseline manufacturer's expected profit function (Equation 4.4.1) such that supplier j operating in molecule market m at time t now receives a fine whenever a shortage happens (positive excess demand):

$$\begin{aligned} \mathbb{E}_{\omega_l} \pi_{jmt}(\cdot) = & \left[1 - F\left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}} \right) \right] p_{jmt} Q_{jmt}^D(\mathbf{p}, \mathbf{X}) \\ & + F\left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}} \right) \left[p_{jmt} q_{jmt} \mathbb{E}\left(\omega_{l(j)t} | \omega_{l(j)} < \frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}} \right) - \text{Penalty} \right] - TC_{jmt} \end{aligned} \quad (15)$$

where $TC_{jmt} \equiv (c_{jlt}(\varphi_l)q_{jmt} + FC_{jmt})$ denotes total cost of firm j . Equation 4.4.1 is unchanged at the exception of the **Penalty** term. I define the fine as a proportion to the missing supplied quantities:

$$\text{Penalty} \equiv \text{PenaltyRate} \underbrace{\left(Q_{jmt}^D(\mathbf{p}, \mathbf{X}) - q_{jmt} \mathbb{E} \left[\omega_l | \omega_l < \frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}} \right] \right)}_{\text{Excess Demand for Supplier } j \text{ (quantities)}} \quad (16)$$

⁶⁶These stylized procurement systems were defined in Taylor and Wiggins (1997) and recently discussed in Schott et al. (2017): "Under the "Japanese" system, buyers motivate sellers to maintain product quality by committing to long-run purchases at a price above sellers' costs. The opposing "American" system, by contrast, has buyers choosing the lowest-cost seller for each order via competitive bidding, and using costly inspection to deter cheaters from shipping low quality".

⁶⁷This was corroborated during a Spring 2022 interview with Todd Ebert, the President of the GPO Association.

Counterfactual Results. I re-simulate the entire model under specification (15), keeping the estimated parameter values constant. Figure (9) displays on the y-axis percentage deviation in equilibrium outcomes from the prevailing baseline, as a function of penalty rates (on the x-axis).

To set penalty rates, I use multiples of average expected firms revenues at the market level (see Equation 16).⁶⁸ I start from minimal penalty rates (0.1 time the average expected firm's revenue in the market) that may not necessarily deter any firm from participating, and increase thresholds to penalty levels that may deter most firms from failing to honor their commitments (2.5 times the average expected firm's revenue in the market).

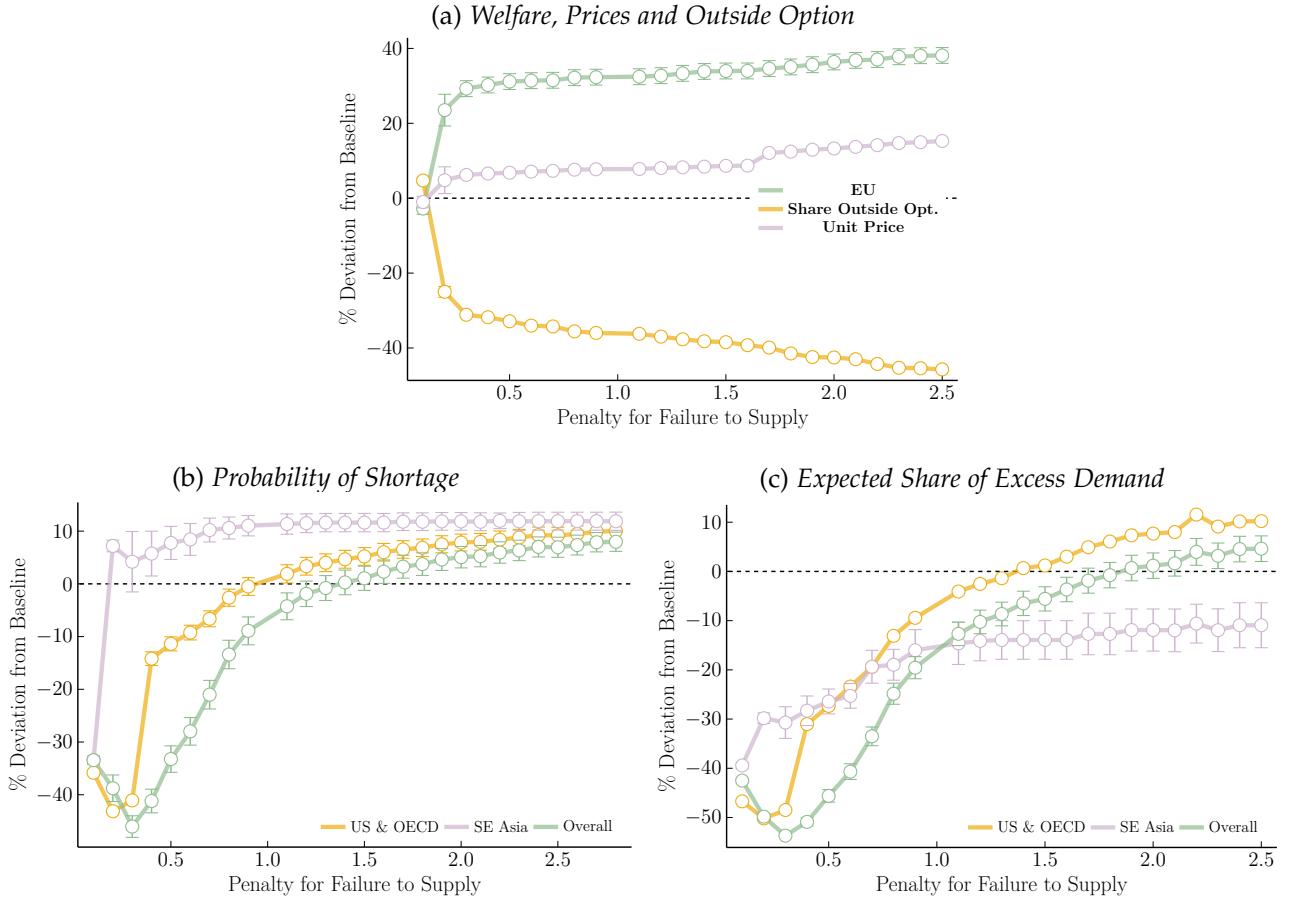


Figure 9: **Counterfactual: Enforcing Penalties for Failures-to-Supply**

Notes: The y-axis displays changes in equilibrium outcomes (as prices, share of excess demand or consumer surplus), measured in percentage deviation from the current baseline measures. The x-axis displays changes in the penalty rate from Equation 16 (the baseline is 0). Panel 9a shows that the average drug's unit prices increases with the penalty value (5 to 15%). This however does not deter consumption, as the outside option share (patients not consuming the drug) decreases by 30 to 45%) and Welfare increases by 30 to 38%. Panels 9b and 9c depict changes in equilibrium shortage outcomes as a function of penalty rates. Penalties approximating 0.5 times the average firm's anticipated market profit would decrease the probability of shortages by 35% and diminish the fraction of unmet demand by 45%. More results can be found in Appendix F3.

The main outcomes of this counterfactual simulation are threefold.

Firstly, higher penalty levels correlate with higher average equilibrium unit prices - in the range of

⁶⁸Alternative economic indicators that could be used to define penalty values include percentages over a firm's contract value or profit margin, or the social cost of supply failure for a given market.

5 to 15% for penalties between 0.5 to 2.5 times the average firm's expected profit (refer to the purple line in Figure 9a).⁶⁹

Secondly, the rise in Consumer Welfare, quantified by Expected Utility (illustrated in green in Panel (9a)) is the most significant among all three counterfactual scenarios examined (reshoring, pricing capacity and "Japanese" contracts). Welfare surges by 30 to 38% relative to baseline levels, translating into pronounced declines in the fraction of patients selecting the outside option (a reduction of 30 to 45% compared to current figures, as shown in yellow in Panel 9a). This suggests that the relative price escalation does not necessarily deter consumption access. Additionally, as shown in Figure 9a, prices, welfare, and patients choices to abstain from consumption remain relatively stable with increasing penalty values. This pattern implies that modest penalty levels could enhance Social Welfare without causing excessive price hikes.

Thirdly, of all counterfactual scenarios, penalty enforcement induces the most substantial reduction in the fraction of unfulfilled demand and the likelihood of shortages, for reasonably small penalty thresholds. Panels 9b and 9c depict equilibrium shortage outcomes against varying penalty rates. Penalties approximating 0.5 times the average firm's anticipated market profit would decrease the probability of shortages by 35% and diminish the fraction of unmet demand by 45%.⁷⁰

Beyond these three main outcomes, penalty enforcement would drive a geographical shift in production, even at modest fine levels (refer to Appendix Figure F3). Imposing larger penalties, at half the average firm's projected revenue, results in a 37% reduction in the proportion of OECD and Asian-based firms relative to current figures. This suggests that under large penalties, the cost of disruptions may become prohibitive for locations experiencing large supply shocks.

6.5. Counterfactuals: Comparison and Takeaways

This section compares the effects of the considered counterfactual experiments on shortage outcomes, prices and Welfare and discusses first and second-best policies.

Main Takeaway: Welfare Increases. My counterfactual experiments show that the net welfare gains of reducing supply shortfall levels may be large and are not offset by the associated increase in drug prices. All three scenarios lead to higher Social Welfare, thus corroborating that the current level of shortages may not be achieving the first-best equilibrium. While under all scenarios, reaching higher welfare levels come at the cost of higher unit prices and an increase in market concentration, this does not result in crowding out effects on the patients side. On the contrary, enforcing penalties results in

⁶⁹I do not differentiate unit prices, welfare and share of consumers choosing the outside option based on production locations, as no pronounced regional differences emerged, unlike the previous counterfactual (6.3).

⁷⁰It is noteworthy that penalties surpassing 1.5 times the average firm's expected revenue fail to mitigate shortage frequency or severity. A plausible explanation is that beyond a specific penalty threshold, the adverse impact on suppliers' profitability becomes too pronounced: disruptions results in big monetary losses and the level of buffers necessary to insure against shortages becomes so important that it reduces profit margins below current levels. . The model indeed forecasts a 30 to 60% surge in supplied quantities for penalties between 1.0 and 2.0 times the firm's expected revenue (refer to Appendix Figures F3), aligning with a spike in market concentration (Herfindahl–Hirschman Index).

a 37% increase in consumption access to drug products and a 30% increase in Welfare, while allowing the market prices to react to excess demand results in a 3% increase in consumption access and a 5% increase in Welfare (see Table 5). Note that targeting a zero-shortage level appears suboptimal: highest welfare levels are reached under equilibria characterized by non-zero levels of excess demand.

Increases in drug prices should remain nominal as they would almost exclusively affect generic medications, most of which currently sell for less than 10% of the original branded prices.⁷¹

Table 5: Counterfactual Changes in Welfare Compared to Baseline

COUNTERFACTUAL EXPERIMENT			
	Reshoring	SR Price Variations	Enforcing Penalties
Social Welfare (Expected Utility)	0.5 to 2%	2 to 9%	20 to 40%
Outside Option Share	-0.1 to -1.3%	-0.1 to -5.5%	-25 to -48%
E[Share Excess Demand]	+13 to +2%	+20 to -90%	-50 to +3%

Comparing Counterfactuals: Underlying Assumptions. Each considered counterfactual deals with one market failure. Subsidizing reshoring addresses regulatory deficiencies outside of the U.S. borders.⁷² Pricing capacity targets short-run price rigidities in procurement contracts. Enforcing penalties for failure to supply addresses the inability of the market to capture and reward reliability in production.

The reshoring scenario rests on the premises that the problem fundamentally lies in the offshoring process, and that the current globalized production structure creates frictions that may not be alleviated. On the contrary, the two following scenarios posit that the current system could be optimally leveraging comparative production advantages—allocating the production of innovative drug products to the North and generic drugs to the South. These scenarios, therefore, aim to tackle the underlying roots of the market failure more directly; specifically, the market’s limited ability to adequately reward production reliability.

Comparing Counterfactuals: How are policies ranked? While heavily emphasized in the recent policy debates (see Appendix F.1), switching back to a re-shored production system does not achieve the first best equilibrium. By implementing changes in procurement contracts that may help aligning price mechanisms with quality assurance to preempt sub-optimal market behaviors, the two other counterfactual experiments allow to reach substantially higher Welfare levels (see Table 5).

Reshoring. A re-shoring subsidy proves to be the priciest option to tackle shortages among all counterfactuals, resulting in a 30% increase in price at the suggested 10% tax break level (in addition to the direct tax subsidy). Its effectiveness in reducing shortage levels is moreover the most contrasted among all three scenarios: while the probability of shortages decreases compared to the baseline situation (by

⁷¹The median unit price of generic drugs in my data is \$1.1, see Table B3 and price increases under all counterfactual experiments maintain generic drug prices below the branded drug levels.

⁷²Namely, differentials in regulatory requirements across locations and inability of the U.S. regulator to enforce inspections abroad, making South locations more prone to adverse supply shocks than North locations, see Section 5.4).

5 to 10%), all levels of reshoring subsidies result in an overall decrease in production quantities (around 10-15%), and higher expected share of excess demand (10%).⁷³ As a result, the increase in consumer surplus is the smallest among all three considered counterfactual (around 1-2%).

Pricing capacity. While improvements in consumer Welfare is more marked than under the reshoring counterfactual (increasing by 2 to 9%), the overall effect of a policy allowing short-run price variations post shock is ambiguous. Anticipated large price increases during shortages may give firms located in places expecting relatively small loss in stocks (such that the U.S.) some incentives to restrict supply. Depending on current price levels, the sensitivity of prices to shortages, and the future production structure (offshored or partially re-shored), allowing prices to adjust upwards when supply is low may in some cases decrease the equilibrium expected share of excess demand. Most values of the price sensitivity parameter however lead to higher welfare levels and lower excess demand, suggesting that laws attempting to prevent generic manufacturers from being able to take any price increase, such that the Inflation Reduction Act of 2022, may be worsening off the current situation.⁷⁴

Enforcing penalties. Changing contracts to move away from an “American Procurement System” and closer to a “Japanese” system (see section 6.4) achieves the highest increase in welfare among all considered counterfactuals, while being the most effective in reducing expected excess demand levels (see Table 5). Implementing a system of penalty for failure to supply would provide higher margins for reliable suppliers (see Figure F3), allowing them to invest back in facilities maintenance and high-quality production processes. This policy automatically leads to an increase in market concentration (higher HHI in Figure F3), purposely driven by an increase in the exit rate of failing manufacturers. Without implementing a reshoring subsidy *per se*, it leads to re-allocation of production plants from South-East Asia to places experiencing less frequent disruptions (the U.S. and OECD). Both the average price increases across injectable markets and the share of reshored production remain inferior to the scenario in which reshoring was directly subsidized.

Anecdotally, a move towards such “Japanese-type” contracts has emerged lately in the U.S., with the stated purpose of fighting drug shortages. A new not-for-profit consortium of hospitals, Civica Rx,⁷⁵ started sourcing directly from contract manufacturers. This new venture negotiates longer-term contracts with trusted manufacturers to guarantee suppliers stable supply levels, in return for higher drug prices, bypassing buyers and the current for-profit procuring system. With some success so far: from 2002 to 2022, Civica Rx achieved a 96% fulfillment of its contractually guaranteed volume, while GPOs only reached 86% (Source: *Biospace*).

⁷³There are two potential explanations behind this outcome. Firstly, high sunk costs of re-shoring production may lead firms to reduce surplus even more in order to minimize additional expenses, magnifying the adverse effects of ‘just-in-time’ production. Secondly, because U.S.-based firms tend to build less buffer than their South-East Asian counterparts (higher production costs means surplus is more expensive), reshoring may actually decreases the overall market capacity.

⁷⁴The Inflation Reduction Act of 2022 (IRA) includes several provisions to lower prescription drug costs for people with Medicare and reduce drug spending by the federal government. This legislation has taken shape amidst *strong bipartisan, public support* for the government to address high and rising drug prices, in the context of collusive behaviors in the non-injectable generic industry (see *Cuddy (2020)*). Scott Gottlieb, former FDA’s commissioner, at CBS News in May 2023:“the administration under the IRA should carve out these old sterile injectable drugs entirely.”

⁷⁵See *The American Hospital Association (AH)*

7. Conclusion

This paper studies the persistent shortages in generic injectable pharmaceutical markets in the United States. Against the backdrop of a tenfold increase in shortages as documented by the FDA, coupled with significant offshoring of production to South-East Asia, this study aims to address core questions about the roots of these shortages and the potential role of offshoring in exacerbating them.

I built for the purpose of this project a unique dataset, which overcomes traceability issues and links drug products to their respective manufacturing plants by leveraging text analysis and probabilistic records linkage techniques. Based on these new data, I find that drugs produced in the South are predisposed to more frequent and longer-lasting shortages. Specifically, once fully offshored, such drugs encounter shortages that are 54% more frequent and last 61% longer.

I then propose a structural model of drug shortages in global markets, which generates insights into the incentive dynamics of supply disruptions and helps rationalize the offshoring correlations I observe in the data. Higher costs of investment in the resilience of their facilities lead firms to lower reliability by choosing higher probability of supply disruption, resulting in lower availability of the drug. The structural model highlights the complex interplay between manufacturer location decisions and the endogenous occurrence of shortages. The model captures a key trade-offs: while investments in better facilities or relocating to stricter regulatory markets can reduce shortage risks, they also raise costs. Offshoring is a way for firms to cut costs in return for a higher probability of manufacturing disruptions. There is a production possibility frontier which trades off manufacturing quality (low risk of disruption) against low production costs. For firms making a generic drug with low margins, moving toward lower-cost production increases profits. Firms would presumably like to cut costs in return for lower resilience within the U.S., but FDA inspection of facilities and the enforcement of minimum quality standards makes low-cost high-failure-probability manufacturing technologies harder to achieve. Hence, when firms offshore production in markets with low margins and high fixed costs, they do so in order to choose lower costs and manufacturing quality than is feasible in the U.S. by taking advantage of lower regulatory thresholds. Because manufacturing is concentrated in very few plants per product,⁷⁶ this in turn leads to more frequent and severe shortages of offshored products.

Disruptions in injectable drug markets turn into shortages because of a set of frictions in the current procurement systems: lack of transparency about the occurrence of future shortages and about suppliers identity and locations, failure of the pricing mechanisms to capture reliability, short-run capacity constraints (in addition to pricing constraints) and high production costs. All these factors make the market especially likely to unravel.

Counterfactual policy analyses suggest that while reshoring production has been a popular policy discussion, its potential to substantially increase welfare may be limited. In fact, reshoring may result in only a modest welfare increase of 1-2% and could hike drug prices by approximately 30%. Conversely, policy measures directly addressing market failures, such as imposing penalties for supply failures, show more promise. Penalizing supply shortcomings could increase welfare by as much as

⁷⁶The market displays strong economies of scale due to high fixed costs of production and technological constraints linked to the ability of plants to quickly switch production lines.

30% and enhance patient access to injectable drug products by 37%. Enhancing welfare in the injectable pharmaceutical market may lie in realigning market mechanisms to incentivize and reward reliability in production.

There is eventually a trade-off between the prices people are willing to pay for a drug, and how reliable they want the supply to be. A sudden spike in demand for drugs, as the one we are experiencing in the current crisis, may well result in shortages no matter where drugs are produced. The major problem may be that the industry focus on cutting costs leaves little room to invest in production capacities that would be strong enough to respond to supply chain disruptions or sudden shocks. Investing in the resilience of drugs supply chain comes at a cost, and solving these issues would most irremediably mean more expensive drug prices in the short-time.

References

- Acemoglu, Daron and Joshua Linn** (Aug. 2004). "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry". In: *The Quarterly Journal of Economics* 119 (3), pp. 1049–1090 (Cited on pages 3, 21).
- Alfaro, L.** et al. (2019). "Internalizing Global Value Chains: A Firm-Level Analysis". In: *Journal of Political Economy* 127 (2), pp. 509–559 (Cited on page 6).
- Antras, P.** and **D. Chor** (2013). "Organizing the Global Value Chain". In: *Econometrica* 81 (6), pp. 2127–2204 (Cited on page 5).
- (2018). "On the Measurement of Upstreamness and Downstreamness in Global Value Chains." In: *World Trade Evolution: Growth, Productivity and Employment* Taylor & Francis Group, pp. 126–19 (Cited on page 6).
- Antras, P., L. Garicano, and E. Rossi-Hansberg** (2006). "Offshoring in a Knowledge Economy". In: *Quarterly Journal of Economics* 121, pp. 31–77.
- Antràs, Pol** and **Robert W. Staiger** (Dec. 2012). "Offshoring and the Role of Trade Agreements". In: *American Economic Review* 102 (7), pp. 3140–83 (Cited on page 6).
- Antràs, Pol** et al. (July 2023). "Exporting, Global Sourcing, and Multinational Activity: Theory and Evidence from the United States". In: Working Paper Series (31488) (Cited on page 6).
- Atal, Juan Pablo, José Ignacio Cuesta, and Morten Sæthre** (Aug. 2022). *Quality Regulation and Competition: Evidence from Pharmaceutical Markets*. Working Paper 30325. National Bureau of Economic Research (Cited on page 9).
- Azghandi, R., J. Griffin, and Mohammad S. Jalali** (Dec. 2018). "Minimization of Drug Shortages in Pharmaceutical Supply Chains: A Simulation-Based Analysis of Drug Recall Patterns and Inventory Policies". In: *Complexity*, pp. 1–14.
- Baldwin, R.** and **F. Robert-Nicoud** (2014). "Trade in goods and trade in tasks: An integrating framework". In: *Journal of International Economics* 92 (1), pp. 51–62. ISSN: 0022-1996 (Cited on page 6).
- Baldwin, Richard** and **Anthony J. Venables** (2013). "Spiders and snakes: Offshoring and agglomeration in the global economy". In: *Journal of International Economics* 90 (2), pp. 245–254 (Cited on page 6).
- Barrot, Jean-Noël** and **Julien Sauvagnat** (2016). "Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks". In: *The Quarterly Journal of Economics* 131 (3), pp. 1543–1592. ISSN: 0033-5533 (Cited on page 7).
- Bernard, Andrew B** et al. (Mar. 2020). "Heterogeneous Globalization: Offshoring and Reorganization". In: Working Paper Series (26854) (Cited on page 6).
- Berndt, Ernst R.** and **Murray L. Aitken** (2011). "Brand Loyalty, Generic Entry and Price Competition in Pharmaceuticals in the Quarter Century after the 1984 Waxman-Hatch Legislation". In: *International Journal of the Economics of Business* 18 (2), pp. 177–201 (Cited on pages 10, 65).
- Bilir, L. Kamran** (July 2014). "Patent Laws, Product Life-Cycle Lengths, and Multinational Activity". In: *American Economic Review* 104 (7), pp. 1979–2013 (Cited on page 6).

- Boehm, Christoph E., Aaron Flaaen, and Nitya Pandalai-Nayar** (Mar. 2019). "Input Linkages and the Transmission of Shocks: Firm-Level Evidence from the 2011 Tōhoku Earthquake". In: *The Review of Economics and Statistics* 101 (1), pp. 60–75. ISSN: 0034-6535 (Cited on page 7).
- Borusyak, K. and X. Jaravel** (2017). "Revisiting Event Study Designs, with an Application to the Estimation of the Marginal Propensity to Consume". In: *Working Paper* (Cited on page 98).
- Bronnenberg, B. et al.** (2015). "Do Pharmacists Buy Bayer? Informed Shoppers and the Brand Premium". In: *The Quarterly Journal of Economics* 130 (4), pp. 1669–1726 (Cited on pages 10, 65).
- Burns, L. R. and J. A. Lee** (July 2008). "Hospital purchasing alliances: utilization, services, and performance". In: *Health Care Manage Rev* 33 (3), pp. 203–215 (Cited on pages 10, 65, 69).
- Caliendo, L and E. Rossi-Hansberg** (2012). "The Impact of Trade on Organization and Productivity". In: *Quarterly Journal of Economics* 127, pp. 1393–1467.
- Capps, Cory, David Dranove, and Mark Satterthwaite** (2003). "Competition and Market Power in Option Demand Markets". In: *The RAND Journal of Economics* 34 (4), pp. 737–763. ISSN: 07416261 (Cited on page 31).
- Carvalho, Vasco M et al.** (Dec. 2020). "Supply Chain Disruptions: Evidence from the Great East Japan Earthquake*". In: *The Quarterly Journal of Economics* 136 (2), pp. 1255–1321. ISSN: 0033-5533 (Cited on page 7).
- Castro-Vincenzi, Juanma** (Nov. 2022). "Climate Hazards and Resilience in the Global Car Industry". In: *Working Paper* (Cited on page 7).
- Chaudhuri, Prabal Ray** (1996). "The contestable outcome as a Bertrand equilibrium". In: *Economics Letters* 50 (2), pp. 237–242. ISSN: 0165-1765 (Cited on page 25).
- Chaudhuri, S., P. Goldberg, and P. Jia** (Dec. 2006). "Estimating the Effects of Global Patent Protection in Pharmaceuticals: A Case Study of Quinolones in India". In: *The American Economic Review* 96 (5), pp. 1477–1514 (Cited on pages 6, 9).
- Ching, Andrew T., Fumiko Hayashi, and Hui Wang** (2015). "Quantifying the Impacts of Limited Supply: the Case of Nursing Homes". In: *International Economic Review* 56 (4), pp. 1291–1322.
- Chopra, S. and M. S. Sodhi** (2004). "Managing risk to avoid supply-chain breakdown". In: *MIT Sloan Management Review* 46 (1), pp. 53–61.
- Conlon, Christopher T. and Julie Holland Mortimer** (Nov. 2013). "Demand Estimation under Incomplete Product Availability". In: *American Economic Journal: Microeconomics* 5 (4), pp. 1–30.
- Conrad, Ryan and Randall Lutter** (Dec. 2019). "Generic Competition and Drug Prices: New Evidence Linking Greater Generic Competition and Lower Generic Drug Prices". In: *U.S. Food & Drug Administration Report* (Cited on page 11).
- Conti, R. and E. Berndt** (2019). "Four Facts Concerning Competition in US Generic Prescription Drug Markets". In: *International Journal of the Economics of Business*, pp. 1–22 (Cited on page 8).
- Conti, R., E. Berndt, and S. Murphy** (2017). "The Landscape of US Generic Prescription Drug Markets, 2004–2016". In: *National Bureau of Economic Research, Working Paper* (No. 23640, July) (Cited on pages 8, 78).
- (2018). "The generic drug user fee amendments: an economic perspective". In: *Journal of Law and the Biosciences* 5 (1), pp. 103–141 (Cited on page 8).

- Conti, Rena M. and Ernst R. Berndt** (2016). "Specialty Drug Prices and Utilization after Loss of U.S. Patent Exclusivity, 2001–2007". In: *Measuring and Modeling Health Care Costs*. National Bureau of Economic Research, Inc, pp. 273–321 (Cited on page 11).
- Costinot, A., J. Vogel, and S. Wang** (2012). "Global Supply Chains and Wage Inequality". In: *American Economic Review Papers and Proceedings* 102 (3), pp. 396–401 (Cited on page 6).
- (2013). "An Elementary Theory of Global Supply Chains". In: *Review of Economic Studies* 80, pp. 109–144 (Cited on page 6).
- Costinot, Arnaud et al.** (May 2019). "The More We Die, The More We Sell? A Simple Test of the Home-Market Effect". In: *The Quarterly Journal of Economics* 134 (2), pp. 843–894 (Cited on pages 3, 21, 22).
- Cuddy, Emily.** (2020). "Competition and Collusion in the U.S. Generic Drug market". In: *Working Paper* (Cited on pages 54, 78).
- Dafny, Leemore S.** (Dec. 2005). "How Do Hospitals Respond to Price Changes?" In: *American Economic Review* 95 (5), pp. 1525–1547 (Cited on page 25).
- Donohue, Julie M. and Derek C. Angus** (2017). "National Shortages of Generic Sterile Injectable Drugs: Norepinephrine as a Case Study of Potential Harm". In: *JAMA* 317 (14), pp. 1415–1417.
- Dubois, P., A. Gandhi, and S. Vasserman** (May 2022). "Bargaining and International Reference Pricing in the Pharmaceutical Industry". In: *Mimeo* (Cited on pages 9, 36, 37).
- Dubois, P. and L. Lasio** (Dec. 2018). "Identifying Industry Margins with Price Constraints: Structural Estimation on Pharmaceuticals". In: *American Economic Review* 108 (12), pp. 3685–3724 (Cited on pages 9, 25, 36).
- Dubois, P. and Morten Saethre** (2018). "On the Effect of Parallel Trade on Manufacturers' and Retailers' Profits in the Pharmaceutical Sector". In: (18-883) (Cited on pages 9, 82).
- Elliott, Matthew, Benjamin Golub, and Matthew V. Leduc** (Aug. 2022). "Supply Network Formation and Fragility". In: *American Economic Review* 112 (8), pp. 2701–47 (Cited on page 7).
- FDA** (Oct. 2019). *Drug Shortages: Root Causes and Potential Solutions - A Report by the Drug Shortages Task Force* (Cited on pages 72, 73, 79).
- Food and Drug Administration** (Sept. 2016). *Self-Identification of Generic Drug Facilities, Sites and Organizations*.
- Fort, Teresa C.** (Aug. 2023). "The Changing Firm and Country Boundaries of US Manufacturers in Global Value Chains". In: *Journal of Economic Perspectives* 37 (3), pp. 31–58 (Cited on page 6).
- Fox, Erin R. and Milena M. McLaughlin** (Nov. 2018). "ASHP guidelines on managing drug product shortages". In: *American Journal of Health-System Pharmacy* 75 (21), pp. 1742–1750.
- Fox, Erin R., Burgunda V. Sweet, and Valerie Jensen** (2014). "Drug Shortages: A Complex Health Care Crisis". In: *Mayo Clinic Proceedings* 89 (3), pp. 361–73 (Cited on page 72).
- Fox, Erin R. et al.** (2009). "ASHP Guidelines on Managing Drug Product Shortages in Hospitals and Health Systems". In: *American Journal of Health-System Pharmacy* 66, pp. 1399–1406 (Cited on pages 2, 12).

- Frank, R., A. Hicks, and E. Berndt** (July 2019). "The Price to Consumers of Generic Pharmaceuticals: Beyond the Headlines". In: *National Bureau of Economic Research, Working Paper* (No. 26120) (Cited on page 8).
- Frank, R. and D. Salkever** (1997). "Generic Entry and the Pricing of Pharmaceuticals". In: *Journal of Economics and Management Strategy* 6(1), pp. 75–90 (Cited on page 8).
- Ganapati, S. and R. McKibbin** (Oct. 2019). "Non Tariff Barriers and Bargaining in Generic and Off-Patent Pharmaceuticals". In: *Mimeo* (Cited on page 9).
- GAO, the United States Government Accountability Office** (2016). *FDA Has Improved Its Foreign Drug Inspection Program, but Needs to Assess the Effectiveness and Staffing of Its Foreign Offices* (Cited on pages 15, 72, 74).
- Goldschmidt, D. and J. Schmieder** (2007). "The Rise of Domestic Outsourcing and the Evolution of the German Wage Structure". In: *The Quarterly Journal of Economics* 132(3), pp. 1165–1217 (Cited on page 98).
- Goodman-Bacon, A.** (2018). "Difference-in-Differences with Variation in Treatment Timing". In: *NBER Working Paper* (Cited on page 98).
- Graham, JR.** (2012). "The shortage of generic sterile injectable drugs: Diagnosis and solutions." In: *Policy Brief, Mackinac Center for Public Policy*.
- Grossman, Gene M., Elhanan Helpman, and Hugo Lhuillier** (2023). "Supply Chain Resilience: Should Policy Promote International Diversification or Reshoring?" In: *Journal of Political Economy* 0(ja), null (Cited on page 7).
- Grossman, Gene M. and Edwin L.-C. Lai** (2008). "Parallel imports and price controls". In: *The RAND Journal of Economics* 39(2), pp. 378–402 (Cited on page 9).
- Grossman, Gene M. and E. Rossi-Hansberg** (2008). "Trading Tasks: A Simple Theory of Offshoring". In: *American Economic Review* 98, pp. 1978–1997 (Cited on page 6).
- (2012). "Task Trade Between Similar Countries". In: *Econometrica* 80(2), pp. 593–629 (Cited on page 6).
- Hernandez, Inmaculada et al.** (2019). "Changes in Drug Pricing After Drug Shortages in the United States". In: *Annals of Internal Medicine* 170(1), pp. 74–76.
- Ho, D. et al.** (2007). "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." In: *Political Analysis* 15, pp. 199–236 (Cited on page 96).
- Ho, W. et al.** (2015). "Supply chain risk management: A literature review." In: *International Journal of Production Research* 53(16), pp. 5031–5069.
- Huang, D. and C. Rojas** (2013). "The Outside Good Bias in Logit Models of Demand with Aggregate Data". In: *Economics Bulletin* 12(1), pp. 198–206.
- (2014). "Eliminating the outside good bias in logit models of demand with aggregate data". In: *Review of Marketing Science* 12(1), pp. 1–36 (Cited on page 37).
- Iacus, S., G. King, and G. Porro** (2008). "Matching for Causal Inference Without Balance Checking". In: *Mimeo* (Cited on page 96).
- ISMP** (2017). "ISMP survey on drug shortages for hospital pharmacy directors or their designees only". In: *ISMP Medication Safety Alert!* 22(17), pp. 5–6.

- Jager, Simon** (2016). "How Substitutable are Workers? Evidence from Worker Deaths". In: *Working Paper* (Cited on page 98).
- Kaakeh, R.** et al. (2011). "Impact of drug shortages on U.S. health systems". In: *m J Health Syst Pharm* 68 (19), pp. 1811–9.
- Kanavos, P.** et al. (2013). "Higher US branded drug prices and spending compared to other countries may stem partly from quick uptake of new drugs". In: *Health Aff (Millwood)* 32 (4), pp. 753–761.
- Kaygisiz, N.** et al. (Dec. 2019). "The Geography of Prescription Pharmaceuticals Supplied to the U.S.: Levels, Trends and Implications". In: *National Bureau of Economic Research, Working Paper* (No. 26524) (Cited on pages 8, 73).
- Khanna, Gaurav, Nicolas Morales, and Nitya Pandalai-Nayar** (Nov. 2022). *Supply Chain Resilience: Evidence from Indian Firms*. Working Paper 30689. National Bureau of Economic Research (Cited on page 7).
- Kim, S. and Fiona M. Scott Morton** (2015). "A Model of Generic Drug Shortages: Supply Disruptions, Demand Substitution, and Price Control". In: *Working Paper*.
- Krugman, Paul** (1980). "Scale Economies, Product Differentiation, and the Pattern of Trade". In: *American Economic Review* 70 (5), pp. 950–59 (Cited on pages 3, 17, 21).
- Le Moigne, Mathilde** (2020). "Exploring the "Fissured workplace": Internal job ladders' fragmentation and its effect on plants and workers". In: *Working Paper* (Cited on page 98).
- Leamer, Edward E. and Michael Storper** (2001). "The Economic Geography of the Internet Age". In: *Journal of International Business Studies* 32 (4), pp. 641–665.
- Lim, Claire S. H. and Ali Yurukoglu** (2018). "Dynamic Natural Monopoly Regulation: Time Inconsistency, Moral Hazard, and Political Environments". In: *Journal of Political Economy* 126 (1), pp. 263–312 (Cited on page 31).
- Link, MP., K. Hagerty, and HM Kantarjian** (2012). "Chemotherapy drug shortages in the United States: Genesis and potential solutions." In: *Journal of Clinical Oncology* 30 (7), pp. 692–694.
- M., Bagnoli and Bergstrom T.** (2005). "Log-concave probability and its applications". In: *Economic Theory* 26 (2), pp. 445–469.
- Melitz, Marc J.** (2003). "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity". In: *Econometrica* 71, pp. 1695–1725.
- Mermelstein, Ben** et al. (2020). "Internal versus External Growth in Industries with Scale Economies: A Computational Model of Optimal Merger Policy". In: *Journal of Political Economy* 128 (1), pp. 301–341 (Cited on page 31).
- Moscelli, Giuseppe, Hugh Gravelle, and Luigi Siciliani** (2021). "Hospital competition and quality for non-emergency patients in the English NHS". In: *The RAND Journal of Economics* 52 (2), pp. 382–414 (Cited on page 11).
- Pichler, Anton** et al. (2023). "Building an alliance to map global supply networks". In: *Science* 382 (6668), pp. 270–272 (Cited on page 7).
- Representatives, U.S. House of** (2011). "Drug shortage crisis: Lives are in the balance.Hearing before the Subcommittee on Health Care, District of Columbia, Census and the National Archives of the Committee on Oversight and Government Reform, House of Representatives, 112 Congress". In:

- Schott, Peter** et al. (2017). "Trade Policy Uncertainty and the Structure of Supply Chains". In: (788) (Cited on page 50).
- Scott Morton, Fiona** (1999a). "Entry Decisions in the Generic Pharmaceutical Industry". In: *RAND Journal of Economics* 30 (3), pp. 421–440.
- Scott Morton, Fiona M.** (1999b). "Entry Decisions in the Generic Pharmaceutical Industry". In: *The RAND Journal of Economics* 30 (3), pp. 421–440 (Cited on page 8).
- (2000). "Barriers to Entry, Brand Advertising, and Generic Entry in the U.S. Pharmaceutical Industry". In: *International Journal of Industrial Organization* 18 (7), pp. 1085–1104 (Cited on page 8).
- Stomberg, C.** (2016). "Drug Shortages, Pricing and Regulatory Activity". In: *National Bureau of Economic Research, Working Paper* (No. 22912, December) (Cited on page 8).
- Sun, Liyang** and **Sarah Abraham** (2020). "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects". In: *Journal of Econometrics* (Cited on page 98).
- Tang, O.** and **S. N. Musa** (2011). "Identifying risk issues and research advancements in supply chain risk management". In: *Journal of Production Economics* 133 (1), pp. 25–34.
- Tang, SY.** and **P. Kouvelis** (2011). "Supplier diversification strategies in the presence of yield uncertainty and buyer competition". In: *Manufacturing & Service Operations Management* 13 (4), pp. 439–451 (Cited on page 27).
- Taylor, Curtis R** and **Steven N Wiggins** (Sept. 1997). "Competition or Compensation: Supplier Incentives under the American and Japanese Subcontracting Systems". In: *American Economic Review* 87 (4), pp. 598–618 (Cited on pages 31, 50).
- Tirole, Jean** (Feb. 1988). *The Theory of Industrial Organization*. Vol. 1. MIT Press Books 0262200716. The MIT Press (Cited on page 25).
- Tucker, E.** et al. (2020). "Incentivizing Resilient Supply Chain Design to Prevent Drug Shortages: Policy Analysis Using Two and Multi Stage Stochastic Programs". In: *IIE Transactions* 52 (4), pp. 394–412.
- Vatter, Benjamin** (Sept. 2022). "Quality Disclosure and Regulation: Scoring Design in Medicare Advantage". In.
- Woodcock, J.** and **M. Wosinska** (2013). "Economic and Technological Drivers of Sterile Injectable Drug Shortages". In: *Clinical Pharmacology and Therapeutics* 93 (2), pp. 170–176 (Cited on page 8).
- Yurukoglu, Ali, E. Lieberman, and D.B. Ridley** (2017). "The Role of Government Reimbursement in Drug Shortages". In: *American Economic Journal: Economic Policy* 9 (2), pp. 348–382 (Cited on page 8).

Appendices

Appendix A - Market Structure and Generic Drugs Supply Chains	64
A.1 The U.S. Market for Generic Drugs	64
A.2 Group Purchasing Organizations (GPOs) and Procurement System	66
A.3 Key Features of Shortages: The Case of Sterile Injectable Markets	69
A.4 Most Injectable Drugs Serve Small Markets	75
A.5 Collusion and Contestability in Generic Sterile Injectable Drug Markets	78
A.6 FDA's Regulatory Procedures on Drug Manufacturing Inspections	80
Appendix B - Additional Reduced-Form Results	81
B.1 Size of the Impacted Market	81
B.2 Pricing	82
B.3 European Shortages	87
B.4 Correlations between Offshoring and Shortages	87
B.5 Additional Results: Instrumental Variable Analysis	93
Appendix C - A Matched Dynamic Differences-in-Differences Framework	94
C.1 Treatment	94
C.2 Matching Procedure	95
C.3 Empirical Framework	96
C.4 Results	98
C.5 Additional Summary Statistics on Matching and Treatment and Control Groups	102
Appendix D - Model Details	104
D.1 Equilibrium Outcomes of the Manufacturer's Profit Maximization Problem: Derivations	104
D.2 Derivatives of main functions	107
D.3 Proof: no minimum is required for the shortage probability to be well-defined	108
D.4 Discussion about buyer's internalizing shortages risk: thinking about reliability in the logit model.	109
Appendix E - Estimation Details	111
E.1 Defining Products and Drug Markets	111
E.2 Prices and Reliability Score Measures	112
E.3 Plant Location-Specific Statistics	113
E.4 Parameter Estimates	114
E.5 Estimation Details	115
E.6 Details on Model Simulation	115
E.7 About reintroducing unobserved heterogeneity ξ_{jmt}	117

Appendix F - Counterfactuals Details	119
F.1 Re-shoring Counterfactual: Details	119
F.2 Short-term Price Variations Counterfactual: Details	121
F.3 Penalizing Failures Counterfactual: Details	124
F.4 Counterfactual 4: Setting "Price Floors"	124
F.5 Counterfactual Outcomes: Parameter Estimates	126
Appendix F - Data Appendix : Mapping the Generic Drug Supply Chain	126
F.1 A Novel Dataset: Details on Data Structure	126
F.2 Probabilistic Record Linkage	130
F.3 Building Sales Measures for Drugs Using Medicare Part B Data	133

Appendix A - Market Structure and Generic Drugs Supply Chains

This section first summarizes the functioning of the U.S. Market for Generic Drugs, before describing the data collected⁷⁷. It then leverages this new dataset to portray the sterile injectable drug markets and underline some key features of this market that make it especially sensitive to shortages.

A.1. The U.S. Market for Generic Drugs

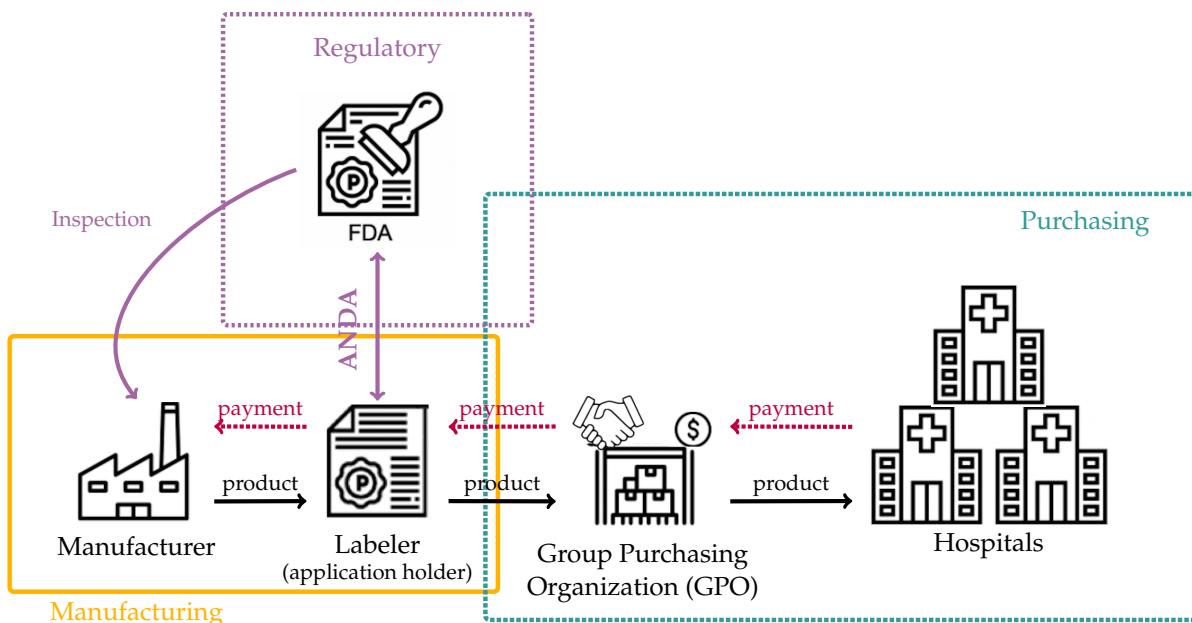


Figure A1: Market Structure

To manufacture a new (branded) drug, one needs a patent. These patents generally have limited life-span (typically twenty years in the United States, but most drugs lose their exclusivity after 7 years,

⁷⁷Figure ?? in Appendix summarizes the main datasets used in this paper

and the first approved generic drug receives 180 days of exclusivity against additional generic entry)⁷⁸. On the day a branded drug's patent expires, numerous generic pharmaceutical companies stand ready to make it. In order to sell a product to the U.S. market, pharmaceutical companies that want to market a generic version of an existing drug (i.e., generic sponsors, or "labelers") submit an Abbreviated New Drug Application ("ANDA") to the FDA. This only requires its sponsor (the "labelers") to establish pharmaceutical and bioequivalence to the brand and compliance with good manufacturing practices. Note that it is not required to establish the safety and efficacy of the generic⁷⁹. The drug sponsor thus manufacture the generic drug, which generally involves several steps, transforming active pharmaceutical ingredients (API)⁸⁰ to finished drug form products (FDF)⁸¹. This manufacturing part may either be made by the labeler's own manufacturing plants, or outsourced to contract manufacturing organizations. This process can be fully or partially offshored. For the remaining of this paper, I will refer to pharmaceutical firms acting as generic sponsors as "labelers", and to the manufacturing plants as "manufacturers".

In the United States, buyers are Group Purchasing Organizations (GPOs), which aggregate demand from hospitals and plays a distribution role. The main purpose of this intermediary is to negotiate volume discounts and achieve lower drug prices than each individual hospital might attain on its own. 96-98% of hospitals would buy drugs through Group Purchasing Organizations (GPOs) and 80% of hospital purchases occur through GPOs. Payers (e.g., insurers, Medicare) then reimburse hospitals for drugs purchased using fixed payments (prospective payments to hospitals, based on weighted average of sale prices in previous period). The GPO buys drug products directly from manufacturers, negotiating prices for bulk products⁸². Burns and Lee (2008) report that 41% of the U.S. providers surveyed belong to more than one GPO, but they route most of their purchases through a single GPO and utilize another only for specific contracts in limited supply areas. See Figure A2 and Section A.2.1 for more details on the procurement process.

There are three main regulatory costs of entering a generic market. First, a labeler needs to hold an abbreviated new drug application (ANDA) approved by the FDA to market a generic drug product, whose estimated cost is \$1-\$5 million (Berndt and Aitken, 2011). Second, since 2012, any company that wants to market generic drugs for the U.S. needs to pay the Generic Drug User Fee Amendments (GDUFA) fees each year, which amount to \$178,799 for fiscal year 2019⁸³. The third and last part, if

⁷⁸Source: <https://www.fda.gov/drugs/development-approval-process-drugs/>

⁷⁹See Bronnenberg et al. (2015) for a discussion of branded versus generic drugs.

⁸⁰An active ingredient is a substance used as a component of a drug to furnish pharmacological activity

⁸¹A FDF is a drug product in the form in which it will be administered to a patient, such as a tablet, capsule, solution, or topical application

⁸²I do not see hospitals (or GPOs) direct purchases. On the demand side, I only have data at the labeler- molecule dosage form product-quarter level for the entire U.S. market, through Medicare Part B and Red Book Prices. Prices I use are thus average prices over all U.S. hospitals for a given drug product

⁸³For fiscal year 2019: "GDUFA II stipulates that user fees should total \$493,600,000 annually, adjusted each year for inflation. For FY 2019, the generic drug fee rates are: ANDA (\$178,799), DMF (\$55,013), domestic API facility (\$44,226), foreign API facility (\$59,226), domestic FDF facility (\$211,305), foreign FDF facility (\$226,305), domestic CMO facility (\$70,435), foreign CMO facility (\$85,435), large size operation generic drug applicant program (\$1,862,167), medium size operation generic drug applicant program (\$744,867), and small business generic drug applicant program (\$186,217)." (Source: [GDUFA fees](#)

sole-sourcing contracts are in place and the one to three year contracts are enforced even during supply disruptions, sales are not guaranteed. See Section A.2.1 for further discussion of the purchasing system.

A.2. Group Purchasing Organizations (GPOs) and Procurement System

This section draws from information acquired through the reading of many health care policy reports, FTC documents considering mergers and market powers in GPO markets, official reports from the Group Purchasing Organization Association, and an interview carried with the President of the Group Purchasing Organization Association, Todd Ebert, on February 22 2022.

A.2.1. Online Procurement Auctions (Request for Proposals)

In practice, GPO's awards drug contracts using an online competitive bidding process. These are *online reverse*⁸⁴, *descending first-price sealed bids auctions*. They generally take only one round. GPOs do not generally post an initial price for a drug contract; it only collects Requests For Proposals (RFPs) from interested vendors.

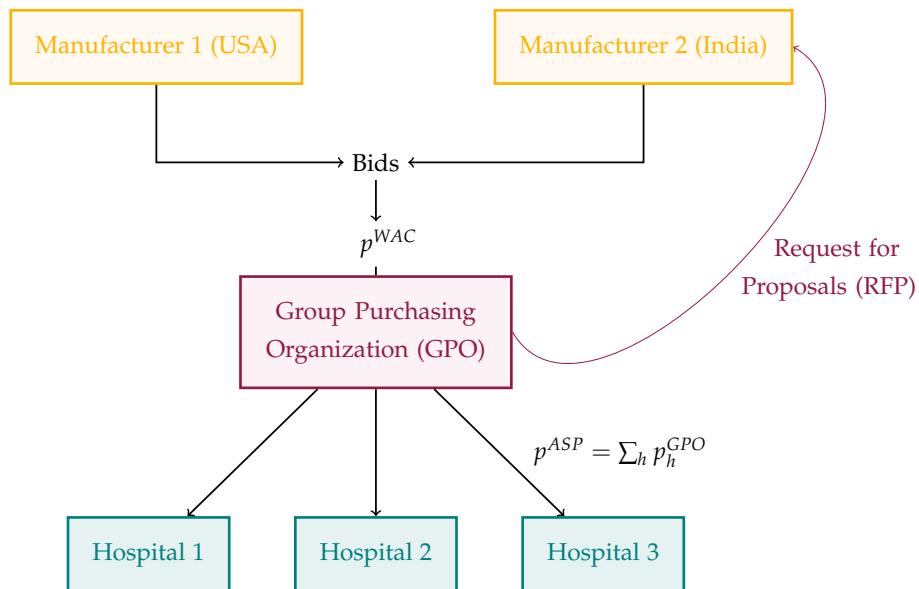


Figure A2: U.S. Procurement System: Tendering

The product's volume and the number of manufacturers in the marketplace generally determine how much pressure the GPO can bring to bear on price. With a first-price sealed bid of one round only, the buyer is given most of the purchasing power. Sellers do not have much opportunity to improve their bids, while buyers have multiple alternatives from different sellers to optimize their purchases.

Under limited circumstances, GPOS may award contracts to vendors without issuing RFPs: for instance if the manufacturer is a monopolist (some GPOs do an auction anyway), or some new manufacturer with innovative product (so not really for generics)

⁸⁴Online reverse auctions (i.e. auctions where a buyer requests several potential suppliers/sellers to make their bids to sell one or more products) are a new frequently used e-commerce approach to purchasing and procuring goods B-2-B in the supply chain.

Contract negotiations for a typical tendering process follows a *3 phases approach*. First, the tendering opens by **issuing a Request for Proposals (RFP)**. During this phase, A GPO notifies vendors of pending requests for proposal, by publicly posting on its website the bid calendars and minimum requirements for vendors. The GPOs' product selection processes generally takes 6 months, and ranges from as short as 1 month to as long as 18 months. In a second phase, the GPO **review proposals**: it scores competitors on how they meet the tendering criteria. The market claims most of the competition occurs on price, conditional on the manufacturers holding an ANDA. In the third and last phase, the GPO **negotiates and awards contracts**. GPOs widely used sole-sourcing for generic drugs, allowing only one supplier can win the bidding process⁸⁵. This practice changed slightly since the beginning of the Covid-19 crisis, with more dual or multi-sourcing from 2020. In situation in which dual or multi-sourcing is chosen, there does not seem to be clear rules about how market shares are allocated. GPOs indicate contracted quantities are awarded on a case-by-case basis and the decision to multi-source is highly "product dependent". The result of the post-bids ranking is privately announced to sellers⁸⁶.

Most contracts between GPOs and labelers would last on average 1 to 3 years. However, member hospitals do not have contracting obligations to purchase from the GPO. Manufacturers contracts can be renegotiated if there is a new entrant (but not all GPOs do so). Contracts that bundle products seem limited. Bundling would allegedly happen rarely and be restricted to products that are used together or otherwise related. Contracts may include failure-to-supply clauses, but enforcement of these clauses is often lacking, and some advocate that penalties are non-existent or too small. All major national GPOs report that their contracting practices have not changed much over time, since the late 1960s.

The practice of sole-sourcing, typically associated with extra rebates and discounts, raises critical questions regarding its potential impact on competition and market contestability. In scenarios where sole sourcing is employed by a Group Purchasing Organization with a large market share, this approach could conceivably cause competitors to withdraw from the market or stifle new entries. In scenarios in which there are several GPOs with similar market shares, and each exclusively contracts with one supplier, potentially considering backups in case of supply issues, this sourcing strategy may not be problematic. The degree of market contestability and the credibility of supplier-side entry threats however remain undetermined. Although these aspects fall outside the remit of the current paper, their importance in shaping market dynamics underscores a compelling need for subsequent research to explore these aspects⁸⁷.

A.2.2. Market Structure and Concentration

GPOs are intermediary formed by hospital banding together in order to negotiate volume discounts for buying drugs (aggregates hospitals' demand to obtain lower prices than each individual hospital

⁸⁵This was confirmed by an interview carried on February 22 2022 with the President of the Group Purchasing Organization Association, Todd Ebert

⁸⁶GPOs consider vendors grievance when they are not awarded a contract; they debrief the vendor on how to make changes to increase their chances of being awarded a contract during the next RFP cycle. Sellers can moreover file a formal grievance

⁸⁷To my knowledge, there is no available data on GPO purchases and contracts, which makes detailed IO studies of the buying side of the market challenging.

might attain on its own). While officially, hundreds of GPOs in the U.S., the GPO market went through many mergers over time and current market shares are concentrated among a few players.

In 2002, there were seven national GPOs with purchasing volumes over \$1 billion that account for more than 85 percent of all hospital purchases nationwide (combined purchasing power of about \$43 billion). The two largest GPOs accounted for about 70% of total GPO purchasing volume for all medical products. One of the two largest GPOs had 1,569 hospital members among the U.S. nation's approximately 6,900 hospitals; the other has 1,469 hospital members.

In 2014, the five largest national GPOs were estimated to hold more than 90% of the market (Amerinet, HealthTrust Purchasing Group, MedAssets, Novation, and Premier). Since 2016-2017, the top four GPOs with the most affiliated staffed beds are Vizient, Premier, Inc., HealthTrust Purchasing Group (HPG), and Premier-ASCEND.⁸⁸

On the purchasing side, GPOs are thus highly concentrated; the largest 3-4 GPOs now account for 90% of the market. The FDA claims that market power of GPOs may have reduced prices for health systems but has also contributed to a "race to the bottom"⁸⁹. The pursuit to offer drugs at the lowest price possible would have decreased generic sponsors' profitability, especially in the case of injectables, which are costly to manufacture. Importantly, because generic drugs are bioequivalent and exchangeable, there is no mechanism in the purchasing system to reward high-quality production, even though the FDA asserts that differences in the quality of manufacturing practices exist and are inextricably linked to shortages. Concentration among intermediaries in the drug purchasing system is a likely factor in driving the prices of some generics so low that generic sponsors do not see them as profitable.

A.2.3. GPOs' members and funding system

GPOs are paid both by member hospitals in the form of an annual membership fees and by drug manufacturers in the form of an administrative fee (CAF) that are capped to 3% of hospitals' drug purchases - aka discounts or rebates (the bulk of rebates is for branded drugs). 35 to 70% would then go back to hospital members.

- CAF are the main source of revenues for GPOs (92% of their revenue in 2012). They are based on a percentage of the purchase price that healthcare providers pay for products obtained through GPO contracts. The fee is only paid when a GPO's provider-member utilizes a GPO contract. Limited to 3% of hospitals' purchases ("safe harbor" regulation of 1996) in order to not be considered a kickback, administrative fees must be disclosed in an agreement between the GPO and each participating member. The agreement must state that the fees are 3 percent or less of the purchase price of the product (Social Security Act, 1986), many talk about "legalized kickbacks" or "pay-to-play" fees). Lots of debates about whether these admin fees charged to suppliers are anti-competitive (some argue this funding structure would create a principal-agent problem and

⁸⁸See Definitive Healthcare, Top 10 GPOs by staffed beds (Feb. 2023).

⁸⁹Note that these claims have to my knowledge never been assessed quantitatively and industry reports do not reach a common ground on the subject

motivates GPOs to search for higher prices, as their compensation increases when prices increase; others argue that GPOs' competition prevent this).

- A percentage of these vendor-fees are often passed to the GPOs' hospital members through discounts or rebates (between 37% and 70% of the fee in 2012 —a total of \$1.6 billion). The extent to which hospitals are reporting this revenue is not known because this has not been reviewed by HHS since 2005, and CMS officials stated that the agency has not specifically identified this as information that should be routinely audited. Medicare payments could be affected if hospitals do not account for revenue they receive from GPOs, which they are required to report as a reduction in costs on their cost reports. [Burns and Lee \(2008\)](#) estimate that GPO membership fees are "non-negligible": \$300-\$600k for a small hospital system anchored around a teaching hospital.

In theory, hospitals can contract with more than one GPO to do this purchasing, and can also purchase directly from drug manufacturers. In practice, it seems it almost never happens (GPO's contracting efficiency) and hospitals generally contract with different GPOs for different products (i.e. source all of a given drug d through a single GPO). Not all GPOs allow their members to belong to other national GPOs (among the 2 largest, one does and the other does not).

A.3. Key Features of Shortages: The Case of Sterile Injectable Markets

A.3.1. Shortages Definition

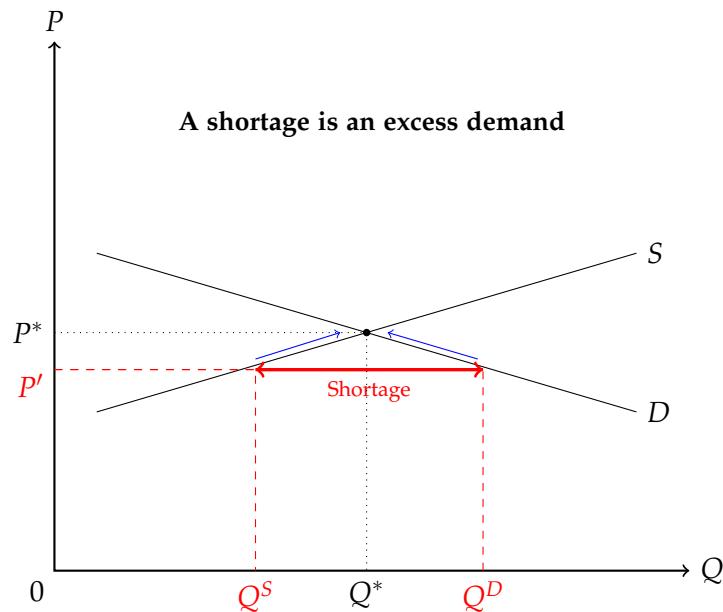


Figure A3: Definition of a Shortage

Notes: A shortage is defined as an excess demand, in contrast with surplus that is defined as excess supply. At price P' , quantity supplied Q^S is below quantity demanded on the market, Q^D . Basic economic theory tells us that, in a perfectly competitive market, shortage spells should be transitory; after a supply disruption (demand exceeds supply), prices should increase and provide an incentive for existing and new suppliers to increase production until there is enough supply to meet demand again (market equilibrium).

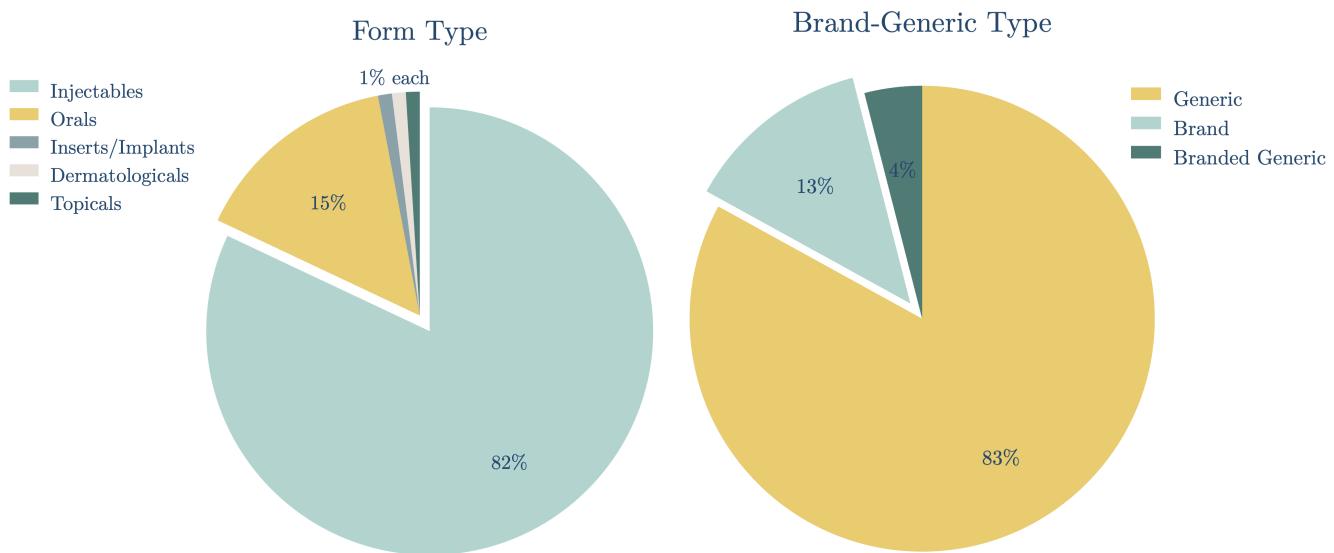


Figure A4: Which Products are the Most Affected by Shortages?

Notes:

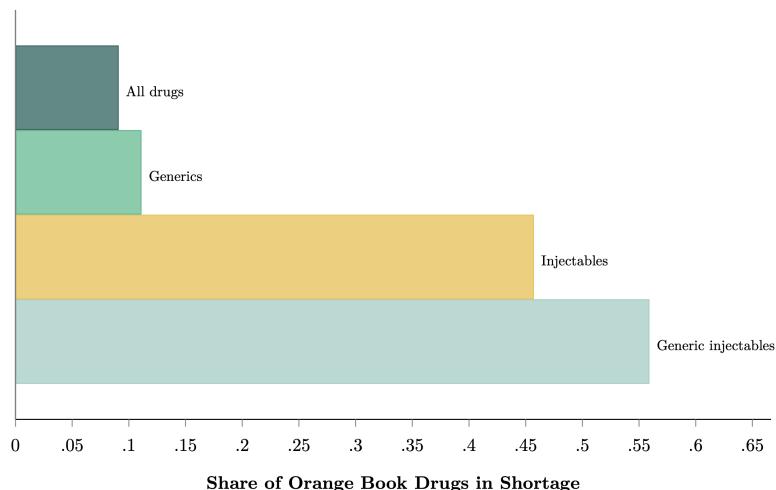


Figure A5: Share of U.S.-Marketed Drugs in Shortage, by Drug Category

Notes: This plot computes the share of U.S.-marketed drugs that went short, across different drug categories. While shortage events only affect 9% of all U.S.-marketed drugs and 12% of all generics over the period 2004-2019, the problem becomes more acute if we focus on smaller market segments. 46% of sterile injectable drug products (defined at the molecule-form level) marketed in the U.S. went short at least once over the period, and this number increases to 56% if we focus on the generic versions of injectable drugs. The major share of generic sterile injectables sold in the U.S. are thus impacted by shortage spells, heavily affecting hospitals' patients relying on drugs within this category. Source: The list of all U.S.-marketed drugs is from the FDA's Orange Book (FDA's list of approved drug products with therapeutic equivalence evaluations). I use both the January 2020 version of the Orange Book and historical Orange Book records to get the list of marketed drugs in each year over the period 2004-2019. Inside the Orange Book data, I extract the list of i) all generic drugs, ii) all sterile injectable drugs, iii) all generic, sterile injectable drugs marketed in the U.S. I then match these lists to the drugs belonging to the exact same category within the UUDIS shortages data. I then compute, within each category of drug, the share of the market affected by shortages over the period 2004-2019.

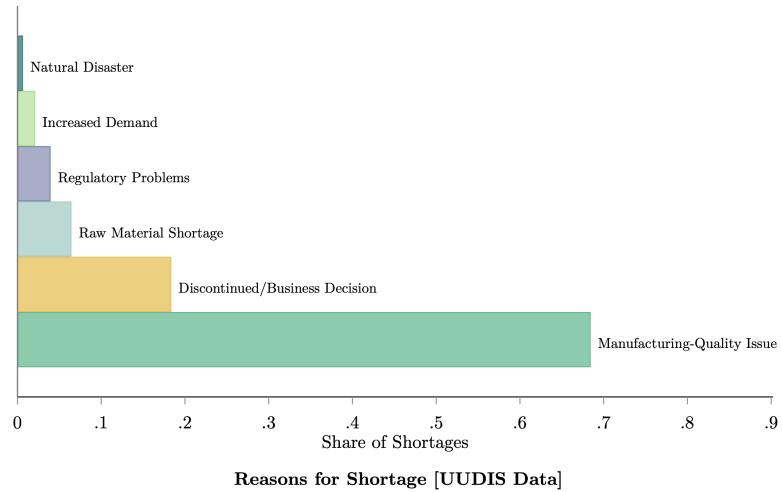


Figure A6: Reasons for U.S. Shortages

Notes: Drug shortages data were provided by Erin Fox from the University of Utah Drug Information Service (UUDIS). Reasons for shortages are provided to the FDA or the American Society of Health-System Pharmacists (ASHP) by pharmaceutical companies or distributors directly, and are verified by the UUDIS. It is not compulsory to report a cause for a shortage and 40% of shortage events do not report a cause. Shortage reasons are generally very detailed in the data and were aggregated into 6 broader categories. Among those shortages that report a cause, 68% are driven by "Manufacturing-quality issues, 18% are due to "discontinuations or business decisions", 6% are due to "raw material shortages", 4% to "regulatory problems", 3% to "increased demand" and 1% to "natural disaster"

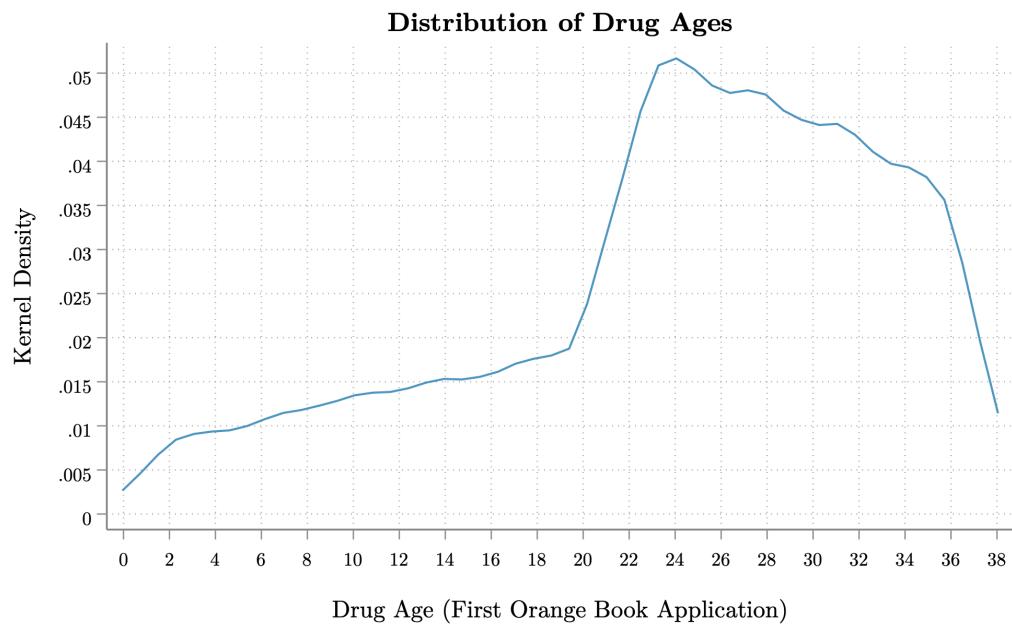


Figure A7: Distribution of Injectable Drug Ages

Notes: This figure plots the distribution of sterile injectable drug ages. Drug age is measured as the number of years since the first marketing application for a drug in the FDA's Orange Book. In the U.S. branded-drug patents generally have limited life-span (typically twenty years, but most drugs lose their exclusivity after 7 years, and the first approved generic drug receives 180 days of exclusivity against additional generic entry). This plot shows that most injectable products are old, off-patents products. Older generic drugs are generally less profitable than newer, more innovative products.

A.3.2. Key Features of Injectable Drug Markets

Drugs in Shortage have High Fixed Costs of Production and Low-Margins. With the institutionalization of near perfect substitutability for generic drugs, buyers (hospitals or pharmacies) consider all generic versions of a drug to be perfect substitutes. The product is generally believed to be of sufficient quality if it is on the market. Because the market cannot directly observe and reward quality, buyers go for the lowest-priced drug⁹⁰.

Sterile injectable drugs display many specific features that seem prone to crystallize the main issues encountered on generic drugs market. Compared to other generic drugs, sterile injectable products must be produced in highly specialized facilities with dedicated lines⁹¹. These injectable drugs are also harder to store, and most need to be refrigerated or are sensible to lights⁹². These high operational costs contribute to increase fixed production costs for the manufacturer. Markups not being high enough to cover fixed costs for this type of drugs may in turn explain the lack of manufacturer investments in facilities. There are indeed a few reported cases in the industry of some of these drugs selling at a loss (negative markups), while some of the newest generic drugs sell for hundreds or thousands of dollars (FDA, 2019). Low-profit generic medicines have been flagged as especially likely to suffer shortages, and many of these generics are sterile parenteral drugs. As an example, the cost of Ondansetron vial (in repeated shortage spells since 2008) is inferior to \$1 for 42mg/2 ML (less than \$2 per vial), which may just not be a sustainable price for a medicine (Fox, Sweet, and Jensen, 2014). This is consistent with estimates I get from my data: I find that the mean price per unit of product across sterile injectables that goes into shortage is \$1.3 (based on quarterly Average Wholesale Prices from the Redbook).

The slim profit margins of generic sterile injectable drugs created low incentives for manufacturers to shell out factory updates in order to avoid supply disruptions. Low margins directly encourages manufacturers to keep costs down by minimizing quality-related investments or by keeping low inventories to avoid losses due to production surplus⁹³. In the case of sterile injectables, there is moreover very little margin for errors in the manufacturing process, contrary to drugs in tablets or capsules. Some of the drug shortages reported in the U.S. have been traced to facilities that have been in operation since the 1960s with almost no factory improvements GAO (2016). Moreover, re-purposing existing factories to manufacture antibiotics or cancer drugs can take years. This would also explain why there is no branded drugs in shortage: brand-name drug makers have the resources to conduct extensive marketing promotions and ad campaigns. They also have incentives to keep

⁹⁰Stakeholders mention that a few phone calls are often enough for wholesalers to switch suppliers and reduce prices. Only relentless cost-cutting could allow to survive such a fierce price competition while making any sort of profit

⁹¹Production must take place in a "clean room" environment to make them perfectly sterile

⁹²These factors may explain why supply disruptions are not easily corrected or why hospital pharmacists may not be easily able to stockpile these medicines in prevision of shortages.

⁹³The FDA Commissioner Gottlieb in July 2018 said that "the low profit margins, and the significant cost of manufacturing these complex drugs, has resulted in consolidation in the industry. The only way to produce these low-margin products profitably is to manufacture them at tremendous scale. This has resulted in fewer and fewer manufacturers for certain key products. (...) It has also resulted in an under-investment in manufacturing, especially of sterile injectable drugs. Unless drug makers are investing in their manufacturing facilities, it can create conditions that give rise to production stoppages in order to fix manufacturing problems".

innovating and investing in manufacturing quality (which allow them to quickly adjust to potential supply or demand shocks).

Concentration and Offshoring

At the same time, high fixed costs of production means that this type of drugs has probably benefited the most from economies of scale and scopes resulting from concentration of specialized factories. At nearly every point in this system, the market has become more concentrated. The FDA estimates that 40% of sterile injectable drugs sold in the US have just one manufacturer, and 2/3 have 3 or less manufacturers (FDA, 2019). These figures are consistent with my data: I find that the median number of manufacturers for sterile injectable drugs, as measured by manufacturing facilities, is 1.9. But the consequence of this concentration is that one glitch on the manufacturing line in the facility that makes a drug can trigger a worldwide shortage. For a note on collusion of generic drugs, see Appendix A.5.

Among manufacturers, concentration is evident among all 3 actors. First, although policy makers have eliminated many barriers to generic entry, there remains high concentration in the supply of generic products. Drug shortages are more likely to occur in markets with only 1 to 3 generic sponsors. Second, because of consolidation of suppliers, competing generic sponsors often rely on a single active ingredient supplier. Third, it is increasingly common for a single contract manufacturer to produce the final dosage forms for all generic sponsors marketing a given product. Market concentration is the underlying reason why markets are so slow in responding to shortages. When production is halted for quality control problems (eg, the sterile injectables produced by the manufacturing facility are non-sterile or contain metal particulates), there is no alternative facility available.

The FDA claims that even though the market power of buyers (GPOs) may have contributed to lower prices for health systems, it also led to a "race to the bottom" which has decreased generic sponsors' profitability, especially in the case of costly-to-manufacture injectables⁹⁴. Crucially, because generic drugs are bioequivalent and considered perfectly exchangeable, there is no mechanism in the purchasing system to reward high-quality production even though FDA asserts differences in the quality of manufacturing practices exist and are inextricably linked to shortages.

Two main factors may deter entry and explain the high concentration of the market. Firstly, compared to most commonly used painkillers tablets and pills, sterile injectable drugs serve relatively small markets (demand for these products is limited to patients in hospitals). Associated with high fixed costs of production and low-profit margins, this may not make it a very attractive market for private manufacturing firms. Secondly, the necessity to have highly-specialized facilities makes entry harder; it seems manufacturers are not able to quickly rotate production of these generics between plants when disruption occurs. These seem to be the two main factors preventing the market to quickly adjust

⁹⁴ Kaygisiz et al. (2019) discuss why concentrations does not lead to higher prices on the market. In addition to the threat of contestability, the current bilateral oligopoly (where both suppliers and buyers are concentrated) makes the market resembles undifferentiated Bertrand competition, where prices are close to marginal costs in spite of there being only a small number of competitors. They argue the intensity of price competition would have helped to establish and maintain downward pressures on generic drug prices.

following disruption (see Appendix A.5 for more details).

Table A1 compares concentration of sterile injectables that went into shortage at least once over the period 2006-2019 to sterile injectables that never went into shortage, using Medicare Part B Spendings data from CMS (see Appendix F.3 for details on using Medicare Part B data to build Medicare Market Shares). Medicare sales are measured for each drug by matching drug names or drug codes (NDCs) to Healthcare Common Procedure Coding System (HCPCS). Sterile injectable drugs that experience shortages over the period 2006-2019 seem more concentrated (they have on average less labelers) and they represent a smaller part of total market sales and shares for sterile injectables.

All these features of the market may encourage manufacturers to keep costs down, either by minimizing quality-related investments or by keeping low-inventory to avoid losses⁹⁵

The problem has been deepened by the re-positioning of factories to India and China, where recalls of poor-quality products are common⁹⁶. Because raw materials and production are generally cheaper in developing countries, manufacturers of active pharmaceutical ingredients (API) have increasingly been based in India or China over time. Based on the FDA dataset on manufacturing facilities location (GDUFA), 60% of finished dosage forms (FDF) and 87% of base ingredients (API) are made at foreign facilities (see Figures ?? and ?? in Appendix). China alone produces 40% of the active components (APIs) of U.S medicines; most of the final manufacturing of the drugs occurs in India (80% of FDFs consumed in the U.S. are imported from India). The FDA says that pharmaceutical imports have more than doubled in the past decade.

The lack of regulatory control of the FDA on foreign-based plants certainly contributes to quality-issues: the FDA inspection records indicate that the agency has been inspecting foreign-based drug manufacturing facilities about once every 14 years, against every two-and-a-half years for U.S.-based facilities⁹⁷. In 2016, an estimated 33% of foreign facilities had never been inspected (as many as a thousand plants) (GAO, 2016). In Figure A8, I report the relative share of warning letters received by each country over the period 2008-2020. After carrying facility inspections, the FDA issues warning letters to drug manufacturers when it identifies a violation, such as poor manufacturing practices. Using restricted warning letters data I obtained from the FDA and adjusting for inspection frequencies, I find that U.S. facilities only account for 16% of warning letters, while this share reaches 40% for Indian and Chinese based-facilities (22% for Indian-based facilities and 18% for Chinese-based facilities). About 60 percent of FDA inspections of pharmaceutical plants in the Asia-Pacific region result in "Form 483s" — flags for findings that might violate U.S. quality standards.

⁹⁵The market has long been characterized by "just-in-time" production. Manufacturers historically never built more than one to two weeks of product stocks.

⁹⁶As an example among many, contamination by a possible cancer-causing chemical was discovered in 2018 in a Chinese factory making Valsartan, the active component in a blood-pressure medication taken by millions of people. That factory made half of the world's supply of Valsartan. This contamination led to the withdrawal of dozens of drugs in at least 22 countries, mostly in the U.S. and Europe.

⁹⁷The FDA has several foreign offices whose purpose is to facilitate inspection, but due to resource constraints, the agency relies on a risk-based approach in order to select plants for inspection. The FDA's Center for Drug Evaluation and Research (CDER) Director Woodcock estimated in 2017 that the FDA would need another \$225 million annually to inspect every foreign drug plant every other year

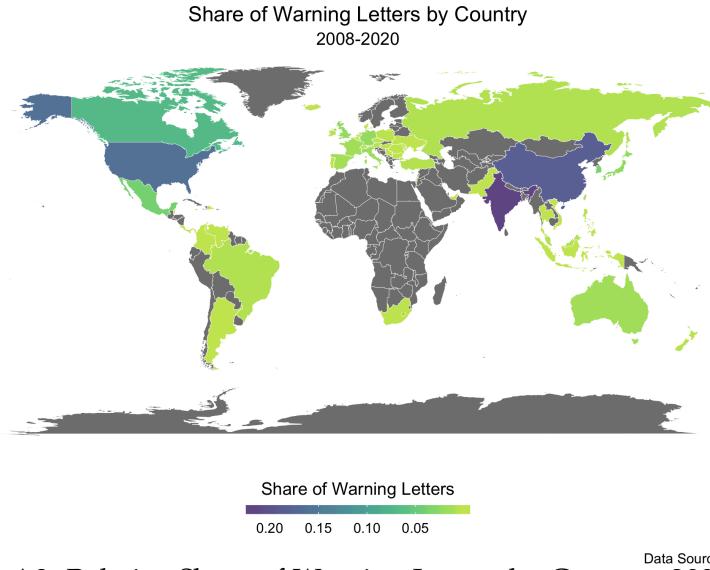


Figure A8: Relative Share of Warning Letters by Country, 2008-2020

A.4. Most Injectable Drugs Serve Small Markets

Using Medicare Part B Spending data from CMS (see Appendix F.3), I compare sales for sterile injectables that went into shortage to sterile injectables that did not. Two main measures of demand are used: 1) annual number of services billed (allowed services) and 2) total annual reimbursements by Medicare (allowed charges). Figure A9 displays that sterile injectables in shortage follow overall a similar pattern of sales compared to similar drugs that never went short (at the exception of the top right panel in which sales seems relatively "flatter" for drugs that went into shortage). If overall demanded quantities for sterile injectable products is low, it thus does not seem to be the main factor driving shortages.

Table A1: Comparing Concentration of Sterile Injectables by Shortage Status

	Labelers per Quarter	Total billable units per NDC	Share billable units per NDC
Shortage Dummy	-4.8*** [.103]	-53619*** [309]	-.0285*** [.00146]
R ²	0.339	0.338	0.005
Observations	170,137	175,057	174,984

Notes: Robust standard errors in parenthesis. This table uses a balanced panel of 333 Generic Sterile Injectable Drugs, 2006 to 2019. 234 drugs went short at least once over 2007-2019; 99 did not. The shortage dummy is equal to one if the drug goes into shortage at least once in my sample. Using the total number of drug units billed each quarter under Medicare Part B, we compute each labeler's total sales and share of sales for a given drug. We also compute the total number of labelers for each drug as measured by unique labeler codes from the Medicare Part B data. This table simply compares difference in means between the two groups, using an OLS regression.

I then regress different measures of shortages on annual Medicare demand for a drug and on share of foreign-based manufacturers. Results are displayed in Table A2. I find that drugs that represent

a higher share of total annual allowed Medicare services are on average less likely to be in shortage at any point in time. Drugs that capture a higher share of demand thus seems to be less likely to go short (note that coefficients on length of shortages and probability that a new shortage arises are not significant). However, when controlling for the share of the drug that is offshored, annual demand for the drug becomes insignificant. In column (2), the coefficient on the share of foreign manufacturers is positive and strongly significant (0.226). This coefficient becomes itself insignificant once we control for the share of foreign facilities that are specifically based in Asian locations. The share of production facilities that is Asian-based thus seems to be the most predictive of shortage probability at the drug level.

Table A2: Probability of shortage on Medicare Demand

<i>Dependent Variable:</i>	P(Shortage=1)			P(New Shortage=1)			Length Shortage (Days)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log Annual Allowed Services [†]	-.03367*** [.00824]	-.0239* [.01291]	-.02869 [.0182]	-.00572 [.00438]	-.00724 [.00683]	-.00893 [.00968]	-27.25 [27.7]	31.89 [39.5]	42.94 [40.98]
Share Foreign Facilities		.2259*** [.07386]	-.04195 [.129]		.05414 [.03745]	-.01482 [.05476]		351.8* [187.1]	-345.7 [419.7]
Share Asian Facilities			.3607** [.1479]			.09219 [.0694]			885.2* [448.1]
St. Dev.	.4962	.5562	.7109	.2636	.2942	.3782	1017	1098	1139
Year & Drug Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Beta Coefficient		-.2614*** [.06398]	-.1856* [.1002]	-.2227 [.1413]	-.06405 [.04906]	-.08116 [.07654]	-.1001 [.1085]	-.09635 [.09795]	.1128 [.1396]
OLS Coefficient		.01229*** [.00208]	.01606*** [.00294]	.01022*** [.00376]	.00114 [.00141]	.00094 [.00188]	-.00281 [.00264]	70.28*** [8.647]	105.2*** [11.86]

See the results for total allowed charges in Table B4

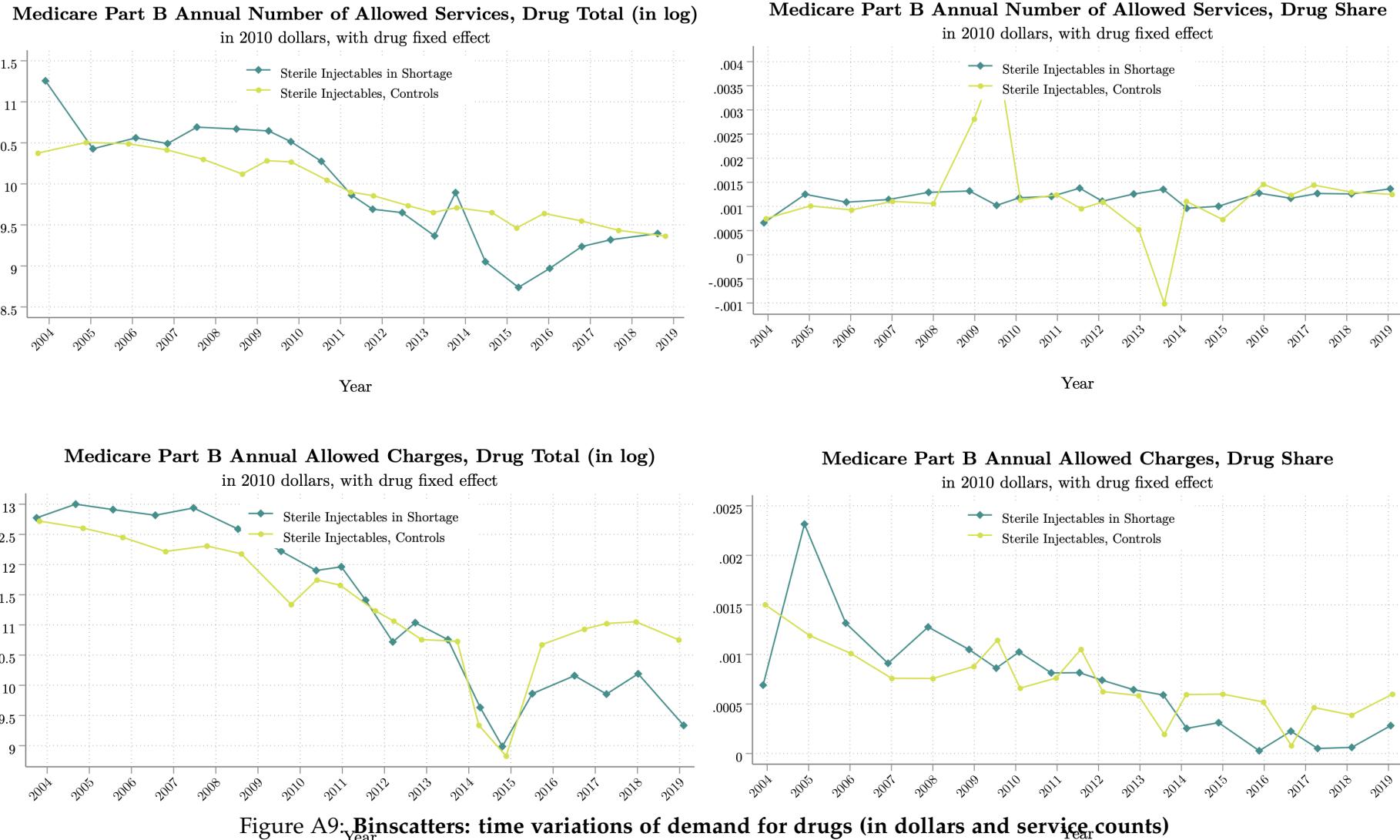


Figure A9: Binscatters: time variations of demand for drugs (in dollars and service counts)

Notes: These binscatter plots regress different measure of drug sales on year and shortage dummies. The top left panel plots the annual number of allowed services for a drug, in log. The top right panel plots the same variable in share (share of annual Medicare Part B Demand represented by the drug). The bottom left panel plots annual allowed charges, in log and the bottom right panel plots the same variable in share of total Medicare demand. The data is from CMS, Medicare Part B National Summary Data File. The sample consists of 333 sterile injectable drugs, 234 of which went into shortage at least once and 99 of which never went into shortage over years 2007-2019.

A.5. Collusion and Contestability in Generic Sterile Injectable Drug Markets

A.5.1. Collusive Practices

Even though generic drug markets are no longer protected by patents, many may not be large enough to attract the number of competitors needed to achieve marginal cost pricing ([Conti, Berndt, and Murphy, 2017](#)). Some players have actually deviated from the low-pricing strategy in recent years, as empirical evidence that generics are not sold at marginal cost in perfectly competitive markets have arisen: since 2010, 20% of U.S. generic or off-patent molecules have experienced a sudden price increase of more than 100% (GAO, 2016b). The price of Pyrimethamine, an anti-parasitic developed in the 1950's, was infamously changed overnight from \$13 to \$750 per pill (Pollack, 2015). Most of the generic drugs under scrutiny for collusive agreements however do not seem to coincide with those that are the most often going into shortages, that is sterile injectable drugs. [Cuddy \(2020\)](#) uses data on collusive firm and supply contract identities from District Court case files from a 2019 indictment. She estimates that among the hundreds of generic drugs accused of collusive conduct in her data, not more than 2 or 3 are injectable drugs. Consequently, a key assumption of her model is that there is a unique drug production technology, for pills (i.e. not injectable drugs). Most of collusive drug markets are focused on "newer" and generally more profitable generic drugs.

While collusion does not appear to be a major feature of the generic sterile injectable drugs market, assuming a competitive or collusive conduct behavior may not be necessary to assess that the average sterile injectable drug displays low-profit margins.

A.5.2. Contracting Practices and Market Contestability

In fiscal year 2012, in response to the peak in shortages and to solve the traceability issue on the market, the FDA enacted the Generic Drug Users Fee Act (GDUFA)⁹⁸. The main goal was for the FDA to know the geographical location of generic drug manufacturer: any foreign-based facilities wishing to produce a finished dosage form or its based ingredients must now register with the FDA and pay an annual fee in order to receive an approved ANDA. But this procedure has several pitfalls. Firstly, if manufacturers are required to provide their exact address, they have however no obligations to list the drugs they produce. Secondly, firms can pay the GDUFA fee without actually producing the drug. This has important implications for market competition: **many generic molecule markets may actually be contestable**, whereby firms hold an option to enter or re-enter. In a contestable market, pricing outcomes may seem indicative of highly-competitive markets, even though the actual number of competitors is small, because the threat of entry disciplines incumbent firms. On the generic drugs market, it means that once a manufacturer receives an approval for its ANDA, it can retain indefinitely even if it temporarily discontinues production. The threat of (re)entry by (temporary) exiting firms might facilitate contestability and discipline prices. More economic research is necessary to investigate

⁹⁸The GDUFA I, enacted in 2013, explicitly mandated that facilities, sites and organizations involved in the manufacture of generic drugs provide self-identification information to the FDA annually between May 1 and June 1. This act had two main purposes, according to the FDA: 1) to assist in setting annual facility fee amounts, and targeting inspections; 2) self-identification was a central component of an effort to promote global supply chain transparency

the potential role of the GDUFA I legislation in 2013 and the GDUFA II in 2017 in affecting temporary or permanent exit.

Eventually, low profit margins of generic drugs may be related to predatory pricing and lack of "failures to supply" clauses in contracts. In its *Drug Shortages Report* ([FDA, 2019](#)), the FDA relays discussions with stakeholders who mentioned that current contracting practices create a high level of business uncertainty, as they generally do not guarantee that a certain volume of products will be purchased at an agreed upon price. Stakeholders mention that contracts often contain clauses that leave manufacturers vulnerable to predatory pricing from competitors that are willing to undercut them to obtain market share, even at unsustainable prices. As a consequence, generic drugs sponsors may not have resources to invest in manufacturing or redundant capacity, which is facilitated by the absence of "failures to supply clauses" in contracts⁹⁹. Such contracts might however be difficult to enforce due to asymmetries of information; they may also encourage the buyer to create a shortage by hoarding, thus receiving a penalty payment from the supplier. Competing suppliers might also voluntarily hoard in order to create a shortage and then supply the product when prices rise.

A.5.3. Entry and Labelers-Contractors Relationships

If supply of products is a problem, especially in Asian-based facilities, why are U.S. generic pharmaceutical companies (the "sponsor" or "labeler" of a drug) not contracting with more than one production facility, or trying to switch manufacturers? Where is the friction? First, it is important to notice than many of these drugs market do not have a U.S.-based labeler, as generic manufacturing firms have increasingly discontinued production of the less profitable drugs in their portfolios.

Here are some quotes from major U.S. generic drug manufacturers: "*Right now our energy is focused primarily on the U.S. oral solids business...It's a unique situation. There are significant pricing declines. At least in the medium term, we don't see a shift to that situation, so we're assessing how best to optimize that, given that dynamic.*" (Comments made during Q4, 2017 earnings call before announcing proposed sale of core U.S. generics business in Sept. 2018) - Novartis (Sandoz) Q4, 2017. Another quote from Teva: "*The overall situation on U.S. generics and pricing ...we had a consolidation on the buyers side and you've had a situation where suppliers may be accepting of lower prices because they used to have a healthy margin...and if you take a race to the bottom on price...the only way out of that negative spiral is of course to stop it ... About 80% of the product we will get out of and they will move to other suppliers and about 20% of the product we will see an increase in price*" - Teva Q1 2018

Many U.S. labelers are foreign-based and owned companies, selling their drug products on the U.S. market. Many generic drugs that received an ANDA approval in the U.S. are thus directly sponsored by foreign companies - contract manufacturing organizations (CMO), labeler or downstream manufacturers¹⁰⁰. This point aside, the high amount of market concentration and high entry and switching costs

⁹⁹If contracts were to impose harsh penalties on manufacturers for failure to supply, then shortages should fall and average prices should rise.

¹⁰⁰There is a large heterogeneity in how and where ANDA holders manufacture generic drugs : in-house for both API and FDF, at facility site same as or different from headquarters, in-house FDF but outsourced API, outsourced both API and FDF, and not an ANDA holder but instead just a contract manufacturer to other firms who were ANDA holders. There is also increased outsourcing to off-shored entities

on the market may give pharmaceutical companies less options to switch suppliers of raw ingredients or finished drug forms. Yet another issue is that it seems hard to purchase for quality on the market, meaning that labelers also do not have information on which facilities are available to produce a particular drug and what is their history of supply disruptions. The FDA makes warning letters and records on more than 483 inspections public, but the names of drugs are redacted. There are no requirements to disclose which company actually makes a product, or manufacturing site. Thus, purchasers (hospitals, pharmacies, but also labelers) cannot always follow the data to take buying decisions.

A.6. *FDA's Regulatory Procedures on Drug Manufacturing Inspections*

Inspections Outcomes.

Upon a comprehensive evaluation of pharmaceutical manufacturing units, the U.S. Food and Drug Administration (FDA) systematically categorizes its findings into three distinctive classifications: "No Action Indicated" (NAI), "Voluntary Action Indicated" (VAI), or "Official Action Indicated" (OAI).

An NAI classification signifies that the inspection found no concerning issues or practices, or if found, they were minor and did not warrant additional regulatory measures. A VAI designation indicates that while there were certain issues discovered, the FDA is not recommending or initiating any administrative or regulatory measures at this time. An OAI classification suggests that the FDA will suggest administrative or regulatory measures.

Post-2012, there has been a discernible amplification in the frequency of inspections, particularly focusing on facilities perceived as high-risk. This increment in oversight led to an uptick in the identification of regulatory infringements, predominantly within overseas facilities. It's pertinent to note that these international facilities, prior to the institution of the FDASIA and GDUFA regulatory frameworks, had received fewer inspections compared to domestic units. Importantly, the majority of these international assessments were executed with prior notification. Yet, irrespective of geographic demarcation, in excess of 90% of evaluations culminated in classifications deemed satisfactory (NAI or VAI) with the sole exception being facilities located in India.

Both international and domestic pharmaceutical manufacturers are uniformly subjected to prevailing regulatory criteria, specifically, adherence to the Current Good Manufacturing Practices (CGMPs). In scenarios where a manufacturing unit demonstrably fails to comply with CGMP standards upon rigorous examination, the FDA possesses a variety of regulatory instruments to induce corrective actions. This encompasses the issuance of warning letters, initiation of import alerts, provision of injunctions, and the invocation of seizures. Should ensuing evaluations confirm sustained non-adherence, the FDA has the authority to implement more stringent measures.

In instances where a foreign manufacturing site shows significant enough issues to receive an OAI classification, the FDA can issue an Import Alert, effectively preventing drugs from that facility from entering the U.S. market. Typically, a facility will be removed from such an Import Alert once a follow-up inspection confirms that the issues have been resolved and the facility is in line with CGMP standards.

Not every pharmaceutical product can be tested before entering the U.S. market. In practice, while

the FDA does perform thousands of tests annually, only about one percent of these drugs do not meet the set quality standards. Considering the sheer volume of drug products (for reference, in 2018, there were almost 186 trillion tablets and capsules in the U.S. market), testing each batch comprehensively is not practical.

Appendix B - Additional Reduced-Form Results

B.1. Size of the Impacted Market

Table B1: Share of Generics and Sterile Injectables in the FDA's Orange Book Data

Generic	Sterile Injectable					
	No		Yes		Total	
	Freq.	Percent	Freq.	Percent	Freq.	Percent
No	3,584.0	19.7	1,642.0	30.3	5,226.0	22.1
Yes	14,653.0	80.3	3,772.0	69.7	18,425.0	77.9
Total	18,237.0	100.0	5,414.0	100.0	23,651.0	100.0

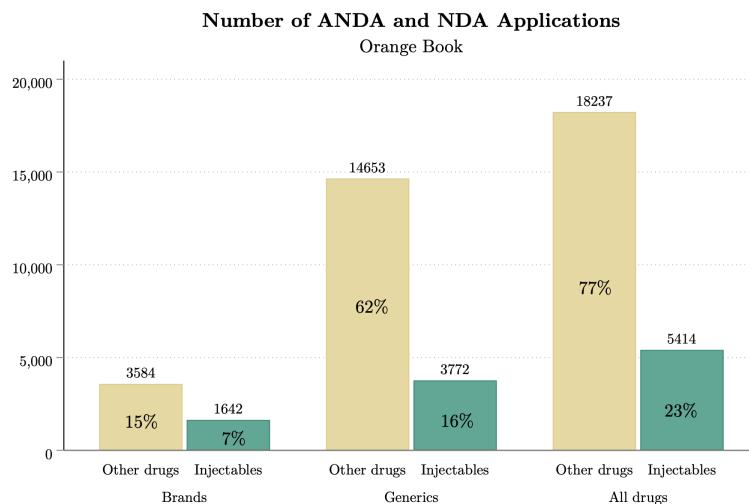


Figure B1: Impacted market size

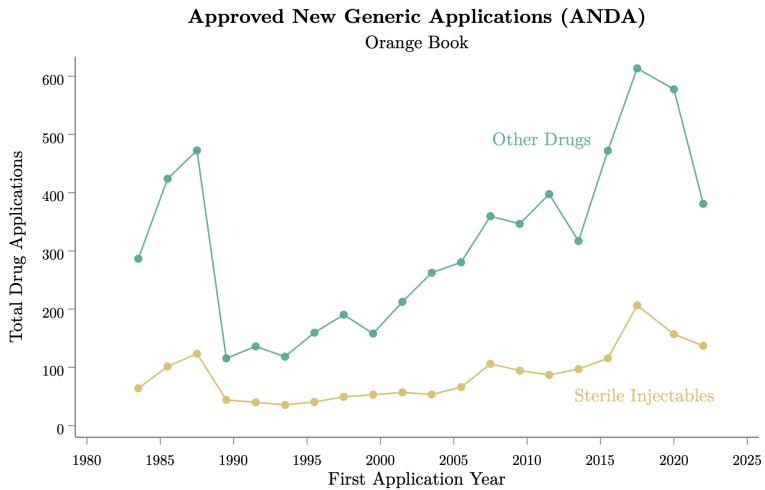


Figure B2: Time Series: Number of Approved Generic Patents (ANDA) in the Orange Book

B.2. Pricing

There is currently no nation-wide regulation of pharmaceutical prices in the United States. There is also no international reference pricing rule linking the U.S. market to other markets, nor any parallel trade of drugs with other countries as there is within Europe (Dubois and Saethre, 2018). For the time frame of this study (2002-2020), drug prices in the U.S. are therefore determined independently from other markets.

B.2.1. Details about the RED BOOK Pricing Data

Prices used in this paper are from the IBM Micromedex® RED BOOK® Historical Pricing Tables, which provides pricing that has been reported to the RED BOOK by drug manufacturers. Only inactive prices are delivered in the Historical Pricing Table. For active prices, I web-scraped the full active RED BOOK databases. The RED BOOK data contains a drug product name and national drug code (NDC), labeler name, effective date, a manufacturer CMSID (a unique identifier for each manufacturer) and a package CMSID (a unique identifier for each package). The main price measure in the RED BOOK is an Average Wholesale Price (AWP)¹⁰¹. This nationally recognized suggested wholesale price corresponds to the price paid by providers (hospitals, pharmacies) to wholesalers¹⁰². This dataset also includes

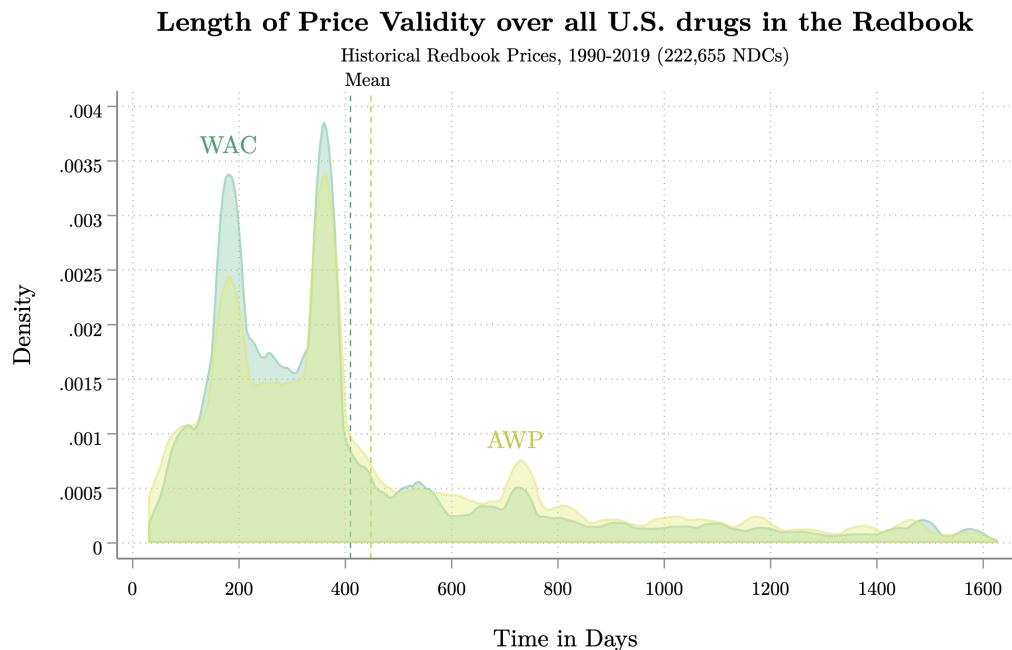
¹⁰¹The Average Wholesale Price (AWP) as published by IBM Watson Health is, in most cases, the manufacturer's suggested AWP and does not reflect the actual AWP charged by a wholesaler. IBM Watson Health bases the AWP data it publishes on the following: (1) AWP is reported by the manufacturer, or (2) AWP is calculated based on a markup specified by the manufacturer. This markup is typically based on the Wholesale Acquisition Cost (WAC) or Direct Price (DIRP), as provided by the manufacturer, but may be based on other pricing data provided by the manufacturer, or (3) Suggested Wholesale Price (SWP) is reported by the manufacturer. When the manufacturer does not provide an AWP, an SWP, or a markup formula from which AWP can be calculated, IBM Watson Health will calculate the AWP by applying a standard 20% markup over the manufacturer-supplied WAC. If a WAC is not provided, the standard markup will be applied to the DIRP.

¹⁰²The term "manufacturer" as used in the REDBOOK includes manufacturers, distributors, repackagers, and private labelers. The term "provider" includes retailers, hospitals, physicians, and others buying from the wholesaler or directly from the manufacturer for distribution to a patient

information about the Wholesale Acquisition Cost (WAC), which is the price paid by wholesalers to manufacturers, and the DIRP or DP (Direct Price) paid by providers (hospitals) to manufacturers (manufacturer pricing to direct-buying retailers)

B.2.2. Length of Price Validity

Even though the RED BOOK quarterly updates prices, most prices display marked price-rigidity. The median length of price validity in the data is 222 days (mean 434 days), using years 1990-2019 for 222,655 NDCs.



B.2.3. Pricing and Shortages

Matching Red Book Prices to Drugs in Shortages. In this section, prices from the Red Book are matched to the UUDIS firm-level shortages (i.e. shortages defined at the NDC 5-4-2 level). Using the historical Red Book data only, 5672 unique NDC 5-4-2 from the shortage data were matched, which represents 70% of unique NDC numbers (2434 NDCs not matched). Using both the historical data and the Red Book Pricing data I web-scraped from the current Red Book files, only 247 NDCs out of 8106 from the shortage data are not matched with prices, so only 3%. 97% of the drugs from the NDC-level shortage list can be linked to pricing data from the Red Book. If we try to match each NDC-year-quarter observation from the shortage data exactly, 393,132 observations over 688,407 are matched with prices, so 57% of the shortage data has exact pricing information at the drug NDC-year-quarter level. Note that a drug, as defined by a molecule-form, may be active under many different NDCs over the years (a NDC 5-4-2 is a National Drug Code that defines a drug at the labeler-dosage-form level). NDCs at the 5-4-2 level may thus not be present for all year-quarters in my data, as firms enter and exit over the period. If during its period of validity, a given NDC-5-4-2 does not display price updates in a given

quarter in the Red Book, I use price information from the last available quarter.

Table B2 provides summary statistics at the manufacturer-drug level (NDC 5-4-2) for all drug products listed in the manufacturer-level shortage data from the UUDIS (for more details about how this information was extracted, see Section F2). 17.4% of NDCs from the shortage list becomes affected by a merger between 2002 and 2020. 42.5% of these NDCs get discontinued. The average NDC goes short 6.52 times from 2002 to 2020, 3 time in median.

Table B2: **NDC-Level Shortages**

	N	p50	mean	sd	min	max	p25	p75
Merger	18,966	0	.174	.379	0	1	0	0
Discontinued	18,966	0	.425	.494	0	1	0	1
Number of Shortages by NDC (5-4-2)	18,966	3	6.52	7.45	1	37	1	8

Notes: NDC level shortages (5-4-2 format), cross-section from January 2002 to April 2020. Data is from the UUDIS. Information on mergers and discontinuation of products has been extracted from Word Files by the author using text mining techniques

Table B3: Monthly Unit Price of Drugs from the Shortage List

	Price type	p50	mean	sd	min	max	p25	p75	N
Brand	AWP	3	28	64	.0051	.64	17	487	238,357
	WAC	3	22	50	.0043	.63	15	415	142,635
Generic	AWP	1.1	7.7	30	.0049	.32	3.7	481	617,110
	WAC	.48	5.4	21	.0041	.12	2.3	362	285,328
Non-Parenteral	AWP	1.2	7.9	31	.0061	.45	3.6	482	395,463
	WAC	.6	7.4	27	.0053	.16	2.9	402	179,455
Parenteral	AWP	1.5	18	51	.0049	.28	9.4	487	459,743
	WAC	1.1	13	39	.0041	.2	7.5	415	248,247
Outside of shortage	AWP	1.3	13	44	.0049	.37	5.4	487	749,564
	WAC	.87	11	35	.0041	.18	5.2	415	380,743
During shortage	AWP	1.3	12	40	.0049	.38	5.5	482	106,122
	WAC	.73	8.1	26	.0041	.18	3.8	402	47,220
Before/After shortage notified	AWP	1.3	13	43	.0049	.37	5.4	487	843,459
	WAC	.86	11	34	.0041	.18	5.1	415	422,002
Month of notified shortage	AWP	1.3	13	41	.0049	.39	5.3	482	12,227
	WAC	.77	10	32	.0041	.17	3.9	402	5,961

Notes: Balanced panel of NDC (5-4-2 format) at the year-month level, from 2002 to 2019. Prices (Average Wholesale Prices, AWP and Wholesale Acquisition Costs, WAC) are from the Red Book (IBM Micromedex). Information about drug shortages are from the UUDIS. Interestingly, unit prices are slightly *lower* during shortages. It may either suggest that only the lowest-price manufacturers survived by the time the shortage starts, so that the ones going short are the one offering higher prices (and thus possibly having higher production costs). The defaulting manufacturers may not be able to ramp up production quickly because their high production costs make it more profitable for them to stay out. Alternatively, that shortages are coinciding with periods during which prices have been driving down (and could be the result of a "race-to-the-bottom" in prices).

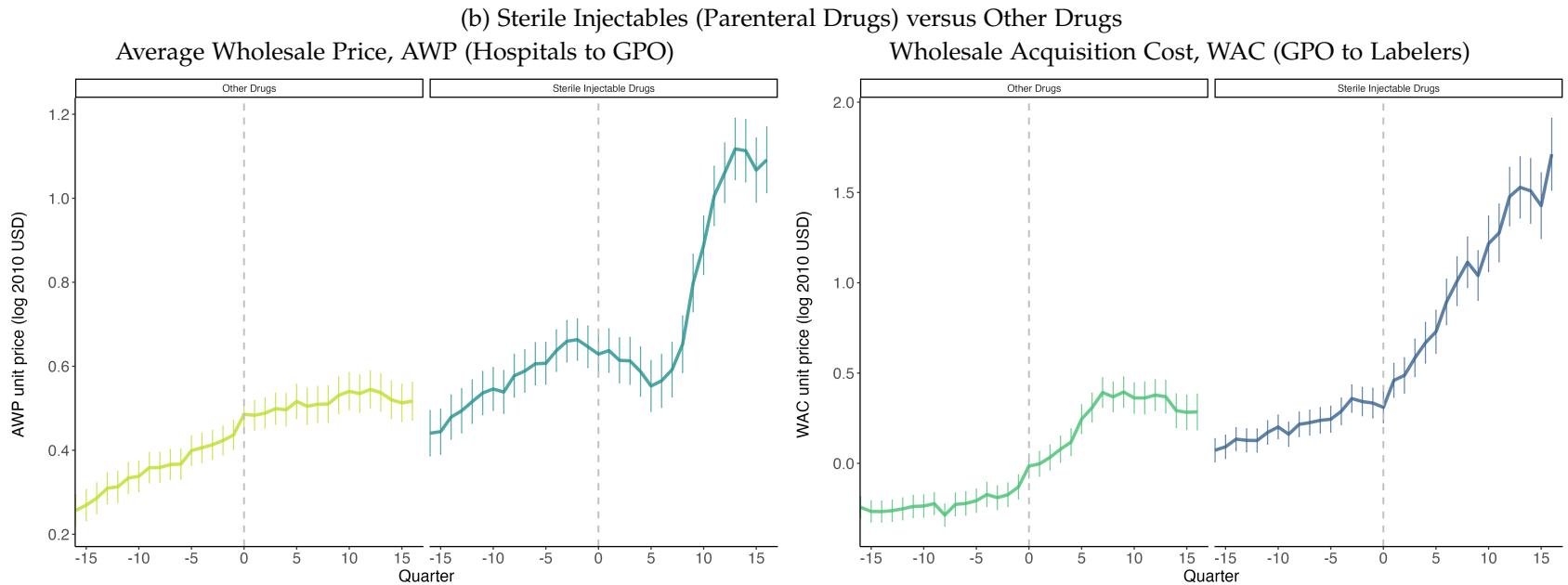
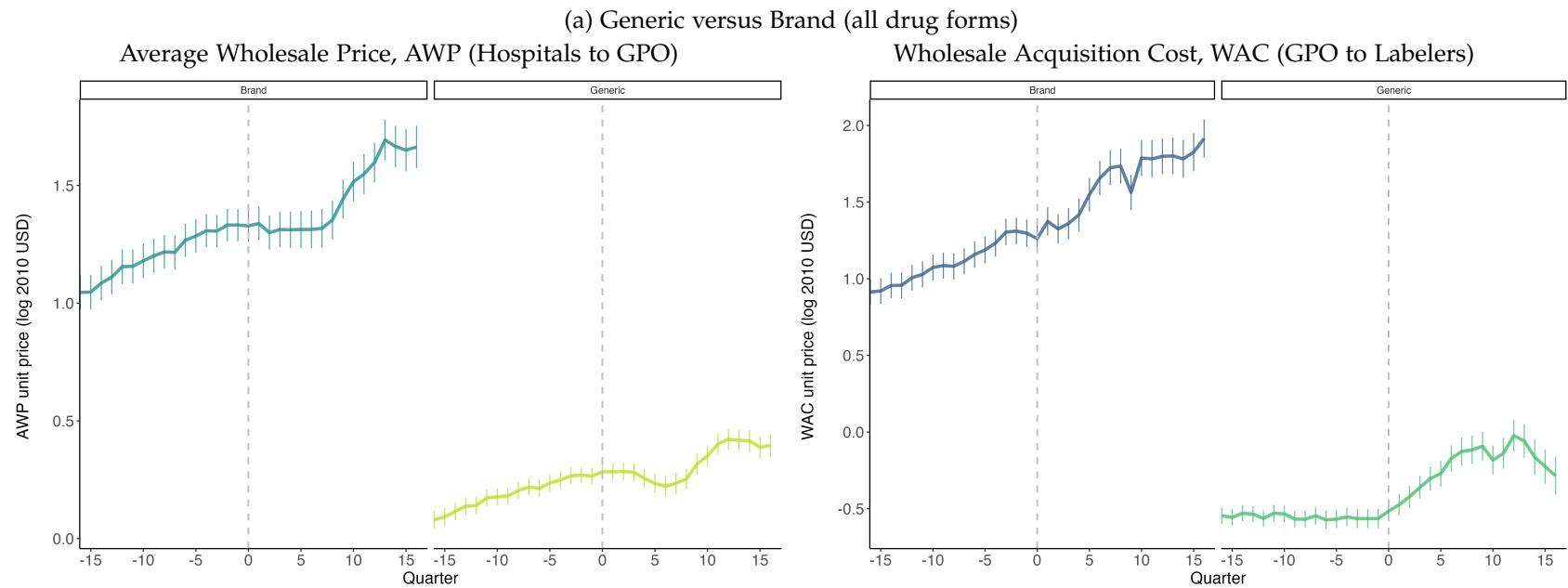


Figure B3: Unit Price Variations (in Logs) Around New Notified Shortages

Notes: Balanced panel of generic injectable drugs in shortage from 2002 to 2019 from the UUDIS. Prices are from IBM Micromedex ("Red Book"). All prices are adjusted for inflation using CPI (base year: 2010). The left panels represent Average Wholesale Price (AWP), which is the price paid by provider hospitals to Group Purchasing Organizations (GPO). The right panels represent Wholesale Acquisition Cost (WAC), which is the price paid by Group Purchasing Organizations (GPO) to Drug Labelers. The top panels compare branded drug products to generic drug products (across all drug forms, not only injectables). The bottom panel compares sterile injectable drugs (also called "parenteral" drugs) to other non-parenteral drugs.

B.3. European Shortages

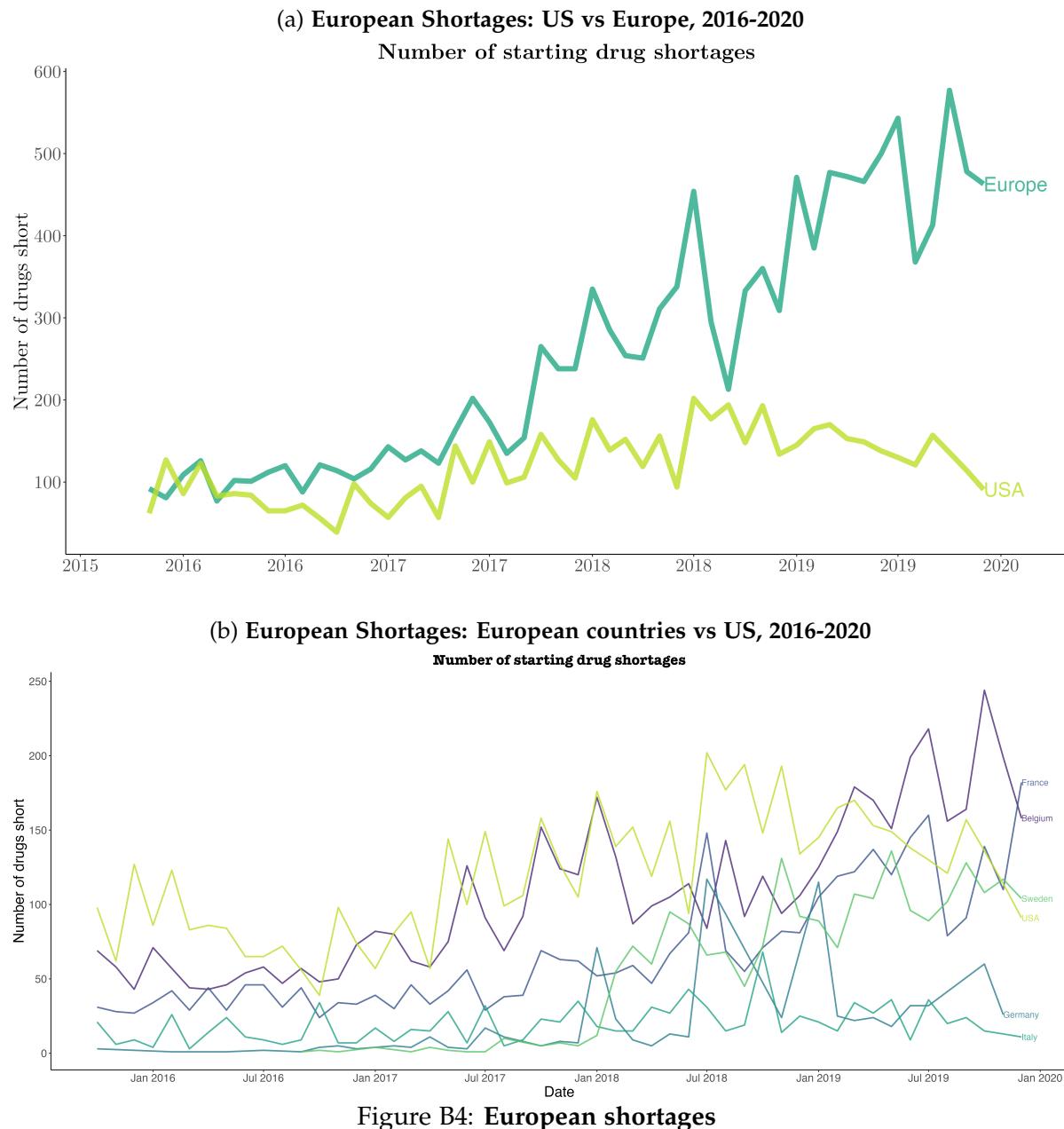


Figure B4: European shortages

B.4. Correlations between Offshoring and Shortages

An offshoring increase is correlated with a higher probability of shortage *ex-post*. Figure B5 then reports the coefficients of a fixed-effects panel regression of shortage on offshoring, with the purpose of assessing one question: is a marginal increase in the share of foreign-based manufacturers for a drug correlated with a higher probability of this drug going short in the *subsequent* period?

To answer, I calculate annual variations in each drug's foreign manufacturer share and categorize it using a dummy variable: 1 signifies an at least 10% increase; 0 signifies a decrease or no change. I

then evaluate changes between $t - (n + 1)$ and $t - n$ for $n \in 1, 3$, just prior to a shortage. Looking at the probability a drug goes into shortage at time t conditional on changes in offshoring rates in years just prior to the shortage allows testing whether variations are influenced by the 2011 peak of offshoring. Because the timing and length of offshoring events vary across drugs, regression coefficients in panel data models can indeed be contaminated by effects from other periods, and correlations can arise solely from heterogeneity in the timing of offshoring (see Section B.5 for a further resolution of these issues). In each case, I regress a shortage dummy on an offshoring dummy, covariates, drug, and time fixed-effects, first using data from 2004 to 2019, then only from 2013 to 2019. The fixed-effects panel regression equivalent is given by the equation:

$$\text{Shortage}_{jt} = \beta \text{Offshored}_{jt} + \alpha_j + \xi_t + x'_{jt}\gamma + \omega_{it} \quad (17)$$

where the dependent variable is an indicator variable equal to 1 if a given drug j is in shortage at time t , Offshored_{jt} is an indicator variable equal to 1 if the drug offshoring rate increases between time $t - n$ and t . x_{jt} are time-varying drug-level covariates (log mean unit price for a drug as measured by average wholesale price, age of the drug, average number of manufacturers), ξ_t are time fixed-effects, α_j are drug fixed-effect. Drug products are defined at the molecule-form level here (not at the manufacturer level).

Figures B5 plot the β coefficients from regression 19. Taking all foreign-based facilities, an increase in the share of foreign manufacturers in the last 1 to 3 years is strongly correlated with a higher probability of the drug being in shortage at time t . The share of facilities located in Asia is overall higher than when taking all foreign facilities together, suggesting a higher correlation when manufacturers choose locations with lower marginal costs of production. Regression coefficients are reported in the appendix Tables B6 and B5.

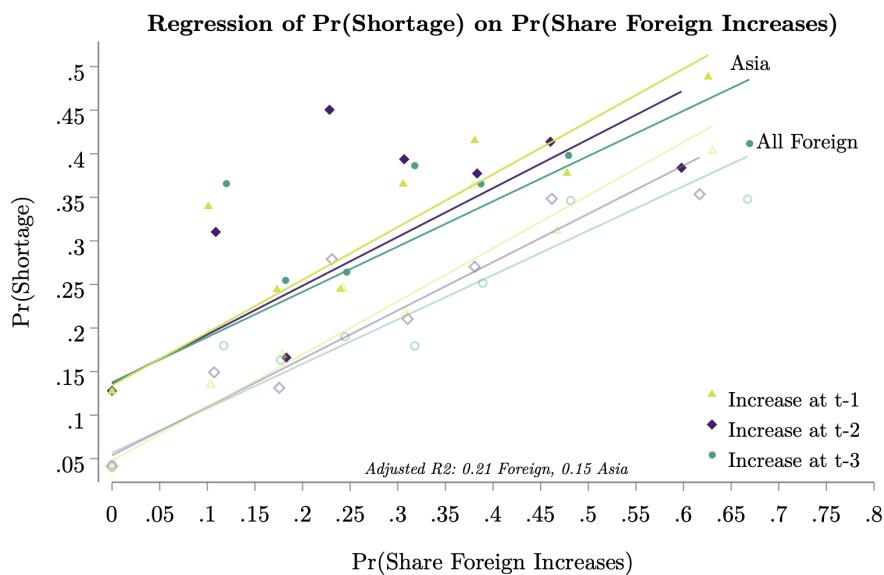


Figure B5: Shortage events are correlated with offshoring increase

Table B4: Probability of shortage on Medicare Demand

Dependent Variable:	P(Shortage=1)			P(New Shortage=1)			Length Shortage (Days)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log Annual Allowed Charges	-.04277*** [.00606]	-.04896*** [.00876]	-.06564*** [.01127]	-.00343 [.00261]	-.00524 [.00379]	-.00726 [.00534]	-81.84*** [15.9]	-65.06*** [22.71]	-58.81*** [21.88]
Share Foreign Facilities		.1977** [.07643]	-.00506 [.1349]		.02696 [.03579]	.03328 [.06414]		323.2 [214.2]	-488 [543.4]
Share Asian Facilities			.3743*** [.1405]			-.00751 [.07777]			1013* [577.1]
St. Dev.	.3598	.3817	.4396	.1552	.1652	.2081	574.6	629.7	606.7
Year & Drug Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Beta Coefficient	-.3627*** [.05135]	-.4152*** [.07427]	-.5565*** [.09558]	-.04204 [.03198]	-.06414 [.04641]	-.08881 [.06533]	-.3161*** [.0614]	-.2513*** [.0877]	-.2271*** [.08448]
OLS Coefficient	-.0011 [.00187]	-.00032 [.00267]	-.00819** [.00357]	.00005 [.00119]	-.00028 [.00158]	-.00309 [.00244]	20.89** [8.216]	53.21*** [11.93]	54.43*** [11.81]

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parenthesis.

Balanced Panel of Generic Sterile Injectable Drugs, 2004 to 2019. All regressions include drug and year fixed effect.

[†]Annual Allowed Charges, Drug Total (in log). Data from CMS, Medicare Part B National Summary Data File, in 2010 dollars

Table B5: Probability of shortage as a function of whether the share of foreign manufacturers increases- 2004 to 2019

	P(Shortage=1)					P(New Shortage=1)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Share Foreign Up at t-1 (t-1 - t-2)	.1817*** [.01354]			.1946*** [.01583]	.1945*** [.01876]	.04698*** [.00848]			.04584*** [.01005]	.03631*** [.01198]
Share Foreign Up at t-2 (t-2 - t-3)		.166*** [.01444]		.153*** [.01546]	.165*** [.01856]		.05554*** [.00918]		.04974*** [.00982]	.04921*** [.01185]
Share Foreign Up at t-3 (t-3 - t-4)			.1802*** [.01515]		.1538*** [.01768]			.05141*** [.00967]		.04384*** [.01129]
Constant	.1605*** [.00577]	.1734*** [.00623]	.1767*** [.0067]	.1457*** [.00722]	.1298*** [.00927]	.0552*** [.00361]	.05903*** [.00396]	.06148*** [.00428]	.0505*** [.00459]	.04627*** [.00592]
Year & Drug Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R ²	0.031	0.026	0.031	0.060	0.093	0.005	0.007	0.006	0.012	0.016
Observations	5555	5017	4488	4245	3077	5555	5017	4488	4245	3077

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parenthesis are clustered at the drug level.

Balanced Panel of Generic Sterile Injectable Drugs, 2004 to 2019. All regressions include drug and year fixed effects.

The dependent variables are dummies equal to 1 if a (new) shortage is notified at t; average values over thus gives the probability of a (new) shortage

Our independent variables are dummies equal to 1 if the share foreign increases between two defined period

Table B6: Probability of shortage as a function of whether the share of foreign manufacturers increases - 2004 to 2019

	P(Shortage=1)					P(New Shortage=1)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Share Foreign Up (t - t-1)	.17*** [.01284]			.1373*** [.01835]	.1453*** [.02172]	.04651*** [.00796]			.03926*** [.01126]	.05064*** [.01345]
Share Foreign Up (t - t-2)		.1744*** [.01328]		.09552*** [.01682]	.09518*** [.02294]		.05276*** [.00832]		.02154** [.01032]	-.00379 [.01421]
Share Foreign Up (t - t-3)			.1799*** [.01467]		.04401** [.02044]			.0637*** [.00923]		.03387*** [.01266]
Constant	.1553*** [.00543]	.1645*** [.00627]	.1816*** [.00735]	.1552*** [.00654]	.1629*** [.00798]	.0521*** [.00336]	.05466*** [.00393]	.05713*** [.00463]	.05048*** [.00401]	.05113*** [.00495]
Year & Drug Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R ²	0.028	0.033	0.035	0.042	0.055	0.006	0.008	0.011	0.008	0.013
Observations	6075	5133	4184	4757	3570	6075	5133	4184	4757	3570

* p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parenthesis are clustered at the drug level.

Balanced Panel of Generic Sterile Injectable Drugs, 2004 to 2019. All regressions include drug and year fixed effects.

The dependent variables are dummies equal to 1 if a (new) shortage is notified at t; average values over thus gives the probability of a (new) shortage

Our independent variables are dummies equal to 1 if the share foreign increases between two defined period

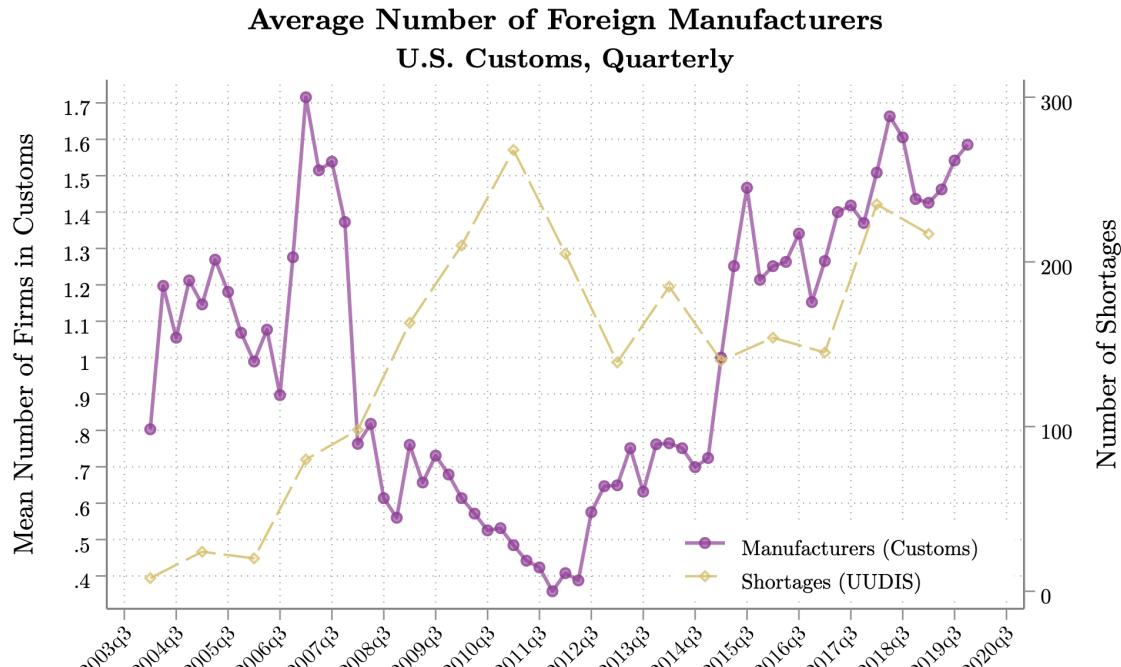


Figure B6: Foreign imports drop during shortages

Notes: This figure plots the average number of manufacturers captured in the U.S. Customs Data (Imports) for sterile injectable drugs that went into shortage (in purple). The peak of U.S. shortages going from 2009 to 2012 corresponds to a sharp drop in the number of foreign manufacturers found in the customs.

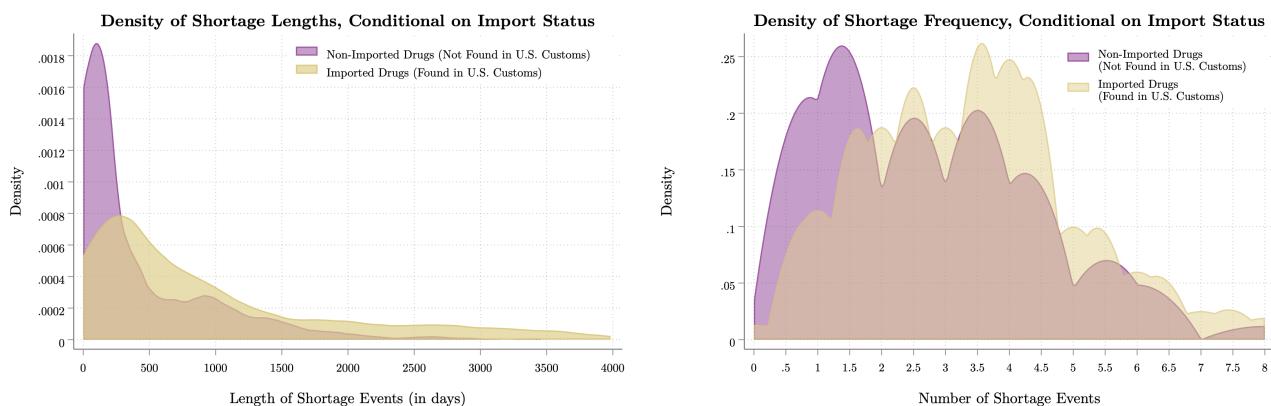


Figure B7: Imported Drugs Face Longer and More Frequent Shortages

Notes: Among all drugs that went into shortage over the 2004-2019 period, imported drugs (as defined by their presence in U.S. imports customs) face longer shortages as measure by the time between the first notification of a shortage and its resolution; 618 days on average (379 median) versus 423 days for non-imported drugs (232 median) on the left panel. On the right panel, imported drugs also face *more frequent* shortage spells compared to non-imported drugs: 3.8 shortages on average (sd 1.7) against 2.1 shortage events (sd 1.4) for non-imported drugs)

B.5. Additional Results: Instrumental Variable Analysis

Table B7: Probability of shortage on share foreign - Instrument: log(PDB)

Dependent Variable:	P(New Shortage=1)	Shortage Days	
Second Stage[†]			
Facilities, Share Foreign	.2088** [.09991]	.2562*** [.06975]	
Facilities, Share Asia		.3828*** [.122]	.4245*** [.08417]
First Stage			
log(PDB) India/China	.32*** [.08532] [.08532]	.2849*** [.06197]	.32*** [.08532]
Year Fixed Effects	✓	✓	✓
Drug Fixed Effects	✓	✓	✓
Std. Dev.	8.286	10.48	5.784
F-stat First-Stage ‡	34.07	41.13	34.07
Beta Coefficient	.3253** [.1556]	.4984*** [.1589]	.6562*** [.1786]
OLS Coefficient	.01973** [.00966]	.07312*** [.0129]	.01324** [.00645]
Observations	16576	16576	16576

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bootstrap SE in parenthesis. [†]Instrument: Predicted Disease Burden (log). F statistics is statistical significant following the "tF" test procedure given by Lee et al.(2020) for weak IVs

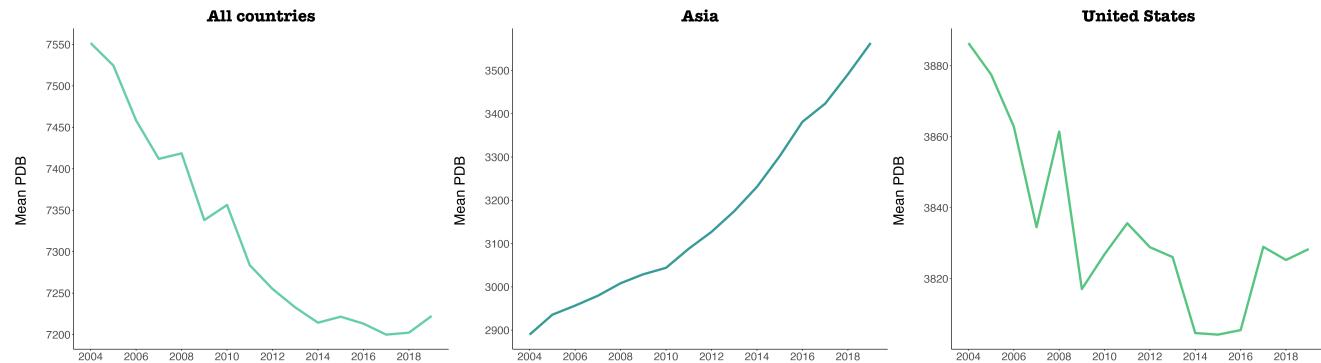


Figure B8: Variations in Predicted Disease Burden Measures

Notes: Variations in the Predicted Disease Burden Measures over time and across countries. The predicted disease burden measure, detailed in equation (1), captures the average country-level disease burden that would be expected given a country's demographic structure. This ratio varies at the country-disease-year level and leverages cross-country and cross-time variations in demographic structures and healthcare access.

Appendix C - A Matched Dynamic Differences-in-Differences Framework

The constellation of evidence presented in Sections 2 to 3 suggests that offshoring of a generic drug manufacturing process is linked to an increase in the drug's shortage probability. These correlations, however, are based on fixed effect panel regressions that may be subject to endogeneity and selection issues and do not establish causal relationships.

In this section, I refine this analysis, using a matched dynamic difference-in-differences (DID) design on a subset of data to estimate the partial equilibrium impact of shifting from domestic to offshore production on drug shortages. In order to address selection of drug products and manufacturing firms into offshoring and the endogeneity of the offshoring decisions, I match sterile injectable drugs whose manufacturing process was offshored with similar drugs that were never produced abroad. This matching, using a Coarsened Exact Matching (CEM) algorithm, controls for selection on observables, while the DID design address potential concerns about selection on unobservables.

The goal of this exercise is to compare closely similar drug products. Previous sections compared drugs with similar administration routes, but these may still vary widely in therapeutic classes, ages, prices, or market concentration. Here, I employ additional observable drug characteristics to create comparable product pairs before offshoring, investigating the pattern of shortages after one drug is offshored and the other is not¹⁰³. The implicit identifying assumption is that, absent the offshoring, each pair of matched drug should have shown similar shortage patterns in the post-period. The DID event study then scrutinizes the evolution of shortages around the offshoring year, and the parallel evolution of treated and control units preceding the increase in foreign production, enabling control for time-invariant heterogeneity across drug products and manufacturers.

C.1. Treatment

Data Sample. While my full dataset consists of a balanced panel of 16 years (january 2004 to december 2019) and 1220 generic sterile injectable drugs (543 of which witnessed at least one shortage event over the period and 677 of which never went short), this difference-in-difference framework relies on a subset of this panel for which I have extended information on observable drug characteristics¹⁰⁴. This subset is constituted of 712 sterile injectable drugs, 543 of which went into shortage over the period. For these 543, I have information on AHFS pharmacologic-therapeutic categories through the UUDIS' drug shortage dataset¹⁰⁵. The remaining 169 drugs are sterile injectable drugs that never went short and that I identified as sterile injectables using the Orange Book data. For this subset of 169 drugs (over 677), I was able to extract therapeutic categories from annual AHFS manuals of therapeutic classifications

¹⁰³For the 543 sterile injectables drugs in shortage in my original dataset, I added 677 sterile injectable drugs that never went short, in order to analyze differences in pricing, concentration and share of foreign manufacturers for drugs with similar characteristics.

¹⁰⁴This analysis will soon be extended to latest version of my dataset, which includes extended information on drug characteristics for all drug products

¹⁰⁵The UUDIS monitors drug shortages reported to the American Society of Health-System Pharmacists

that reports classes for 1761 drugs¹⁰⁶, then matching therapeutic categories back to my dataset based on drug names, active ingredients and route of administration.

Defining Treatment and Control Groups. In order to estimate the causal impact of an increase in offshoring of drug manufacturing facilities on the probability of the drug going into shortage, I define treatment and control groups based on yearly changes in their share of foreign manufacturers¹⁰⁷. Here, the data structure is characterized by a balanced panel of information on treated and non-treated observations, in which treatments have varying start dates and duration.

I considered different definition of treatment, at the drug level, for robustness. The two options I consider are to either i) identify drugs that were never offshored or ii) identify drugs that were always mostly produced in U.S.-based facilities. In particular, my base specification identifies control drug units as product that were never produced abroad ("never offshored") and define treated units as any drug product that was "offshored" over the period. This results in 477 control drugs and 743 treated drugs. For robustness, I also specify alternative groups based on the maximal offshoring share for each drug (defined as a drug's maximal share of foreign-based manufacturers as observed in the data). I use two different specifications based on different offshoring thresholds for the control group: (i) drugs that never had more than 10% of offshored manufacturers over the period 2004-2019 (489 control drugs versus 731 treated) and (ii) Drugs that never had more than 20% of offshored manufacturers over the period 2004-2019 (506 control drugs versus 714 treated).

Defining an Offshoring Event. Offshoring events correspond to situations during which a given drug product, initially manufactured in the U.S. (at "Home"), then starts to be manufactured in a foreign-based locations ("Abroad"). I define a time dummy for the start of the offshoring event (the "treatment") and test several specifications in order to identify the "main" offshoring year for a drug. In the base specification, I compute for each drug the annual growth rate in offshoring shares and select the year displaying the highest growth rate as the main offshoring event. This is the time at which the treatment "starts". In a second specification, I select the second year with the highest offshoring growth rate as the start of the treatment. Eventually, I consider scenarios defining the start of the treatment based on a offshoring threshold: treatment "starts" the year offshored production reaches either 30, 40, 50 or 60% of total production.

Based on these different thresholds, I generate a time dummy equals to one for the treated group in the post-treatment period. This dummy always take a value of zero for controls, and for treated units in the pre-treatment period.

C.2. Matching Procedure

In order to generate an appropriate control group for offshored (i.e. treated) drug that holds the parallel trends assumption, I apply a Coarsened Exact Matching (CEM) on the pre-treatment period.

¹⁰⁶The full restricted AHFS classification covers almost all marketed drugs in the U.S. and includes more than 200,000 National Drug Codes (NDCs).

¹⁰⁷Note that a future version of this work will weight these shares by volumes of production, using imported volumes from the U.S. Customs Data and inferring volumes produced by U.S.-based facilities using demand and manufacturers' market shares from Medicare Part B data

The principle of CEM is to temporarily coarsen each variable used for the match into substantively meaningful groups, exact match on these coarsened data and then only retain the original (uncoarsened) values of the matched data for the regression analysis. CEM places observations into strata that are the foundations for calculating the treatment effect. CEM allows of choosing the coarsening *ex ante* and thus to better control the amount of imbalance in the matching solution. Matching inescapably comes with trade-offs: larger bins (more coarsening) will result in fewer strata. Fewer strata will result in more diverse observations within each strata and, thus, higher imbalance. Matching involves trimming observations that have no close matches on pre-treatment covariates in both the treated and control groups. Among advantages of CEM is typically less model-dependence, lower bias, and (by removing heterogeneity) increased efficiency ([Iacus, King, and Porro, 2008](#); [Ho et al., 2007](#)).

I start with a balanced panel of 19,520 drugs \times year events and identify drugs which were never produced abroad over the study period as my main control group specification. I match drugs on a number of variables. Firstly, I exact match each offshored drug to a "counterfactual" set of drugs in i) the same route of administration (sterile injectables) and ii) the same 2-digits AHFS therapeutic category¹⁰⁸ (e.g.: cardiovascular drugs, antiprotozoals, blood derivatives, hormones...). In levels, I then impose treated and control drugs to belong to: (i) the same quartile of annual number of manufacturers in the three years before the event, (ii) the same decile of age, as measured by the first year of a drug's approval in the Orange Book and (iii) the same decile of log mean unit price in each of the three years before the event (using Average Wholesale Prices from the Red Book, after adjusting prices based on the 2010 CPI). I find for each treated drug at least one control drug in one of the associated control categories.

Across my different specifications for treatment groups and time of treatment dummies, I obtain an average total of 79 treated drugs and 104 controls, each treated drug having on average 1.3 associated control drugs. I do not restrict my drug-level matching to be one-to-one and keep all control plants drugs that meet the matching criteria to a treatment¹⁰⁹. To make sure I only compare drugs that are comparable, I drop treated drugs that do not have common support in the control groups for some observed characteristics. Table C2 compares moments of treated drugs and their matched controls before and after coarsened matching and confirms nonparametrically that treated and control units share similar characteristics.

C.3. Empirical Framework

It is not obvious that drugs whose manufacturing processes have been offshored to foreign-based facilities would have higher probability of shortages compared to U.S.-based manufactured drugs. Consider the following basic shortage equation:

$$\text{Shortage}_{ijt} = \beta \text{Offshored}_{jt} + z'_{jt}\gamma + x'_{jt}\delta + \epsilon_{jt} \quad (18)$$

¹⁰⁸There are 31 two-digits therapeutic categories in total.

¹⁰⁹A unique counterfactual drug for each treatment could have been selected based on either a propensity score or a one-to-one Coarsened Exact Matching (CEM), thereby eliminating the necessity of data weighting for balanced samples. However, due to sample size constraints, all potential controls were retained.

where Shortage_{jt} is the shortage status of drug i at time t . Offshored_{jt} is an indicator variable taking a value of 1 if the drug is offshored and 0 if not. z_{jt} are drug manufacturer characteristics, x_{jt} are drug characteristics and ε_{jt} is an error term. Drug characteristics that may affect shortage probability includes drug quality ("purity" of the product, absence of contaminations...), manufacturing facilities locations (e.g. in places more prone to natural disasters), concentration of the manufacturing process (in terms of number of manufacturers or production lines)...Pharmaceutical firm characteristics that may affect disruptions include portfolio of (competing or new) products, financial stability or size. Whether a drug is offshored or produced in domestic-based facilities may affect the probability of supply shortages for various reasons. If markets are not perfectly competitive and generic drug quality is difficultly observable, then offshoring may allow for lower-standard manufacturing practices. For example, offshoring could lead to more frequent disruptions due to costs cutting on production lines and lack of facility updates, which could be allowed by less stringent enforcement of regulations and controls by regulatory agencies outside of the national territory.

This equation could be estimated using OLS, but offshoring status is likely correlated with firm and drug characteristics. Although panel data regressions may help control for drug characteristics via drug fixed effects, I miss information on firms characteristics (as for instance product portfolios). In section ??, I relied on drug fixed effects regressions with panel data to estimate differences in shortage probabilities when the share of foreign-based manufacturers of a drug increase, following the below specification:

$$\text{Shortage}_{jt} = \beta \text{Offshored}_{jt} + \alpha_i + \gamma_t + x'_{it}\delta + \varepsilon_{it} \quad (19)$$

where Offshored_{jt} is an indicator variable taking a value of 1 if the drug offshoring rate increases between time $t - 1$ and t (or $t - 2$ and t or $t - 3$ and t according to specifications) and 0 if not. Using this method, we identify the impact of offshoring on shortages using the movement of drugs between domestic and foreign status, rather than through the timing of offshoring. The downside of this approach is that there is more potential for selection into which drug becomes offshored.

To solve these pitfalls, I provide alternative estimates of the effect of offshoring using a dynamic event-study design around offshoring events. My dynamic event-study framework exploits my full history of data for the treatment and control groups by estimating regression models of the form:

$$\text{Shortage}_{jt} = \sum_{k=-4, k \neq -1}^{10} \beta_k \mathbb{1}(t = t^* + k) \times \text{Offshored}_j + \gamma_t + \text{Offshored}_j + \alpha_{match} + x'_{it}\delta + \varepsilon_{it} \quad (20)$$

where Shortage_{jt} is an indicator variable for shortage status for drug i in year t and Offshored_j is an indicator for whether the drug was offshored in year t^* . ξ_t are year fixed-effects to control for year-level shocks that could affect all drugs and firms, x_{jt} are drug-level time varying control and ε_{jt} is an error term. α_{match} is a strata fixed-effect (or "matched group" fixed effect), which ensures that I am estimating the effect of offshoring only within a group of matched treated and control drugs. The indicator for time, $D := 1\{t = t^* + k\}$ for unit i being k periods away from treatment time t^* at calendar time t . In this specification, for never-treated units, the indicator is set to 0 for all k and all t . Standard errors are clustered at the matched strata level¹¹⁰. In this specification, identification comes solely from

¹¹⁰Clustering at the drug level yields similar standard errors

differences between control and treated units within a narrow matching group, and not from units coming in and out of treatment.

Matching on route of administration, therapeutic class, age, price and number of manufacturers allows to find drugs that would plausibly exhibit common trends in the absence of offshoring activity. This matching strategy is similar to a number of recent papers implementing a dynamic difference-in-differences research design (Jager, 2016; Goldschmidt and Schmieder, 2007; Borusyak and Jaravel, 2017; Le Moigne, 2020). Using a *matched* control group that is never treated prevents the analysis to suffer from the identification issues that arise in traditional event-study designs with never-treated units Borusyak and Jaravel (2017) or difference-in-differences designs with staggered timing Goodman-Bacon (2018); Sun and Abraham (2020). In particular, researchers often use two-way fixed effects regressions that include leads and lags of the treatment to estimate the dynamic effects of treatment. In settings with variation in treatment timing across units, lead and lag coefficient can be contaminated by effects from other periods, and apparent pre-trends can arise solely from treatment effects heterogeneity.

In the matched specification, identification comes solely from differences in treated and never-treated units over time. In order to avoid multicollinearity (either among the relative period time indicators or with the fixed effects), I exclude some relative periods from the dynamic specification. Excluding relative periods close to the initial treatment is common in practice, so that I normalize relative to the period prior to treatment (exclude $t - 1$).

C.4. Results

The raw means are informative in the matched sample, but there is still possibly selection and time-varying variables at the drug level that affect the changes over time. The matched event study design allows to control for this type of selection by estimating equation (4) controlling for matched drugs fixed effects as well as year-level shocks.

Table C1 reports DID coefficients from equation (4), using as definition of a control group all drugs that were never offshored during the period 2004-2019¹¹¹. Appendix tables ?? and ?? reports robustness checks using different definition of control groups (controls are respectively defined as drugs that never had more than 20% or 30% of offshored manufacturers over the period 2004-2019). The coefficient of interest β can be interpreted as the causal effect of offshoring on the probability of drug shortage in response to a marginal increase in a drug's share of foreign-based manufacturing plants, under the assumption that drug outcomes in the treatment ("offshored") and control group ("not offshored") would have evolved similarly in the absence of offshoring. Figure C1 displays the dynamic event study.

Comparing coefficients from the different columns of Table C1, my differences-in-differences coefficient seems robust to different specifications for the time of treatment dummy. All post-treatment coefficients are positive and strongly correlated with shortages. Offshored drugs are on average 12% more likely to experience shortages following the change in manufacturing locations compared to similar drugs whose production remained U.S.-based. There is also a 6% higher probability of a new shortage being notified just after offshoring for these drugs compared to their counterparts. Moreover, drugs

¹¹¹Meaning their share of foreign-based manufacturing facilities is zero percent over the period

experiencing an offshoring event are 39% more likely than control drugs to be in the panel of drugs that went at least once into shortage over the whole period 2004-2019. Eventually, offshored drugs are strongly correlated with longer shortage events compared to non-offshored drugs - on average 125 days longer across specifications. These results are robust to alternative definitions of treatment groups (see Appendix Tables ?? and ??).

Figure C1 plots the β_k coefficients of specification (4). It shows the difference in shortage outcomes between drugs that were never offshored and drugs that were offshored with a matched-group, pre and post the offshoring event and normalizing to zero at time $t - 1$. The top-left panel regresses shortage on offshoring, for each matched group of drugs, and plots the average differential shortage outcome of treated units. The top-right panel does the same for shortage length (measured in days as the difference between the first reported date of a shortage and the date at which it was officially resolved). In both cases, we see some long-lasting positive effect of offshoring on the probability a drug goes into shortage. This is robust to the specification of the control group: the bottom left and right panels reproduce these plots for alternative definition of a control group (either drugs that were never offshored, or drugs with less than 10 or 20% of offshored production). Independently of the definition of a control group, the probability of a shortage pre and post an offshoring event evolves in similar fashion. In all cases, we see a clear increase in the marginal number of days of additional shortages for treated drugs post offshoring.

Table C1: Average effect of offshoring events on drugs' shortage outcomes
 Control group: Drugs that were never offshored

TREATMENT TIME SPECIFICATION	Highest Offshoring	2^d Highest Offshoring	30% Offshoring	40% Offshoring	50% Offshoring	60% Offshoring
<i>Dependent Variable: Dummy for Shortage</i>						
Post x Treatment	.134*** [.0464]	.121*** [.0361]	.1169*** [.0335]	.114*** [.0319]	.114*** [.03189]	.118*** [.0323]
Adjusted R^2	0.0671	0.0499	0.0722	0.0682	0.0682	0.0714
Mean (treatment)	.115*** [.0082]					
<i>Dependent Variable: Dummy for New Notified Shortage</i>						
Post x Treatment	.057*** [.01786]	.056*** [.0157]	.062*** [.0144]	.058*** [.0129]	.0583*** [.0129]	.060*** [.0131]
Adjusted R^2	0.0262	0.0229	0.0433	0.0382	0.0382	0.0399
Mean (treatment)	.0526*** [.0058]					
<i>Dependent Variable: Dummy for Drug Ever Being in Shortage over 2004-2019</i>						
Post x Treatment	.389*** [.0987]	.382*** [.1087]	.433*** [.0889]	.423*** [.0873]	.423*** [.0873]	.435*** [.0874]
Adjusted R^2	0.1574	0.1401	0.2793	0.2654	0.2654	0.2759
Mean (treatment)	.471*** [.0129]					
<i>Dependent Variable: Shortage Length in Days</i>						
Post x Treatment	154.87** [71.98]	129.95* [68.81]	108.77** [48.89]	105.77** [47.97]	105.77** [47.97]	110.87** [42.93]
Adjusted R^2	0.1574	0.0496	0.0565	0.0532	0.0532	0.0567
Mean (treatment)	125.59*** [9.126]					
Observations	1,648	1,648	1,648	1,648	1,648	1,648
Year & Strata Fixed Effects	✓	✓	✓	✓	✓	✓

Notes: $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parenthesis are clustered at the matched strata level. Balanced Panel of Generic Sterile Injectable Drugs, 2004 to 2019. Treatment is defined as any drug that was "offshored" and control units are defined as drugs that were "never offshored". Five different specifications are reported for the time dummy marking the start of the offshoring event. Each column reports a different specification for the time of treatment dummy. In column (1) I compute annual growth in offshoring shares for each drug and select the year with the highest growth rate as the main offshoring event (i.e. the start of treatment). I repeat this and select the second highest growth year as the start of the treatment year in column (2). For columns (3) to (6), I define the start of treatment for a drug as the first year at which the offshoring rate reaches either 30, 40, 50 or 60% (offshoring rate is defined as the share of foreign-based facilities)

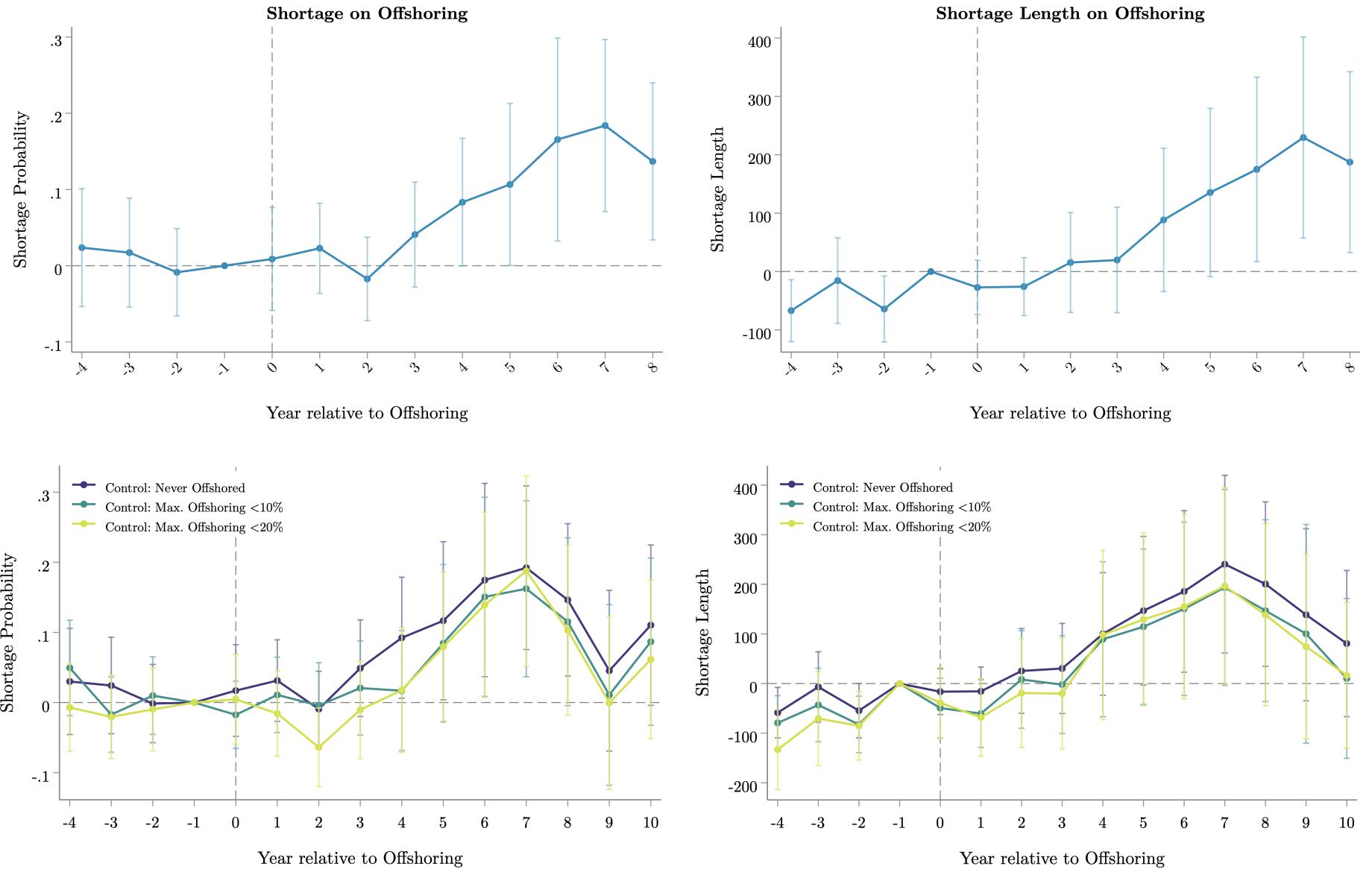


Figure C1: DiD Event Studies : Effect of Offshoring on Shortage Spells

Notes: These figures plot the β_k coefficients of specification 20, using as dependent variable either an indicator equal to one if a shortage occurs (left panels) or the length of shortage events, measured in days as the difference between the first reported date of a shortage and the date at which it was officially resolved (right panels). The top-left panel regresses the probability of shortage on offshoring and plots the average differential shortage outcome of treated units. The top-right panel performs the same regression using shortage length in days as a dependent variable. The bottom left and right panels reproduce these plots for alternative definition of a control group (either drugs that were never offshored, or drugs with less than 10 or 20% of offshored production).

C.5. Additional Summary Statistics on Matching and Treatment and Control Groups

Table C2: **Summary Statistics: treated and control drugs pre and post matching**

	Pre-Matching		Post-Matching	
	Treated	Control	Treated	Control
Mean AWP Unit Price (in log)	2.511 [2.9073]	2.894 [2.9016]	2.473 [1.972]	2.543 [2.036]
Yearly Number of Manufacturers	1.91 [6.8580]	5.3873 [3.0357]	2.0875 [3.0133]	3.04 [2.94]
First Year of Approval, Orange Book	1989.353 [11.8573]	1990.389 [10.3248]	1989.804 [10.3452]	1990.103 [10.2386]
Obs.	6,461	1,728	816	832

Notes: Standard deviation in parenthesis. This table compares moments of treated drugs and their matched controls before and after coarsened matching, in order to confirm nonparametrically that treated and control units share similar characteristics. The data is a balanced panel of drugs-year from 2004 to 2019. Results on matching (number of individual drugs matched in each group, number of strata matched and global imbalanced as measure by the \mathcal{L}_1 statistics can be found in Table C3.

Table C3: Summary Statistics on Coarsened Exact Matching Results

OFFSHORING, TREAT. TIME SPEC.	Highest	2^d Highest	30%	40%	50%	60%
<i>Control: Never Offshored</i>						
\mathcal{L}_1 Imbalance Before Match ^a	.71137956					
\mathcal{L}_1 Imbalance	3.537e-15	9.836e-16	3.929e-15	2.910e-15	1.963e-15	4.562e-16
Matched Strata	31	31	37	36	35	31
Matched Treatments	49	44	79	69	66	58
Matched Controls	67	55	75	72	72	54
<i>Control: Offshoring Max 10%</i>						
\mathcal{L}_1 Imbalance Before Match	.64980728					
\mathcal{L}_1 Imbalance	2.652e-15	3.452e-16	1.076e-15	3.316e-15	1.919e-15	1.815e-15
Matched Strata	31	37	41	40	38	33
Matched Treatments	45	46	79	70	66	54
Matched Controls	76	79	87	86	86	78
<i>Control: Offshoring Max 20%</i>						
\mathcal{L}_1 Imbalance Before Match	.59691182					
\mathcal{L}_1 Imbalance	2.585e-15	3.244e-16	8.838e-16	3.813e-15	2.230e-15	3.166e-16
Matched Strata	37	42	45	43	44	40
Matched Treatments	43	40	81	70	69	55
Matched Controls	92	94	104	100	101	89

Notes: This table displays results on matching (number of individual drugs matched in each group, number of strata matched and global imbalance as measure by the \mathcal{L}_1 statistics^b

^aIt is based on the \mathcal{L}_1 difference between the multidimensional histogram of all pre-treatment covariates in the treated group and that in the control group.

^bIt is based on the \mathcal{L}_1 difference between the multidimensional histogram of all pretreatment covariates in the treated group and that in the control group.

Appendix D - Model Details

D.1. Equilibrium Outcomes of the Manufacturer's Profit Maximization Problem: Derivations

At a given time t , manufacturer j 's ex-ante profit maximization problem is:

$$\begin{aligned}
 \max_{q_j, p_j} \mathbb{E}_{\varepsilon_l} [\pi_j(\cdot)] &= \underbrace{\left[1 - F\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right) \right]}_{\Pr(\text{no excess demand})} \underbrace{p_j Q_j^D(\mathbf{p}, \mathbf{X})}_{\mathbb{E}(\text{Revenue} \mid \text{no excess})} + \\
 &\quad \underbrace{F\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right) p_j q_j \mathbb{E}\left[\varepsilon_{l(j)} \mid \varepsilon_{l(j)} < \frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right]}_{\Pr(\text{excess demand}) \quad \mathbb{E}(\text{Revenue} \mid \text{excess demand})} - \underbrace{(c_{jl} q_j + F_j^c)}_{\text{Total Production Cost}} \\
 &= \left[1 - F\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right) \right] p_j Q_j^D(\mathbf{p}, \mathbf{X}) + p_j q_j \int_0^{\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}} y f(y) dy - c_{jl} q_j - F_j^c
 \end{aligned}$$

where the expected location-specific capacity yield ε_{jl} due to the ex-post exogenous supply shock, conditional on expectation about positive excess demand is

$$\mathbb{E}\left[\varepsilon_{l(j)} \mid \varepsilon_{l(j)} < \frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right] = \frac{1}{F\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right)} \times \int_0^{\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}} y f(y) dy$$

and the probability that a shortage occurs is defined as the probability that the realized supply of firm j is below demanded levels (excess demand), which is defined as $\Pr(q_j \varepsilon_{l(j)} \leq Q_j^D(\mathbf{p}, \mathbf{X})) = F\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right)$ with $\varepsilon_{l(j)} \sim F(\mu_{\varepsilon_l}, \sigma)$.

The firm's first-order conditions of the firm's problem are:

$$\begin{aligned}
 (1) \quad \frac{\partial \mathbb{E}[\pi_j(\cdot)]}{\partial q_j} &= 0 \Leftrightarrow \\
 &-p_j Q_j^D(\mathbf{p}, \mathbf{X}) \frac{\partial F\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right)}{\partial q_j} + p_j \int_0^{\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}} y f(y) dy + p_j q_j \frac{\partial \int_0^{\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}} y f(y) dy}{\partial q_j} - c_{jl} &= 0 \\
 \Leftrightarrow \quad c_{jl} &= p_j \frac{Q_j^D(\mathbf{p}, \mathbf{X})^2}{q_j^2} f\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right) \\
 &+ p_j \left[\int_0^{\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}} y f(y) dy - f\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right) \frac{Q_j^D(\mathbf{p}, \mathbf{X})^2}{q_j^2} \right]
 \end{aligned}$$

$$\Leftrightarrow \frac{Q_j^D(\mathbf{p}, \mathbf{X})^2}{q_j^2} f\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right) + \int_0^{\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}} yf(y)dy - f\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right) \frac{Q_j^D(\mathbf{p}, \mathbf{X})^2}{q_j^2} = \frac{c_{jl}}{p_j}$$

$$\Leftrightarrow (\text{FOC 1}) \quad p_j^* = \frac{c_{jl}}{\int_0^{\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}} yf(y)dy}$$

$$(2) \quad \frac{\partial \mathbb{E}[\pi_j(\cdot)]}{\partial p_j} = 0 \Leftrightarrow$$

$$\begin{aligned} & -\frac{1}{q_j} f\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right) \frac{\partial Q_j^D(\mathbf{p}, \mathbf{X})}{\partial p_j} \cdot p_j Q_j^D(\mathbf{p}, \mathbf{X}) + \left[1 - F\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right)\right] \left[Q_j^D(\mathbf{p}, \mathbf{X}) + p_j \frac{\partial Q_j^D(\mathbf{p}, \mathbf{X})}{\partial p_j}\right] \\ & + q_j \int_0^{\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}} yf(y)dy + p_j q_j f\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right) \frac{\partial Q_j^D(\mathbf{p}, \mathbf{X})}{\partial p_j} \frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j^2} = 0 \\ \Leftrightarrow & \left[1 - F\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right)\right] \left[Q_j^D(\mathbf{p}, \mathbf{X}) + p_j \frac{\partial Q_j^D(\mathbf{p}, \mathbf{X})}{\partial p_j}\right] + q_j \int_0^{\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}} yf(y)dy = 0 \end{aligned}$$

I can re-write the second FOC using the fact that the partial derivative of the logit demanded quantities with respect to price is $\frac{\partial Q_j^D(\mathbf{p}, \mathbf{X})}{\partial p_j} = -\alpha Q_j^D(\mathbf{p}, \mathbf{X}) (1 - s_j)$:

$$\begin{aligned} & \left[1 - F\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right)\right] \left[Q_j^D(\mathbf{p}, \mathbf{X}) - \alpha p_j Q_j^D(\mathbf{p}, \mathbf{X}) (1 - s_j)\right] = -q_j \int_0^{\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}} yf(y)dy \\ \Leftrightarrow (\text{FOC 2}) \quad & \left[1 - F\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right)\right] \frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j} [1 - \alpha p_j (1 - s_j)] = - \int_0^{\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}} yf(y)dy \end{aligned}$$

Combining FOCs (1) and (2):

$$(1) \quad \int_0^{\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}} yf(y)dy = \frac{c_{jl}}{p_j^*} \quad \text{or} \quad p_{jt}^* = \frac{c_{jlt}}{\int_0^{\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}} yf(y)dy}$$

$$(2) \quad \left[1 - F\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j^*}\right)\right] \frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j^*} [1 - \alpha p_j^* (1 - s_j)] = - \int_0^{\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j^*}} yf(y)dy$$

$$(1) + (2) \quad \underbrace{\left[1 - F\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right)\right]}_{\Pr(\text{no shortage}) = P(q_j \epsilon_l \geq Q_j^D(\cdot))} \underbrace{\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}}_{\text{DSR} \in [0, 1]} \underbrace{\left[1 - \underbrace{\alpha p_j (1 - s_j)}_{\equiv \eta_j(\text{price elasticity})}\right]}_{\text{price-cost ratio}} = - \frac{c_{jl}}{p_j}$$

where the second term is the Demand-to-Supply Ratio (DSR $\in [0, 1]$, with $q_j \geq Q_j^D(\mathbf{p}, \mathbf{X})$). The higher the DSR, the more demand exceeds supply, so the higher price p_j should be by the law of supply and demand¹¹². The share of “excess surplus” can be computed as (1-DSR).¹¹³

From the FOCs, I can rewrite the equilibrium q_j^* as a function of p_j^* , s_j , $Q_j^D(\mathbf{p}, \mathbf{X})$ and c_j :

$$\text{Equilibrium supply} \Rightarrow \underbrace{-\frac{p_j^*}{c_{jl}}}_{\text{price elasticity}} \left(1 - \alpha p_j^* (1 - s_j)\right) Q_j^D(\mathbf{p}, \mathbf{X}) \left[1 - F\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j^*}\right)\right] = q_j^*$$

$$\underbrace{\frac{p_j^*}{c_{jl}}}_{\text{price-cost ratio}} \underbrace{\left[1 - F\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j^*}\right)\right]}_{\Pr(\text{no excess demand})} Q_j^D(\mathbf{p}, \mathbf{X}) = q_j^*$$

and using the fact that $\partial Q_{jmt}^D(\mathbf{p}, \mathbf{X}) / \partial p_{jmt} = -\alpha Q_{jmt}^D(\mathbf{p}, \mathbf{X}) (1 - s_{jmt})$:

$$\Rightarrow \underbrace{-\frac{(1 + \eta_j)}{\text{price elasticity}}}_{\frac{\Pr(\text{no excess demand})}{1 - F\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j^*}\right)}} \left[\frac{\Pr(\text{no excess demand})}{\int_0^{\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j^*}} y f(y) dy} \right] Q_j^D(\mathbf{p}, \mathbf{X}) = q_j^*$$

$\int_0^{\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j^*}} y f(y) dy$ can be interpreted as the expected share of production that survives a supply shock given that this share is less than $Q_j^D(\mathbf{p}, \mathbf{X}) / q_j$ (scenario under which the surviving capacity post-supply shock is insufficient to meet demand). This integral thus represents the expected surviving capacity (as a proportion of initial capacity) in the event of a supply shock that leads to a shortage.

The ratio $\left[1 - F\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j^*}\right)\right] / \int_0^{\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j^*}} y f(y) dy$ in turn represents the comparison of the probability of there being no shortage (no excess demand) to the expected surviving capacity in the event of a supply shock that leads to a shortage. The numerator is the probability of there being no excess demand, which is the likelihood that the surviving capacity after a supply shock is greater than the ratio of demand to supply. The denominator is effectively a measure of the expected shortage in the case of a supply shock. Hence, this ratio is comparing the likelihood of there being no shortage to the expected severity of a shortage, in the event of a supply shock. If this ratio is high, that means shortages are less likely, or that when shortages occur, they are less severe on average. If the ratio is low, that could indicate that shortages are more likely, or that when they do occur, they tend to be

¹¹²Markets with higher DSR are generally expected to display less surplus and higher growth than low DSR markets

¹¹³For instance, if demand $Q_j^D(\mathbf{p}, \mathbf{X}) = 100$ and $q_j = 110$, the DSR=0.9, and surplus $\left(1 - \frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right)$ is 10%. If demand $Q_j^D(\mathbf{p}, \mathbf{X}) = 100$ and $q_j = 150$, the DSR falls to 0.66, and surplus is around 33%.

more severe.

I can rewrite this FOC to get the classical price equal markup over marginal cost form:

$$p_j^* = \underbrace{-\left(\frac{1}{1+\eta_j^*}\right)c_{jl}}_{\text{markup over mc}} \underbrace{\left(\frac{q_j^*}{Q_j^D(\mathbf{p}, \mathbf{X}) \left[1 - F\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j^*}\right)\right]}\right)}_{\text{equilibrium supply adjusted by } 1/\Pr(\text{no shortage})} \quad (21)$$

with $\eta_{jt} = 1 - \alpha p_{jt}^* (1 - s_j)$. If I want to express the optimal supply in terms of the Lerner index :

$$\begin{aligned} L_j &\equiv \frac{p_j - c_j}{p_j} = 1 - \frac{c_j}{p_j} \\ \Rightarrow -\frac{p_j}{c_j} &= \frac{1}{L_j - 1} \end{aligned}$$

I can then rewrite

$$q_j^* = \left[1 - F\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j^*}\right)\right] \frac{1 + \eta_j^*}{L_j - 1}$$

D.2. Derivatives of main functions

(1) In general form, $h(x) \equiv \int_0^x yf(y)dy \Rightarrow h'(x) = xf(x)$

Here, defining $h\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right) \equiv \int_0^{\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}} yf(y)dy$, I thus have:

$$(i) \quad \frac{\partial h(.)}{\partial \frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}} = \frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j} f\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right)$$

$$(ii) \quad \frac{\partial h(.)}{\partial q_j} = -f\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right) \times \frac{Q_j^D(\mathbf{p}, \mathbf{X})^2}{q_j^3} \quad \text{By Leibniz's rule}$$

$$(iii) \quad \frac{\partial h(.)}{\partial p_j} = f\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right) \times \frac{\partial Q_j^D(\mathbf{p}, \mathbf{X})}{\partial p_j} \times \frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j^2} \quad \text{By Leibniz's rule}$$

(2) Differentiating the cumulative distribution function $F(.)$ of the supply shock ε_l with respect to manufacturer j 's choice variables, q_j and p_j :

$$(i) \quad \frac{\partial F\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right)}{\partial q_j} = -f\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right) \times \frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j^2}$$

$$(ii) \quad \frac{\partial F\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right)}{\partial p_j} = f\left(\frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}\right) \times \frac{\partial Q_j^D(\mathbf{p}, \mathbf{X})}{\partial p_j} \times \frac{1}{q_j}$$

(3) Derivative of the demand function from the logit model of demand.

I have from the logit model of demand: $Q_j^D(\mathbf{p}, \mathbf{X}) = M_t \cdot s_{jt} = M_t \cdot \frac{\exp(X_{jt}\beta - \alpha p_{jt} + \xi_{jt})}{1 + \sum_{k=1}^J \exp(X_{kt}\beta - \alpha p_{kt} + \xi_{kt})}$

So differentiating the demand function with respect to j 's own-price gives $\frac{\partial Q_j^D(\mathbf{p}, \mathbf{X})}{\partial p_j} = M_t \cdot \frac{\partial s_{jt}}{\partial p_{jt}}$

Price-elasticities from the logit model are:

$$\eta_{jkt} \equiv \frac{\partial \delta_{jt}}{\partial p_{kt}} \cdot \frac{p_{kt}}{s_{jt}} = \begin{cases} -\alpha p_{jt}(1 - s_{jt}) & \text{if } j = k (\text{own-price elasticity}) \\ \alpha p_{kt}s_{kt} & \text{if } j \neq k (\text{cross-price elasticity}) \end{cases}$$

Thus derivatives of market shares with respect to prices give:

$$(1) \quad \frac{\partial s_{jt}}{\partial p_{jt}} = -\alpha s_j(1 - s_j)$$

$$(2) \quad \frac{\partial s_{jt}}{\partial p_{kt}} = \alpha s_j s_k$$

Hence

$$\begin{aligned} \frac{\partial Q_j^D(\mathbf{p}, \mathbf{X})}{\partial p_j} &= -M_t \alpha s_{jt}(1 - s_{jt}) \\ &= -\alpha Q_j^D(\mathbf{p}, \mathbf{X})(1 - s_{jt}) \end{aligned}$$

D.3. Proof: no minimum is required for the shortage probability to be well-defined

This section provides a proof that I can get rid of the minimum in $F \left(\min \left\{ \frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}, 1 \right\} \right)$ by using a simple proof by contradiction that the manufacturer never chooses $q_j < Q_j^D(\cdot)$. I demonstrate here that, when the equilibrium supply choice q_j^* is to produce below demanded quantity $Q_j^D(p_j^*, p_{-j}^*, X_j, X_{-j})$, there exists another pair of price and quantity that brings a higher expected profit than the current equilibrium, such that the manufacturer has incentives to deviate for sure and $q_j^* < Q_j^D(p_j^*, p_{-j}^*, \mathbf{X})$ never optimal

Recall my original formula for the expected profit of manufacturer j :

$$\begin{aligned} \mathbb{E}_{\varepsilon_l} \left[\pi \left(p_j, q_j; Q_j^D(\mathbf{p}, \mathbf{X}), c_{jl}, \varepsilon_l \right) \right] &= \left(1 - F \left(\min \left\{ \frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}, 1 \right\} \right) \right) p_j Q_j^D(\mathbf{p}, \mathbf{X}) \\ &\quad + F \left(\min \left\{ \frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}, 1 \right\} \right) p_j q_j \mathbb{E} \left[\varepsilon_l \mid \varepsilon_l < \min \left\{ \frac{Q_j^D(\mathbf{p}, \mathbf{X})}{q_j}, 1 \right\} \right] - c_{jl} q_j - F_j^c \end{aligned}$$

with $\varepsilon_l \in [0, 1]$ with $\varepsilon_l \sim F(\mu_{\varepsilon_l}, \sigma)$.

Claim: For any price vector \mathbf{p}_{-j} , observable vector \mathbf{X} , any cost c_{jl} , any fixed cost F_j^c , and any distribution function F with support on $[0, 1]$, manufacturer j choosing (p_j, q_j) such that $q_j < Q_j^D((\mathbf{p}_{-j}, p_j), \mathbf{X})$ is a dominated strategy.

Proof: Fix $(\mathbf{p}_{-j}, \mathbf{X}, c_{jl}, F_j^c, F)$. Choose an arbitrary (p_j, q_j) such that $q_j < Q_j^D((\mathbf{p}_{-j}, p_j), \mathbf{X})$. Since, $q_j < Q_j^D((\mathbf{p}_{-j}, p_j))$, there will always be a shortage, i.e. $F\left(\min\left\{\frac{Q_j^D((\mathbf{p}_{-j}, p_j), \mathbf{X})}{q_j}, 1\right\}\right) = 1$. Thus, the I can write the expected profit function as

$$\mathbb{E}_{\epsilon_l} \left[\pi(p_j, q_j; Q_j^D((\mathbf{p}_{-j}, p_j), \mathbf{X}), c_{jl}, \epsilon_l) \right] = p_j q_j \mu_{\epsilon_l} - \underbrace{(c_{jl} q_j + F_j^c)}_{\text{Total cost}}$$

where $\mu_{\epsilon_l} = \mathbb{E}_{\epsilon_l}[\epsilon_l]$. Now consider an alternative choice of $p'_j > p_j$ such that $Q_j^D((\mathbf{p}_{-j}, p'_j), \mathbf{X}) = q_j < Q_j^D((\mathbf{p}_{-j}, p_j), \mathbf{X})$. For such a choice of p'_j to exist, it is sufficient to show both that Q_j^D is monotonically decreasing in p'_j , for all \mathbf{p}_{-j} and all \mathbf{X} , and that $\lim_{p \rightarrow \infty} Q_j^D((\mathbf{p}_{-j}, p), \mathbf{X}) = 0$, for all \mathbf{p}_{-j} and all \mathbf{X} . Both of these conditions should hold as long as Q_j^D is a well-defined demand function.

Note that it is still the case that $F\left(\min\left\{\frac{Q_j^D((\mathbf{p}_{-j}, p_j), \mathbf{X})}{q_j}, 1\right\}\right) = 1$, and thus the expected profit function is the same as above, i.e. there will also always be a shortage with this new choice of p'_j .

Now compare the expected profit from choosing (p_j, q_j) to the expected profit from choosing (p'_j, q_j) :

$$\begin{aligned} & \mathbb{E}_{\epsilon_l} \left[\pi(p'_j, q_j; Q_j^D((\mathbf{p}_{-j}, p'_j), c_{jl}, \epsilon_l)) \right] - \mathbb{E}_{\epsilon_l} \left[\pi(p_j, q_j; Q_j^D((\mathbf{p}_{-j}, p_j), c_{jl}, \epsilon_l)) \right] \\ &= [p'_j q_j \mu_{\epsilon_l} - c_{jl} q_j - F_j^c] - [p_j q_j \mu_{\epsilon_l} - c_{jl} q_j - F_j^c] = (p'_j - p_j) q_j \mu_{\epsilon_l}. \end{aligned}$$

Since $q_j > 0, \mu_{\epsilon_l} > 0$, and $p'_j > p_j$

$$\mathbb{E}_{\epsilon_l} \left[\pi(p'_j, q_j; Q_j^D((\mathbf{p}_{-j}, p'_j), c_{jl}, \epsilon_l)) \right] - \mathbb{E}_{\epsilon_l} \left[\pi(p_j, q_j; Q_j^D((\mathbf{p}_{-j}, p_j), c_{jl}, \epsilon_l)) \right] > 0$$

Therefore, I can always find an alternative (p'_j, q_j) that dominates (p_j, q_j) when $q_j < Q_j^D((\mathbf{p}_{-j}, p_j), \mathbf{X})$, which is what I were aiming to prove. As a result of the preceding result, it must always be the case that if firms are optimizing, $F\left(\min\left\{\frac{Q_j^D(\mathbf{p}^*, \mathbf{X})}{q_j}, 1\right\}\right) = F\left(\frac{Q_j^D(\mathbf{p}^*, \mathbf{X})}{q_j}\right)$, and the min function can be ignored.

D.4. Discussion about buyer's internalizing shortages risk: thinking about reliability in the logit model.

X_j captures a static buyer's reputation, or quality score. In reduced-form analysis, this can capture a seller's ability to satisfy on-time delivery, which can be based on a measure of past-shortages.

Without adding dynamics into the problem, the question of keeping X_j static in the model is mostly a Heuristics issue. I have to make it clear that the current *behavioral* assumption is that buyers do not internalize shortages when choosing among sellers. X_j is now thought as an exogenous measure of quality (may be information given by the FDA about manufacturer j).

If buyers were to internalize shortages, they should know that a seller offering low prices is linked to higher supply q_j , and that a higher demand for manufacturer j 's products $Q_j^D(\cdot)$ means more severe shortages if the realization of ε_l is bad. Buyer's demand function does not incorporate q_j so it does not directly incorporate shortage risks beyond what is captured by p_j through $Q_j^D(p_j, p_{-j}, \mathbf{X})$. Fundamentally, a buyer should NOT value past shortage history beyond the information this history is providing about future shortage risks, i.e. what the buyer really values is for instance $\mathbb{E}[X_{jt} | X_{jt-1}, X_{jt-2}]$ or written otherwise $\mathbb{E}[X_{jt} | \mathcal{H}_t(\mathbf{X})]$. Thus, a buyer should only value shortage history to the extent it is expected to predict future shortages: $\Pr(\text{Shortage}_t | \text{Shortage}_{t-1})$. So $\text{Cov}(X_t, X_{t-1})$ is what matters?

For now, I assume there is only one buyer (one GPO buying for the entire U.S. market). If I assume instead that I have a continuum of small buyers (each individual hospital is small), then it is reasonable to assume that each individual hospital cannot "move" the market by itself. Thus, if each buyer i is small, they reasonably do not think that an increase in their own demand for products manufactured by j $Q_{ij}^D(\cdot)$ will create a shortage.

This is thus a behavioral assumption that buyers choose manufacturers based on prices and static quality score → this should be kept constant when doing counterfactuals.

Aside note: in my logit demand model, ξ_{jt} may capture risky firms. I may also write a more complicated version in which I include shortage in demand directly (i.e. I allow buyer to directly internalize shortage risks), instead of making it dynamic, by setting $\mu_{\varepsilon_l} = G(X_j, \xi_j)$, that is by mapping the supply shock to one-time delivery.

Appendix E - Estimation Details

E.1. Defining Products and Drug Markets

The final dataset is a balanced panel dataset at the fiscal quarter level for the period January 2002 to December 2019. A unit of observation is a drug manufacturer (defined by a manufacturing plant), molecule-form product, fiscal quarter. Drug products are defined at the FDA's National Drug Code (NDC) 5-4 level (labeler and molecule-dosage-form level, not taking into account commercial package size). We compute for each drug-product the average price over package size. Note that sales (in USD and number of services) displayed in the Medicare Part B data are reported at the HCPCs level, which are service codes that can be linked with NDC 5-4 directly. One HCPCs code generally corresponds to one administration of a drug.

Table E1: Number of NDCs 5-4 and Labelers by Drug market

		p50	mean	sd	min	p25	p75	p95	max	N Markets
Molform	Number of NDCs-54	3	7.55334728	11.31705285	1	1	8	34	70	956
Market	Number of Labelers	2	3.575313808	3.807622199	1	1	5	12	22	956
ATC-4	Number of NDCs 5-4	14	27.96697626	33.45622747	1	5	42	103	182	969
Market	Number of Labelers	8	10.29132231	8.816399539	1	3	16	30	38	969
ATC-5	Number of NDCs 5-4	3	7.150219298	9.338635912	1	1	9	28	61	914
Market	Number of Labelers	2	3.692560175	3.525002142	1	1	5	12	18	914

Notes: The FDA's National Drug Code (NDC) identify drugs using a unique, three-segment number, under the format "5-4-2". The first set of numbers in the NDC identifies the labeler, such as the drug manufacturer, repackager, or distributor. The second set of numbers is the product code, which identifies the specific strength, dosage form (i.e, capsule, tablet, liquid) and formulation of a drug for a specific labeler. Finally, the third set is the package code, which identifies package sizes and types. The NDC 5-4 level thus defines a drug product at the labeler-molecule-dosage-form level, without taking into account the commercial package size. The first two rows of the table define a drug market at the molecule-form level, or "Molform" (a market is thus composed of many labelers). To form markets, I leverage NDC 5-4 drug names. Each unique identifier at the NDC 5-4 level displays several generic and branded drug names across FDA datasets (Orange Book, NDC Directory, FDA listings and Drug Shortages data, RxMix). I create crosswalks in which all NDCs 5-4 that can be found in at least one this dataset under the same drug name are part of the same molecule-form markets. The second four rows of the dataset define a drug market at the Anatomical Therapeutic Chemical (ATC) levels. This classification, created by the World Health Organization, divides active drug substances into different groups according to the organ or system on which they act and their therapeutic, pharmacological and chemical properties. The ATC-4 level groups drugs belonging to the same chemical subgroup (example: the ATC-2 therapeutic subgroup A10 refers to drugs used in diabetes; the ATC-4 chemical subgroup A10BA

E.2. Prices and Reliability Score Measures

Table E2: Correlations between Prices and Reliability Score

	Fraction of Time Not Short in Past Quarter	3 Categories Score
Unit Price (AWP, 2010 USD)	0.0301	0.0332
Unit Price (WAC , 2010 USD)	0.0286	0.0316

E.3. Plant Location-Specific Statistics

Table E3: Shortages, Prices and Sales by plant Location

Plant location	Variable	Median	Mean	Std	N Obs
OECD	Fraction of Time Short by Quarter	0.000	0.071	0.247	59,064.000
	Shortage Dummy	0.000	0.089	0.284	59,064.000
	Never Short Dummy	0.000	0.386	0.487	59,064.000
	Unit Price (AWP, 2010 USD)	6.768	123.155	424.670	59,064.000
	Unit Price (WAC, 2010 USD)	5.858	104.130	363.898	53,586.000
	Package Price (AWP, 2010 USD)	271.270	675.988	1,116.624	58,922.000
	Package Price (WAC, 2010 USD)	216.760	562.212	920.184	53,586.000
	Allowed Services, Medicare B	129,693.000	6217636.336	19673338.652	32,645.000
	Allowed Charges, Medicare B	359,800.156	9130123.612	29437215.085	32,645.000
	Payments, Medicare B	285,100.125	7621867.040	25412284.679	32,645.000
South-East Asia	Fraction of Time Short by Quarter	0.000	0.138	0.329	130,621.000
	Shortage Dummy	0.000	0.169	0.375	130,621.000
	Never Short Dummy	0.000	0.183	0.387	130,621.000
	Unit Price (AWP, 2010 USD)	2.173	43.021	173.338	130,621.000
	Unit Price (WAC, 2010 USD)	0.441	27.076	127.359	128,018.000
	Package Price (AWP, 2010 USD)	168.314	446.732	669.476	130,621.000
	Package Price (WAC, 2010 USD)	57.319	159.507	361.903	128,018.000
	Allowed Services, Medicare B	751,304.600	7690111.939	23180861.470	93,056.000
	Allowed Charges, Medicare B	974,533.562	14905242.461	43466409.027	93,056.000
	Payments, Medicare B	776,760.750	12334883.005	39091898.607	93,056.000
U.S.	Fraction of Time Short by Quarter	0.000	0.115	0.309	76,062.000
	Shortage Dummy	0.000	0.136	0.342	76,062.000
	Never Short Dummy	0.000	0.295	0.456	76,062.000
	Unit Price (AWP, 2010 USD)	0.917	66.115	362.792	76,062.000
	Unit Price (WAC, 2010 USD)	0.637	44.715	268.170	57,817.000
	Package Price (AWP, 2010 USD)	97.928	344.659	852.468	76,044.000
	Package Price (WAC, 2010 USD)	70.075	252.711	648.128	57,817.000
	Allowed Services, Medicare B	223,228.000	2988682.442	17326919.340	37,059.000
	Allowed Charges, Medicare B	188,267.094	13562407.301	60764743.574	37,107.000
	Payments, Medicare B	159,123.156	12158999.404	59253764.036	37,107.000

E.4. Parameter Estimates

Table E4: Parameters Estimates of the Structural Model

	Parameter	Description	Estimate	SE
Demand	α	Price Sensitivity	-0.6106	0.2552
	β	Taste for Reliability	-0.00418	0.00744
	λ	Brand Dummy	-0.01293	0.00035
	μ_{ε_F}	Mean - Supply Shock Distribution (Foreign, South)	-0.9995	0.006
	μ_{ε_H}	Mean - Supply Shock Distribution (Domestic)	-0.62125	0.004
	$\mu_{\varepsilon_{OECD}}$	Mean - Supply Shock Distribution (Foreign, OECD)	-0.44385	0.009
	σ_{ε_F}	Variance - Supply Shock Distribution (Foreign, South)	0.0395	0.0028
	σ_{ε_H}	Variance - Supply Shock Distribution (Domestic)	0.00434	0.0043
	$\sigma_{\varepsilon_{OECD}}$	Variance - Supply Shock Distribution (Foreign, OECD)	0.00690	0.0055
Supply	$\bar{\varepsilon}_F$	Mean Capacity Yield (Foreign)	0.4284	0.011
	$\bar{\varepsilon}_H$	Mean Capacity Yield (Domestic)	0.5603	0.038
	$\bar{\rho}_F$	Mean Shortage Probability (Foreign)	0.2246	0.0041
	$\bar{\rho}_H$	Mean Shortage Probability (Domestic)	0.179	0.039
	\bar{c}_F	Mean Marginal Cost (Foreign)	6.33	0.0027
	\bar{c}_H	Mean Marginal Cost (Domestic)	9.401	0.051
	μ_{γ_F}	Mean - Firm TFP (Foreign, South)	0.4172	0.103
	μ_{γ_H}	Mean - Firm TFP (Domestic)	0.0114	0.008
	$\mu_{\gamma_{OECD}}$	Mean - Firm TFP (Foreign, OECD)	0.6579	0.0223
	$\sigma_{\gamma_H}^2$	Variance - Firm TFP	0.014	0.002
Location Choice	μ_{FC_F}	Mean - Fixed Cost Distribution (Foreign, South)	2.10602	0.0841
	μ_{FC_H}	Mean - Fixed Cost Distribution (Domestic)	2.83722	0.0121
	$\mu_{FC_{OECD}}$	Mean - Fixed Cost Distribution (Foreign, OECD)	2.5585	0.0121
	$\sigma_{FC_H}^2$	Variance - Fixed Cost Distribution (Foreign, South)	0.016	0.0032
	$\sigma_{FC_H}^2$	Variance - Fixed Cost Distribution (Domestic)	0.01	0.0016
	$\sigma_{FC_{OECD}}^2$	Variance - Fixed Cost Distribution (Foreign, OECD)	0.02	0.005
	\bar{FC}_F	Mean Fixed Cost (Foreign)	8.1607	0.0231
	\bar{FC}_H	Mean Fixed Cost (Domestic)	12.1724	0.049
	μ_{F_N}	Markup (Foreign, OECD)		
	μ_{F_S}	Markup (Foreign, South)		
	μ_H	Markup (Domestic)		

E.5. Estimation Details

1. Estimating the Distribution of Supply Shocks: Nested Fixed Point Algorithm

The parameters of the distribution of the shock and the equilibrium supplied quantities are jointly determined, creating a "nested fixed point problem". The outer problem is the estimation of the parameters of the distribution of the shock μ_{ε_l} and σ_{ε_l} , while the inner problem is the computation of the equilibrium quantities $q_{jmt}(\mu_{\varepsilon_l}, \sigma_{\varepsilon_l})$ given these parameters.

I solve this problem using an iterative or "inner-outer" estimation procedure:

- Initialization:** I start with some initial guesses for the parameters of the distribution of the shock $\varepsilon_{l(j)t}$.
- Inner Loop (Equilibrium Computation):** Given these parameters, and previously estimated demand parameters, I compute the equilibrium quantities $q_{jmt}(\mu_{\varepsilon_l}, \sigma_{\varepsilon_l})$ using my fixed point iteration method. Equilibrium quantities are given by the first-order condition of the firm's problem:

$$\frac{-\underbrace{(1 + \eta_{jjt})}_{\text{price elasticity}}}{\underbrace{\int_0^{\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}^*}} \epsilon f(\epsilon) d\epsilon}_{\text{expected share of surviving capacity in the event of a shortage}}} = q_{jmt}^*$$

Pr(no excess demand) $\left[1 - F\left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}^*}\right) \right]$ contracted quantities $\overbrace{Q_{jmt}^D(\mathbf{p}, \mathbf{X}, \xi)}$

- Outer Loop (Parameter Estimation):** Given the computed equilibrium quantities $q_{jmt}(\mu_{\varepsilon_l}, \sigma_{\varepsilon_l})$ from the first step, I estimate the parameters of the distribution of the shock using GMM with moment conditions equating the model-predicted probability of shortage with its empirical counterpart, the fraction of time short within a quarter for each firm in the data.
- Convergence Check:** I eventually check if the estimated parameters have converged (i.e., the change in the parameter estimates is below a certain tolerance). If not, I update the parameter guesses with the new estimates and go back to step 2. I iterate between these two steps until convergence, i.e., until the estimates of the parameters and the equilibrium quantities stop changing significantly between iterations.

2. Estimating the Distribution of Productivity Shocks using a Simulated Method of Moments

Under Revision

E.6. Details on Model Simulation

The process involves solving the optimization problem for each firm sequentially, taking into account the choices of other firms in the market in order to find a Nash equilibrium. I can then iterate over all

firms until we reach a stable equilibrium where each firm's best response function is consistent with the choices of other firms (i.e. until the changes in prices, quantities, and location choices (thus in X_{jmt}) are below a specified tolerance level, at which point I can consider the equilibrium to be reached.). This iterative process aims to capture the strategic interactions among firms in a competitive market setting.

1. **Initialization:** Set initial guesses for location, prices, and reputation (X_j) for each firm in the market. This will allow you to compute initial market shares for all firms using the logit market share formula and then compute demanded quantities (Q_j) based on prices and reputation.
2. **Optimization preparation:** For each firm j , set initial guesses for prices (p_j) and supplied quantities (q_j), taking the current choices of other firms' prices and reputation as given.
3. **Optimization:** Solve the profit maximization problem for each firm j , given the current state of the market (i.e., other firms' prices and reputation). This involves:
 - (a) Computing firm j 's market shares and demanded quantities (Q_j) based on the current values of prices and reputation.
 - (b) Plugging these values, along with the supplied quantities (q_j), into the profit function and optimizing using an optimization library like Optim.jl.
4. **Update:** Iterate through all firms in the market and update their equilibrium prices, supplied quantities, location, and reputation (X_j) based on the optimization results. For each firm, compare the profit in both locations and choose the location that provides the highest profit. This will determine the location-dependent reputation (X_j) for each firm.
5. **Market shares update:** Update market shares for all firms using the logit market share formula, taking into account the updated prices, reputation, and demanded quantities (Q_j).
6. **Convergence check:** Check whether the market shares, prices, and supplied quantities have converged. If not, go back to step 3 with the updated values and repeat steps 3-5.

This iterative approach allows each firm to optimize its profit given the current state of the market (other firms' prices and reputation). It also reflects the real-world situation where firms do not jointly optimize their profits but instead compete with each other by optimizing their individual profits.

Solution approach

- Use a systematic grid of starting points for guesses, based on Latin Hypercube Sampling (LHS). LHS divides each dimension of the search space into equally probable intervals and samples points so that each interval is represented once, ensuring better coverage of the search space. This method is particularly useful for high-dimensional problems.
- Add sensitivity analysis, varying initial values and parameter values within plausible ranges and checking that the iterative process converges and a fixed point is reached.

- Global optimization, using Particle Swarm Optimization¹¹⁴

E.7. About reintroducing unobserved heterogeneity ξ_{jmt}

The estimation of my logit demand model does not incorporate a Berry-style unobserved heterogeneity term, ξ_{jmt} . Instead, I rely on the rich structure of my panel data to add market fixed effects that controls for unobserved heterogeneity across molecule-quarter markets.

Compared to a standard logit demand model, the necessity to solve for endogenous market shares and censoring add here several estimation issues. I can easily recover the ξ_{jmt} parameters using the standard logit inversion procedure for all firms not being shorts (ξ_{jmt} is simply the residual from the logit model, and captures the difference between the observed market shares and the logit predicted shares). However, I cannot do the inversion for firms being short. In this case indeed, the observed shares resulting from the observation for sold quantities do not correspond to the classical demanded shares; these shares are endogenously affected by supply shocks. This is a typical censoring issue; I only observe the "logit shares" in absence of shortages, and I observe a truncation of these shares for firms that are short, due to conditions of excess demand.

I discuss here a solution that could be implemented to recover the ξ_{jmt} terms inside logit models affected by censoring issues. This solution would involve parametrizing the distribution of the ξ_{jmt} and carrying a Heckman-type correction to estimate the deviation from the "true" ξ_{jmt} during shortage spells. I could assume that unobserved heterogeneity follows a normal distribution $\xi_{jmt} \sim \text{Normal}(\mu_\xi, \sigma_\xi)$ (which I can recover in the absence of shock via logit inversion). During shortage spells however, we would only observe $\xi_{jmt} + r_{jmt}$ where r_{jmt} is some extra shortage-induced costs (for instance, a restocking cost):

$$\xi = \begin{cases} \xi_{jmt} & \text{when } q_{jmt}\varepsilon_{lt} \geq Q_{jmt}^D(\mathbf{p}, \mathbf{X}) \text{ [no shortage]} \\ \xi_{jmt} + r_{jmt} & \text{when } q_{jmt}\varepsilon_{lt} < Q_{jmt}^D(\mathbf{p}, \mathbf{X}) \end{cases}$$

And we can compute conditional expectation easily in closed form for normally distributed variables, as a function of the inverse Mills Ratio:

$$\mathbb{E}[X|X < Y] \quad \text{when } X \sim \text{Normal}(\mu, \sigma) \text{ is:} = \mu + \sigma \underbrace{\left(\frac{\phi((Y - \mu)/\sigma)}{1 - \Phi((Y - \mu)/\sigma)} \right)}_{\text{Inverse Mills Ratio}} \quad \text{with } \phi \text{ PDF and } \Phi \text{ CDF of the Normal}$$

I would thus observe $\mathbb{E}[\xi_{jmt}]$ absent of shortages and $\mathbb{E}[\xi_{jmt}|\text{Short}_{jmt}, \text{params}]$ when there is a shortage.

Implementing a Heckman correction version would then follow two steps:

Step 1: Selection Equation First, we estimate a model that predicts the probability of observing ξ_{jmt} (i.e. the probability that there is no shortage). This is directly given in my case by the probability

¹¹⁴Particle Swarm Optimization is a computational method that optimizes a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality. It solves a problem by having a population of candidate solutions, here dubbed particles, and moving these particles around in the search-space according to simple mathematical formula over the particle's position and velocity.

of shortage:

$$Pr(\text{shortage})_{jmt} = Pr\left(q_{jmt}\varepsilon_{lt} < Q_{jmt}^D(\mathbf{p}, \mathbf{X})\right) = Pr\left(\varepsilon_{l(j)t} < \frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}}\right) = F\left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}}\right)$$

where $F(\cdot)$ is the CDF of ε which follows a Logit Normal distribution with mean μ_ε and variance normalized to 1.0.

From this model, I must calculate the Inverse Mills Ratio (IMR) for each observation. The IMR is normally given by the ratio of the PDF to the CDF of the standard normal distribution, evaluated at the predicted values from the selection model. But in my case, ε follows a Logit Normal distribution, not a standard normal distribution. The IMR for a non-standard normal distribution (or any other distribution) is not as straightforward to calculate as it is for a standard normal distribution. In the case of a Logit Normal distribution, there is not a simple closed-form expression for the IMR. Instead, I would need to compute the ratio of the PDF to the CDF numerically for each observation, using the specific parameters of the Logit Normal distribution:

$$\lambda(Z_{jmt}) = \frac{f\left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}}\right)}{F\left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}}\right)}$$

Step 2: Outcome Equation

Next, I must include the IMR as an additional variable in the logit model. The modified logit model would look something like this:

$$v_{ijt} = X_{jmt} * \beta - \alpha * p_{jmt} + \lambda g_{jm} + \xi_{jmt} + \delta \cdot \lambda(Z_{jmt}) + \epsilon_{ijt}$$

where $\lambda(Z_{jmt})$ is the IMR calculated in step 1, and δ is the coefficient on the IMR, which measures the extent of selection bias.

By including the IMR in the model, I could control for the selection bias that arises due to the observation of ξ_{jmt} only when there is no shortage.

Appendix F - Counterfactuals Details

F.1. Re-shoring Counterfactual: Details

1. Political and Institutional Context of Reshoring Policies

From 1976 to 2006, the Internal Revenue Code exempted from taxation corporate income generated in U.S. territories. The tax break was estimated to have allowed drug companies to save the 2022 equivalent of \$235 million per year, which represents 70 cents per American. The end of the tax break in 2006 incidentally coincided with the start of generic drugs' offshoring.

Recent public news have emphasized the need to reshore to such U.S. territories, and in particular to Puerto Rico to alleviate U.S. drug shortages (*Washington Post, January 2023*). As a result of the pre-2006 tax break, Puerto Rico indeed has idle plants, a trained pharmaceutical workforce and relatively low wages compared to the mainland United States. Because Puerto Rico is on U.S. land, the U.S. regulator (the FDA) also does not meet the same impediments as abroad when it comes to implementing inspections.¹¹⁵ There are however a couple of issues related to "near-shoring" to Puerto Rico specifically. Firstly, Puerto Rico is subject to natural disaster: in September 2017, Hurricane Maria hit Puerto Rico, causing more than 3,000 fatalities and millions of dollars in damage.¹¹⁶ The Hurricane destroyed most drug manufacturing plants that remained in Puerto Rico and pushed firms to offshore entirely production. Secondly, Puerto Rico has been subject to many recent power failures and the fragility of the electricity infrastructures makes it difficult and risky for manufacturers to install their plants on the island. While re-shoring to Puerto Rico may help in reducing the frequency of shocks while keeping costs relatively lower than in the mainland U.S., it thus also introduces a potential for supply disruptions that may be rarer, but bigger in magnitude. I thereby here consider policies encouraging re-shoring to less risky locations within the mainland United States.

Since 2020, the U.S. government has been discussing potential tax deductions of up to 10 percent of an organization's revenue for reshoring production. Republican Brad Wenstrup first carried the *Domestic Medical and Drug Manufacturing Tax Credit*, a legislative proposal that would lower the tax rate on the income from the domestic manufacturing and sales of active pharmaceutical ingredients and medical countermeasures. The bill would provide manufactures a credit of 10.5 percent of the net income from the sale of important medical products manufactured domestically, thus effectively cutting their corporate tax rate in half. In March 2021, President Joe Biden's American Jobs Plan further promised to invest \$300 billion for U.S. manufacturing, including \$50 billion to strengthen supply chains for critical goods and \$30 billion to protect from future pandemics.¹¹⁷ The initial Health

¹¹⁵In particular there is no need for visa and the FDA's inspectors do not need to announce their visits months in advance as they do not require foreign approval and firms' sponsorship to cross borders.

¹¹⁶Several pharmaceutical companies with manufacturing and distribution operations on the island were hit hard, including Pfizer, Amgen and Bristol-Myers Squibb.

¹¹⁷The White House's June 2021 100-day review of America's Supply Chains: "Building Resilient Supply Chains, Revitalizing American Manufacturing and Fostering Broad-Based Growth"

and Human Services response to the executive order is the proposed formation of a public-private consortium for advanced manufacturing and onshoring of domestic essential medicines production, as well as a commitment to invest \$60 million from the Defense Production Act to develop novel platform technologies to increase domestic manufacturing capacity for API.

In France, President Emmanuel Macron announced in June 2023 that 50 crucial medicines would be reshored to France, in Ardeche, with the explicit hope to battle shortages of items like antibiotics and paracetamol. Under Macron's plan, \$173 million in public money will go to support eight of the new reshoring projects, including for amoxycillin – produced by British pharma giant GSK in northwestern France – as well as anaesthetics, painkillers and cancer drugs. Drug companies reshoring to France would also be able to apply for a share of a further 50 million euros in funding.

Note that these public proposals to reshore do not come free of frictions. It may take up to 2 years to *transfer* a single product to a new site (which includes processing the transfer, regulatory approvals, filing, etc.), while building a *new* facility can top \$2 billion and take 5 years. Furthermore, note that changing the cost of offshoring via a subsidy to produce in the home location may not necessarily result in less shortages. If firms tend to insure more against disruptions by building buffer more abroad, because it is cheaper to do so than at home, then disruptions may get worse once we oblige firms to reshore.

2. Additional Equilibrium Outcomes

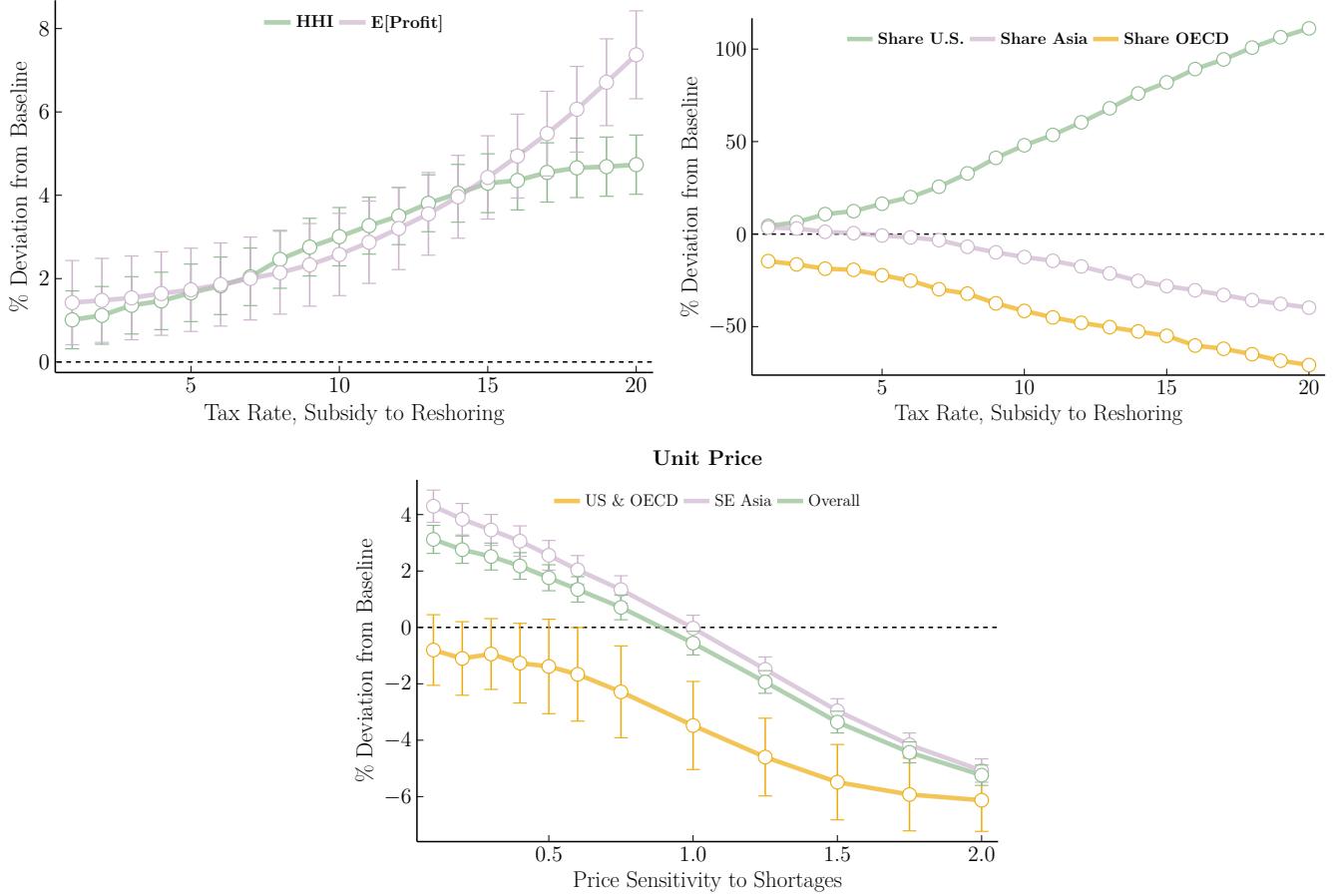


Figure F1: **Reshoring counterfactuals**

F.2. Short-term Price Variations Counterfactual: Details

1. Details: Changes in Manufacturers' Expected Revenues

Firm j 's expected revenue now depends on whether j is short and whether $k \geq 1$ other firms are short in the market. Three situations may arise. Firstly, if no firm is expected to be short in the market, firm j 's expected revenue remains unchanged and equals to

$$\left[1 - F\left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}}\right)\right] Q_{jmt}^D(\mathbf{p}, \mathbf{X}) p_{jmt}$$

Secondly, if firm j is not short but there are $k \geq 1$ firms in the market short, firm j 's expected revenue becomes $\left[1 - F\left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}}\right)\right] Q_{jmt}^D(\mathbf{p}, \mathbf{X}) \tilde{p}_{jmt}$ where

$$\tilde{p}_{jmt} = \left(p_{jmt} + \gamma^{\text{excess}} \sum_{k \neq j}^{J_m} \max \left\{ \delta_{jmt}^{\text{excess}}, 0 \right\} p_{jmt} \right)$$

The coefficient γ^{excess} represents *price sensitivity* - the rate at which firm j 's price increases in proportion to the amount of excess demand. The maximum function ensures that price increases only occur when

there is *positive* excess demand, such that firms whose expected supply is above their own buyer's demand are not counted.¹¹⁸ Larger γ^{excess} coefficients mean larger price increase post shortage.

Finally, if firm j is short and $k \geq 1$ other firms are also short, firm j 's expected revenue becomes

$$\left[1 - F\left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}}\right)\right] Q_{jmt}^D(\mathbf{p}, \mathbf{X}) \left(1 + \gamma^{\text{excess}} \sum_{j=1}^{Jm} \max\left\{\delta_{jmt}^{\text{excess}}, 0\right\}\right) p_{jmt}$$

We can summarize the new pricing scheme by combining all three cases together.¹¹⁹ The only difference among these three scenarios is whether the new price scheme enters a manufacturing firm's expected profit function through the "non-short" part or under the "short" part.¹²⁰ We can thus rewrite:

$$\tilde{p}_{jt} = \begin{cases} p_{jmt} & \text{if no firms short} \\ p_{jmt} + \gamma^{\text{excess}} \sum_{k=1}^{Jm} \max\left\{\delta_{kmt}^{\text{excess}}, 0\right\} p_{jmt} & \text{if } k \geq 1 \text{ firms short} \end{cases}$$

¹¹⁸Recall that we assume for now surplus disposal. Firms mostly benefits from competitors' shortages through the increase in prices of their own products

¹¹⁹Under the case where no firm is short the excess demand terms sum down to zero. Similarly, under the scenario where j is not short but one or more other firms are, the excess demand part for j will remained capped to 0

¹²⁰Recall that baseline firm j 's expected revenue is defined as a weighted sum of time short and not short: $\mathbb{E}[R_{jmt}] \equiv$

$$\overbrace{\left[1 - F\left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}}\right)\right] p_{jmt} Q_{jmt}^D(\mathbf{p}, \mathbf{X})}^{\text{No excess demand}} + \overbrace{p_{jmt} q_{jmt} \int_0^{\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}}} e f(e) de}^{\text{Excess demand}}$$

2. Additional Equilibrium Outcomes

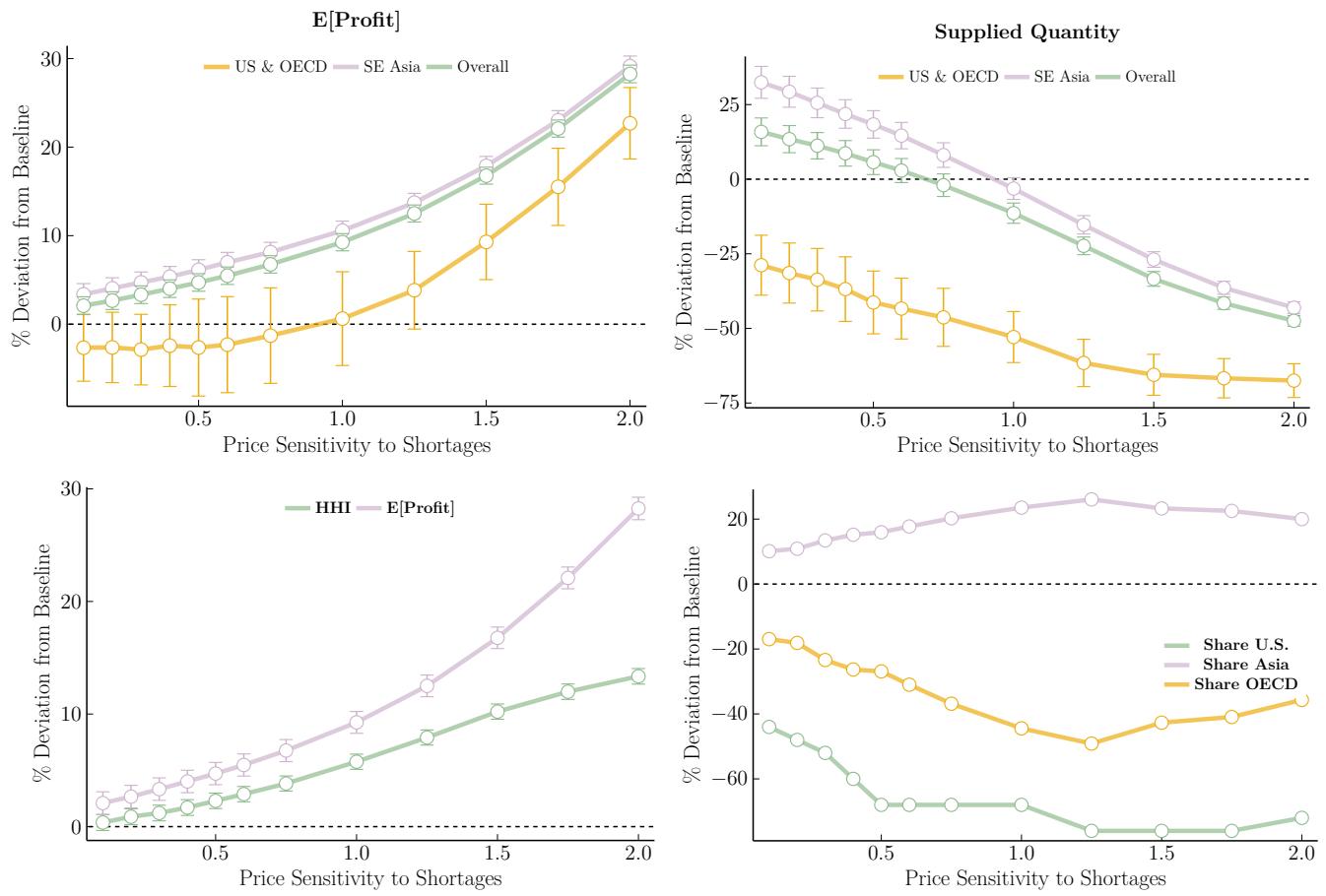


Figure F2: Allowing Short-term price Variations Post Shock

F.3. Penalizing Failures Counterfactual: Details

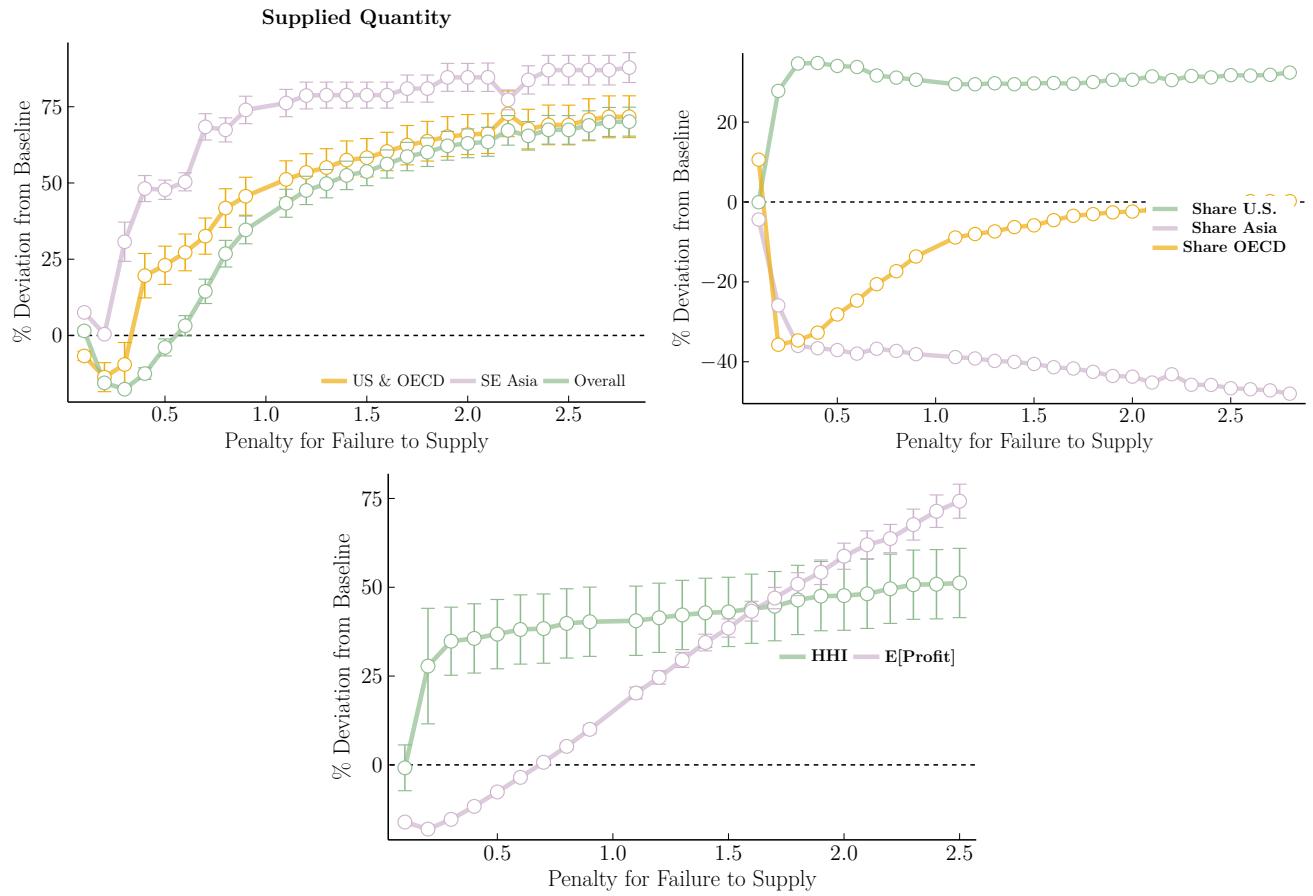


Figure F3: Counterfactual: Enforcing Penalties for Failures-to-Supply

Notes: Both market concentration and the expected profit of manufacturers increase with price sensitivity to shortage (by 2 to 10% for HHI and 3 to 23% for Expected suppliers' profit). Compared to today, allowing price variations post-shock lead to higher offshoring levels, with 10 to 20% more production offshored to Asian and a sharp 50 to 70% decrease in the share of U.S.-based firms. The expected profit of Foreign plants located in the South increases indeed more on average with higher price sensitivity to shortage compared to U.S. and OECD-based firms (8-10 percentage point higher), while the observed supplied quantities of U.S. based plants fall more compared to baseline than Foreign, South-based plants (by around 40% on average over price sensitivity levels)

F.4. Counterfactual 4: Setting "Price Floors"

A key characteristics of generic markets is their high level of contestability. The patent expiration of the branded product the generic copies generally results in the entry of multiple generic manufacturers ready to manufacturer the drug at reduced-costs compared to the innovator. Under the market institutionalization of perfect substitutability between generics, a procurement system based on online, reverse auctions creates favorable conditions for the market to unravel. Multiple market observers and policymakers underline the current dynamics of “race-to-the bottom” to be a key driver of shortages, creating an environment where only the lowest-cost producer can survive (and be willing to operate in the market).

In this counterfactual scenario, I consider how imposing minimum price levels may help mitigating the current “race-to-the-bottom” dynamics. Price-floors could play two roles here. First, reasonable price levels may help suppliers make positive profit margins, hereby deterring product discontinuation and cost-cutting. Second, it may help firms selling above minimum market prices “signal” production quality.

Minimum Price Framework. In the absence of any price floor, each firm maximizes profit with respect to own price, which yields our baseline unconstrained optimal price, which depends on marginal costs c_{jlt} , supply shock $\omega_{l(j)}$, supply quantities q_{jmt} and other firms’ prices and characteristics through the demand function $Q_{jmt}^D(\mathbf{p}, \mathbf{X})$:

$$p_{jmt}^* = \frac{c_{jlt}}{\int_0^{\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}}} \epsilon f(\epsilon) d\epsilon}$$

The constraint imposed by a price floor \underline{p}_t implies that each firm j ’s new equilibrium price is given by:

$$\tilde{p}_{jmt} = \max\{p_{jmt}, \underline{p}_t\}$$

The price floor *binds on firm j* if $p_{jmt} < \underline{p}_t$. The minimum expected profit that firm j can expect to generate under this framework thus becomes:

$$\begin{aligned} \mathbb{E}_{\omega_l} [\pi_{jmt}(\underline{p}, q_{jmt}; Q_{jmt}^D(\mathbf{p}, \mathbf{X}), c_{jlt}, \omega_{l(j)})] &= \left[1 - F\left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}}\right) \right] \underline{p} Q_{jmt}^D(\mathbf{p}, \mathbf{X}) \\ &\quad + \underline{p} q_{jmt} \int_0^{\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}}} \epsilon f(\epsilon) d\epsilon - c_{jlt} q_{jmt} - F_{jmt}^c \end{aligned} \quad (22)$$

For each given molecule market m , the regulator would like to choose a minimum price level that minimize the probability of market-wide shortages, or alternatively minimizes the share of excess demand in the market. I study two alternative specifications for price floor levels.¹²¹ Under the first specification, the regulator could set, for each market m , minim price levels that minimize the probability that shortages occurs:

$$\underline{p}_t = \arg \min \sum_{j=1}^{J_{mt}} \Pr(q_{jmt} \omega_{l(j)} < Q_{jmt}^D(\mathbf{p}, \mathbf{X})) = \arg \min \sum_{j=1}^{J_{mt}} F\left(\frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jmt}}\right) \quad (23)$$

If we instead want to minimize the share of expected *positive* excess demand in the market, the regulator could set:

$$\underline{p}_t = \arg \min \sum_{j=1}^{J_{mt}} \left(1 - \frac{q_{jmt} \mathbb{E}[\omega_{l(j)t} | \omega_{l(j)t} < \frac{Q_{jmt}^D(\mathbf{p}, \mathbf{X})}{q_{jm}}]}{Q_{jmt}^D(\mathbf{p}, \mathbf{X})} \right) \quad (24)$$

¹²¹Another natural candidate for the price floor could be the percentile of the observed distribution of prices in the data. One could think about whether the price floor must be the same for all manufacturing firms or whether it should vary based on factors as location, reliability index, average market price, concentration...

Re-simulating the estimated model under this counterfactual pricing scenario indicate that the establishment of price floors naturally lead to bunching at the minimum price \underline{p}_t . One might question the incentive for a firm to set its price above the minimum threshold, $p_{jmt} > \underline{p}_t$? Pricing above the established minimum could serve as an implicit indicator of superior production quality. In contrast to the prevailing situation, which assumes complete homogeneity among generic products and therefore, perfect substitutability among suppliers, it may allow prices to better capture differences in manufacturing quality among suppliers.

Simulation results suggest a key advantage of the price floor policy. Compared to the existing price equilibrium, any regulatory policy that sets a minimum price within a range of $+/- 30\%$ of the optimal minimum price \underline{p}_t yields higher expected utility for consumers. While pinpointing the precise optimal price floor for every drug market may pose challenges for regulators, this indicates some flexibility or “margin for error” in determining those levels.

There are potential constraints to this counterfactual policy that need further exploration. For instance, if firms merely increase their profits due to the price floor without improving manufacturing quality, it negates the intended benefit.

F.5. Counterfactual Outcomes: Parameter Estimates

To be updated/filled soon

Appendix F - Data Appendix : Mapping the Generic Drug Supply Chain

F.1. A Novel Dataset: Details on Data Structure

Each record contains information for one shortage event, and includes the drug name, an indicator for whether the shortage remains active, the date of first reported shortage, the ending date of the shortage if it is no longer active, reason for shortage if reported, the type of drug (injectable versus non-injectable) and AHFS therapeutic drug classification (American Hospital Formulary Service classification).

Drug Shortage Data. Data on drug shortages for the United States were provided by Erin Fox from the University of Utah Drug Information Service (UUDIS), which compiles information on current and past shortages that were reported to the American Society of Health-System Pharmacists (ASHP) and to the Food and Drug Administration (FDA)¹²². Each record contains information for one shortage event, and includes the drug name, an indicator for whether the shortage remains active, the date of first reported shortage, the ending date of the shortage if it is no longer active, reason for shortage if reported, the type of drug (injectable versus non-injectable) and AHFS therapeutic drug classification

¹²²The UUDIS data are richer than the FDA and ASHP data: firstly, they do not only focus on drug considered as medically necessary, but on all drugs and medical supplies in shortage; secondly, they go farther away in past history of shortages, with data as far as 1993

(American Hospital Formulary Service classification). A second-version of this dataset, providing detailed firm-drug level shortage information, was then used to define shortages at the manufacturer level and identify which firms exactly are driving shortages, which alternative National Drug Codes (NDCs) are available, and what are the reasons given by each manufacturer for shortages. This information was stored under thousands of Word Documents, which I digitized using automatic text processing and text mining techniques to extract data from text. More details and an overview of these documents can be found in Appendix F.2. A main limitation of this data is that it does not quantify the missing product quantities when a shortage occurs. This will be estimated from my model.

FDA's Orange Book Data. In order to obtain main drug characteristics, I am then linking the drug shortage data to the FDA's Approved Drug Products with Therapeutic Equivalence Evaluations (so called "Orange Book data"). This dataset records all approved drug products (both branded and generics) available for consumption in the United States in a year, along with their active manufacturers (i.e. the NDA or ANDA holders that applied for manufacturing the drug and is legally responsible for it - it does not need to be the actual manufacturing site due to outsourcing). It also indicates the number of years since the earliest approval of a manufacturer for a drug, route of administration, dosage forms and National Drug Codes (NDCs). This dataset allows to find all approved generic drugs for a branded product and to link them to potential substitutes. I use the Orange Book data to extract the full list of generic sterile injectable drugs sold on the U.S. market, along with their key characteristics. Indeed, the drug shortage data only records sterile injectable drugs that went into shortage at least once over the 2001-2019 period. In order to compare sterile injectable drugs that went into shortage to those that did not, and thus to identify what causes differences in shortage spells across similar drugs, I need the full list of sterile injectable drugs for the U.S.

U.S. Drug Pricing Data. I also obtain historical price data for all U.S. marketed drugs from the "Red Book", a private dataset released from IBM Micromedex. This data gives current clinical and pricing information about pharmaceuticals and drugs. The Historical Pricing Tables provide pricing that has been submitted to the Red Book by drug manufacturers, from year 1980 to today, but it only contains inactive prices are delivered in the ref Pricing Table. For active prices, I web-scraped the full active Red Book databases. Details on Pricing can be found in Appendix . The "Red Book" provides the Average Wholesale Price (AWP) for each NDC (National Drug Code, a unique numeric identifier given to medications at the labeler-product-dosage form level), as published by IBM Watson Health. AWP is the price charged by the wholesaler to the hospitals, which was the price based on which providers were reimbursed for drugs under Medicare Part B until 2005¹²³. While there is a payment "cap" for drugs sold under Medicare Part B, this cap is based on average sales price for the preceding period, so it should pick up changes in the underlying market price for the drug.

FDA's Manufacturing Facilities Locations Data. Economic research on drug shortages was long impeded by the lack of available data disclosing manufacturer names and facilities locations. This paper relies on a recently released FDA dataset which provides the geographical location of foreign and domestic drug manufacturing facilities. To implement the Generic Drug User Fee Act in fiscal year 2013 (GDUFA I and II), the FDA began collecting self-reported information on generic drug manufacturing

¹²³Most injectable drugs are administered in hospitals and thus potentially target a reasonably large population of elderly

locations including domestic and foreign active pharmaceutical ingredients (API) and finished dosage form (FDF) facilities. Implementation of the GDUFA required that every company which wishes to manufacture a drug for the U.S. market self-registers (including full name and address) and pays a fee to the FDA. The first dataset released by the FDA in 2016 included all manufacturers who self-registered (many of whom never completed the application process). In late 2019, the FDA published a new dataset including only those facilities who actually paid the GDUFA fee, and are thus theoretically allowed to produce a drug or its main components for the United States. For each registrant, the data contains: firm name, facility full address, a unique Facility Establishment Identifier (FEI) and the type of drugs produced (API, FDF, both or CMO). I combine these data to new confidential data obtained from the Center for Drug Evaluation and Research (CDER) at the FDA, which provides the geographical location of foreign and domestic drug manufacturing facilities for all years 2000-2020.

There are two issues with this data: first, paying the GDUFA fee is not tied to actually producing and marketing the drug. The FDA found in 2019 that among all generic drugs with approved applications, only 39 percent were observed to be marketed... As a consequence, the FDA still does not know exactly who makes which drug¹²⁴. Secondly, the FDA data identifies the address of the site and the organization paying the Generic Drug User Fee, but no information is available on the products actually manufactured at the site nor their prices or volumes.

U.S. and Foreign Customs Data. I solve these issues by linking the FDA datasets to U.S. and Foreign customs data obtained through a private company, Descartes Datamyne. Datamyne collects and organizes shipping manifests and customs records from government agencies and private companies, for the U.S and most countries in the world. They have U.S. imports data from 2004 to 2020 for all pharmaceutical goods. The data is at the shipment level, and includes shipments of all imported drugs from point-to-point in the world. Most publicly available trade datasets such as the Comtrade data aggregate trade flows at the sector level, and do not allow to track imports/exports of individual goods. For each good imported to the U.S. via freights or air, the data provide importer and exporter names and locations, port of arrival and departure, final destination, precise description of the shipped good, quantity and weight. Shipments data are at the shipment level, i.e. each shipment imported to the U.S. is indexed by the day it entered U.S. customs. For each drug extracted from the shipment description (see Appendix ??), I aggregate the total quantity and weight of products imported to the U.S. at the manufacturer-month level. To my knowledge nobody used this data to track pharmaceutical imports¹²⁵. This dataset solves two pitfalls of the newly disclosed FDA data. Firstly, it allows me to link each drug product to its list of foreign manufacturers. Secondly, it allows me to check which of the manufacturers present on the GDUFA list of payers actually exports products to the United States, from which date and in which quantity. This empirical contribution is of first-order given that until now, the lack of a reliable registry of manufacturing plants has made it difficult to estimate the true market share of foreign-made products, and any potential link with supply disruptions.

¹²⁴This was confirmed by a call I had with FDA economists from the Drug Shortage Team

¹²⁵My talks to the FDA confirmed they do not have access to it and/or did not know about it. Other economists and research centers I contacted also do not have trade data other than the Comtrade data

Several other datasets were used in this paper, among which Medicare Part B data to obtain quarterly sales quantities (which I will use in my model, as it allows to compute Medicare Market Shares) and firm-level data web-scraped based on my list of NDCs from *Cortellis Clarivate*. This firm provides data on labeler firm characteristics, such as firm's size, annual revenue, number of employees. I eventually use FDA's data on plants inspection, and another dataset providing warning letters sent to plants post-inspection to signal manufacturing-quality issues, both obtained through FOIAs.

The final data is a balanced panel of 1,220 sterile injectable drug (defined at the molecule-form level) aggregated at the year-month-manufacturer level. After matching the U.S. Customs to the shortage data and the FDA list of manufacturing facility locations, I obtain the list of all shipments, for all years 2004 to 2019, for which we find U.S. imports of sterile injectable drugs. I then match back the data to the shortage dataset, so that for each sterile injectable drug, we get a full list of manufacturers, quantity of products imported and indicators providing information about the status of shortages for all year-month dates between January 2004 and December 2019.

F.2. Probabilistic Record Linkage

1. Overview of Data Structure

Benztropine injection	
Name	
	Benztropine injection_012914_current.docx
	Benztropine injection_030414_current.docx
	Benztropine injection_050814_current.docx
	Benztropine injection_061114_current.docx
	Benztropine injection_071813_current.docx
	Benztropine injection_073114_current.docx
	Benztropine injection_082813_current.docx
	Benztropine injection_091814_current.docx
	Benztropine injection_092613_current.docx
	Benztropine injection_102913_current.docx
	Benztropine injection_120413_current.docx
	Benztropine injection_121613_current.docx

Benztropine Injection

18 September 2014

Products Affected - Description

Benztropine injection, American Regent
1 mg/mL, 2 mL vials, NDC 00517-0785-05

Benztropine injection, Fresenius Kabi
1 mg/mL, 2 mL vials, NDC 63323-0970-02

Benztropine injection, Nexus
1 mg/mL, 2 mL vials, NDC 14789-0300-02

Reason for the Shortage

- American Regent has benzotropine injection on back order due to manufacturing quality issue
- Fresenius Kabi USA recalled benzotropine injection due to potential for glass particles in the vials. Product may have been under APP or Nexus labels

Available Products

Benztropine injection, Akorn
1 mg/mL, 2 mL ampules, NDC 17478-0012-02

Benztropine injection, West-Ward
1 mg/mL, 2 mL ampules, NDC 00143-9729-05

Estimated Resupply Dates

- American Regent has benzotropine injection on back order and the company cannot estimate a release date.
- Fresenius Kabi USA has benzotropine injection on long-term back order and the company cannot estimate a release date. Both labels from Fresenius Kabi and Nexus are affected.

Updated

September 18, 2014; July 31, 2014; June 11, 2014; May 8, 2014; March 4, 2014;
January 29, 2014; December 16, 2013; December 4, 2013; October 29, 2013;
September 26, 2013; August 28, 2013; July 18, 2013, University of Utah, Drug
Information Service. Copyright 2014, Drug Information Service, University of Utah, Salt
Lake City, UT.

2. Matching Techniques

An issue with these new data is that no common identifier exists, which makes it challenging to join corresponding observations from different datasets. Building a usable dataset requires significant cleaning and matching efforts, as most data are in text and vary substantially in terms of format, both within a dataset and across datasets. We often have drug names rather than drug codes, or different level of drug codes that do not match directly from one to another, or manufacturer names or addresses rather than facility numbers. Two examples among many, the GDUFA dataset of manufacturing facility locations lacks consistency in reported manufacturer names over years, as well as in manufacturer addresses. The U.S. Customs data requires fuzzy string matching on a vector of pre-defined drug names in order to extract drug products and dosage forms from the description of shipments. As mentioned, the void of quantitative economic and health research on drug shortages can be directly inputed to this lack of comprehensive and unified dataset. In particular, there is no direct mapping from drug products to manufacturer facilities, an issue that even the FDA faces when trying to identify risky supply chains.

To give a made-up example, one manufacturer in India is called "aarti pharmaceuticals limited". It appears into the FDA data under this name, but also under "aarti pharmaceuticals", "aarti pharma ltd", "india Aartis drug", or with typos "street24aaarti". This manufacturer has three different plants, each of which has a different address, and each of these addresses may also be reported under different format and/or with typos. The goal is to get, for each manufacturer, a unique name and a unique address for each of its separate plant, which allows to improve matching rates between datasets. Given the data cover most countries in the world, the nature of the discrepancies in names and addresses is quite diverse. This kind of text problem appears even more acutely in the U.S. customs data (shipment descriptions with drug names and characteristics may be several paragraph long and contain many typos).

To overcome these issues and build a comprehensive dataset that provides both information on drug products and on their supply chains, I rely on probabilistic record linkage techniques. Instead of detailing the exact processes used to extract information for each individual dataset and to link them all together, I describe the general technique used to clean and build data. Probabilistic record linkage mainly involves four steps:

1. The first step involves pre-processing and text cleaning of each dataset in order to improve subsequent matching rates. The preprocessing step ensures that the datasets to match have the same formats and that chosen fields are meaningful in matching. This first step consists of two sub-steps: 1) parsing a field into the relevant subcomponents and 2) standardizing common character strings.
2. A second step, specific to this project, involves geocoding all drug manufacturing facilities. For this purpose, I build an algorithm that automatically queries the Google Maps API¹²⁶ and returns a unique longitude, latitude and normalized Google address for each facility. This step is applied to both facilities listed by the FDA as generic drug manufacturers for the U.S. market (GDUFA)

¹²⁶The function queries Google maps servers, giving it a string that includes a first cleaned version of manufacturing firm names and addresses, and return selected geocoded elements of the API

and CDER datasets) and to shipping firms from the U.S. Customs Imports Data. In these three datasets, I originally have firm names and addresses, but their formats are non-normalized (they vary both within a dataset and across datasets). This geocoding step allows me to create cross-walks of manufacturing company and facility names and improves my fuzzy matching algorithm in step 3.

3. In a third step, I apply a fuzzy matching algorithm to merge each data entry on its closest match. To match manufacturing facilities together, I first match plants based on their cleaned names. I then use geocoded locations (longitudes and latitudes) to match manufacturing plants that did not obtain a perfect match using the first round of fuzzy matching. For those last set of facilities I do not match exactly based on normalized names and locations, I use a measure of distance based on longitude and latitude (imperfectly matched facilities with similar names and zip code which are less than 10 kilometers away from one another are grouped together)¹²⁷.
 - I also use these fuzzy matching techniques to match drugs across datasets (each drug may be recorded under different names or different drug codes; drug codes are not always linked to drug names so that I first need to match drug to their National Drug Codes (NDCs) before matching these NDCs to their different drug names).
4. A last step involves clerically review of matched pairs with low scores and non-matched entries. Although the pair-similarity scores are correlated with correct matches, they are an imperfect metric. Typical record-linking processes require several runs, during which I try different combinations of fields, criteria for choosing candidates and their associated weights.

The fuzzy matching algorithm used in this paper is based on the Levenshtein distance between strings, which matches strings based on a similarity score. This score is based on the number of changes necessary to go from one string to the other. In particular I use a variation of this Levenshtein distance metrics called a *bigram score*, which is computed from the ratio of the number of common letters in the two strings and their average length minus one, giving higher scores to consecutive strings matching.

3. Matching Results

Geocoding of firm addresses is achieved with a very high success rate. For example, after several round of cleaning and querying Google Maps, only 2 Indian manufacturers were not geocoded out of originally 5,800 unique firm names. I have been able to geocode 98% facilities across the U.S. Customs and FDA's datasets.

¹²⁷Some manufacturing plants in Asia have very different formatted addresses (some adding many details like fax numbers or ownership status), but are obviously the same plant to a human observer. In that cases, Google may return different longitude and latitude points, or some "NAs". To solve this, I proceed in three step. First, I extract the most frequent zipcode for each manufacturer. Second, I get the most detailed address for the main zipcode, and corresponding longitude/latitude. Third, I compute the shortest geographical distance between each geocoded plant location and the main geocoded location for this manufacturer (using 'Haversine' great circle distance). If distance is below 10 kilometers, I replace coordinates by those of the main plant

"Shippers" in the U.S. customs data corresponds to the manufacturing facilities from the origin country (the plant or its administrative office). This is verified by matching the FDA list of manufacturing facilities (GDUFA and CDER datasets) to the U.S. customs data. I am able to link 84% of plant names referenced in the FDA data to at least one corresponding foreign manufacturing firm in the U.S. Customs Imports data. The remaining 16% may either : i) not actually export to the U.S. or ii) not appear in the U.S. customs data. A firm may not appear in the U.S. customs data if i) it chooses not to divulge its name to Customs (around 15% of "shippers" in the U.S. customs dataset are "not declared"), ii) it exports drug products by air instead of maritime shipments (but around 90% sterile injectable drugs are shipped by sea; drugs shipped by air are generally the most expensive and perishable drugs, as for instance vaccines).

I also link 63% of shortage events to foreign production over the years 2004-2019, by looking at the description of shipments in the U.S. customs data.

F.3. Building Sales Measures for Drugs Using Medicare Part B Data

I estimate total US demand of each drug based on Medicare Part B data. Using Medicare part B, I can obtain an estimated measure of 1) the market share of each manufacturer (labeler) for a given drug and 2) the share of total Medicare demand for a drug. I use CMS data from 2006 to 2019 and adjust prices for inflation to year 2010 dollars.

I first use Medicare Part B reimbursement data from the CMS Part B National summary files. The key variables are total reimbursements by Medicare (allowed charges or payments) and number of services billed (allowed services) for a Healthcare Common Procedure Coding System (HCPCS) code and year. HCPCS codes are used by providers to bill Medicare for procedures. A typical HCPCS code represents one administration of a drug. The same drug, defined at the molecule-form level, can be listed under multiple HCPCS codes representing different dosages. Allowed services are defined as the count of the number of services performed for a specific Part B procedure minus the denied services. The allowed charge is the Medicare approved amount for the Part B procedure submitted by the physician or supplier. Medicare usually pays about 80% of the total allowed charge and the other 20% is the coinsurance share, which is paid by the beneficiary. The data also includes information about drug package size (amount in one item), package quantity (number of items in the NDC), the number of billable units per package and billable units per 11-digit NDC¹²⁸.

Drugs are defined in the CMS data at the HCPCS code level. In order to combine it with my dataset of drug shortage, which is defined at the drug name level (and contains for each drug multiple National Drug Codes, NDCs), the data go through several cleaning and merging steps. I start by extracting, cleaning and merging all yearly files for all HCPCS codes beginning with J (codes J0000 - J0849 indicate "Drugs other than Chemotherapy" and codes J8521 to J9000 indicate "chemotherapy Drugs"). Some drugs have multiple dosages with different HCPCS J codes. I obtain 762 unique HCPCS J codes for a total of 652 unique drug names (at the molecule-form level, i.e. for instance "acyclovir

¹²⁸Recall that a NDC is a drug code that uniquely identifies a product and is composed of three parts at the 11-digits level: a labeler code, a product code and a dosage-form code.

injection"). The average HCPCS J code contains 16.29 National Drug Codes (NDCs).

I then extract CMS crosswalks linking drug names to HCPCS codes from the Medicare Part B Pricing Data. I clean drug names in these crosswalks, following the normalization of names used in my main dataset. I then merge these cleaned crosswalks to the Medicare demand data based on HCPCS and year. This steps permits to get drug names and drug descriptions for each HCPCS in Medicare Part B data. I eventually fuzzy match these drug names to my dataset of sterile injectable drugs, following similar matching techniques as described in Appendix F.2. I eventually match a total of 333 of my sterile injectable drug to at least one HCPCs code, 234 of which went into shortage at some point over the period 2006-2019.