Ramsey Pricing of Pharmaceuticals: A Theoretical and Machine Learning Analysis*

Federico Nutarelli[a†]

[a] IMT School for Advanced Studies Lucca, Italy

Abstract

International price differentiation in pharmaceuticals based on Ramsey prices has been widely discussed in the literature while empirical studies have assessed its validity indirectly by estimating the relationship between prices and income. The present study evaluates the Ramsey pricing of pharmaceuticals by directly analyzing whether pharmaceutical prices vary inversely with the price elasticities of demand and derives a pricing rule in case of interdependent demand and price distortions via insurance coverage. In this regard, it first develops a theory to analyze the Ramsey pricing of pharmaceuticals by considering important confounding factors such as insurance, income, and cross-prices. Then it identifies and estimates the price elasticities of demand for 33 molecules in 34 countries for the period 2008-2020 by using recently developed double/debiased machine learning methods. The results overall support the inverse elasticity rule with nuances. Within national markets, the evidence of Ramsey pricing is moderate with strong cross-elasticity effects. On the other hand, Ramsey prices for generics are prevalent across countries.

Keywords: Ramsey pricing; pharmaceutical industry; price elasticity; differential pricing, debiased machine learning
JEL Codes: I10, L41, L51, L65, K21.

1 Introduction

Cross-national pricing of pharmaceuticals has long been discussed as one of the most prominent cases of international price discrimination (Maleug and Schwartz, 1994; Danzon, 1997; Jack and Lanjouw, 2005; Towse et al., 2015; Towse et al., 2018; Danzon, 2019). On the one hand, some authors argue that remunerative drug prices are necessary because manufacturers need funding for pharmaceutical research and development (R&D) to bring new breakthrough products to market (Danzon, 1997). Others point to equity aspects and argue that medicines cannot be treated as normal market goods because they are essential for public health; therefore, they should be accessible to low-income populations (Outterson, 2005). In theory, differential pricing can reconcile these two objectives: higher prices in higher-income countries support R&D activities, whereas lower prices in lower-income countries improve access to medicines (Sherer and

*This research was supported in part by MSD's Project on Price Transparency. The empirical analysis uses data obtained under license from the IQVIA MIDAS™ database. We would like to thank these sponsors for making the research possible. The conclusions and views expressed herein are not necessarily those of MSD or IQVIA.

†Corresponding author: federico.nutarelli@imtlucca.it, IMT School for Advanced Studies Lucca, Lucca, Italy.

Watal, 2002; Danzon, 2018)). In essence, differential pricing increases utilisation and enhances both static efficiency and vertical equity with respect to uniform pricing. As a result, differential pricing has been a widely accepted strategy with a long tradition in pharmaceutical markets.

Among the various theoretical approaches used to provide support for price differentiation in pharmaceutical markets – i.e, price discriminating monopoly, Ramsey pricing, and value-based differential pricing – Ramsey pricing has been widely discussed in the literature, in part because of its theoretical appeal (Danzon, 2018).

Originally, Ramsey pricing was developed to solve the problem of achieving a certain level of tax revenue with an optimal set of tax rates that maximizes social welfare (Ramsey, 1927). Ramsey's solution was to set prices that change inversely to the price elasticities of demand, known as the inverse elasticity rule. In the 1970s, Ramsey theory experienced a renaissance following the influential work of Baumol and Bradford (1970), who used the inverse elasticity rule to find an optimal set of prices for a monopoly firm to recover joint/sunk costs that cannot be causally assigned to specific customer groups. Subsequently, Ramsey pricing has been applied primarily to utility firms in industries where joint costs are high and marginal costs are lower, such as telecommunications, airlines, railroads, and healthcare.

Danzon (1997) argues that pharmaceutical markets create the conditions for Ramsey pricing because R&D sunk costs are essential to bring innovative and effective drugs to market and cannot be allocated to a specific group of consumers or countries. This is because R&D costs are incurred before a product reaches the market and therefore cannot be captured by marginal costs. Danzon (1997) also theoretically shows that Ramsey pricing is welfare superior to marginal cost pricing when R&D costs are inherently high. Other theoretical studies support the welfare superiority of Ramsey pricing over marginal cost pricing and uniform pricing (Danzon, 1998; Danzon and Towse, 2003a; Barros and Martinez-Giralt, 2008).

Despite strong theoretical support, there is little empirical evidence for Ramsey prices in pharmaceutical markets. The existence of Ramsey prices has so far been tested indirectly by examining the relationship between prices and income. The basic idea is that in the presence of Ramsey prices, one should observe a high positive correlation between pharmaceutical prices and per capita income (GDP), assuming that GDP is a good indicator of the price elasticity of demand (Yadav, 2010a). However, empirical evidence on the relationship between drug prices and income does not necessarily predict high positive correlations (Maskus, 2001; Sherer and Watal, 2002; Lichtenberg, 2011), while some report no correlations at all (Waning et al., 2009b; Morel et al., 2011; Petrou and Vandoros, 2016). To the best of our knowledge, there is no study that directly tests the validity of the inverse elasticity rule for Ramsey prices in international pharmaceutical markets. In other words, there is a lack of empirical evidence on whether drug prices vary inversely with the price elasticity of demand across countries. Within countries, the price elasticity of demand for pharmaceuticals has been estimated in a series of patient-level studies using data from natural experiments at the individual level resulting from changes in the copayment shares of health insurance plans (Manning et al., 1987; Goldman et al., 2004; Landsman et al., 2005; Chandra et al., 2014). Accordingly, these estimates refer only to the insured population within a country, generally the USA, and none of them addresses Ramsey pricing.

The present study aims to fill the above-mentioned theoretical and empirical gaps by assessing the validity of the inverse elasticity rule through a theoretical and empirical cross-country analysis of pharmaceutical demand. In the theoretical part, we start from a within-country, within-molecule model with non-homothetic preferences and uneven insurance coverage to show that income shifts do not need to move own-price elasticities when drug baskets behave as near-necessities for dominant, well-insured buyers; hence, GDP per capita is a poor proxy for elasticity, especially when coverage reshapes effective prices. Building on this, we compare representative high-income countries (HICs) and medium and low-income countries (MLICs) to show that the deadweight loss in total welfare is generically larger for MLICs where demand is more elastic, rationalizing why the inverse-elasticity rule tends to appear stronger across HICs with broader insurance coverage and dampened price sensitivity. We then extend Ramsey–Boiteux pricing to interdependent demands under insurance, obtaining a closed-form markup system in which optimal prices load on both own- and cross-price elasticities and decompose into a Standard Ramsey Component and an Insurance-Distortion Amplifier that scales with coverage.

We believe that these theoretical results together may explain why cross-country "prices vs. elasticities" correlations are often weak, why monopoly distortions bite harder in poorer, more elastic environments, and how insurance design and substitution patterns jointly determine and influence Ramsey markups together with own and cross elasticities.

The empirical part of the study investigates the validity of the hypotheses formulated in the theoretical part by estimating own and cross price elasticities of pharmaceuticals and tests whether the inverse elasticity rule holds in pharmaceutical markets. In doing so, we draw on an IQVIA database of national pharmaceutical markets that contains detailed information on product names and brands, launch date, quantity sold, sales, molecules, therapeutic classes, manufacturers, and other relevant covariates. The sample consists of 34 countries, including 10[1] MLICs, 24[2] HICs [3], and 33 widely used molecules[4] for the period 2008-2020. The price elasticity of demand is estimated by using a recently developed double/debiased machine learning (DDML) method with heterogeneous treatment effects (Chernozhukov et al., 2018a; Semenova et al., 2021). DDML combines sample splitting and Neyman orthogonality to provide an asymptotically normal debiased estimator of treatment effect parameters in a low-dimensional and cross-sectional setting, assuming no unobserved unit effects. Semenova et al. (2021) extend this framework by proposing an estimator that performs better in a panel setting with a large number of heterogeneous treatment effects and with an even larger number of potential covariates, and unobserved unit heterogeneity[5]. The ability to use a very large number of covariates allows us to control for cross-price elasticity and reduces the likelihood of standard bias from omitted variables, which has been a major drawback of traditional demand estimates. The DDML model takes into account previous realizations of the demand system, such as price lags and sales lags, and thus accounts for unobserved heterogeneity and serial correlation. These aspects of the demand model allow us to better approximate the true own-price elasticities of pharmaceuticals.

The results show that all countries except the US have a negative average price elasticity, ranging from the highest (-1.42) for Colombia to the lowest (-0.43) for China, while the molecules have a wide range of elasticities from -2.04 (diclofenac) to -0.21 (mesalazine). To test the validity of the inverse elasticity rule in pharmaceutical markets, we first analyze the relationship between drug prices and the price elasticities of pharmaceutical demand. We find that country-level prices vary inversely with price elasticities in both HICs and MLICs, indicating the existence of Ramsey prices in cross-country pharmaceutical markets. As a second investigation of the inverse elasticity rule in pharmaceutical markets, we analyze the relationship between GDP per capita in purchasing power parity (PPP) and the price elasticities of pharmaceutical demand. We find evidence for the inverse relationship between price elasticities and GDP per capita in HICs and MLICs, consistent with the Ramsey hypothesis.

The empirical analysis is repeated by using only generic products. Given that generics are no longer

[1] Algeria, Bulgaria, China, Colombia, Egypt, India, Romania, Russia, Serbia, Turkey

[2] Australia, Austria, Belgium, Canada, Croatia, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Japan, Netherlands, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, United Kingdom, United States

[3] According to the World Bank's classification of countries by income.

[4] Acetaminophen, acetylcysteine, acetylsalicylic acid, aciclovir, allopurinol, amlodipine, betamethasone, bisoprolol, budesonide, carbamazepine, ciprofloxacin, clindamycin, dexamethasone, diclofenac, finasteride, fluconazole, furosemide, ibuprofen, ketoconazole, lactulose, levothyroxine sodium, loratadine, mesalazine, methylprednisolone, metoprolol, metronidazole, omeprazole, ranitidine, risperidone, salbutamol, simvastatin, testosterone, valproic acid.

[5] See Appendix for further details

under patent protection and can be produced at costs close to marginal costs, there are more substitutes for generics than branded products, which might lead to higher price sensitivity. Indeed, the estimates show that generic pharmaceuticals are highly price elastic and exhibit Ramsey principles.

The present study makes the following key empirical contributions to the literature. First, it presents the first empirical analysis to examine Ramsey pricing in international pharmaceutical markets by directly testing the validity of the inverse elasticity rule.[6] Second, it provides the first cross-country estimates of drug price elasticities at the molecule level. Thirdly, it enriches the theoretical understanding of Ramsey pricing in pharmaceutical markets by showing that (i) Income may not be a good proxy for price elasticity of pharmaceuticals (ii) Under Ramsey–Boiteux pricing, the share of deadweight loss in total welfare is generically larger in MLICs than HICs, theoretically underpinning why the inverse-elasticity rule tends to manifest stronger across HICs. (iii) We extend the Ramsey rule to settings with interdependent demands and insurance-induced price distortions, delivering a closed-form markup system in which optimal prices depend on both own- and cross-price elasticities, a standard Ramsey component, and an "insurance-distortion amplifier" term, providing a unified theory of the prior preeminent works by Höffler (2006) and Barros and Martinez-Giralt (2008).

The remainder of the paper is structured as follows. Section 2 gives an overview of the related literature. Section 3 develops the theoretical model. Section 4 presents the empirical model with data source (Section 4.1), methodology (Section 4.2), and results (Section 4.3). Finally, Section 5 discusses the main findings and concludes.

2 Literature Review

It is well known that some structural features of pharmaceutical markets violate some criteria of perfect competition, *e.g.* imperfect information (Kalo et al., 2013), monopoly rents from patented innovations and sunk costs of R&D (Danzon, 1997). Therefore, drug prices may exceed marginal costs. In this context, international differences in drug prices have long been studied in the literature (Maleug and Schwartz, 1994; Danzon, 1997; Jack and Lanjouw, 2005, Towse et al., 2015; Towse et al., 2018; Danzon, 2019). Researchers have focused primarily on the welfare effects of various pricing strategies, including marginal cost pricing, differential pricing[7], as well as uniform prices. Danzon (1997) argues that marginal cost pricing in pharmaceuticals may not be optimal because R&D is an essential part of the production process and accounts for about 30% of total costs. Since the development of new drugs undoubtedly serves all customers around the world, R&D costs should be considered a global joint cost (Danzon and Towse, 2003a).A crucial aspect is that pharmaceutical R&D costs are fixed. In other words, R&D costs do not depend on the number of countries or people that buy the drug and therefore cannot be simply allocated as a marginal cost (Danzon, 1997), so pricing at marginal cost is not the first-best pricing strategy. Based on this consideration, Danzon (1997) suggests Ramsey pricing as the most efficient option. Given a certain profit level that covers all costs, including R&D, Ramsey principles yield the most efficient price differentiation[8] (Danzon, 1997; Danzon, 2000; Danzon and Towse, 2003a).

Ramsey pricing was introduced by mathematician Frank Plumpton Ramsey in 1927. He developed a theory to analyze how to most efficiently collect a given level of tax revenue from different goods. Ramsey's

[6]It should be noted that the methodology used also allows the direct calculation of elasticities at the drug level, as these are available in the dataset. However, for reasons of data protection, the analysis focuses on molecules.

[7]Differential pricing (price discrimination, tiered pricing) means charging different prices for the same drug in different markets. It can occur both in low-income and high-income countries and in one country between different payers (Danzon, 2018)

[8]The literature provides other rules for drug price differentiation, such as value-based price differentiation (Danzon, Towse, and Mestre-Ferrandiz, 2015a) and indication-based price differentiation (Towse et al., 2018),

theory experienced a renaissance after the seminal work of Baumol and Bradford (1970). They built on Ramsey's theory and applied it to a model of utility pricing that they called *the second-best pricing*.

The crucial idea behind Ramsey pricing is the rule of inverse elasticity, which "requires that each price be set so that its percentage deviation from marginal cost is inversely proportionate to the item's price elasticity of demand" (Baumol and Bradford, 1970, p. 267). Ramsey pricing leads to static efficiency by maximizing consumer surplus and total output consumed when marginal costs are lower than average costs, implying fixed costs. Therefore, it has been applied mainly to utilities in industries where joint costs are high and marginal costs are lower, including telecommunications (Ford et al., 1993), airlines (Martin-Cejas, 1997), railroads (Sanchez-Borras et al., 2010), electricity (Bigerna and Bollino, 2016), health care (Melnick et al., 1992), and other deregulated industries.

Price differentiation based on Ramsey principles suggests that drug prices should vary inversely with the elasticity of demand (Danzon, 1997; Danzon and Towse, 2003a), assuming that the absolute price elasticity of demand is negatively correlated with net income. In other words, the markup on marginal costs increases when the price elasticity of demand decreases and vice versa. This means that high-income countries pay higher prices sufficient to cover the cost of medicines, while low-income countries pay lower prices to improve public access to medicines. Moreover, demand in high-income countries is more inelastic to price changes than in low-income countries.

Both customers and firms are better off with Ramsey prices than with uniform prices (Danzon, 1998). If a uniform price applies to all, regardless of the elasticity of demand, price-sensitive customers will limit their consumption more than price-insensitive customers. Moreover, highly price-sensitive customers might exit the market, reducing aggregate output and thus social welfare. Uniform prices could also hinder pharmaceutical R&D in the long run. If differentiated prices are charged based on the Ramsey principle, companies can exploit the willingness to pay of different income groups. Since differential pricing varies with the elasticity of demand, companies can raise more funds for R&D by keeping in the market customers whose markup is below the uniform price but above marginal cost.

The literature theoretically confirms the superiority of Ramsey prices in terms of welfare, while empirical evidence is sparse. Danzon (1997) conducts a theoretical analysis with several firms and products to compare the welfare effects of Ramsey prices and marginal cost prices in the US and the EU. She shows that price differences in the U.S. can be explained by the price elasticity of demand, which is consistent with Ramsey pricing. This is mainly because the drug market in the U.S. is relatively competitive thanks to various health plans that can reflect patient preferences. In the EU, however, the market is structured differently. Here, the government usually acts as the single-payer and exploits its monopsonistic position. Therefore, prices may differ from the optimal Ramsey prices. Danzon and Towse (2003a) emphasize that marginal cost prices cover short-run production costs but cannot generate sufficient funds to finance pharmaceutical R&D; therefore, patent protection guarantees prices above marginal costs, ensuring sustainable incentives for R&D activities. They argue that patents do not necessarily lead to higher prices if manufacturers use differential pricing based on Ramsey principles. Barros and Martinez-Giralt (2008) incorporate insurance costs into Ramsey pricing to investigate how different insurance coverage rates change optimal Ramsey prices. They show theoretically that higher insurance coverage leads to higher Ramsey prices.

In the empirical literature, the existence of international Ramsey prices for drugs has previously been tested indirectly by examining the relationship between drug prices and income. The basic idea is that if Ramsey prices exist, one should observe a high positive correlation between drug prices and per capita income (GDP), assuming that GDP is a good index of the price elasticity of demand (Yadav, 2010a). However, the empirical evidence to date does not necessarily show strong correlations. Sherer and Watal (2002) examine whether pharmaceutical companies applied Ramsey principles to the prices of HIV/AIDS drugs in the 1990s. They use multiple regressions to determine whether there is a strong relationship between prices and gross national product (GNP) per capita. The results suggest a weak relationship: an additional 1000$ of per capita income increases relative prices by 0.018. Similarly, Maskus (2001) uses international data on 20 major molecules for the period 1994-1998 to analyze correlations between molecule prices and per capita income. According to the results, 17 of 20 correlations are positive and six are above 0.5, suggesting a positive but not very strong relationship between drug prices and income. Lichtenberg (2011) runs multiple regressions to analyze price differences across countries using data on 6500 medicines in 38 countries in the third quarter of 2008. He finds a positive correlation between prices and per capita income. However, he highlights the large variability within low-income countries, such that lower-income citizens in some countries pay higher prices than higher-income citizens.

In addition, several studies have found that drug prices do not necessarily vary inversely with per capita income, implying that some low-income countries may pay higher prices (Waning et al., 2009a). Using data from 14 middle-income countries and 3 high-income countries from 1999 to 2008, Morel et al. (2011) computes Laspeyres and Paasche's indices to examine whether drug prices vary systematically with income. They find no such variation; instead, they report high variability in prices across countries. Petrou and Vandoros (2016) calculates the Laspeyres index in an analysis of drug price differences in 11 European countries. They find no evidence of any relationship between prices and income levels.

Despite the theoretical underpinnings, there is little consensus on whether Ramsey prices exist in pharmaceutical markets. The literature shows that differential pricing is limited to some categories of drugs, i.e., HIV/AIDS drugs, antimalarials, and vaccines. Essential drugs, on the other hand, generally remain exempt from differential pricing (Yadav, 2010a). A key contradiction is that observed price differentials within and/or across countries do not approximate Ramsey's price levels (Danzon and Towse, 2003a) for two main reasons. First, drug prices are highly influenced by third-party public buyers who exert monopsony power in price negotiations (Kalo et al., 2013). In such a case, public buyers often resort to international reference pricing to control and compare drug prices. Reference pricing is considered one of the practices that affect differential pricing (Schmidt et al., 2001a; Danzon, 2018). Second, price differences across countries provide an incentive for cross-border pharmaceutical arbitrage or parallel trade (Schmidt et al., 2001a), which may reduce the incentive to invest in pharmaceutical R&D and innovation and to adopt differential pricing strategies (Danzon and Towse, 2003a; Jack and Lanjouw, 2005; Reisinger et al., 2019). Based on these premises, the debate about the existence of Ramsey prices in pharmaceutical markets remains inconclusive.

2.1 Empirical Evidence of the Price Elasticity of Demand

This section reviews the literature on the price elasticity estimates as a reference point for our analysis. To our knowledge, no empirical study has tested the inverse elasticity rule by analyzing the relationship between international prices and price elasticities of pharmaceuticals. Thus, none of the studies reviewed in this section considers Ramsey pricing.

Estimating price elasticity for pharmaceuticals is challenging given the complexity of the market and the multitude of factors that influence patient and prescriber preferences. Accordingly, estimates are inconsistent to some degree due to differences in calculation methods, demand definitions, price measurements, data sources, and study periods (Li and Chollet, 2006).

A number of studies have estimated the price elasticity of demand for drugs at the patient level by exploiting single-country individual-level data from natural experiments of copayment shares of insurance plans (health plans). Such studies analyze the price response (own-price elasticities) of different groups of drugs or the number of prescriptions in relation to changes in insurance copayments. Accordingly, these studies cover only the insured population, thus, can be referred to as *coinsurance elasticity*, as in Manning et al. (1987). The RAND Health Insurance Experiment, conducted in the U.S. in the 1970s, is the best-known study in this area because of its randomized sample. The price elasticity of demand is estimated to be -0.2 with respect to all health care, including prescription drugs (Manning et al., 1987). Other studies estimate the price elasticity of demand with respect to total drug costs to range from -0.33 to -0.12 (Gilman and Kautter, 2008; Chandra et al., 2014; Brot-Goldberg et al., 2017). Some studies have measured price elasticity by drug class. Goldman et al., 2004, for example, estimates elasticities with respect to the number of drug days supplied and finds the following: NSAIDs (-0.45), antihistamines (-0.44), antihyperlipidemics (-0.34), antiasthmatics (-0.32), antihypertensives (-0.26), antidepressants (-0.26), and antidiabetics (-0.25). Landsman et al. (2005) measure elasticities with respect to number of prescriptions and finds lower elasticities for drugs for asymptomatic conditions such as ACE inhibitors, statins (from -0.10 to -0.16) and higher elasticities for drugs for symptomatic conditions such as COX -2 inhibitors, NSAIDs, SSRIs (from -0.24 to -0.60).

However, Yeung et al. (2018) criticize the existing practice because, in a natural experiment with changes in copayments, cross-price elasticities must also be taken into account, since drugs are likely to have complements and substitutes whose prices are also affected by the same change in copayment. Consequently, own-price elasticities cannot be estimated accurately because *ceteris paribus* is violated. Yeung et al. (2018) propose a value-based formulary –which includes both the prices of the drug itself and its complement and substitute products– to overcome such an endogeneity problem when using data from natural experiments. They report an overall price elasticity for drugs of -0.16. Another line of studies uses the instrumental variables method to generate exogenous variation and estimate price elasticities ranging from -0.14 to -0.4 (Ellis et al., 2017; Einav et al., 2018).

To our knowledge, the only cross-country empirical study of demand elasticities is that of Danzon, Mulcahy, et al. (2015). They estimate cross-country income elasticities for drugs used to treat HIV/AIDS, malaria, and tuberculosis, for both originator and generic drugs. They find a positive but overall low income elasticity (0.2) for all countries, 0.4 for originator drugs and 0.6 for generics. For MLIC, the average income elasticity is 0.15 and is neither significant nor negative for originator drugs.

Controlling for cross-price elasticities is crucial for estimating drug price elasticities. The method proposed in this study adequately accounts for the prices of substitutes and complements to provide reliable estimates of cross-price and own-price elasticities in the same regression framework. To the best of our knowledge, the present work is the first to control for the cross-price elasticity of each drug by simultaneously considering the national market and the time dimension. This gap in the literature is mainly due to the lack of an appropriate method to estimate own-price elasticities by taking into account the prices of all other drugs. The difficulty lies mainly in two aspects: the high granularity and the lack of regularization. In other words, the analysis of drugs grouped in many molecule-country pairs requires the construction of an extensive simultaneous equation model that includes in each equation the price of all other products in all other countries, potentially introducing a large number of confounding factors into the model. This makes the estimation infeasible both because of the huge computational costs and the biases in the estimates. These problems are solved with the method DDML which uses regularization techniques (LASSO) to remove the bias due to potentially high-dimensional confounders and avoid computational inefficiencies thanks to its optimized design[9].

3 Theory

In principle, cross-country differential pricing based on Ramsey principles can only be implemented if the price elasticities of demand exhibit some degree of variation across countries. One major reason for price

[9] See Semenova et al. (2021) for more details.

elasticities to vary is cross-country income inequalities. The marginal utility of money is expected to be higher in MLICs, which causes higher price sensitivity, *i.e.,* price elastic demand, than it is expected to be in HICs. As a result, much of the literature has used GDP as a proxy for own-price elasticity. In our view, this correlation rests on an oversimplified assumption that overlooks the complexity of factors shaping consumption decisions and inter-product relationships, including cross-price elasticities and the share of the population with insurance in a given country. For instance, in HICs, broad insurance coverage reduces consumers' price sensitivity, whereas in MLICs, out-of-pocket payments may dominate, raising price elasticity among uninsured consumers. If the insured share consists largely of wealthy citizens with private insurance, a monopolist may choose to serve only this segment, excluding poorer consumers and thereby undermining the expected response of elasticity to a GDP shock. Further details on this example are provided in the Appendix. Moreover, as briefly overviewed in Section 2, the literature provides mixed results on the relationship between pharmaceutical prices and GDP per capita income, with positive (Lichtenberg, 2011), negative (Waning et al., 2009a), and neutral results (Petrou and Vandoros, 2016). In this regard, we first analyze whether the latter relationship is credible in a complex economy (Section 3.1). We then move to studying the Ramsey pricing of pharmaceuticals in different income levels across countries (Section 3.2).

Section 3.3 complements Section 3.2 by examining within-country differential pricing under the same framework. In this setup, unlike in the first part of the theory, the assumption of independent demand no longer applies. The demand for branded and generic drugs is interdependent due to cross-elasticities. As a result, the Ramsey problem is adjusted accordingly.

3.1 On Using GDP per capita as a Proxy of Own-Price Elasticity

The following assumptions hold for a representative MLIC, L and a representative HIC, H.

T.1 $\rho_B + \rho_G = 1$, meaning total pharmaceutical consumption is normalized to 1 and is split between branded (ρ_B) and generic (ρ_G) drugs (i.e. ρ_B, ρ_G represent shares of a consumer's drug basket allocated to branded and generic products, respectively.).

T.2 There is a numeraire good, y, consumed alongside drugs and whose consumption is denoted by C_y.

T.3 Insurance reduces the effective price of drugs by covering part of the cost, leaving consumers' income largely unaffected. Let δ_i denote the proportion of drug costs covered by insurance for a consumer group l, where $l \in \{L_R, L_P\}$ in L and $l \in \{H_R, H_P\}$ in HICs, representing the rich and poor segments in the country. Price distortions are allowed due to the presence of insurance. In other words, moral hazards are present following Martínez Giralt and Barros (2006). [10].

T.4 Preferences are non-homothetic[11]. Heterogeneity in health expenditure across income countries is a well-established fact in the pharmaceutical economics literature (Blazquez-Fernandez et al., 2014; Wagstaff, 2002; Adebayo et al., 2015). We accordingly modelled the utility function as reflecting non-homothetic preferences according to income status. Other evidence can be found, for instance in Adebayo et al. (2015) who found that, among others, education and income status influence insurance coverage and demand of drugs in medium and low-income countries.

The notation of Appendix A is followed. Furthermore, our focus is firstly within country and molecule[12].

[10]The reader is referred to the formal derivation of the Appendix A.

[11]The analysis naturally extends and simplifies in case of homothetic preferences (Armstrong and Vickers, 2018).

[12]Notice that if a within-country, between-molecule analysis, rather than between-product analysis within the same molecule as we are doing, is conducted, the results are simplified because the cross-elasticities between molecules are assumed to be zero, and the classical Ramsey solution applies in this case.

To provide the reader with some context for the consumer decision, we ask how people split their drug spending between the branded (B) and the generic version (G) of that same drug. We assume they're gross substitutes. Insurance matters because people face effective prices after copay: with the same copay rate on branded and generic, their relative price doesn't change and the split is driven by posted prices and tastes; with different copays, the cheaper effective option gets a bigger share. To avoid mechanically inducing branded–generic substitution via coverage, we assume a common copayment across types. Preferences are non-homothetic (income-sensitive): in the isoelastic case, as groups become better off, they may reweight toward the drug type whose appeal grows more with income (e.g., slightly more branded if its taste weight rises with income). Conceptually it's a two-step problem: the broader decision of "how much to spend on drugs vs other goods" sits outside; inside this molecule we just decide the branded against generic split based on effective relative prices and income-driven tastes.
Following Hanoch (1975) preferences are represented by a non-homothetic CES utility aggregator:

$$1 = \sum_{k\in\{B,G,y\}} (f_k(U))^{\frac{1}{\sigma}} C_{k,l}^{\frac{\sigma-1}{\sigma}},$$

where $k \in \{B,G,y\}$ indexes branded drugs (B), generic drugs (G), and a numeraire good that represents the consumption of all other goods (y), $f_k(U)$ is an income-dependent taste parameter, and $\sigma > 1$ represents the elasticity of substitution and $l \in \{H_R, H_P, L_P, L_R\}$. We acknowledge here an implicit assumption of the CES model: the substitution elasticity is common, but Marshallian price elasticities are allowed to differ across groups through budget shares.
The budget constraint for each consumer group l is:

$$\tilde{P}_{B,l}\,\rho_B Q_l \;+\; \tilde{P}_{G,l}\,\rho_G Q_l \;+\; C_{y,l} = I_l, \tag{1}$$

$$\rho_B + \rho_G = 1, \qquad Q_l \geq 0 \tag{2}$$

ρ_B and ρ_G are the shares of the drug basket devoted to branded and generic products, respectively; Q_l is the scale (total quantity or expenditure) of drug consumption for group l. The level-form variables are recovered via $C_{B,l} = \rho_B Q_l$ and $C_{G,l} = \rho_G Q_l$. We keep the $C_{k,l}$ notation in the Appendix A.1 because the algebra of the CES aggregator, elasticities, and Ramsey pricing is far cleaner when everything is expressed directly in quantities—shares and scale would have to be re-combined at every step. y is priced at 1 (normalized). I_l is the income of consumer group l with $I_H >> I_L$, and the effective prices with insurance are:

$$\tilde{P}_{B,l} = (1-\delta_{l,B})P_B, \quad \tilde{P}_{G,l} = (1-\delta_{l,G})P_G$$

with $\delta_{l,B} \leq \delta_{l,G}$ and $\delta_{l,B} + \delta_{l,G} \leq 1$. This assumption implies that insurers cover a fixed amount, which at most entirely covers the cost of branded and generic drugs purchased by the consumer. The latter amount is generally covered unevenly between generics and branded drugs. The share covered for generics is higher than that for branded drugs, as insurers have greater incentives to cover generics—all else being equal in terms of therapeutic efficacy. Thus, for a given income I_l, insurance coverage affects the allocation between ρ_B and ρ_G based on the effective prices $\tilde{P}_{B,l}$ and $\tilde{P}_{G,l}$, influencing the choice of branded versus generic drug consumption differently in H and L. Indeed, in high-income country H, widespread insurance coverage results in high effective insurance rates across all social strata, substantially reducing out-of-pocket costs for both branded and generic drugs i.e. $\delta_{H,k} > \delta_{L_R,k} > \delta_{L_P,k}$ $k \in \{B,G\}$ (van Hees et al., 2019).

In this context, let E_l total expenditure (income) of group l and define the budget share $s_{k,l} \equiv \tilde{P}_{k,l}C_{k,l}/E_l$

(see Appendix A). As derived in the Appendix, in this CES demand system, the Marshallian own-price elasticity is $\varepsilon_{kk,l} = -\sigma(1-s_{k,l})$, and the income (expenditure) elasticity is

$$\eta_{E_l}^{C_{k,l}} = \sigma + (1-\sigma)\frac{\mathrm{ei}_k}{\bar{\mathrm{ei}}_l}, \qquad \bar{\mathrm{ei}}_l \equiv \sum_{m\in\{B,G\}} s_{m,l}\,\mathrm{ei}_m,$$

where ei_k is the drug-specific income-elasticity parameter and $\bar{\mathrm{ei}}_l$ is its share-weighted average for group l. Differentiating $\varepsilon_{kk,l} = -\sigma(1-s_{k,l})$ with respect to $\ln E_l$ and using $\frac{d\ln s_{k,l}}{d\ln E_l} = \eta_{E_l}^{C_{k,l}} - 1$ yields

$$\frac{d\,\varepsilon_{kk,l}}{d\ln E_l} = \sigma\, s_{k,l}\big(\eta_{E_l}^{C_{k,l}} - 1\big),$$

so, holding other prices fixed, higher income moves the demand elasticity, specifically making it less elastic (own elasticity less negative) for luxuries and more elastic for necessities. The magnitude scales with σ and the current budget share $s_{k,l}$. In a within-country group l, within-molecule framework, whether income changes induce changes in elasticity hinges on two conditions: (i) $s_{k,l} \neq 0$ and (ii) $\eta_{E_l}^{C_{k,l}} \neq 1$. Accordingly, GDP per capita is a poor proxy for own elasticities whenever most of the drug basket is a near necessity within group l in that country and within that molecule. In that case, the budget share of a given molecule/category changes little as income varies – so $\eta_{E_l}^{C_{k,l}} \approx 1$ and the derivative above is close to zero.

A further subtle but important point concerns the rationale for approximating own elasticity with GDP per capita: high-income individuals are assumed to have inelastic demand for pharmaceuticals, whereas low-income individuals are assumed to have more elastic demand. A plausible mechanism is wealth and the availability of private insurance. However, the extent to which the insured share of the population "depends on income" is crucial for the validity of this approximation. Appendix A.4 elaborates: in settings where the rich are insured and the poor are not – an eventuality that typically verifies in MLICs with the absence of a public insurance scheme (Hooley et al., 2022) –, the rich may continue to treat drugs as near-necessities (so $\eta_{E_l}^{C_{k,l}} \approx 1$), while the poor substitute toward the outside good y, driving their share $s_{k,l}$ to zero i extreme cases [13]. The result is an overall lack of relationship between GDP per capita and elasticity. Another consequence of this case – as shown in Appendix A.4 – is that prices do not shift as a consequence of GDP shocks.

Although the analysis above is within a country–group and a single molecule, the approximation of elasticities via GDP per capita is more often seen in cross-country studies. This entails moving from a within–country group, within-molecule setting to a cross-country, within-molecule comparison. We will do so in a simplified scenario. In particular, let countries be indexed by c, molecules by $k \in \{B,G\}$, and within-country groups by l. For group l in country c, Appendix A already delivers the own-price elasticity (with respect to the *effective* price $\tilde{P}_{k,l,c} = (1-\delta_{l,k,c})P_{k,c}$) which we will rewrite here adding the country subscript for underlying that it refers to a specific country c

$$\varepsilon_{kk,l,c} = -\sigma\big(1 - s_{k,l,c}\big), \qquad \frac{d\,\varepsilon_{kk,l,c}}{d\ln E_{l,c}} = \sigma\, s_{k,l,c}\big(\eta_{E_{l,c}}^{C_{k,l,c}} - 1\big),$$

where $s_{k,l,c} \equiv \tilde{P}_{k,l,c}C_{k,l,c}/E_{l,c}$ and $\eta_{E_{l,c}}^{C_{k,l,c}} = \sigma + (1-\sigma)\,\mathrm{ei}_k/\bar{\mathrm{ei}}_{l,c}$. To turn those group elasticities into a country-level number, the idea is to take a weighted average, giving more weight to groups that buy more of that molecule. This averaging only makes "apples-to-apples" sense if three things hold. Namely, we assume that

[13] Appendix A.4 considers only negative shocks to GDP per capita. However, in our simplified example, if GDP per capita increases, the group L_R would still treat drugs as *near necessities*, whereas whether s_{k,L_P} tends to zero depends on whether a simultaneous price response (or other forces) offsets the budget gain of L_P.

everyone is playing by the same substitution rule σ, so a one-point change in price means the same kind of behavioral pull everywhere, otherwise, mixing elasticities would blend different behaviors. Sensitivity must be with respect to the same notion of price (i.e, effective price after copay), so we're not averaging reactions to posted prices in one place and net prices in another. Finally, when we average, we hold the weights (who buys how much of what) fixed, so changes in the average aren't just driven by who happens to buy more this year; they reflect genuine differences in price sensitivity, not shifting mix. A simple example may clarify the latter point. Consider a single molecule k in country c with two groups $l \in \{R, P\}$ (Rich, Poor). Let the within–country, within–molecule own-price elasticities (with respect to the effective price. Notice however that copayment rates are fixed) be $\varepsilon_{kk,R,c} = -0.8$ and $\varepsilon_{kk,P,c} = -1.6$. Denote the group weights within molecule k by $\omega_{l|k} \equiv C_{k,l,c}/\sum_j C_{k,j,c}$. If composition is allowed to vary by country, the market elasticity for molecule k in country c is the share-weighted average $\varepsilon^{\text{mkt}}_{kk,c} = \omega_{R|k}\,\varepsilon_{kk,R,c} + \omega_{P|k}\,\varepsilon_{kk,P,c}$. Suppose Country A has $(\omega_{R|k}, \omega_{P|k}) = (0.8, 0.2)$ and Country B has $(\omega_{R|k}, \omega_{P|k}) = (0.2, 0.8)$. Then $\varepsilon^{\text{mkt}}_{kk,A} = 0.8(-0.8) + 0.2(-1.6) = -0.96$, $\quad \varepsilon^{\text{mkt}}_{kk,B} = 0.2(-0.8) + 0.8(-1.6) = -1.44$. The difference $(-0.96 \text{ vs. } -1.44)$ is largely *compositional* (who buys how much), not behavioral (how sensitive each group is). Assumption A.3 below removes this confounding by fixing the weights to a common reference $\{\bar{\omega}_{l|k}\}$ (e.g., a base-country or base-year mix) and defining the comparable aggregate.
Formally, to compare elasticities across countries, we impose three additional restrictions:

A.1 Common curvature: the substitution parameter is invariant across countries, $\sigma_c = \sigma$.

A.2 Aligned price base: all elasticities are expressed with respect to the *same* price notion. We keep the effective-price basis used above and take $\delta_{l,k,c}$ as policy-determined and controlled for in cross-country comparisons.[14] Results are invariant after converting $\varepsilon^P_{kk,l,c} = (1-\delta_{l,k,c})\,\varepsilon_{kk,l,c}$.

A.3 Composition fixed: when aggregating, the quantity weights are held constant across countries. Within molecule k, the group weights are $\omega_{l|k} \equiv C_{k,l,c}/\sum_j C_{k,j,c}$ (taken as fixed in c); across molecules, the country weights are $\pi_k \equiv Q_{k,c}/\sum_j Q_{j,c}$ with $Q_{k,c} \equiv \sum_l C_{k,l,c}$ (also fixed).

It is moreover important to notice that moving from *group*-level to *country*-level (market) objects will use quantity additivity (and therefore does not necessitate of particular functional forms for the utility). This is exactly in accordance with the derivation in Appendix A which assumes linear aggregation of demand and elasticities across subgroups, that the posted prices P_j, P_k and costs c_j are common across all consumer segments l (so derivatives are taken with respect to the same price), constant marginal costs, and independence of demand responses between segments.

Under A.1–A.3, the market own-price elasticity for molecule k in country c is the quantity-weighted average

$$\varepsilon^{\text{mkt}}_{kk,c} = \sum_l \omega_{l|k}\,\varepsilon_{kk,l,c},$$

and the country-level average own elasticity is

$$\bar{\varepsilon}_c = \sum_k \pi_k\,\varepsilon^{\text{mkt}}_{kk,c} = \sum_k \sum_l \pi_k\,\omega_{l|k}\,\varepsilon_{kk,l,c}.$$

Relating $\bar{\varepsilon}_c$ to GDP per capita Y_c requires an income–GDP scaling. Let

$$\frac{d\ln E_{l,c}}{d\ln Y_c} = \kappa_l \quad \text{with} \quad \kappa_l > 0,$$

[14] Equivalently, one can convert to posted-price elasticities via $\varepsilon^P_{kk,l,c} = (1-\delta_{l,k,c})\,\varepsilon_{kk,l,c}$; all statements below then carry through with $\varepsilon^P_{kk,l,c}$ and the same weights.

so that a change in Y_c proportionally shifts group l's expenditure. Differentiating and using Appendix A gives the cross-country slope

$$\frac{d\,\bar{\varepsilon}_c}{d\ln Y_c} = \sum_k \sum_l \pi_k\, \omega_{l|k}\, \sigma\, s_{k,l,c}\, \big(\eta^{C_{k,l,c}}_{E_{l,c}} - 1\big)\, \kappa_l.$$

This delivers a direct, model-consistent link between the *country* average own elasticity and income.

Notice that even under A.1-A.3 and the scaling above, the derivative is a weighted sum of terms $\sigma\, s_{k,l,c}\, (\eta^{C_{k,l,c}}_{E_{l,c}} - 1)\, \kappa_l$. It will be small whenever (i) dominant buyers treat the basket as *near necessity* ($\eta^{C_{k,l,c}}_{E_{l,c}} \approx 1$); (ii) relevant molecule shares are low ($s_{k,l,c}$ small); and/or (iii) positive (luxury) and negative (necessity) contributions offset across (k,l) under the fixed weights $\pi_k \omega_{l|k}$. In those (empirically plausible) cases, $\frac{d\,\bar{\varepsilon}_c}{d\ln Y_c}$ is close to zero, yielding a weak (essentially flat) cross-country relationship between average own-price elasticity and GDP per capita despite the within-group mechanics above.

3.2 Ramsey Pricing with Different Income Levels

The following theoretical model explores the efficiency of Ramsey pricing in countries with different income levels, i.e., HIC and MLIC, by focusing on minimizing deadweight loss (DWL) relative to total welfare [15].

A single global firm supplies the product in both markets with constant marginal cost and no binding capacity constraints. Markets are separable: demand in H does not directly affect L, and vice versa.
A supra-national Ramsey regulator sets posted prices by country with the goal of maximizing social welfare across H and L. A single planner with a single break-even constraint is initially adopted as a benchmark because of the possibly confounding roles of heterogeneous shadow prices. Our goal is to study how Ramsey pricing allocates markups across countries. A single supra-country regulator that sets country-specific prices subject to one firm-wide break-even constraint delivers a transparent cross-market Ramsey rule with a common shadow cost (μ) that can be easily compared to the case if each country had its own regulator (and its own cost-coverage or budget constraint). The latter case is equivalent to the case of a single planner facing asymmetric budget pressure (or places different social weights) by country. This would imply that each market would generally have a different shadow cost (μ_H, μ_L). As we will see considering this furrther source of heterogeneity could flip the results with respect to the benchmark case of a single shadoww cost μ.

In each country, market demand is downward sloping and summarizes both clinical need and financial access. Insurance coverage and cost-sharing shape observed demand: in H, broader insurance and lower cost-sharing make demand less price sensitive; in L, heavier reliance on out-of-pocket payments makes demand more price sensitive. We treat these insurance wedges as already embedded in each country's demand since our aim in this Section is more on the efficiency of Ramsey pricing in H and L.

Our analysis focuses solely on regulated outcomes under the supra-national planner. Unregulated monopoly behavior is used as a benchmark. Efficiency is evaluated within each country using the share of deadweight loss in total surplus, computed at the regulated prices. This metric lets us compare how much of the potential surplus is lost to price-induced distortions in H versus L, abstracting from market size.

[15] A similar measure of efficiency in the pharmaceutical market was employed by Guell and Fischbaum (1995) and Guell and Fischbaum (1997) to evaluate various policy strategies aimed at reducing the monopoly power of pharmaceutical firms in the U.S. market. However, their analysis focuses on a single market context. In this framework, a regulator sets prices for a multiproduct monopolist to maximize social welfare while ensuring cost coverage. See also Cutler and Marzilli Ericson (2010).

T2.1 For each country $c \in \{H, L\}$, the market demand for the molecule as a function of the *posted* price P is isoelastic:

$$Q_c(P) = A_c P^{-\varepsilon_c}, \qquad A_c > 0, \;\; \varepsilon_c > 1.$$

(Insurance wedges are absorbed into the scale parameter: if effective demand is $\widehat{Q}_c(\tilde{P}) = \widehat{A}_c \tilde{P}^{-\varepsilon_c}$ with $\tilde{P} = (1-\delta_c)P$, then $Q_c(P) = A_c P^{-\varepsilon_c}$ with $A_c \equiv \widehat{A}_c (1-\delta_c)^{-\varepsilon_c}$.)

T2.2 A single supra-country Ramsey planner sets posted prices $P_{R,c}$ to maximize a weighted sum of country welfares subject to one firm-wide break-even constraint (cost coverage).
The regulator sets posted prices $P_{R,c}$ satisfying $\frac{P_{R,c}-MC}{P_{R,c}} = \frac{\mu}{\varepsilon_c}$ with a common $\mu \in (0,1)$.

T2.3 We assume that most of the drug basket is not a near necessity within group of income people in that country and within that molecule, so that according to Section 3.1 we can safely assume that $\varepsilon_L > \varepsilon_H > 1$. The latter is partly confirmed by the literature (Baltagi and Moscone, 2010). A consequence of $\varepsilon_L > \varepsilon_H > 1$ is that $P_{R,H} > P_{R,L}$ due to the presence of the super-country regulator. These observations also align with the evidence that countries having higher co-insurance rates usually face higher prices under Ramsey pricing (Martínez Giralt and Barros, 2006).

T2.4 A brief study is dedicated also the case when there is not a regulator but an unregulated monopolist firm. In that case, the monopolist's marginal cost is constant and common across countries: $MC > 0$. Namely, the marginal cost (MC) curve is horizontal, reflecting constant marginal costs. For simplicity, we assume that the unregulated monopolist regulator exploits economies of scale, meaning that the cost of producing the last unit is equal to the cost of producing previous units (Cameron et al., 2009; Chaudhuri, 2012; Mackintosh et al., 2015).

In the unregulated monopolist case, the posted price would be equivalent to the monopoly price denoted as $P_{m,c}$ for distinguishing it from the regulated posted price $P_{R,c}$. Namely, for each country $c \in \{H, L\}$, let (effective) market demand be isoelastic with the form specified in Assumption T2.1. Let inverse demand be $P_c = P_c(Q)$ and define total revenue $TR_c(Q) = P_c(Q)\,Q$. The (Marshallian) price elasticity is

$$\eta_c \;\equiv\; \frac{dQ_c}{dP_c}\,\frac{P_c}{Q_c} \;<\; 0, \qquad \varepsilon_c \;\equiv\; |\eta_c| \;>\; 0.$$

Differentiating TR_c and using $\frac{1}{\eta_c} = (dP_c/dQ_c)\,(Q_c/P_c)$ yields

$$MR_c \;\equiv\; \frac{dTR_c}{dQ_c} \;=\; P_c + Q_c\,\frac{dP_c}{dQ} \;=\; P_c\left(1+\frac{1}{\eta_c}\right) \;=\; P_c\left(1-\frac{1}{\varepsilon_c}\right).$$

Under constant marginal cost $MC > 0$, a profit-maximizing monopolist sets $MR_c = MC$, so

$$P_{m,c}\left(1-\frac{1}{\varepsilon_c}\right) = MC \quad\Longleftrightarrow\quad \frac{P_{m,c}-MC}{P_{m,c}} = \frac{1}{\varepsilon_c} \quad\Longleftrightarrow\quad \frac{P_{m,c}}{MC} = \frac{\varepsilon_c}{\varepsilon_c - 1}.$$

If instead prices are chosen by a Ramsey planner with a binding revenue (or cost-coverage) constraint, the first-order condition implies the inverse-elasticity rule

$$\frac{P_{R,c}-MC}{P_{R,c}} = \frac{\mu}{\varepsilon_c}, \qquad \mu \in (0,1) \;\; \text{common across markets},$$

so that $P_{R,c}/MC = \varepsilon_c/(\varepsilon_c - \mu)$; the monopoly formula is recovered as the special case $\mu = 1$. In our framework, we will denote $PS_c = (P_{R,c} - MC)\, Q_c(P_{R,c})$ as the market-level operating surplus earned by the single global firm from country c at the regulator-set price—i.e., margin times quantity (before fixed/R&D costs). The Ramsey planner's single break-even constraint applies to the sum $\sum_c PS_c$ (not to each PS_c individually), which yields a common shadow cost μ and the cross-country inverse-elasticity pricing rule.

Proposition 1:

Under Assumptions (T2.1–T2.3), define consumer surplus (CS), producer surplus (PS), and deadweight loss (DWL) in country c by

$$CS_c = \int_{P_{R,c}}^{\infty} Q_c(P)\, dP, \qquad PS_c = (P_{R,c} - MC)\, Q_c(P_{R,c}), \qquad DWL_c = \int_{MC}^{P_{R,c}} Q_c(P)\, dP - (P_{R,c} - MC)\, Q_c(P_{R,c}).$$

Under isoelastic demand $Q_c(P) = A_c P^{-\varepsilon_c}$ and Ramsey pricing $\dfrac{P_{R,c}}{MC} = \dfrac{\varepsilon_c}{\varepsilon_c - \mu}$, define

$$t_c \equiv \frac{P_{R,c}}{MC} = \frac{\varepsilon_c}{\varepsilon_c - \mu}.$$

Then the regulated and competitive quantities are

$$Q_{R,c} \equiv Q_c(P_{R,c}) = A_c P_{R,c}^{-\varepsilon_c} = A_c\, t_c^{-\varepsilon_c} MC^{-\varepsilon_c}, \qquad Q_c^* \equiv Q_c(MC) = A_c MC^{-\varepsilon_c}.$$

Closed-form expressions:

$$CS_c = \frac{A_c}{\varepsilon_c - 1}\, P_{R,c}^{1-\varepsilon_c} = \frac{A_c}{\varepsilon_c - 1}\, t_c^{1-\varepsilon_c} MC^{1-\varepsilon_c},$$

$$PS_c = (P_{R,c} - MC)\, Q_{R,c} = A_c (t_c - 1) t_c^{-\varepsilon_c} MC^{1-\varepsilon_c},$$

$$DWL_c = \frac{A_c}{\varepsilon_c - 1}\Big(MC^{1-\varepsilon_c} - P_{R,c}^{1-\varepsilon_c}\Big) - (P_{R,c} - MC) A_c P_{R,c}^{-\varepsilon_c} = A_c MC^{1-\varepsilon_c}\left[\frac{1 - t_c^{1-\varepsilon_c}}{\varepsilon_c - 1} - (t_c - 1) t_c^{-\varepsilon_c}\right].$$

Hence the efficiency ratio

$$\frac{DWL_c}{CS_c + PS_c} = \frac{\dfrac{1 - t_c^{1-\varepsilon_c}}{\varepsilon_c - 1} - (t_c - 1) t_c^{-\varepsilon_c}}{\dfrac{t_c^{1-\varepsilon_c}}{\varepsilon_c - 1} + (t_c - 1) t_c^{-\varepsilon_c}}, \qquad t_c = \frac{\varepsilon_c}{\varepsilon_c - \mu}.$$

For any fixed $\mu \in (0, 1]$, this ratio is strictly decreasing *in $\varepsilon_c > 1$ (more elastic demand $\Rightarrow$ lower DWL share). Consequently, if $\varepsilon_L > \varepsilon_H$,*

$$\frac{DWL_L}{CS_L + PS_L} < \frac{DWL_H}{CS_H + PS_H}.$$

Notice that Proposition 1 is not necessarily true for every functional form of the demand. For instance, in the linear case, we should rely on a parallel-shift assumption for the Proposition to apply [16]. Thus Proposition

[16] To illustrate this point, let country c have linear inverse demand $P = \alpha_c - \beta_c Q$ and constant MC. Then

$$Q_c^* = \frac{\alpha_c - MC}{\beta_c}, \qquad Q_{m,c} = \frac{\alpha_c - MC}{2\beta_c} \quad \Rightarrow \quad \Delta Q_c \equiv Q_c^* - Q_m^c = \frac{\alpha_c - MC}{2\beta_c}.$$

(1) implies that, with a unique shadow price μ the share of DWL relative to total welfare is higher in the country L, where demand is more elastic.
This result is however not definitive as, more realistically, the regulator would account for the heterogeneity of H and L facing asymmetric budget pressure. In this case Assumption (T2.2) should be restated as follows:

T2.2' A supra-country planner sets posted prices $P_{R,c}$ to satisfy the country-specific Ramsey rule

$$\frac{P_{R,c}-MC}{P_{R,c}} = \frac{\mu_c}{\varepsilon_c}, \qquad \mu_c \in (0,1], \;\; \mu_c < \varepsilon_c,$$

where μ_c summarizes the planner's (possibly asymmetric) budget pressure / social weights for country c. With one supra-country regulator and *one firm-wide break-even constraint*, there is a single Lagrange multiplier (budget tightness)[17].

Proposition 2:

Let Assumptions (T2.1),(T2.2'),(T2.3),(T2.4) *hold. Thus, Ramsey rule is country-specific:*

$$\frac{P_{R,c}-MC}{P_{R,c}} = \frac{\mu_c}{\varepsilon_c}, \qquad \varepsilon_c > 1, \;\; 0 < \mu_c < \varepsilon_c, \;\; c \in \{H, L\}.$$

Retain all definitions from Proposition 1 *and set*

$$t_c \equiv \frac{P_{R,c}}{MC} = \frac{\varepsilon_c}{\varepsilon_c - \mu_c} > 1, \qquad R_c \equiv \frac{DWL_c}{CS_c + PS_c}.$$

For each c,

$$R_c + 1 = \frac{\varepsilon_c - \mu_c}{\varepsilon_c(1+\mu_c) - \mu_c}\left(\frac{\varepsilon_c}{\varepsilon_c - \mu_c}\right)^{\varepsilon_c}, \qquad \textit{equivalently} \qquad R_c = \frac{\varepsilon_c - \mu_c}{\varepsilon_c(1+\mu_c) - \mu_c}\left(\frac{\varepsilon_c}{\varepsilon_c - \mu_c}\right)^{\varepsilon_c} - 1.$$

where on the domain $\{\varepsilon_c > 1,\; 0 < \mu_c < \varepsilon_c\}$,

$$\frac{\partial R_c}{\partial \varepsilon_c} < 0 \qquad \textit{and} \qquad \frac{\partial R_c}{\partial \mu_c} > 0.$$

i.e. more elastic demand lowers the DWL share; bigger shadow prices μ_c raises it. Define

$$F(\varepsilon, \mu) \equiv \frac{\varepsilon - \mu}{\varepsilon(1+\mu) - \mu}\left(\frac{\varepsilon}{\varepsilon - \mu}\right)^{\varepsilon}.$$

Hence ΔQ_c depends on both the intercept gap $(\alpha_c - MC)$ and the slope β_c. Knowing only that $D'(p) < 0$, that L is "flatter" ($\beta_L < \beta_H$), or that $\varepsilon_L > \varepsilon_H$ at $P_{m,c}$ does not determine the sign of

$$\Delta Q_L - \Delta Q_H = \frac{\alpha_L - MC}{2\beta_L} - \frac{\alpha_H - MC}{2\beta_H}.$$

A sufficient condition for $\Delta Q_L > \Delta Q_H$ is

$$\frac{\alpha_L - MC}{\alpha_H - MC} > \frac{\beta_L}{\beta_H},$$

or, e.g., a parallel-shift assumption $\alpha_L = \alpha_H$ (with common MC), in which case the flatter demand implies $\Delta Q_L > \Delta Q_H$. Absent such restrictions, the inequality cannot be asserted from $D'(p) < 0$ and $\varepsilon_L > \varepsilon_H$ alone.

[17] Here, $\mu_c \in (0,1]$ is not an arbitrary free parameter; it encapsulates the common break-even multiplier and the planner's social weight on country c. Algebraically we can treat μ_c as given (subject to $0 < \mu_c < \varepsilon_c$); economically the collection $\{\mu_c\}$ must be jointly consistent with the single constraint and the weights. That consistency matters for identification, not for the closed-form formulas below.

Then $R_c + 1 = F(\varepsilon_c, \mu_c)$. Hence, for countries H and L,

$$\frac{DWL_L}{CS_L + PS_L} > \frac{DWL_H}{CS_H + PS_H} \quad \Longleftrightarrow \quad F(\varepsilon_L, \mu_L) > F(\varepsilon_H, \mu_H).$$

If $\varepsilon_L > \varepsilon_H$, there exists a unique $\mu_L^* \in (0, \varepsilon_L)$ solving $F(\varepsilon_L, \mu_L^*) = F(\varepsilon_H, \mu_H)$, and therefore

$$\frac{DWL_L}{CS_L + PS_L} > \frac{DWL_H}{CS_H + PS_H} \iff \mu_L > \mu_L^*, \qquad \frac{DWL_L}{CS_L + PS_L} < \frac{DWL_H}{CS_H + PS_H} \iff \mu_L < \mu_L^*.$$

Intuitively, Proposition 2 states that once we allow country-specific tightness, the ranking of DWL in H and L is no longer pinned down by elasticities alone. The DWL share in country c depends on two forces: (i) how *elastic* demand is (ε_c), which *reduces* inefficiency as it rises, and (ii) how *tight* the planner's budget/coverage pressure is (μ_c), which *raises* inefficiency as it rises because it forces a higher markup $P_{R,c}/MC$. As seen in Proposition 1, with $\varepsilon_L > \varepsilon_H$, the usual prediction (with equal tightness) is

$$\frac{DWL_L}{CS_L + PS_L} < \frac{DWL_H}{CS_H + PS_H}.$$

But once tightness differs by country, the ranking is no longer mechanical: if the low-income country faces sufficiently higher tightness μ_L than the high-income country's μ_H, its higher markup can more than offset its greater elasticity, making the *inefficiency share higher in L than in H*. Formally, holding (ε_H, μ_H) fixed and given $\varepsilon_L > \varepsilon_H$, there exists a *unique* threshold $\mu_L^* \in (0, \varepsilon_L)$ such that the two DWL shares are equal at μ_L^*. If $\mu_L > \mu_L^*$, then L's DWL share exceeds H's; if $\mu_L < \mu_L^*$, it remains below. This threshold comes directly from the strict monotonicities we proved: the DWL share *falls* with ε and *rises* with μ.

Consequently, whether Ramsey implies more inefficiencies in H or L depends on μ_c which however is unobserved in our data.

In our analysis we found more evidence of Ramsey pricing in HICs than in MLICs. In our framework, the country-specific tightness μ_c summarizes fiscal capacity, insurance coverage, and distributional weights. Hence, the fact that Ramsey pricing is more evident in HICs than in MLICs is consistent with $\mu_H < \mu_L$: better insurance and financing in HICs (i.e. low μ_H) keep markups aligned with elasticities and the DWL share small, whereas weak coverage and unequal incomes in MLICs (i.e. high μ_L) raise effective tightness, push up markups for posted prices paid out-of-pocket, and—through exclusion—raise the DWL share despite higher elasticities.

3.3 The Role of Own- and Cross-Price Elasticities in Pricing Dynamics of Generics and Branded Drugs

In this section, we turn to analyzing own- and cross-price elasticities within countries to better understand their direct relationship with prices. Fix a molecule m and a country–time cell; all objects refer to this (m,c,t) cell. Let $\mathscr{I}_m$ be the set of products (brand and generic) for molecule m, with $n_m = n_{B,m} + n_{G,m}$, $\mathscr{B}_m = \{1, \dots, n_{B,m}\}$ and $\mathscr{G}_m = \{n_{B,m} + 1, \dots, n_m\}$, so $\mathscr{I}_m = \mathscr{B}_m \cup \mathscr{G}_m$ and $\mathscr{B}_m \cap \mathscr{G}_m = \emptyset$. Indices i, j denote individual products (brand or generic) in the molecule's set $\mathscr{I}_m$. Cross-molecule elasticities are set to zero, so all sums $\sum_{j \neq i}$ below are over $j \in \mathscr{I}_m \setminus \{i\}$.Since the analyses below are within a molecule we avoid repeating that the products that we refer to are $\in \mathscr{I}_m$ in order to ease up the notation. To be more transparent on the subsequent paragraphs, when using subscripts B and G we are referring to a product i having the attribute of being branded or generic respectively, e.g.:

$$p_i = \begin{cases} p_G & \text{if } i \in \mathscr{G}_m, \\ p_B & \text{if } i \in \mathscr{B}_m, \end{cases} \qquad \mathrm{ep}_{ii} = \begin{cases} \mathrm{ep}_{GG} & \text{if } i \in \mathscr{G}_m, \\ \mathrm{ep}_{BB} & \text{if } i \in \mathscr{B}_m, \end{cases} \qquad MC_i = \begin{cases} MC_G & \text{if } i \in \mathscr{G}_m, \\ MC_B & \text{if } i \in \mathscr{B}_m. \end{cases}$$

Production is governed by a common, subadditive cost function $C(\mathbf{q})$, reflecting economies of scale over the relevant range of output. Given such cost subadditivity, a social planner aiming to maximize total consumer surplus CS would find it optimal to allocate production to a single firm in order to minimize costs while internalizing the cross-price dependencies among branded and generic drugs.

This subsection focuses specifically on price dynamics within a single country, which is why the country subscript c is removed from the equations of this section. For a cross-country analysis the reader is referred to Barros and Martinez-Giralt (2008). Conducting the analysis at the country level allows us to examine the relation of both own-price and cross-price elasticities of generics and branded drugs on prices, a topic of significant relevance in the literature as detailed in Section 1. This kind of analysis is particularly important in understanding how generics compete both with each other and with branded drugs once patents expire. As highlighted by the literature, the competition among generics becomes especially fierce post-patent expiration, which has critical implications for pricing strategies and welfare. In this context, we theoretically analyze three types of relationships—natural extensions of the Ramsey pricing framework and the substitution effects between branded and generic drugs – under the assumption that generic pharmaceuticals exhibit higher own-price elasticities than branded pharmaceuticals due to their greater substitutability. (i) First, we examine whether the own-price elasticity of generics is more strongly associated with their own prices than the own-price elasticity of branded drugs is with theirs. (ii) Second, we investigate whether – under a Ramsey policy – cross-price elasticities among generics exert a stronger influence on generic prices than cross-price elasticities among branded drugs do on branded prices. In other words, both in (i) and (ii) we are asking whether observed branded and generic prices behave as if they were chosen optimally (in the Ramsey sense) given the (estimated) elasticities. The theoretical analyses described below will be based on the following assumptions: (e.1) based on Conrad and Lutter (2019), we assume that – within the same molecule, country, and time period as in the present case – generic prices are, on average, approximately 73% lower than branded prices; (e.2) the price of branded is, on average, higher than the price of generics, i.e. $p_B > p_G$; (e.3) generics are, on average, more "own-elastic" than branded $|ep_{GG}| > |ep_{BB}|$ (Yeung et al., 2018; Frank et al., 2021); (e.4) $|\mathrm{ep}_{kG}| > |\mathrm{ep}_{kB}|$ and $|\mathrm{ep}_{jG}| < |\mathrm{ep}_{jB}|$; (e.5) The marginal cost of branded drugs (MC_B) is higher than the one of generics (MC_G). Based on Frank et al. (2021) and Frech et al. (2023), we will assume that the marginal cost of branded drugs is 2 to 3 times higher than that of generics; (e.6) Similarly, $M_B > M_G$.

A discussion on the above assumptions is compelled. Generally, pharmacological research has demonstrated that generics are as effective as branded drugs, particularly when critical factors—such as bioequivalence and the role of excipients—are properly accounted for in clinical trials (Gallelli et al., 2013; Desai et al., 2019; Jarusriwanna et al., 2024; Guru et al., 2025). However, exceptions do exist (Borgheini, 2003). Switching from branded drugs to generics is not a widespread trend among patients. For instance, concerns such as contamination in psychiatric medications have been reported (Desmarais et al., 2011), and both physicians and retailers have noted complaints, often expressing skepticism about whether generics offer the same efficacy as originator brands (Dunne et al., 2014). Indeed, physicians frequently avoid prescribing generics due to perceived quality issues and uncertainties related to pharmacy-driven substitution (Premanath and Kulkarni, 2024).[18] Acceptance of generics tends to be higher among consumers with greater educational attainment, whereas patients from lower socioeconomic backgrounds—often with less formal education—tend to exhibit greater mistrust toward generics. Additionally, many in the general population perceive the switching process itself as costly or burdensome (Steinman et al., 2007; Dunne and Dunne,

[18] For completeness, it is worth noting that in the U.S., insurance formularies generally favor generic drugs over branded alternatives. For example, Dusetzina et al. (2020) found that 84% of Medicare Part D plans covered only the generic version of a drug, while fewer than 1% covered only the brand-name version. However, in certain specialty drug contexts, patients may actually face higher out-of-pocket costs for generics due to rebate structures and the design of Medicare's coverage gap (Dusetzina et al., 2019).

2015). Interestingly, the reverse phenomenon—switching from generics back to branded drugs—can also occur, even in countries with strong generic promotion policies, such as Japan (Hamada et al., 2020). Thus we conclude that, while $\mathrm{ep}_{kG} > \mathrm{ep}_{kB}$ may be observed among more informed segments of the population, heterogeneous perceptions and switching frictions suggest that, on average, $\mathrm{ep}_{kB} > \mathrm{ep}_{kG}$—driven by physician recommendations and the perceived high switching costs associated with common pathologies. Accordingly we assume that, on average, $\mathrm{ep}_{jG} < \mathrm{ep}_{kB}$, i.e. people generally prefer to switch from a branded drug to another than to switch from a branded drug to a generic drug.
It is, hence, easy to see that, under the above justified assumptions that $p_B > p_G$, $|ep_{GG}| > |ep_{BB}|$, that $|\mathrm{ep}_{kG}| > |\mathrm{ep}_{kB}|$ and $|\mathrm{ep}_{jG}| < |\mathrm{ep}_{jB}|$.

As we will see, in this specific exercise, we focus at the product level and assume interdependent demands of goods. This is done for the sake of completeness. As specified in Section 4.2 the cross-elasticities between goods belonging to different molecules are assumed to be zero. However, in many empirical specifications, researchers can control for molecule fixed effects, allowing the analysis to focus within molecules, where cross-elasticities are not zero. In other words, the following theoretical specification is conducted within a country and within a molecule, thus the assumption of independent demands typical of Ramsey settings falls.
Ramsey pricing can apply in pharmaceutical markets where both branded and generic drugs are present, provided that a centralized price-setting entity—such as a regulator or a multi-product monopolist—controls pricing across interdependent goods and internalizes both own- and cross-price elasticities, which is the case of the present setting [19]

Specifically the following result – extending previous works on Ramsey pricing to the more complex scenario of interdependent demands in presence of price distortions and heterogeneous goods– proved in the Appendix A.3 holds

Proposition 3 (Optimal Ramsey markups with insurance and interdependencies in demand): *Consider two drug types $k \in \{B, G\}$ and income groups l, with (i) separable preferences yielding inverse demand in effective prices $\tilde{P}_{k,l}$, (ii) interdependent demands, (iii) per-unit coverage $\delta_{l,k}$ and fiscal distortion ξ, and (iv) a regulator maximizing W as in Eq. (A.5) subject to the break-even constraint $\sum_k (P_k - c_k) q_k = F$ with multiplier λ, where $q_k = \sum_l q_{k,l}$.*
Let $\mu_i := \frac{P_i - c_i}{P_i}$, then the Ramsey rule applies as follows:

$$\frac{P_k - c_k}{P_k} = \sum_{j \neq k} \mu_j M^{(j \to k)}_{\textit{Höffler}} + \sum_{j \neq k} M^{\delta,(j \to k)}_{\textit{Höffler-corrected}} + IDA + SRC\,.$$

where

$$M^{(j \to k)}_{\textit{Höffler}} := -\frac{q_j}{q_k} \frac{\bar{\varepsilon}_{jk}}{\bar{\varepsilon}_{kk}} \frac{P_j}{P_k}, \qquad M^{\delta,(j \to k)}_{\textit{Höffler-corrected}} := \frac{1+\xi}{1+\lambda} \frac{q_j}{q_k} \frac{\bar{\varepsilon}^{\delta}_{jk}}{\bar{\varepsilon}_{kk}} \frac{P_j}{P_k}.$$

and

$$IDA := \frac{\xi\,\delta_k + (1+\xi)\bar{\varepsilon}^{\delta}_{kk}}{(1+\lambda)\,\bar{\varepsilon}_{kk}}, \qquad SRC := -\frac{\lambda}{(1+\lambda)\,\bar{\varepsilon}_{kk}}.$$

[19] Theoretical foundations for this approach under interdependent demand are laid out in Höffler (2006), and extended to regulated and multi-product environments by Schmidt et al. (2001b) and Danzon and Towse (2003b) Notice that we work in an institutional regime. Throughout this subsection we adopt an administered-price regime: a centralized price setter (regulator/single payer) chooses prices subject to a break-even constraint, so Ramsey–Boiteux conditions apply to both branded and generic drugs within a country. In jurisdictions where generics face unconstrained Bertrand competition and prices are not administered, generic prices tend toward marginal cost and the Ramsey rule ceases to describe their equilibrium (Danzon, Towse, and Mestre-Ferrandiz, 2015b); our analysis should then be interpreted as (i) applying to branded drugs with residual market power, and (ii) providing a normative counterfactual for generics. Empirically, when generics are effectively competitive, one may set $M_G \approx 0$, which collapses the generic-side cross terms accordingly.

In our data we do not have explicit information about health insurance directly in each country. Therefore we cannot measure its distortion directly unless we consider it to be absorbed by the country-time fixed effects. Setting insurance/distortion to zero in Proposition 3 ($\xi = \delta = 0$) and identify notation $\mu_i := \frac{P_i - c_i}{P_i} \equiv M_i$, $P_i \equiv p_i$, $c_i \equiv MC_i$, and $\bar{\varepsilon}_{jk} \equiv \mathrm{ep}_{jk}$, the Ramsey–Boiteux markup condition within a country reduces to the one in Höffler (2006)

$$M_i \;=\; -\frac{\lambda}{1+\lambda}\frac{1}{\mathrm{ep}_{ii}} \;-\; \sum_{j\neq i} M_j \frac{\mathrm{ep}_{ji}}{\mathrm{ep}_{ii}}\frac{q_j}{q_i}\frac{p_j}{p_i}. \tag{3}$$

where λ is the Lagrange multiplier, M_i is the relative markup of price over marginal cost, ep_{ii} is the own-price elasticity of demand for good i, and the second term aggregates cross-price elasticities ep_{ji} with $j \neq i$.[20]

Define

$$DR \;:=\; \frac{\lambda}{1+\lambda}\frac{1}{\mathrm{ep}_{ii}}, \qquad CI \;:=\; \sum_{j\neq i} M_j \frac{\mathrm{ep}_{ji}}{\mathrm{ep}_{ii}}\frac{q_j}{q_i}\frac{p_j}{p_i},$$

where DR is the (direct) Ramsey inverse-elasticity component and CI captures cross-product interactions induced by demand interdependence. Using $M_i = (p_i - MC_i)/p_i$, the *exact* price identity is

$$p_i \;=\; \frac{MC_i}{1+DR+CI}. \tag{4}$$

For exposition, when $|DR+CI|$ is sufficiently small we use the first–order approximation

$$p_i \;\approx\; MC_i\left[1-(DR+CI)\right] \;=\; MC_i\left[1-\frac{\lambda}{1+\lambda}\frac{1}{\mathrm{ep}_{ii}} - \sum_{j\neq i} M_j \frac{\mathrm{ep}_{ji}}{\mathrm{ep}_{ii}}\frac{q_j}{q_i}\frac{p_j}{p_i}\right], \tag{5}$$

which truncates the geometric series $(1+X)^{-1} = 1 - X + X^2 - \cdots$ at $X := DR + CI$. The approximation error equals $MC_i \cdot \frac{(DR+CI)^2}{1+DR+CI}$, so accuracy is high when $|DR+CI| \ll 1$ (and $1+DR+CI > 0$).[21] In typical pharmaceutical settings, $\mathrm{ep}_{ii} < 0$ implies $DR < 0$, and for substitutes $\mathrm{ep}_{ji} > 0$ together with $\mathrm{ep}_{ii} < 0$ and $M_j \geq 0$ imply $CI < 0$, so $DR + CI$ is negative and often small in magnitude. Hence, in what follows we report comparative statics using the linear approximation (5). As a further argument in favor of the $|DR + CI| < 1$ case, notice that DR will be likely negative and modest in magnitude (we maintain in any case that $1 + DR + CI > 0$ so the exact formula is also well-defined) given that, on average, own elasticites of drugs are negative and $0 < \frac{\lambda}{1+\lambda} < 1$. Moreover, CI is negative for substitutes. Even in the case of complements, CI is positive but likely contained such that $DR + CI$ remains lower than 1. For instance, (Duggan and Morton, 2010) show that even substantial insurance-driven changes in drug coverage, such as under Medicare Part D, produce only moderate utilization responses—indirectly suggesting that real-world cross-demand elasticities, including those arising from complementarity, are limited in magnitude. [22]

$$p_i = MC_i\left[1 - \frac{\lambda}{1+\lambda}\cdot\frac{1}{\mathrm{ep}_{ii}} - \sum_{j\neq i} M_j \cdot \frac{\mathrm{ep}_{ji}}{\mathrm{ep}_{ii}}\cdot\frac{q_j}{q_i}\cdot\frac{p_j}{p_i}\right] \tag{6}$$

The following analysis will involve $\frac{\partial p_i}{\partial \mathrm{ep}_{ii}}$. Importantly, $\frac{\partial p_i}{\partial \mathrm{ep}_{ii}}$ is framed as a theoretical policy counterfactual, without any causal claim. The estimated elasticities ep_{ii}, obtained ex post from a prior demand

[20] We do not adopt Appendix A.3 here since it incorporates insurance-induced distortions; to keep the exposition focused we work in the no-insurance case. The expression above is standard; see Höffler (2006) for derivation. It also nests the special case in Appendix A when there is a single income group and $\xi = \delta_{l,k} = 0$.

[21] From $M_i = (p_i - MC_i)/p_i$ we get $p_i = MC_i/(1-M_i)$. Substituting $M_i = -(DR+CI)$ yields $p_i = MC_i/(1+DR+CI)$, i.e. Eq. (4). Eq. (5) follows from the first–order expansion of $(1+DR+CI)^{-1}$.

[22] Similar results, available upon request, are however reached for the case $DR + CI > 1$.

model, are treated as descriptive summaries of demand responsiveness. Our goal is not to estimate the effect of elasticity on price, but to examine whether observed prices are consistent with Ramsey-Boiteux pricing logic. In this sense, the derivative is a diagnostic tool: it indicates how prices should respond to elasticity under normative assumptions, without implying endogeneity or reverse causality. To enforce this argument we highlight that when answering (i) and (ii) we will make use of the envelope theorem. For instance, in line with the envelope theorem, the derivative $\frac{\partial p_i}{\partial \mathrm{ep}_{ii}}$ will isolate the direct sensitivity of optimal price to own-price elasticity, holding cross-price elasticities constant. Since our setting allows for the estimation of both own and cross elasticities –as detailed in the methodological section– this comparison reflects a local policy counterfactual under Ramsey logic abstracting from indirect effects through re-optimization.

We first focus on answering question (i). We provide here, in particular, a formal explanation for the contrasting results encountered in the literature when trying to address the effect of own elasticity on own price. To understand the effect of own-price elasticity on price, we take the partial derivative of p_i with respect to ep_{ii}. Differentiating term by term, the first term yields $\frac{\partial}{\partial \mathrm{ep}_{ii}}\left(-\frac{\lambda}{1+\lambda}\frac{1}{\mathrm{ep}_{ii}}\right) = \frac{\lambda}{1+\lambda}\frac{1}{\mathrm{ep}_{ii}^2}$, and the second term contributes $\frac{\partial}{\partial \mathrm{ep}_{ii}}\left(-\sum_{j\neq i} M_j \frac{\mathrm{ep}_{ji}}{\mathrm{ep}_{ii}}\frac{q_j}{q_i}\frac{p_j}{p_i}\right) = \sum_{j\neq i} M_j \frac{\mathrm{ep}_{ji}}{\mathrm{ep}_{ii}^2}\frac{q_j}{q_i}\frac{p_j}{p_i}$. Combining these, we study a within-country, drug-by-drug comparative static: ep_{ii} perturbs the i^{th} Ramsey FOC while holding $\{p_j\}_{j\neq i}$ and $\{\mathrm{ep}_{ji}\}_{j\neq i}$ fixed. Because CI_i contains p_j/p_i, the total derivative includes a feedback factor $\left(1 - MC_i \sum_{j\neq i} M_j \frac{\mathrm{ep}_{ji}}{\mathrm{ep}_{ii}}\frac{q_j}{q_i}\frac{p_j}{p_i^2}\right)^{-1}$. A full multi-product re-optimization would add cross-price adjustments $dp_j/d\mathrm{ep}_{ii}$; under moderate cross-interactions (Assumptions (e.4)–(e.6)) these are second-order, so our ranking results for (i) are unaffected[23]. We obtain

$$\frac{\partial p_i}{\partial \mathrm{ep}_{ii}} = \frac{MC_i\left[\frac{\lambda}{1+\lambda}\frac{1}{\mathrm{ep}_{ii}^2} + \sum_{j\neq i} M_j \frac{\mathrm{ep}_{ji}}{\mathrm{ep}_{ii}^2}\frac{q_j}{q_i}\frac{p_j}{p_i}\right]}{1 - MC_i \sum_{j\neq i} M_j \frac{\mathrm{ep}_{ji}}{\mathrm{ep}_{ii}}\frac{q_j}{q_i}\frac{p_j}{p_i^2}}. \tag{7}$$

Without loss of generality, we can hence define the sets $\mathscr{G}'$ and $\mathscr{B}'$ as follows: if product $i \in \mathscr{G}$ (i.e., i is a generic good), then $\mathscr{G}' = \mathscr{G}\setminus\{i\}$ denotes the set of *other generic goods*; otherwise, if $i \in \mathscr{B}$, we let $\mathscr{G}' = \mathscr{G}$. Analogously, if $i \in \mathscr{B}$, we define $\mathscr{B}' = \mathscr{B}\setminus\{i\}$ to denote the set of *other branded goods*; otherwise, if $i \in \mathscr{G}$, we set $\mathscr{B}' = \mathscr{B}$.

Notice that such partitions derive from simply splitting the sum over j into $\mathscr{G}'$ and $\mathscr{B}'$ is therefore just a relabelling of the Höffler-type cross terms by product class in Eq.(3).

These definitions allow us to partition the cross-elasticity terms into contributions from branded and generic products, conditional on the identity of product i as follows:

$$\sum_{j\neq i} M_j \frac{\mathrm{ep}_{ji}}{\mathrm{ep}_{ii}^2}\frac{q_j}{q_i}\frac{p_j}{p_i} = \sum_{k\in\mathscr{G}'} M_k \frac{\mathrm{ep}_{ki}}{\mathrm{ep}_{ii}^2}\frac{q_k}{q_i}\frac{p_k}{p_i} + \sum_{j\in\mathscr{B}'} M_j \frac{\mathrm{ep}_{ji}}{\mathrm{ep}_{ii}^2}\frac{q_j}{q_i}\frac{p_j}{p_i}$$

[23] As aforementioned we adopt here a comparative-static scope. In an administered-price regime, the regulator chooses $p = (p_1,\ldots,p_N)$ by solving all Ramsey FOCs jointly, $G(p;\varepsilon) = 0$. The exact response to a perturbation in the own elasticity ε_{ii} is $dp/d\varepsilon_{ii} = (I-H)^{-1}u$, where $H := \partial G/\partial p$ is the Jacobian and $u := \partial G/\partial \varepsilon_{ii}$ has support only on coordinate i. Our drug-by-drug experiment holds $\{p_j\}_{j\neq i}$ and $\{\varepsilon_{ji}\}_{j\neq i}$ fixed; because CI_i contains p_j/p_i, the ith equation features an own-equation feedback $(1-\Theta_i)^{-1}$ with $\Theta_i := MC_i \sum_{j\neq i} M_j \frac{\varepsilon_{ji}}{\varepsilon_{ii}}\frac{q_j}{q_i}\frac{p_j}{p_i^2}$. Off-diagonal effects (how p_k feeds into the ith FOC) enter only via loop terms of order H^2 and higher; under moderate cross-interactions (Assumptions (e.4)–(e.6)) these are second-order and do not affect the branded–generic ranking in (i).

This leads to:

$$\Theta_i := MC_i \left[\sum_{k \in \mathscr{G}'} M_k \frac{\text{ep}_{ki}}{\text{ep}_{ii}} \frac{q_k}{q_i} \frac{p_k}{p_i^2} + \sum_{j \in \mathscr{B}'} M_j \frac{\text{ep}_{ji}}{\text{ep}_{ii}} \frac{q_j}{q_i} \frac{p_j}{p_i^2} \right].$$

$$\frac{\partial p_i}{\partial \text{ep}_{ii}} = \frac{MC_i \left[\frac{\lambda}{1+\lambda} \frac{1}{\text{ep}_{ii}^2} + \sum_{k \in \mathscr{G}'} M_k \frac{\text{ep}_{ki}}{\text{ep}_{ii}^2} \frac{q_k}{q_i} \frac{p_k}{p_i} + \sum_{j \in \mathscr{B}'} M_j \frac{\text{ep}_{ji}}{\text{ep}_{ii}^2} \frac{q_j}{q_i} \frac{p_j}{p_i} \right]}{1 - \Theta_i}. \tag{8}$$

Assumption (e.3) implies that the own-price elasticity term $\frac{\lambda}{1+\lambda} \frac{1}{\text{ep}_{ii}^2}$ is smaller for generics because $|\text{ep}_{GG}|$ is larger than $|\text{ep}_{BB}|$. Now, the interesting aspect lies in comparing $\frac{\partial p_i}{\partial \text{ep}_{ii}}$ in the case where i is a generic good (G) versus when i is a branded drug (B). However, the latter comparison remains ambiguous even under assumptions (e.1)–(e.6) due to two opposing forces: one reflecting strong generic–generic competition, and the other reflecting branded–branded substitution—each influencing price sensitivity in opposite directions.[24] Also, because $(1 - \Theta_i)^{-1}$ may be larger for generics (lower p_i) yet their M_k are smaller, the net impact on $\frac{\partial p_i}{\partial \text{ep}_{ii}}$ remains theoretically ambiguous.

Hypothesis 1 (H1): *The following holds:*

H1.1 When controlling for cross-elasticities, generic pharmaceuticals exhibit higher own-price elasticities than branded pharmaceuticals in absolute value.

H1.2 Under the Ramsey pricing framework within a country, whether branded or generic drug prices are more sensitive to changes in own-price elasticity remains theoretically ambiguous. This ambiguity stems from the trade-off between a larger direct elasticity component for branded drugs—due to lower own-price elasticity in absolute terms—and stronger cross-substitution effects among generics, both of which are weighted by market shares and relative price–quantity terms. The net effect depends on the relative magnitudes of these opposing forces and cannot be signed without additional empirical information.

Hence, the relationship between price and own-price elasticity remains ambiguous. Branded drugs, with less elastic demand, are more affected by direct markup adjustments because Ramsey pricing allows higher prices when consumers are less sensitive to price changes. As a result, even small shifts in elasticity can lead to notable price changes (we refer to the term $\frac{\lambda}{1+\lambda} \cdot \frac{1}{\text{ep}_{ii}^2}$ here). In contrast, generics operate in a highly competitive environment where many close substitutes exist. This makes their prices more responsive not just to their own elasticity, but also to how they interact with other generics—i.e., cross-substitution effects (we refer here to the sums in ep_{ji} and ep_{ki}). These effects amplify price responses, since changing the price of one generic can shift demand significantly across others. (Laffont and Tirole, 2001). These two forces pull in opposite directions, and without precise data on their relative strength, the overall pricing response to elasticity changes cannot be determined a priori. Coherently, as shown in Appendix D, the relationship between

[24] Namely, the following relationships remain unknown to literature

$$\sum_{k \in \mathscr{G}'} M_k \cdot \text{ep}_{kB} \cdot \frac{q_k}{q_B} \cdot \frac{p_k}{p_B} \gtreqless \sum_{k \in \mathscr{G}'} M_k \cdot \text{ep}_{kG} \cdot \frac{q_k}{q_G} \cdot \frac{p_k}{p_G}$$

$$\sum_{j \in \mathscr{B}'} M_j \cdot \text{ep}_{jB} \cdot \frac{q_j}{q_B} \cdot \frac{p_j}{p_B} \gtreqless \sum_{j \in \mathscr{B}'} M_j \cdot \text{ep}_{jG} \cdot \frac{q_j}{q_G} \cdot \frac{p_j}{p_G}$$

as well as whether they compensate with the term $\frac{\lambda}{1+\lambda} \cdot \frac{1}{\text{ep}_{ii}^2}$.

price and own-price elasticity across molecule-country pairs –when controlling for cross-price elasticities– appears ambiguous when considering all pharmaceutical products (Figure D.1). To further understand the drivers of pricing behavior under Ramsey logic, we investigated the special case of an economy composed solely of generic pharmaceuticals (Figure D.2). We focused on this scenario because generics constitute a more homogeneous and competitive segment of the pharmaceutical market, where own-price elasticities tend to be higher in absolute value ($|\mathrm{ep}_{GG}| \gg |\mathrm{ep}_{BB}|$) and prices are substantially lower than those of branded drugs. This makes generics a theoretically cleaner environment to isolate the relationship between elasticity and price, free from brand-related frictions. In such a setting, Ramsey pricing theory predicts that if demand becomes less elastic—i.e., ep_{GG} becomes less negative—then the price should increase, as the inverse elasticity term $\frac{1}{\mathrm{ep}_{GG}^2}$ rises. Our empirical analysis (Figure D.2) confirms this theoretical intuition. Higher prices are consistently associated with less negative elasticities (i.e., demand becomes less elastic), indicating that the direct term $\frac{1}{\mathrm{ep}_{GG}^2}$ dominates the overall expression for the partial derivative $\frac{\partial p_G}{\partial \mathrm{ep}_{GG}}$, i.e. it overcomes the influence of other generics in the economy on drug $i = G$ (this aligns with the expectations in the literature (Schmidt et al., 2001b; Danzon and Towse, 2003b; Danzon, Towse, and Mestre-Ferrandiz, 2015a)). Namely, the contribution from the cross-elasticity terms, while present, does not overturn the sign of the relationship.

Secondly, we focus on question (ii). To address this, it is more convenient to rewrite Equation (6) as follows:

$$p_i = MC_i \left[1 - \frac{\lambda}{1+\lambda} \cdot \frac{1}{\mathrm{ep}_{ii}} - \sum_{k \in \mathscr{G}'} M_k \cdot \frac{\mathrm{ep}_{ki}}{\mathrm{ep}_{ii}} \cdot \frac{q_k}{q_i} \cdot \frac{p_k}{p_i} - \sum_{j \in \mathscr{B}'} M_j \cdot \frac{\mathrm{ep}_{ji}}{\mathrm{ep}_{ii}} \cdot \frac{q_j}{q_i} \cdot \frac{p_j}{p_i} \right]$$

where $k \in \mathscr{G}'$ (other generics) and $j \in \mathscr{B}'$ (other branded).
The latter decomposition, makes it easier to distinguish the cases where the effect is taken with respect to a branded or a generic product:

$$\frac{\partial p_i}{\partial ep_{ki}} = \frac{-MC_i\, M_k\, \dfrac{1}{ep_{ii}}\, \dfrac{q_k}{q_i}\, \dfrac{p_k}{p_i}}{1 - MC_i \displaystyle\sum_{j \neq i} M_j\, \dfrac{ep_{ji}}{ep_{ii}}\, \dfrac{q_j}{q_i}\, \dfrac{p_j}{p_i^2}} \tag{9}$$

$$\frac{\partial p_i}{\partial ep_{ji}} = \frac{-MC_i\, M_j\, \dfrac{1}{ep_{ii}}\, \dfrac{q_j}{q_i}\, \dfrac{p_j}{p_i}}{1 - MC_i \displaystyle\sum_{\ell \neq i} M_\ell\, \dfrac{ep_{\ell i}}{ep_{ii}}\, \dfrac{q_\ell}{q_i}\, \dfrac{p_\ell}{p_i^2}}\,. \tag{10}$$

The interested reader is referred to the Appendix D for the technical details leading to the above equalities.
The derivatives in (9)–(10) imply that—under Ramsey pricing with interdependent demands—when goods i and k (or j) are *substitutes* ($ep_{ki}, ep_{ji} > 0$) the optimal price of good i *increases* with stronger cross-substitution: $\frac{\partial p_i}{\partial ep_{ki}} > 0$ and $\frac{\partial p_i}{\partial ep_{ji}} > 0$ (given nonnegative markups and a positive denominator). Intuitively, greater substitutability means rival prices shift more demand toward i, loosening the revenue–welfare trade-off on i and supporting a higher Ramsey price. The magnitude of this adjustment scales with rivals' profitability ($\pi_k = p_k q_k$, $\pi_j = p_j q_j$), the own elasticity and market size of i (ep_{ii}, q_i, p_i), and the strength of the cross-demand terms. Hence, for substitutes, the comparative statics are strictly positive irrespective of whether i is branded or generic.

Subsection D.1 in the Appendices recover criterion (D.8) from Eq. (9) (symmetrically for equation (10)):

$$\mathscr{S}_G > \mathscr{S}_B \iff \frac{\sum_{k\in\mathscr{G}'} M_k}{D_G} > \frac{\sum_{j\in\mathscr{B}'} M_j}{D_B} \iff \frac{\sum_{k\in\mathscr{G}'} M_k}{\sum_{j\in\mathscr{B}'} M_j} > \frac{D_G}{D_B}. \tag{11}$$

where $\mathscr{S}_G(\mathscr{S}_B)$ represent is a simple index of how much, in total, other generics (branded) can move a generic (branded) drug's price once you account for the price brake D_i on that drug. Specifically,

$$D_i = 1 + MC_i \sum_{m\neq i} M_m \frac{|\mathrm{ep}_{mi}|}{|\mathrm{ep}_{ii}|} \frac{q_m}{q_i} \frac{p_m}{p_i^2} \; > \; 1.$$
$$=: 1 + MC_i K_i, \qquad K_i := \sum_{m\neq i} M_m \frac{|\mathrm{ep}_{mi}|}{|\mathrm{ep}_{ii}|} \frac{q_m}{q_i} \frac{p_m}{p_i^2}.$$

The left-hand side of (11) is the total "markup mass" among generic rivals; the right-hand side is the extra dampening on the focal generic relative to the focal brand. Generics "win" (their cross-elasticities move prices more) when generic rivals' total markups are large compared with the generic's dampening. Brands "win" when brand rivals' total markups are so large that they outweigh the brand's stronger dampening.

We now try to understand the direction of the inequality. To do so we exploit assumptions (e.3)–(e.6). Under (e.5), with MC_B being 2–3 $\times MC_G$, we typically have $D_B > D_G$ (brands are more damped); and under (e.3), because $|\mathrm{ep}_{GG}| > |\mathrm{ep}_{BB}|$, the ratios $|\mathrm{ep}_{mi}|/|\mathrm{ep}_{ii}|$ are smaller for the generic focal good, which tends to reduce K_G and hence D_G relative to D_B. These forces often tilt the comparison toward $\mathscr{S}_G > \mathscr{S}_B$, but the conclusion is ultimately empirical because it depends on the realized sums of markups. In particular the direction is determined mainly by competition forces in the ATC markets of the products under scrutiny. As an example, in the limit of perfect competition in the generic segment (generic markups $\to 0$), generic–generic cross-price elasticities have no effect on generic prices p_G: is pinned to marginal cost. Hence our comparison is meaningful only when at least one side has positive markups (e.g., branded segment or regulated/tendered generics). If brands retain positive markups, brand–brand cross elasticities affect branded prices while the generic-side effect is zero; if both segments are competitive, both effects are zero.

Hypothesis 2 (H2): *For substitutes and $D_i > 1$, the branded (resp. generic) price is more responsive to within–category cross–price elasticity shifts iff the corresponding within–category sensitivity index is larger:*

$$\frac{\partial p_B}{\partial \mathrm{ep}_{jB}} \gtrless \frac{\partial p_G}{\partial \mathrm{ep}_{kG}} \iff \mathscr{S}_B \gtrless \mathscr{S}_G \iff \frac{\sum_{j\in\mathscr{B}'} M_j}{D_B} \gtrless \frac{\sum_{k\in\mathscr{G}'} M_k}{D_G} \iff \frac{\sum_{j\in\mathscr{B}'} M_j}{\sum_{k\in\mathscr{G}'} M_k} \gtrless \frac{D_B}{D_G}.$$

In words: brand–brand (resp. generic–generic) cross elasticities move prices more than generics when the cumulative rival markup on branded (generic) drugs exceeds the cumulative rival markup on generic (branded) drugs after accounting for the side–specific price "brake" D. In the limit of perfect competition for generics ($M_k \to 0$), $\mathscr{S}_G = 0$ and the brand side weakly dominates.

Intuitively, Hypothesis 2 shows that which side's prices move more –brands with other brands or generics with other generics – comes down to two forces: (i) the total markup at stake among that side's rivals (we call it the "push") and (ii) the side-specific price "brake" D that dampens any adjustment. In multi-product Ramsey with interdependent demand, markups are tied together through cross-elasticities, so a side with larger cumulative rival markups has more scope to reallocate rents, while a bigger brake – i.e. if the focal product is costly to make or small price changes trigger big quantity feedbacks – mutes the response.

Thus, brand–brand cross elasticities move branded prices more when brand markup mass beats the brand brake; generic–generic cross elasticities move generic prices more when generic markup mass beats the (typically lighter) generic brake. In the limit of perfect competition for generics (generic markups $\rightarrow 0$), the generic index collapses and the brand side dominates mechanically. It is worth mentioning that this logic mirrors multiproduct Ramsey with interdependent demands (Höffler, 2006) and the pass-through principle that price responses scale with markups and a system "denominator," while applying Ramsey reasoning to the branded-vs-generic margin in pharma (Danzon and Towse, 2003b).

In many mature markets, generic shares are high and markets are fragmented, yet brand loyalty persists (Berndt and Aitken, 2011; Panchal et al., 2012; Costa-Font et al., 2014), so branded products often retain positive markups, while generics operate with thin margins. In those settings, the brand side typically dominates (larger markup mass despite a heavier brake). By contrast, early post-patent or limited-entry/tender settings can leave meaningful margins among generics and a lighter brake, making the generic side more responsive. Empirically, these patterns align with (i) the post–Hatch–Waxman rise of generics alongside enduring brand loyalty/segmentation , and (ii) evidence that increases in generic use can outpace decreases in brand use due to segment effects and regulatory structures (Blankart and Vandoros, 2024) –both of which support the plausibility of our criterion.

4 Empirical Analyses

We present below our empirical analysis. We acknowledge here that the analyses performed in this Section are designed to detect the presence of Ramsey-like patterns, rather than to structurally estimate the full model.

However the former is an interesting problem per se as the major challenge faced by the literature to detect Ramsey-like structures within and between countries has been to identify and estimate own and cross-price elasticities which is the main empirical effort of the Section.

4.1 Data

To estimate the own and cross price elasticities of pharmaceuticals, in line with the theoretical part above, we utilise pharmaceutical data from IQVIA[25], formerly IMS Health and Quintiles. The original database comprises an unbalanced panel dataset. More precisely, it comprises 1.3 million product-level observations covering about 45,000 different products in 34 countries over a varying number of quarters (from a minimum of 2 to a maximum of 48) in the period from 2008 to 2020. The data contain various product-level information, such as the active ingredient of the drug, the company that manufactured it, therapeutic information, the generic name along with packaging information, sales information and other product-specific covariates. In order to obtain a balanced dataset – which is required for the correct application of the Double Debiased Machine Learning algorithm – we selected a subset of molecules that have been sold in many countries over a sufficiently long period of time. The resulting dataset has a molecule-wise panel structure and covers 48 quarters for each molecule in each country in a comparable manner. The dataset provides quarterly sales in USD and quarterly standard units in thousands for each product containing the selected molecules as active ingredients. In this way, we can calculate the average ex-manufacturer prices for each product in each country. Other variables in the database are the year of launch of the products, the number of products for each molecule, the manufacturer, the number of companies dealing with each molecule as an indicator of competition in the market, and the anatomical therapeutic class of the product. According to the anatomical

[25] https://www.iqvia.com

therapeutic classification (ATC), drugs are categorized into different groups based on the organ or system they act on and their chemical, therapeutic and pharmacological properties.

Our sample comprises 34 countries — including 10 MLICs and 24 HICs— and 33 molecules for the period 2008-2020[26], summing up to 53,856 observations at the molecule-country-quarter level.The current version of the DDML algorithm works better with balanced panel data and sufficiently long time series (Semenova et al., 2021). Therefore, we created a sample with an even distribution of molecules to obtain a balanced panel, resulting in a lower number of molecules and countries than available in the dataset. Therefore, we focus our work on widely used products that have been available in many countries for a long time.

The dependent variable for the estimation of the elasticities is the total demand for the molecule G, in country c, in quarter j measured in log standard units. The treatment variable is the average log price of the molecule G, in country c, in quarter j. We compute the price of a standard unit by taking the average of the prices of drug based on molecule G weighted by the standard units sold in quarter j. The covariates are the number of firms that produce a product inclusive of molecule G in the country c in quarter j, the percentage of generic products among all the products inclusive of molecule G in the country c in quarter j, yearly log GDP per capita based on purchasing power parity (PPP), Gini index, fixed effects for the quarter, country, molecule and ATC at the first-level. In addition, previous realizations of the demand system in the form of 48 lagged variables of quarterly molecule prices and sales are included as covariates.

The number of firms controls the level of competition in the market. If there are relatively more firms producing drugs inclusive of molecule G, some degree of price competition is expected, which may lead to lower prices and vice versa. The percentage of generic products also controls for the level of competition, but also for the degree of novelty of molecule G. Originator drugs are patented after their invention. Hence firms cannot produce the same drug during its patent life, leading to a monopolistic market structure in which originator drugs can be sold at pretty high prices. After the originator drug becomes off-patent, other firms can produce the same drug (generators), and consequently, the price tends to decrease. The percentage of generator drugs accounts for this process. GDP per capita controls the level of economic development. Gini index is included in the model to account for income inequalities in countries. The literature on pharmaceutical pricing underlines that prices might be lower in low-income countries. Moreover, some studies argue that pharmaceutical firms tend to apply different prices for different income groups to benefit the income inequalities in low and middle-income countries. We thus control for GDP per capita and Gini index to partly address these pricing strategies. GDP PPP and Gini index data are obtained from the World Development Indicators database of the World Bank.

[26] Observations start in the third quarter of 2008 and end in the third quarter of 2020

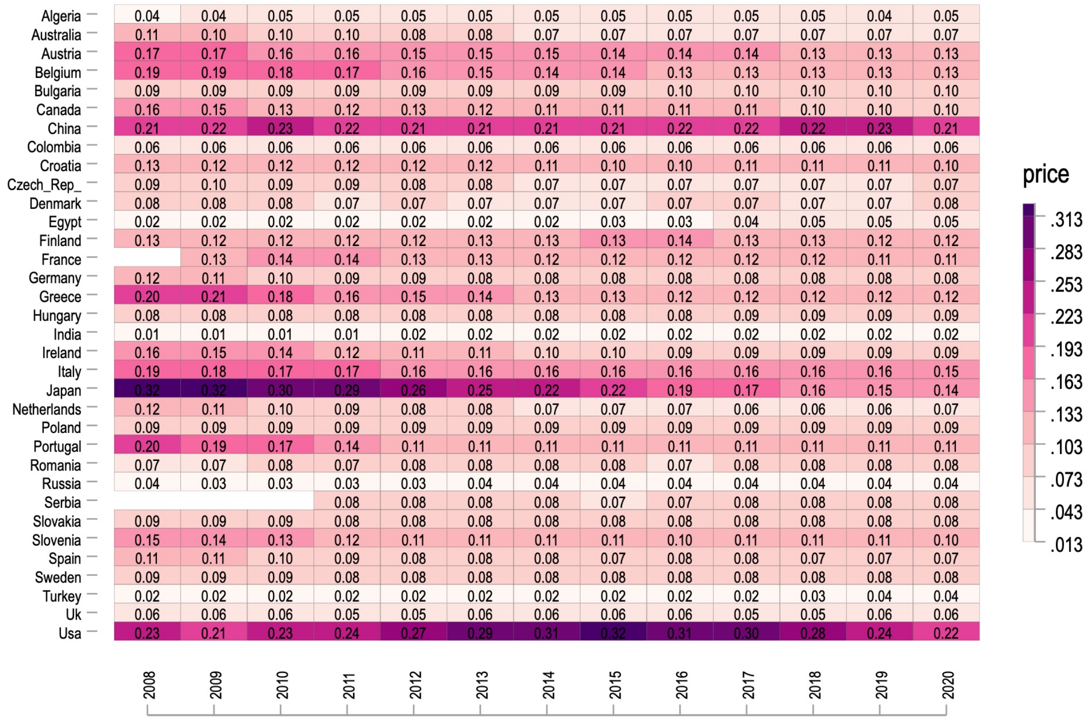


Figure 1. Average Yearly Price of Pharmaceuticals by Country. Each cell represents the average price of pharmaceutical products, measured in standard units, per country and year.

Figure 1 shows the average drug prices per standard unit over the years and markets.[27] The price data are measured in US dollars from 2008 to 2020, averaged across 33 molecules and weighted by standard units. The US, China, Japan and Italy have the highest average drug prices, while India, Turkey and Russia have the lowest average prices. The figure clearly shows that the average prices in some HICs are lower than the prices in some MLICs. For example, the United Kingdom and the Netherlands have lower average prices than Serbia and Romania.

[27] Values are missing in some countries because the range of molecules in these countries did not include a sufficient number of products after the balanced dataset was created. This inadequacy prevented a reliable estimate of their average price.

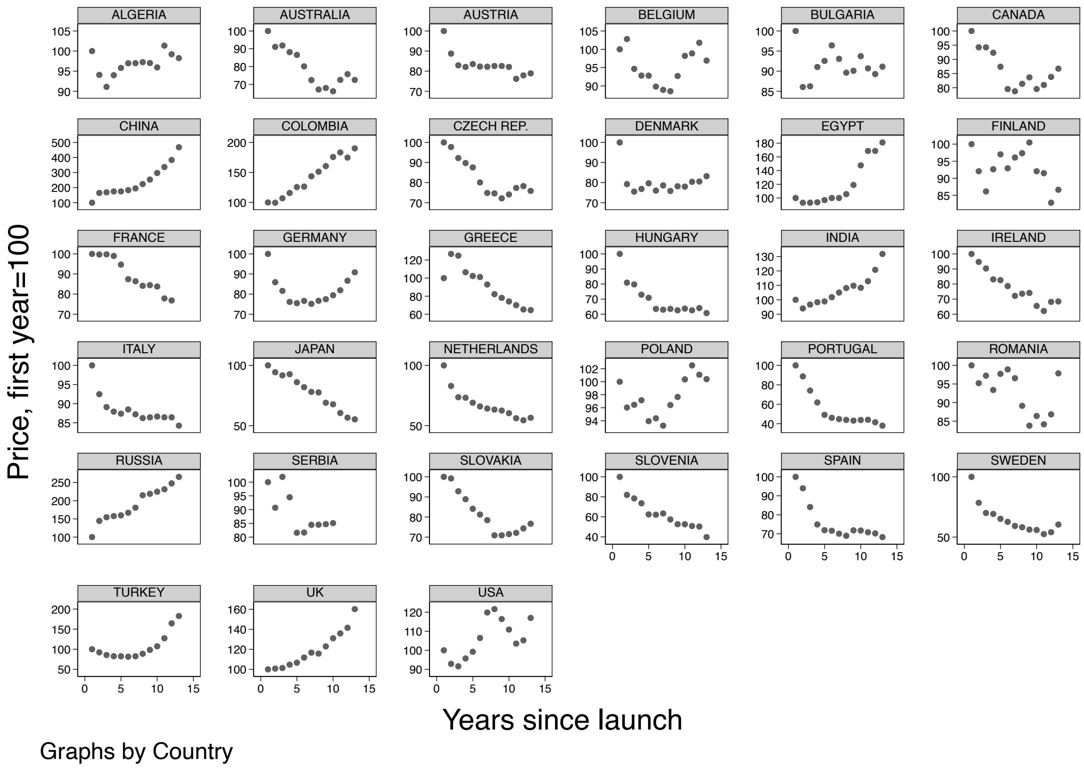


Figure 2. Normalised Price Trends of Molecules. Each subfigure displays the price trends of molecules from their launch year in a given country.

Figure 2 shows the price shifts of the molecules over the next 5, 10 and 15 years from the year of launch to give an indication of the price trends in the countries of the dataset. The year of launch is denoted by 0, and the normalized starting price by 100. The year of launch is the year in which the first product of the molecule G was sold on the national market of country c. The normalized initial price represents the average price of the products of the molecule G in the country c in the year t. Figure 2 shows that the price of the molecule tends to change linearly in most countries, *i.e.* it either rises or falls. This revealed pattern signals the need to take price trends into account when calculating the price elasticities of demand. If this were not the case, one could argue that the quantities of a molecule sold in a country tend to increase (decrease) as prices naturally tend to decrease (increase). Thus, estimates of the price elasticity of demand based on the price evolution of molecules in national markets turn out to be biased in absolute terms. We avoid these concerns by accounting for price lags in the demand model. A detailed explanation of where and how price lags are incorporated into the analysis is provided in the next section.

4.2 Methodology

The methodological approach we follow in this paper – debiased/double machine learning (DDML) with heterogeneous treatment effects in high-dimensional panels -— is based on Chernozhukov et al. (2018a) and Semenova et al. (2021). Section 4.2.1 adapts the DDML method described in Appendix B to our scenario and postulates the demand model we estimate. Section 4.2.2 briefly comments on how DDML deals with potential endogeneity when estimating the demand model.

4.2.1 Double Machine Learning with Heterogeneous Treatments

The increasing availability of big data and sophisticated machine learning techniques has raised interest in modelling consumer demand over the past decade (Bajari et al., 2015). Previously, estimating demand with a large number of observations and covariates was challenging. Thanks to the recent advancements in machine learning methods, the ability to use a high-dimensional set of covariates decreases the probability of standard omitted variable bias, which was an essential drawback of conventional demand estimates. The double/debiased machine learning (DDML) framework of Chernozhukov et al. (2018a) builds on sample splitting [28] and Neyman-orthogonality[29], automates model selection to allow large number of controls and interactions (classical DDML). In DDML, machine learning is not applied to estimate the treatment effect directly, but to predict the plug-in parameters composing the Average Treatment Effect (ATE). Semenova et al. (2021) extend this framework to a panel setting by allowing (i) high-dimensional and sparse treatment effect parameters and (ii) unobserved unit effects (extended DDML). A further reason we reverted to DDML is its ability to perform nonparametric estimation of the demand function without the typical binding assumptions on its functional form found in the literature.

The description of the method in this section follows Semenova et al. (2021), with a summary of the key elements of interest provided in Appendix B for reference. Below is reported a scenario in which the price elasticity of demand is estimated by exploiting a partially linear panel model. The linear part is the heterogeneous treatment effect that consists of the baseline treatment effect (BTE) and treatment modification effect (TME). The nonlinear part is the effect of confounding factors on the outcome.

Consider a demand model of N heterogeneous products indexed by $i \in \{1,2,...,N\}$, each associated with a unique molecule $g \in \{1,2,...,G\}$ and observed across counties over time. We introduce a partition Ξ_c such that $\Xi_c : \{1,2,\ldots,N\} \to \{1,2,\ldots,G\}$, where $\Xi_c(i)$ denotes all products within the same molecule g and ATC-1 as product i in country c. For the sake of clarity we will define $\Xi_c(-i)$ as being all products within the same molecule g and ATC-1 as product i in country c *except* from product i. In this setup, cross-price elasticity is assumed to be zero between products belonging to different molecules, i.e., the cross-price elasticity is zero if $\Xi_c(i) \neq \Xi_c(j)$.

The log sales of product i in molecule g and county c at time t, denoted by $Y_{igc,t}$, are modeled as follows:

$$\begin{aligned} &Y_{igc,t} = D'_{igc,t}\theta_o + g_o(X_{igc,t}) + \nu_i + \mu_g + \eta_c + \nu_{i,g} + \nu_{i,c} + \mu_{g,c} + U_{igc,t}, \\ &E[U_{igc,t} \mid (P_{igc,t}, X_{igc,t}, \phi_{igc,t})_{j\in\Xi_c(i)}] = 0 \end{aligned} \tag{12}$$

where $D_{igc,t} = \left[P_{igc,t}(1,K_{igc,t}), P_{\Xi_c(-i),t}(1,K_{igc,t})\right]$ is the treatment vector capturing the price of product i belonging to molecule g and country c as well as the average price of other products within the same molecule and country of i (but different from i). $P_{igc,t}$ represents the log price of product i in molecule g, country c, and time t and $P_{\Xi_c(-i),t} = \frac{\sum_{j\in\Xi_c(i), j\neq i} P_{jgc,t}}{|\Xi_c(i)|-1}$ denotes the average log price of all other products within the same

[28] Sample splitting, performed on the time dimension in our setting, ensures the absence of overfitting.

[29] Neyman-orthogonality implies the robustness of the treatment effect to approximation errors in the estimate of the treatment propensity score and the conditional mean outcomes for treated and controls (Huber, 2021).

cluster $\Xi_c(i)$ as product i. Moreover, ν_i is a product-level fixed effect, μ_g are molecule and ATC-1 specific fixed effects, η_c are country specific fixed effect. $\nu_{i,g}, \nu_{i,c}, \mu_{g,c}$ are pairwise interactions of the fixed effects which we could include thanks to the reduced risk of multicollinearity given by regularization. Finally, $U_{igc,t}$ is a disturbance term satisfying the conditional sequential exogeneity condition.

The log price $P_{igc,t}$ is modeled as:

$$P_{igc,t} = m_o(X_{igc,t}) + \xi^P_{igc} + V_{igc,t}, \quad E[V_{igc,t} \mid X_{igc,t}, \phi_{igc,t}] = 0 \tag{13}$$

where $m_o(X_{igc,t})$ is a flexible function capturing the influence of control variables $X_{igc,t}$ on price. Here, ξ^P_{igc} denotes an unobserved product-specific effect, and $V_{igc,t}$ is an error term with the conditional mean-zero assumption.

The vector $X_{igc,t}$ includes control variables such as lagged sales $Y_{igc,t-1}$, lagged prices $P_{igc,t-1}$, product characteristics, and exogenous controls. The functions $g_o(X_{igc,t})$ and $m_o(X_{igc,t})$ are unrestricted, allowing for complex, nonlinear relationships known as nuisance parameters. The variable $\phi_{igc,t} = \{(X_{igc,s}, D_{igc,s}, Y_{igc,s})_{s=1}^{t-1}; \xi^E_{igc}, \xi^P_{igc}\}$ denotes the historical information available up to time t, comprising both observed covariates and unobserved unit effects ξ^E_{igc} and ξ^P_{igc}.

In this framework, $D_{igc,t}$ represents the interaction of the product's own price $P_{igc,t}$ and the molecule-level average price $P_{\Xi_c(-i),t}$ with observable characteristics $K_{igc,t}$, a subset of $X_{igc,t}$. The causal interpretation of θ_0, the vector of treatment effect parameters, is facilitated by the sequential exogeneity assumption for $U_{igc,t}$ in Equation (12) which follows from Semenova et al. (2021)'s work.

The main equation of interest is (12). The parameter of interest is θ_o[30] —the vector of heterogeneous treatment effects— that we interpret as the price elasticity of demand. In other words, θ_o is identified as the causal effect of price on demand under the assumption that $U_{igc,t}$ is mean independent of the information set ϕ_{it} (Semenova et al., 2021). In particular, for the identification of θ_o we assume that –after controlling for all pre-determined variables–$U_{igc,t}$ is mean independent of all information $(P_{igc,t}, X_{igc,t}, \phi_{igc,t})_{j \in \Xi_c(i)}$ about members of the i^{th} cluster in country c[31] Then,

$$\theta_o = (\theta_o^{\text{own}}, \theta_o^{\text{cross}}) \tag{14}$$

where θ_o^{own} represents the vector of own-price elasticities, and θ_o^{cross} represents the vector of cross-price elasticities.[32] For clarity with respect to Section 3.3, note that our demand is estimated in log–log form. Hence the parameter vector θ_o identifies the structural elasticity map: baseline coefficients and interactions that map observed characteristics K_{igct} into observation-specific elasticities. In particular, θ_o^{own} maps K_{igct} to the *own-price* elasticity and θ_o^{cross} maps K_{igct} to the cross-price elasticities. The realized (observation-specific) elasticities are therefore

$$\text{ep}_{ii} \;=\; (1, K_{igct})'\theta_o^{\text{own}}, \qquad \text{ep}_{ji} \;=\; (1, K_{igct})'\theta_o^{\text{cross}}(j \to i),$$

which coincide with the derivatives $\partial \log Q_{igct} / \partial \log P_{igct}$ and $\partial \log Q_{igct} / \partial \log P_{jgct}$, respectively. These objects are unit-free elasticities and are the quantities that enter the Ramsey conditions in Section 3.3.

[30] θ_o plays the same role as β_0 in the Appendix B

[31] Notice also that an implicit assumption is that, once we controlled for fixed effects, no residual time-varying unobserved heterogeneity across countries occurs. Namely $E[U_{igc,t}, U_{igc',t'}] = 0$ for $c \neq c'$ and $t \neq t'$. However, since Semenova et al. (2021)'s context allows for heteroskedasticity, we can release the latter assumption.

[32] The target parameter θ_o is Neyman orthogonal with respect to the nuisance parameters, which is why the DDML framework is applicable for estimation and inference on θ_o. In this context, θ_o has dimension $2p+2$, comprising both the own-price and cross-price elasticities, with baseline terms and interactions for each of the p product characteristics. This result assumes that each component of θ_o^{own} and θ_o^{cross} includes a scalar baseline elasticity and p coefficients corresponding to observable product characteristics $K_{igc,t}$.

Demand is affected by changes in own price ($\Delta P_{igc,t}$) as indicated in Equation (15), and by changes in the average cluster price ($\Delta P_{\Xi_c(-i),t}$) as shown in Equation (16).

$$\Delta D'_{igc,t}\theta_o = \Delta P_{igc,t}(1, K_{igc,t})'\theta_o^{\text{own}} = \alpha_{0,1}^{\text{own}} + K'_{igc,t}\gamma_{0,-1}^{\text{own}} \tag{15}$$

$$\Delta D'_{igc,t}\theta_o = \Delta P_{\Xi_c(-i),t}(1, K_{igc,t})'\theta_o^{\text{cross}} = \alpha_{0,1}^{\text{cross}} + K'_{igc,t}\gamma_{0,-1}^{\text{cross}} \tag{16}$$

where $\alpha_{0,1}^{\text{own}}$ and $\alpha_{0,1}^{\text{cross}}$ denote the Baseline Treatment Effect (BTE), while $K'_{igc,t}\gamma_{0,-1}^{\text{own}}$ and $K'_{igc,t}\gamma_{0,-1}^{\text{cross}}$ represent the Treatment Modification Effect (TME). Equations (15) and (16) illustrate that the causal effect of a unit change in the base treatment $P_{igc,t}$ on the outcome $Y_{igc,t}$ is the sum of the BTE and the TME (Semenova et al., 2021). The treatment vector is highly sparse as it includes interactions between product-specific and cluster-average prices and product characteristics. Specifically, $\dim(D_{igc,t}) = \dim(\theta_o) \gg NT$ but only a small subset $s \ll NT$ has non-zero effects, so that $\|\theta_o\|_0 = s$. Furthermore, the sub-Gaussianity condition holds. Namely, residuals $V_{igc,t}$[33] and $U_{igc,t}$ exhibit sub-Gaussian tails conditional on covariates $X_{igc,t}$, so that $P(|V_{igc,t}| \geq \varepsilon \mid X_{igc,t}) \leq 2\exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right)$ and $P(|U_{igc,t}| \geq \varepsilon \mid X_{igc,t}, \phi_{igc,t}) \leq 2\exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right)$ for some finite constant σ^2. This imposes that $V_{igc,t}$ and $U_{igc,t}$ have light tails, reducing the likelihood of extreme values, which is essential for ensuring the consistency of the estimators used in Semenova et al. (2021). Notice that the sub-Gaussian tails assumption is reasonable in contexts with aggregated data like sales or prices, where extreme values are less common due to averaging effects. As we will see later in our context, product sales are aggregated at the molecule level as well as prices.

Given these assumptions and premises, the algorithm of Semenova et al. (2021) described in Appendix B is applicable and ensures identification of the coefficient of interest θ_o.

In our setting, with 48 quarters and 33 molecules, we estimate a partially linear, sequentially exogenous dynamic panel model for the quarterly demand for pharmaceuticals. We conduct the analysis at the molecule-country level because product names vary by country, making it impossible to identify specific products consistently across countries. Additionally, the panel of products is unbalanced, as some products are only available in certain countries. This approach is also necessary for privacy reasons. The base treatment, therefore, becomes:

$$P_{gc,t} = \sum_{l=1}^{48} P_{gc,t-l}\alpha_l^P + \bar{M}'_g\gamma_0 + \rho_q^P + a_{gc}^P + V_{g,t} \tag{17}$$

where $P_{gc,t}$ represents the log price of molecule g in country c at time t, where time is represented by quarters. Notice that –as a consequence of the shift of dimensionality from product level to molecule level– Ξ is a mapping such that $\Xi : \{1, 2, \ldots, G\} \to \mathscr{P}(\{1, 2, \ldots, C\})$, where $\mathscr{P}(\{1, 2, \ldots, C\})$ denotes the power set of countries. In this context, $\Xi(g)$ denotes all molecules within the same country and similarly from the more general form above, $P_{\Xi(-g),t} = \frac{\sum_{k \in \Xi(g), k \neq g} P_{kc,t}}{|\Xi(g)|-1}$. In other words, $P_{\Xi(-g),t}$ consists of the average price of all the other molecules in the same country of molecule g[34]. The term $\sum_{l=1}^{48} P_{gc,t-l}\alpha_l^P$ captures the lagged effect of prices over the past 48 quarters. $\bar{M}'_g\gamma_0$ represents molecule g-specific characteristics. The term ρ_q^P where $q = t \bmod 4$ indicating each quarter in every year. The latter denotes quarterly fixed effects to capture seasonality, while a_{ig}^P is an unobserved effect specific to the molecule, and country. Lastly, $V_{gc,t}$ is a stochastic error term, assumed to satisfy sequential exogeneity.

[33] $V_{igc,t} = \left[P_{igc,t}(1, K_{igc,t}), P_{\Xi_c(-i),t}(1, K_{igc,t})\right] - d_{i0}(X_{igc,t})$ where $d_{i0}(X_{igc,t})$ represents the fitted value of the treatment vector $D_{igc,t}$ based on the covariates $X_{igc,t}$. This residual $V_{igc,t}$ captures the component of $D_{igc,t}$ that is not explained by $X_{igc,t}$.

[34] A similar model where $P_{gc,t}$ represents the log price of molecule g in country c at time t comprising product i and $P_{\Xi(-g),t}$ is the average price of molecule g excluding product i can be easily constructed.

Accordingly, we build a dynamic panel model for the outcome variable (log sales) as follows:

$$
\begin{aligned}
&S_{gc,t} = D'_{gc,t}\theta_o + g_o(X_{gc,t}) + \nu_g + \eta_c + \mu_{g,c} + U_{gc,t} \\
&E[U_{gc,t} \mid (P_{kc,t}, X_{kc,t}, \phi_{kc,t})_{j\in\Xi(g)}] = 0, \\
&D_{gc,t} = \left[P_{gc,t}(1, K_{gc,t}), P_{\Xi_{(-g),t}}(1, K_{gc,t})\right], \\
&\theta_o = (\theta_o^{\text{own}}, \theta_o^{\text{cross}})
\end{aligned}
\tag{18}
$$

where $K_{gc,t}$ is a vector of (heterogeneity relevant) observable covariates for molecule g in country c at time t, including variables such as market characteristics and average product features [35]. The variables $\nu_g + \eta_c + \mu_{g,c}$ are defined similarly to those in equation (12). the parameter of interest is the average heterogeneous effect, which encompasses the (overall) price elasticity of demand:

$$
\varepsilon_{gc,t} = (1, K_{gc,t})'\theta_0 \tag{19}
$$

Given the plausibility of the sub-gaussianity assumption –favored by the higher order aggregation at the molecule-country level–and of the sparsity assumption –given the large number of covariates at the molecule-country level in time– we employed the Double/Debiased Machine Learning (DDML) framework with heterogeneous treatment effects to estimate the demand model specified above. Given the above assumptions, the model is guaranteed to correctly identify $\varepsilon_{gc,t}$ (Semenova et al., 2021).

As discussed in Appendix B, the estimation strategy is implemented in two stages. In the first stage, we define the reduced forms of equations (17) and (18). Equation (20) displays the unit-specific treatment reduced form of equation (17), namely:

$$
\widehat{p}_{g0}(X_{gc,t}) = \sum_{l=1}^{48} P_{gc,t-l}\widehat{\alpha}_l^P + \bar{M}_g'\widehat{\gamma}_0 + \widehat{\rho}_q^P + \widehat{a}_{gc}^P \tag{20}
$$

where $X_{gc,t} = (M_{gc,t}, \bar{M}_g, P_{gc,t-1}, \ldots, P_{gc,t-48})$ is a vector of pre-determined control variables and fixed effects. $\sum_{l=1}^{48} P_{gc,t-l}\widehat{\alpha}_l^P$ represents the influence of 48 quarters of lagged prices on the current price, with $\widehat{\alpha}_l^P$ as the estimated coefficients. The term $\bar{M}_g'\widehat{\gamma}_0$ captures molecule-specific characteristics, $\widehat{\rho}_q^P$ denotes quarterly fixed effects to account for seasonality, and $\widehat{a}_{gc}^P$ represents unobserved unit-specific effects for each molecule-country combination.

Similarly, the reduced form prediction for the outcome variable, $\widehat{l}_{g0}(X_{gc,t})$, is derived from the sales model as follows:

$$
\widehat{l}_{g0}(X_{gc,t}) = \widehat{p}_{g0}(X_{gc,t}) \cdot (1, K_{gc,t})'\widehat{\theta}_o + g_o(X_{gc,t}) + \widehat{\nu}_g + \widehat{\eta}_c + \widehat{\mu}_{g,c} \tag{21}
$$

where $\widehat{p}_{g0}(X_{gc,t})$ represents the predicted treatment values from equation (20).
Following Semenova et al. (2021), equations (20) and (21) are estimated using Lasso[36], obtaining the residuals for the treatment (equation (22)) and the outcome (equation (23)) as follows:

$$
\tilde{P}_{gc,t} = P_{gc,t} - \widehat{p}_{g0}(X_{gc,t}) \tag{22}
$$

[35] To ensure dimensional consistency in compact form, note that $D_{gc,t} = \left[P_{gc,t}(1, K_{gc,t}), P_{\Xi(-g),t}(1, K_{gc,t})\right]$ is structured as a $2(p+1)\times 1$ vector, where p is the dimension of the covariate vector $K_{gc,t} = (1, K_{gc,t,1}, K_{gc,t,2}, \ldots K_{gc,t,p})$. Consequently, $\theta_o = (\theta_o^{\text{own}}, \theta_o^{\text{cross}})$ must also be $2(p+1)\times 1$, comprising two components: θ_o^{own} and θ_o^{cross}, each with dimension $(p+1)\times 1$. This structure allows $D'_{gc,t}\theta_o$ to represent the own- and cross-price effects in a dimensionally consistent manner.

[36] Belloni and Chernozhukov (2013) suggest that using Lasso at the first stage enhances the fit of the model in the second stage.

$$\tilde{S}_{gc,t} = S_{gc,t} - \widehat{l}_{g0}(X_{gc,t}) \tag{23}$$

In the second stage, we regress $\tilde{S}_{gc,t}$ on $\tilde{P}_{gc,t}$ to capture the remaining causal relationship between the two.

Furthermore, to ensure robustness in estimation and inference, particularly given the allowance for heteroskedasticity, we cluster standard errors at the molecule-country level to properly account for cross-sectional dependencies and any remaining within-cluster correlation over time. Specifically, given the longitudinal nature of our dataset, where the same molecules are observed over time in specific countries, clustering standard errors at the molecule-country level is essential to account for within-cluster correlations due to repeated observations (Abadie et al., 2023). This approach addresses the potential for correlated shocks within each molecule-country pair, as demand shocks or other unobserved factors may persist across quarters. By clustering at this level, we align with recommended practices that suggest clustering based on the structure of the data's assignment or sampling mechanism, rather than solely on the error structure. Specifically, with respect to Appendix B we employ a clustered variance-covariance matrix of residuals:

$$\hat{Q}^{\text{cluster-robust}} = \frac{1}{NT} \sum_{c=1}^{C} \left(\sum_{g \in c} \sum_{t=1}^{T} \hat{V}_{gt} \hat{V}_{gt}' \hat{u}_{gt}^2 \right),$$

where clustering adjustments ensure robustness to both heteroskedasticity and within-cluster dependence. This matrix is then inverted and symmetrized to obtain $\hat{\Omega}^{\text{cluster-robust, CLIME}}$, used to construct the debiased Lasso estimator [37], allowing for reliable inference on heterogeneous treatment effects [38]

The covariates ($X_{gc,t}$) in both stages include the number of firms producing a product containing a molecule g in country c at time t, the percentage of generic products among all products containing molecule g in country c at time t, yearly log GDP per capita based on purchasing power parity, Gini index, as well as fixed effects for quarter, country, molecule, and ATC level. Additionally, the model includes previous realizations of log price and log standard units (48 lags for 48 quarters) to account for the effects of price and sales history. The average non-self unit price accounts for the prices of substitutes and complements.

Based on the methodology and demand model above, we estimate the price responsiveness of pharmaceutical demand at the country-molecule level. We perform three different second-stage estimations, using OLS[39], cross-validated Lasso, and Debiased Lasso, as appropriate. Furthermore, we estimate the demand model separately for generic pharmaceuticals to account for product heterogeneity, which may contribute to variations in price elasticities.

4.2.2 Endogeneity and DDML

Endogeneity is an essential problem for demand estimates. Notably, two sources of endogeneity —omitted variable bias and simultaneity— are of major concern. As already mentioned, DDML allows for a large number of covariates, including product fixed effects, lags of prices and sales. In a standard regression model, using a large set of covariates is an actual problem leading to poor parameter estimates and multicollinearity (Bajari et al. (2015)). Accordingly, variable selection becomes an integral part of the estimation strategy. DDML automates variable selection in the sense that large numbers of covariates included in the model are

[37] See $\hat{\Omega}^{\text{CLIME}}$ in Appendix B

[38] As a robustness-check we also adopt the Relevant One-step Selective Inference for the Lasso (ROSI) approach by Taylor and Tibshirani (2018) and Liu et al. (2018) to obtain the p-values. ROSI enables post-selection inference for Lasso and provides an updated method to recover standard errors based on the "relevant conditioning event" (Liu et al., 2018), in comparison to the Huber-White criterion. Results are available upon request.

[39] That is inference on CATE was done by using a FE estimator.

regularised, *i.e.* covariates with zero coefficients are removed from the model. This process decreases the possibility of standard omitted variable bias.

Another major source of endogeneity, simultaneity, occurs if producers and consumers know the unobserved unit characteristics —such as product quality, fewer side-effects, and higher effectiveness— and producers set prices based on them. In this case, unobserved unit characteristics are correlated with prices and quantities (Berry et al., 1995). A simultaneity problem might lead to upward-biased elasticity estimates if unobserved characteristics are not accounted for. Introducing unobserved unit effects in the extended DDML algorithm alleviates the overfitting problem since they account for molecule-country-quarter specific characteristics that are difficult to quantify. They also capture quantifiable molecule-country-quarter characteristics not present in the data set we use. Notice that unobserved unit effects are unit-specific disturbances estimated with unit-specific expectation functions (see Section 4.2.1). Afterwards, their effects have been removed from sales and prices during the orthogonalisation step. Consequently, unobserved characteristics of pharmaceutical products cannot be correlated with prices and sales in the second stage of estimation. In other words, the orthogonalisation step of DDML removes the potential endogeneity.

In what follows, we instrumented quarterly pharmaceutical prices with the average pharmaceutical prices in other countries. More precisely, the instrument for the price of a drug inclusive of molecule G in country c at time t (the baseline treatment), is the average price of drugs inclusive of the same molecule G in other countries at time t. We also re-estimate the demand model postulated above without the instrument[40]. The results are pretty similar to the main findings, indicating the absence of a strong endogeneity bias.

Since the instrumental variable approach is the one carried on throughout the analysis, we formalize it below following Chernozhukov et al. (2018b) and Syrgkanis et al. (2021). Following Semenova et al. (2021), we estimate a partially linear dynamically exogenous panel model for pharmaceuticals demand, incorporating sieve-based IV for robust estimation. Our sieve-based IV estimator extends the two-stage least squares (2SLS) methodology by employing a sieve basis for the treatment, covariates, and outcome variables. Formally, we acknowledge the potential endogeneity of the baseline treatment $P_{gc,t}$.

To simplify the notation we will denote as i molecule g in country c, so that for instance, the logarithm of price for molecule g in country c at time t, $P_{gc,t}$ becomes simply p_{it}. Similarly the logarithm of sales for molecule g in country c at time t, $S_{gc,t}$ becomes s_{it}. Moreover, for the sake of clarity in the notation, we will focus for the rest of the section on the within-molecule equation and without fixed effects, though the method can be generalized to our scenario (see Chernozhukov et al. (2018b)).
Specifically, the reduced compact form of Eq.(18) and Eq.(17) are respectively

$$\begin{aligned} s_t &= \Upsilon_0(L,t)p_t + \Upsilon_U(L)U_t \\ p_t &= \Upsilon_1(L)V_t \end{aligned} \tag{24}$$

where $\Upsilon_0(L,t) := (I - \alpha_S(L))^{(-1)}[b_0 I + b_K D(K_t) + \alpha_P(L)]$, $\Upsilon_U(L) := (I - \alpha_S(L))^{(-1)}$, $\Upsilon_1(L) := (I - \Phi(L))^{(-1)}$. Namely, $\alpha_S(L)$ and $\alpha_P(L)$ are the degree-48 lag polynomials from the sales equation, $\Phi(L)$ is the degree-48 lag polynomial from the price equation $p_t = \Phi(L)p_t + V_t$, $D(K_t)$ is the diagonal matrix with K_t on the diagonal. $\alpha_S(L), \alpha_P(L), Phi(L)$ are degree-48 strict-lag polynomials (no L^0 term); $b_0, b_K^D(K_t)$ capture contemporaneous price effects. The interested reader is referred to Appendix C. The compact form serves to ensure that we can write p_t and s_t consistently from now on.

Let's now introduce the sieve method as follows:

[40]The detailed results are available in Buyukyazici (2022). The identification guarantees are provided in Section 4.2.1 and in Appendix B.

$$s_t \;=\; \sum_{d=1}^{d^{p_t}} \sum_{k=1}^{d^{X_t}} \Pi'^{s_t}_{0,d,k}\; \psi_d(p_t)\,\rho_k(X_t) \;+\; \gamma(X_t, W_t) \;+\; \varepsilon_t \tag{25}$$

$$p_t \;=\; \sum_{d=1}^{d^{\overline{p}}} \sum_{k=1}^{d^{X_t}} \Pi^{p_t}_{1,d,k}\; \phi_d(\overline{p}_{-c,t})\,\rho_k(X_t) \;+\; \delta(X_t, W_t) \;+\; \nu_t \tag{26}$$

The analytical framework incorporates a selection of basis functions for modelling the variables in question. Similarly to Chernozhukov et al. (2018b), we define $\{\psi_d\}$ as basis functions in the *treatment* dimension (price), i.e., they act on p_t and are indexed by degree d^{p_t}. Concurrently, $\{\phi_d\}$ are basis functions applied to the *instrument* $\overline{p}_{-c,t}$, with degree $d^{\overline{p}}$. The covariates X_t are represented through basis functions $\{\rho_k\}$, each bearing a polynomial degree d^{X_t}. X_t includes the 48 lags of prices and sales, fixed effects, and market covariates. This means we rely on sequential exogeneity of the lags and instrument only contemporaneous p_t.

Within this context, $\overline{p}_{-c,t}$ is the instrumental variable, defined as the average price of drugs with the same molecule G in other countries at time t. The pair (X_t, W_t) denotes the covariates (allowing for correlated error structures). Each function ψ_d maps p_t to $\mathbb{R}$; the ρ_k map X_t to $\mathbb{R}$; and the ϕ_d map $\overline{p}_{-c,t}$ to $\mathbb{R}$. The terms ε_t and ν_t are equation errors.

Notice that, since we adopt a *linear first stage* for 2SLS, in Eq. (26) we set $d^{\overline{p}} = 1$. In the second stage, instead, we remain flexible to avoid misspecification, hence $d^{p_t} > 1$ in general, and Π'_0 need not equal Π_0.

In particular, in our empirical implementation we use a linear basis in the first stage. Rather than using p_t directly in estimation, we employ its fitted values from a linear projection on the instrument and controls:

$$\hat{p}_t \;=\; \pi_1\,\overline{p}_{-c,t} \;+\; \delta(X_t, W_t) \;+\; \omega_t, \tag{27}$$

where $\hat{p}_t$ denotes the fitted value of p_t and ω_t is the first-stage residual. The second stage then estimates s_t using the sieve in $\psi_d(\cdot)$, with p_t orthogonalized via $\hat{p}_t$. More formally, following Chernozhukov et al. (2018b), our target is

$$\theta_0(p_{t,0}, p_{t,1}, \mathbf{x}_t) \;=\; \sum_{d=1}^{d^{p_t}} \sum_{k=1}^{d^{X_t}} \Pi'^{s_t}_{0,d,k}\; \rho_k(\mathbf{x}_t)\,\big(\psi_d(p_{t,1}) - \psi_d(p_{t,0})\big). \tag{28}$$

Operationally, we first estimate the conditional expectations $\mathbb{E}[\psi_d(p_t) \mid X_t, \overline{p}_{-c,t}, W_t]$ by (linear) projection of $\psi_d(p_{t,i})$ on the feature set $\{\phi_d(\overline{p}_{-c,it})\rho_k(x_{it})\}$ and (x_{it}, w_{it}) (see Eq. (27)). We then project s_{it} on the estimated functions and (x_{it}, w_{it}) to obtain $\widehat{\Pi}'^{s_t}_{0,d,k}$, from which we construct θ_0.

Identification To ensure identification of the CATE coefficients, we employ a sieve approximation to flexibly model treatment–outcome relationships. The Neyman-orthogonal score used in the second stage yields robustness to first-stage estimation error; endogeneity is addressed by the IV conditions. In particular, let the structural function $g(X, T)$ represent the relationship between sales $y = s_t$, covariates X_t, and treatment $T = p_t$. We write $s_t = g(X_t, W_t, p_t)$ and—for identification—omit W_t henceforth.

Following Newey and Powell (2003), identification of $g(X_t, p_t)$ is based on the Fredholm integral equation

$$\mathbb{E}[s_t \mid \overline{p}_{-c,t}, X_t] \;=\; \int g(X_t, p)\, dF\big(p \mid \overline{p}_{-c,t}, X_t\big),$$

where $F(\cdot \mid \cdot)$ is the conditional distribution of p_t given the instrument and covariates (Wu et al., 2022)[41].

In our sieve specification,

$$g(X_t, p_t) \;=\; \sum_{d=1}^{d^{p_t}} \sum_{k=1}^{d^{X_t}} \Pi'^{s_t}_{0,d,k}\, \rho_k(X_t)\, \psi_d(p_t),$$

so that

$$\mathbb{E}[s_t \mid \overline{p}_{-c,t}, X_t] \;=\; \int \left(\sum_{d=1}^{d^{p_t}} \sum_{k=1}^{d^{X_t}} \Pi'^{s_t}_{0,d,k}\, \rho_k(X_t)\, \psi_d(p) \right) dF\big(p \mid \overline{p}_{-c,t}, X_t\big)\,.$$

As the number of sieve basis functions increases, the approximation error vanishes under standard smoothness and richness conditions on g and the bases.

The kernel in this integral equation is the conditional law $F(p \mid \overline{p}_{-c,t}, X_t)$. For solvability and correct identification of $g(X_t, p_t)$, we impose:

Assumption 1 (Continuity). For almost every $(\overline{p}_{-c,t}, X_t)$, the conditional law of P_t is continuous in p; equivalently, if a density $f(p \mid \overline{p}_{-c,t}, X_t)$ exists, it is continuous on a compact support $\mathscr{P}$.

Assumption 2 (Support/Monotonicity). For almost every $(\overline{p}_{-c,t}, X_t)$, the conditional distribution $F(\cdot \mid \overline{p}_{-c,t}, X_t)$ is strictly increasing on $\mathscr{P}$ (no point masses), and its density—when it exists—is bounded above and away from zero on $\mathscr{P}$.

Assumption 3 (Exogeneity, Relevance, and Completeness). (i) *Exogeneity/Exclusion:* $\mathbb{E}[\varepsilon_t \mid \overline{p}_{-c,t}, X_t] = 0$. (ii) *Relevance:* $\mathrm{Var}(\mathbb{E}[P_t \mid \overline{p}_{-c,t}, X_t]) > 0$. (iii) *Completeness (injectivity):* for any square-integrable h,

$$\mathbb{E}\big[h(P_t) \mid \overline{p}_{-c,t}, X_t\big] = 0 \text{ a.s.} \;\Rightarrow\; h(P_t) = 0 \text{ a.s.}$$

Given these conditions, $\mathbb{E}[s_t \mid \overline{p}_{-c,t}, X_t]$ and $F(p_t \mid \overline{p}_{-c,t}, X_t)$ can be consistently estimated from the data, and the integral equation inverted (numerically) to recover $g(X_t, p_t)$. Consequently, the CATE coefficients $\theta_0(p_{t,0}, p_{t,1}, \mathbf{x}_t)$ are identified via the two-stage sieve-based IV procedure above.

4.3 Results

4.3.1 Overall Results

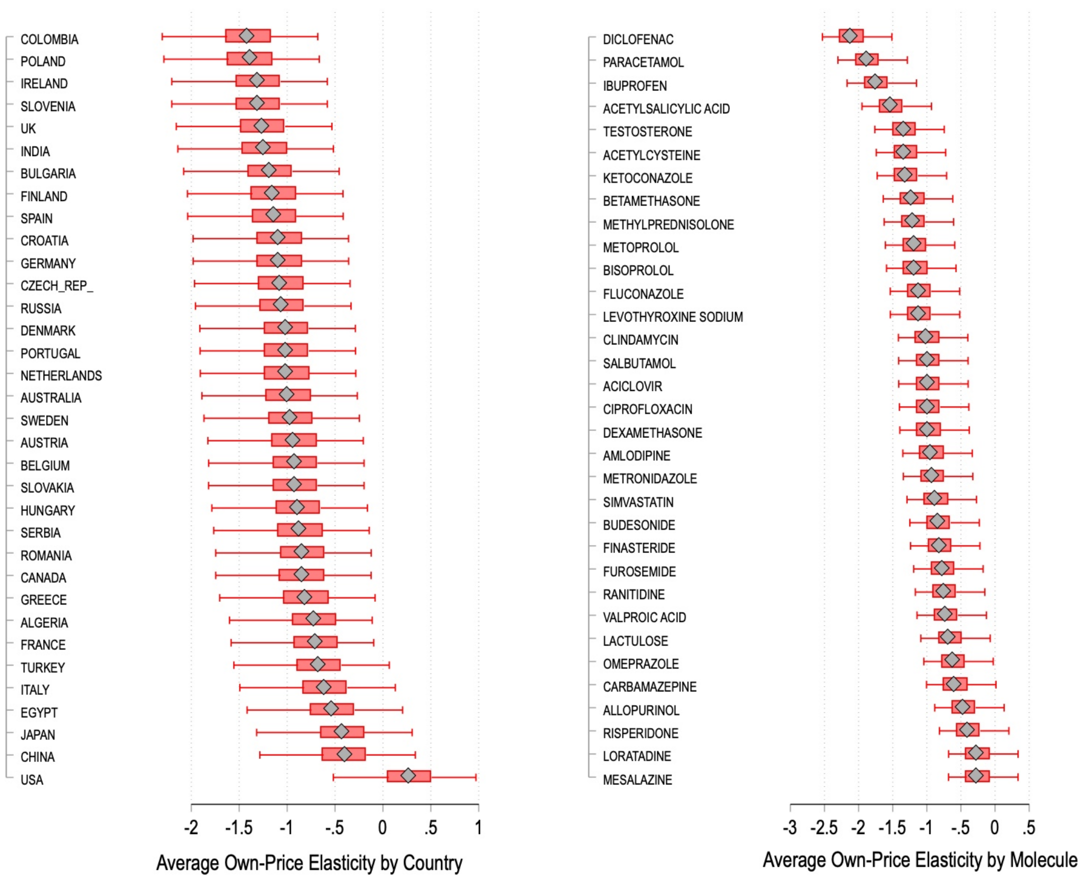
Figure 3 shows the overall results estimated with Debiased Lasso[42]. Given that we only estimate own-price elasticities at this step, the dimension of treatment is relatively low, hence Lasso, Debiased Lasso, and OLS produce very similar results (see Figure F.1). Figure 3(a) displays the own-price elasticities of pharmaceuticals by country, averaged on molecules and the period 2008-2020. Figure 3(b) exhibits the own-price elasticities by molecules, averaged on countries and the period 2008-2020. For the sake of completeness, we have also included the within-country results as displayed in Appendix D.

The results show that all countries have average price elasticities less than zero, except for the USA (0.25), ranging from the highest (-1.42) for Colombia to the lowest (-0.43) for Chile, while molecules

[41] The function $g(X_t, p_t)$ maps prices and covariates to expected sales. The CATE $\theta_0(p_{t,0}, p_{t,1}, x_t) = g(x_t, p_{t,1}) - g(x_t, p_{t,0})$ is the conditional effect of moving price from $p_{t,0}$ to $p_{t,1}$.

[42] Notice that in the case of Lasso a check on sparsity has been performed ensuring that the number of non-zero coefficients (≈ 200) is strictly lower than $\sqrt{N}$ (≈ 230) in order to adopt plug-in λs and avoid first order impacts of the first stage estimation on θ_0.

have a higher variation from the highest (-2.04) to the lowest (-0.21). Interestingly, the most price-elastic molecules are those traditionally adopted as painkillers, *i.e.* analgesics. The average price elasticity of painkillers is -1.8, in line with previous estimates (Björnerstedt and Verboven, 2012; Chevalier et al., 2003). The least elastic molecules are those curing mainly chronic diseases. A possible reason is that drugs like analgesics – and painkillers in general – are likely to be adopted for shorter amounts of time and have more substitutes in the market, hence being more prone to price variations. Instead, medicines curing chronic diseases are employed by patients for a longer amount of time and have a lower degree of substitutability, thus being less sensitive to price variations.



((a)) **((b))**

Figure 3. Price Elasticities by Country and Molecule.

While the literature agrees on the fact that molecules employed for the treatment of chronic diseases should have a low elasticity (Hernández-Izquierdo et al., 2019), there is no solid consensus on the behaviour of the analgesics' price elasticity. On the one hand, Gatwood et al. (2014) find the NAIDs and opioids (among which Diclofenac) to be only moderately responsive to price. On the other hand, Soni (2019) detects a significant and high elasticity for prescribed opioids. Halme et al. (2009) elaborate a natural experiment to assess patients' preferences between over-the-counter generics and branded products adopting Ibuprofen. In this study, they find that half of the respondents are strongly price sensitive, while the other half rely on the opinion of the physician. Chevalier et al. (2003) estimate the price elasticity of painkillers as -1.9. Finally, Björnerstedt and Verboven (2012) report the price elasticity of analgesics as -2.7. Overall, the literature seems to be more inclined to accept the high price responsiveness of painkillers. A possible explanation for the diversity in the results of the different studies might be found in Harris (2022) which indicates the geographical heterogeneity as the root cause for such diversity.

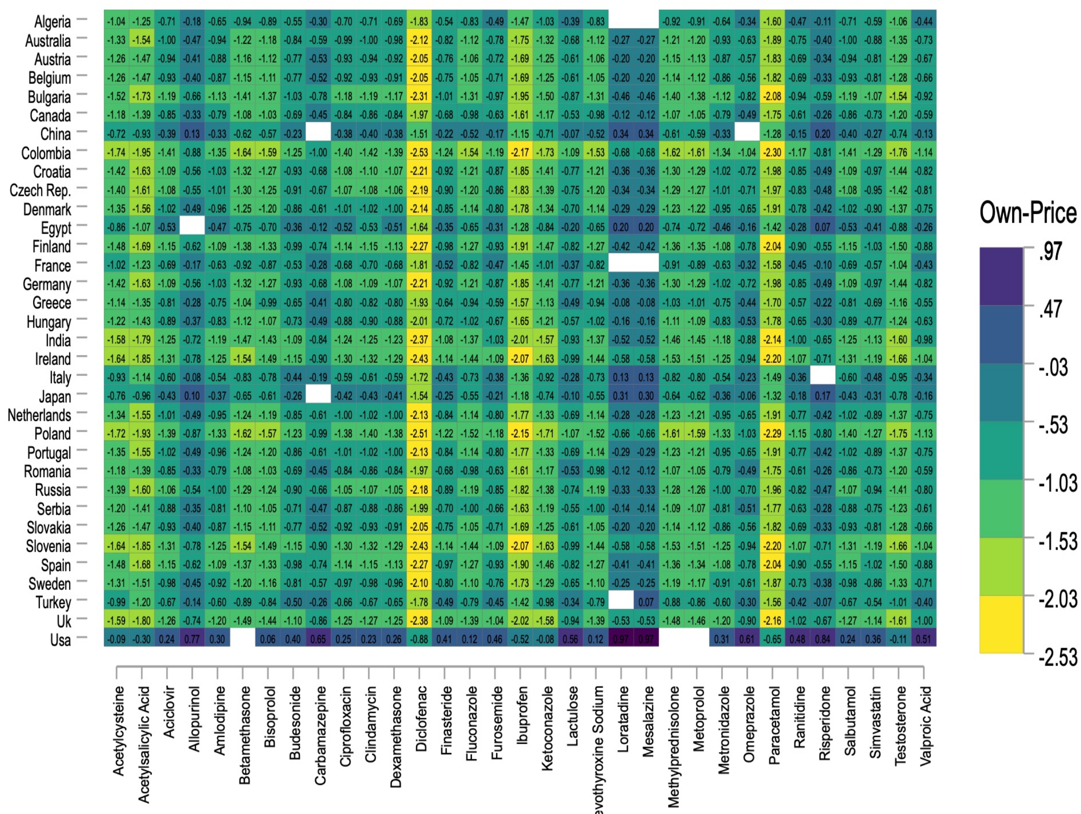


Figure 4. Price Elasticities by Molecule-Country Pairs. Each cell represents the unit-specific own-price elasticity estimate averaged on the period 2008-2020. Each estimate is significant at the 95% confidence interval. Insignificant unit estimates are left blank.

Figure 4 indicates how price elasticities vary by country-molecule pairs. Some molecules have significantly elastic demand —*i.e.* Diclofenac, Paracetamol— across all countries, represented by lighter-coloured columns; while others have relatively inelastic demand —i.e. Loratadine, Mesalazine— represented by darker-coloured columns. Interestingly, several molecule-country pairs have positive elasticities, which could be a signal of the Giffen Paradox in pharmaceuticals (see Yeung et al. (2018)). Nevertheless, almost all molecules show some degree of variation across countries. For instance, the own-price elasticity of Dexamethasone is 0.26 in the USA, while it is -1.38 in Poland. Similarly, each country exhibits varying elasticities across molecules despite some have a more elastic average, such as Colombia, India and Poland, represented by lighter-coloured rows. In contrast, China, Japan, and the USA have a rather inelastic average with substantial within-variation.

Differences between Figures 3 and 4 underline the importance of high-granular level estimates. Figure 3 reports the highest (lowest) elasticity estimate to be -1.42 (-0.42) for countries and -2.04 (-0.21) for molecules. Figure 4 shows that elasticities can be more elastic (-2.53) than average levels, even positive (0.97).

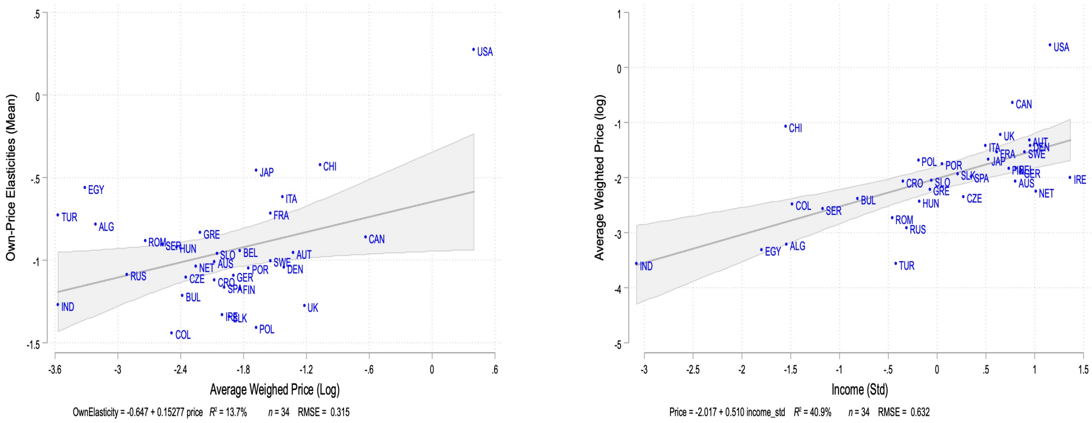
Our results support the criticism of Yeung et al. (2018) who argue that the demand elasticities of pharmaceuticals are biased to be inelastic, *i.e.* close to zero, when prices of substitutes and complements are not taken into account, which was the case in previous elasticity estimates as mentioned in Section 2. The methodology and demand model we use allow us to incorporate non-self prices (equation 4) into the estimates; thus, we can rule out the cross-price effects and approximate true own-price elasticities. Correspondingly, our estimates are more elastic than the previous studies that range from -0.12 to -0.6, even though they cannot be easily compared to our case in terms of data type, model selection, and aggregation level as discussed above.

4.3.2 Ramsey Pricing? Price and Price Elasticity Relation

Ramsey pricing is built on the inverse elasticity rule, which states that prices should vary inversely with the price elasticity of demand. Accordingly, the existence of Ramsey prices can be tested by looking at the relationship between prices and the price elasticities of demand. As an example, consider two markets, m_1 and m_2, with different prices and price elasticities. Let the average price be 25$, and elasticity be -0.8 in m_1. Suppose now that the price is 22$ and the elasticity is -1.3 in m_2. If Ramsey pricing exists in these two markets, the price should be higher in the market with the lower elasticity of demand price, - smaller in magnitude, indicating a lower price response - , *i.e.* m_1. Indeed, m_1 has a lower price elasticity and higher price than m_2, meaning that the inverse elasticity rule holds and Ramsey pricing exists in these markets.

Motivated by this reasoning, we analyse the relationship between pharmaceutical prices and the price elasticities of pharmaceutical demand to test whether the inverse elasticity rule holds in national (see Appendix D) and international pharmaceutical markets. Figure 5(a) exhibits the predictions for price elasticity from a linear regression of price elasticity on average prices in logarithms. The resulting line of the price-own elasticity relationship has a strong positive slope, indicating a negative association between price elasticity and price at the country level. Specifically, countries where pharmaceutical prices are typically lower tend to be more sensitive to price changes compared to countries with higher average prices of drugs. Appendix E deepens the analysis, considering HICs and MLICs separately. The negative relation between price elasticity and price persists despite an increase in the degree of slope for HIC (Figure E.3) and a decrease for MLICs (Figure E.4), meaning that the inverse elasticity rule is stronger in HICs[43]. These results validate that the inverse elasticity rule holds, thus supporting the hypothesis that differential pricing based on Ramsey principles exists in cross-national pharmaceutical markets.

[43]This result might be driven by the fact that our sample comprises more HICs than MLICs. It is likely that the distance between the poorest and the richest country among the HICs is much higher than the same distance in the MLICs sample since the division has been operated based on the OECD classification.



((a)) **((b))**

Figure 5. Relationship Between weighted price and (a) Price elasticity (b) Income as measured by GDP Per Capita (PPP). The own-price elasticity of pharmaceuticals is averaged on 33 molecules and the period 2008-2020, for each country. The x-axis of (a) displays the log price of pharmaceuticals, averaged on molecules and years, weighted by standard units. The x-axis of (b) displays the standardized GDP per capita averaged for the period 2008-2020.

Taken together, Figure 5(a) confirms the inverse–elasticity relation (higher prices where demand is less elastic), while Figure 5(b) shows a positive price–income correlation. These facts are not contradictory: if, on average, $|\varepsilon|$ declines with income and other fundamentals co-move with income, Ramsey's mapping $(P-MC)/P = 1/|\varepsilon|$ implies that prices can rise with income even as they rise with inelasticity. Hence, income is an imperfect proxy for $|\varepsilon|$, and direct tests of $P \propto 1/|\varepsilon|$ are more informative than price–income regressions alone (Danzon, Towse, and Mestre-Ferrandiz, 2015b).
In particular, cross-country differential pricing based on Ramsey principles can only be implemented if the price elasticities of demand exhibit some degree of variation across countries[44]. One major reason for elasticities to vary is cross-country income inequality. The marginal utility of money is plausibly higher in MLIC, implying greater price sensitivity (larger $|\varepsilon|$) than in HIC. Under traditional Ramsey/Lerner (but as we have seen similarly in more complex versions of it), markups satisfy $(P-MC)/P = 1/|\varepsilon|$. Prior empirical work therefore often expects—and sometimes finds—a *positive* price–income slope when richer markets are less elastic (e.g. Maskus, 2001; Sherer and Watal, 2002; Lichtenberg, 2011), with the mechanism that $|\varepsilon|$ tends to decline with income (Danzon, Mulcahy, et al., 2015). In our data, Figure 5(a) exhibits the inverse-elasticity relation: the fitted line is *positive* because the ordinate is ε (typically negative); equivalently, the slope is *negative* in $|\varepsilon|$. However, Figure 5(b) presents the relationship between pharmaceutical prices and GDP per capita (PPP) as being positively related. Contrary to the longstanding assumption of the literature, the inverse elasticity rule holds even though prices do not inversely vary with income. Hence, the comparison of Figures 5(a) and 5(b) is important as the literature has traditionally employed GDP per capita as a proxy for price elasticities due to the lack of consistent estimates of the latter (Danzon, Towse, and Mestre-Ferrandiz, 2015b).

[44] Prices, for instance, will vary with each country's per-capita income (wealthier countries may have higher ICER thresholds and thus higher drug prices), consistent with Ramsey optimal pricing.

A further explanation for differences in the application of Ramsey pricing across countries can be due to unobserved price distortions caused by different insurance regimes in different countries (Barros and Martinez-Giralt, 2008). A potential limitation is that, although we controlled for country fixed effects, we did not include cross-country fixed effects in order to preserve degrees of freedom for efficient estimation. This may have left some unobserved heterogeneity unaccounted for in the model.

We further analyse this pattern by looking at the relationship between own-price elasticities and income, which has not been empirically tested so far. Figure 6 provides limited evidence of the inverse relation between the price elasticities of pharmaceuticals and GDP per capita income as indicated by the slightly positive fitted line, which is further affirmed by E.1 and E.2.

Overall, the results presented in this study suggest that assessing Ramsey pricing by looking at the relationship between pharmaceutical prices and income, as in prior studies, is very likely to be misleading, as we show that the inverse relation of prices-price elasticities and price elasticities-income exists despite the positive relation between prices and income. This evidence is in line with the findings in the Theory part where we show that in MLICs, pharmaceutical companies focus mostly their attention on the rich share of the population, which on average is less price sensitive than the middle and lower classes.

It is worth mentioning here some interesting results emerging from Appendix E where we observe that when GDP decreases in low-income countries, drug prices either remain stable or increase despite Ramsey pricing and a corresponding rise in price sensitivity. This counterintuitive lack of correlation between heightened price sensitivity and drug prices is particularly intriguing for several reasons. First, the mechanism may be partially explained by the presence of insurance, which mitigates individual price sensitivity. However, this effect is primarily evident in HICs, as suggested by Danzon, Towse, and Mestre-Ferrandiz (2015b). Second, the observation supports the hypothesis that, in MLICs, pharmaceutical companies tend to target wealthier, less price-sensitive segments of the population. This raises a compelling question: why are wealthier individuals in low-income countries less price-sensitive? Possible explanations include higher education levels, which make them more health-conscious and likely to insure themselves, significant income disparities that allow the wealthy to remain indifferent to drug price changes, or a wealth effect, where greater financial resources inherently reduce price sensitivity, as discussed in Section 3 in a more formal way.

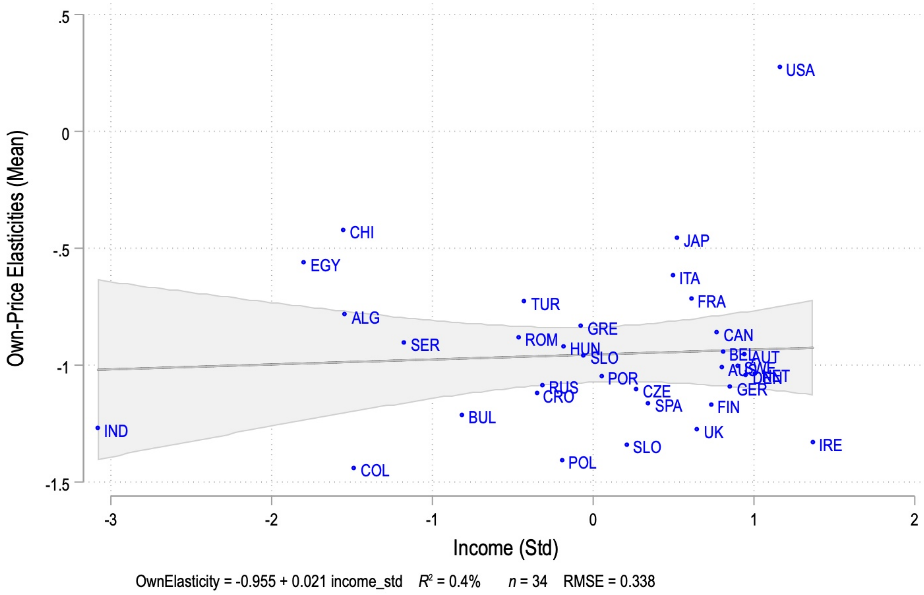


Figure 6. Relationship Between GDP per capita PPP and price elasticity. Y-axis displays the own-price elasticity of pharmaceuticals averaged on 33 molecules and period 2008-2020, for each country. X-axis displays the standardised GDP per capita PPP averaged on the period 2008-2020.

4.3.3 Heterogeneity in Treatment Effect: Generic Pharmaceuticals

Danzon, Mulcahy, et al. (2015) argue that generic markets are more competitive than originator markets because originators' price has to cover the fixed cost resulting from pharmaceutical R&D alongside the marginal cost. Moreover, originators might have patent protection that prevents new entries into the market and allows firms to price discriminate across and within countries. On the contrary, generics' prices generally reflect the marginal production costs that tend to be uniform across countries. These aspects are likely to introduce different patterns of price elasticities for generic products. Therefore, we replicate the previous analysis only for generic pharmaceuticals products.

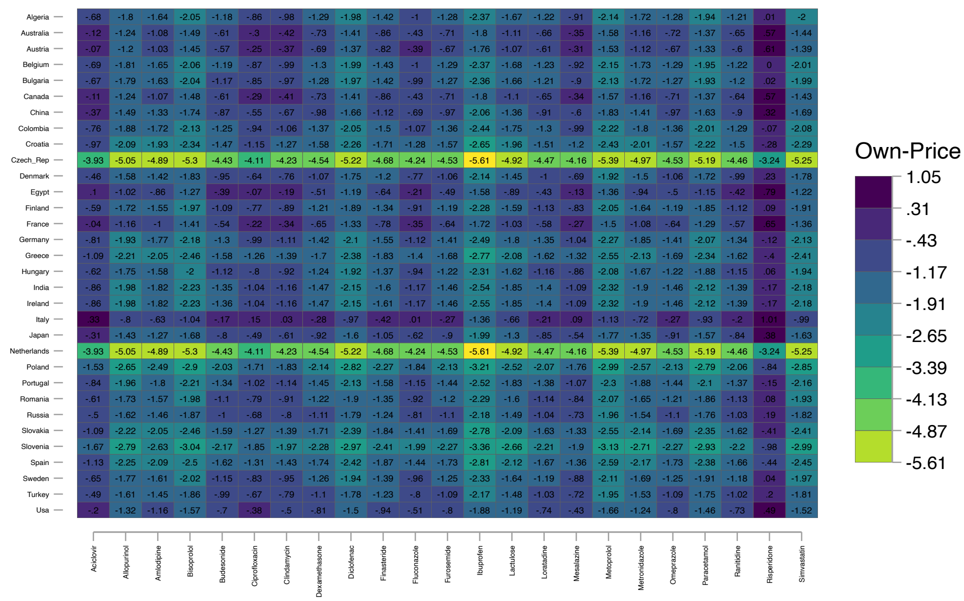


Figure 7. Price Elasticities by Molecule-Country Pairs (generics). Each cell represents the unit-specific own-price elasticity estimate averaged on the period 2008-2020. Each estimate is significant at the 95% confidence interval.

As before, we begin the analysis with an overview of the elasticities for molecule-country pairs. Figure 7 presents the unit-specific own-price elasticity estimate averaged on the period 2008-2020 in a range of -5.61 to 1.01. The majority of the elasticities (97%) are negative as expected.

Figure 8 displays the results with substantial differences across countries and molecules. Figure 8(a) shows that the average price elasticities for generic products exhibit substantial dissimilarities[45] between the most (Czech Republic, Netherlands, and Slovenia) and least price-elastic (Italy, Egypt, and France) countries. The most price-elastic countries are expected to have either a higher share of generics —probably because of favouring the generic entrance into the market through their legislation, *i.e.* Poland (Wouters et al., 2017) and Slovenia— or particularly low generic prices which increase the access of a higher share of the low-income population as in Spain and Netherlands (Wouters et al., 2017). The case of the Czech Republic is self-standing. The Czech Republic has a moderate share of generics, which would not explain its first position in the ranking. One potential explanation is the Czech Republic government's policy to ease the generic entrance into the market by formulating *ad hoc* legislation to favour generics, thereby reducing the drug prices since the generic industry has a big potential for low-cost drugs due to the simultaneous expiration of patents (Lőrinczy, 2013). This policy has intensified after the financial crisis of 2011, supporting our estimate.

[45] Also detected with an insignificant Kendall τ of ranking similarity.

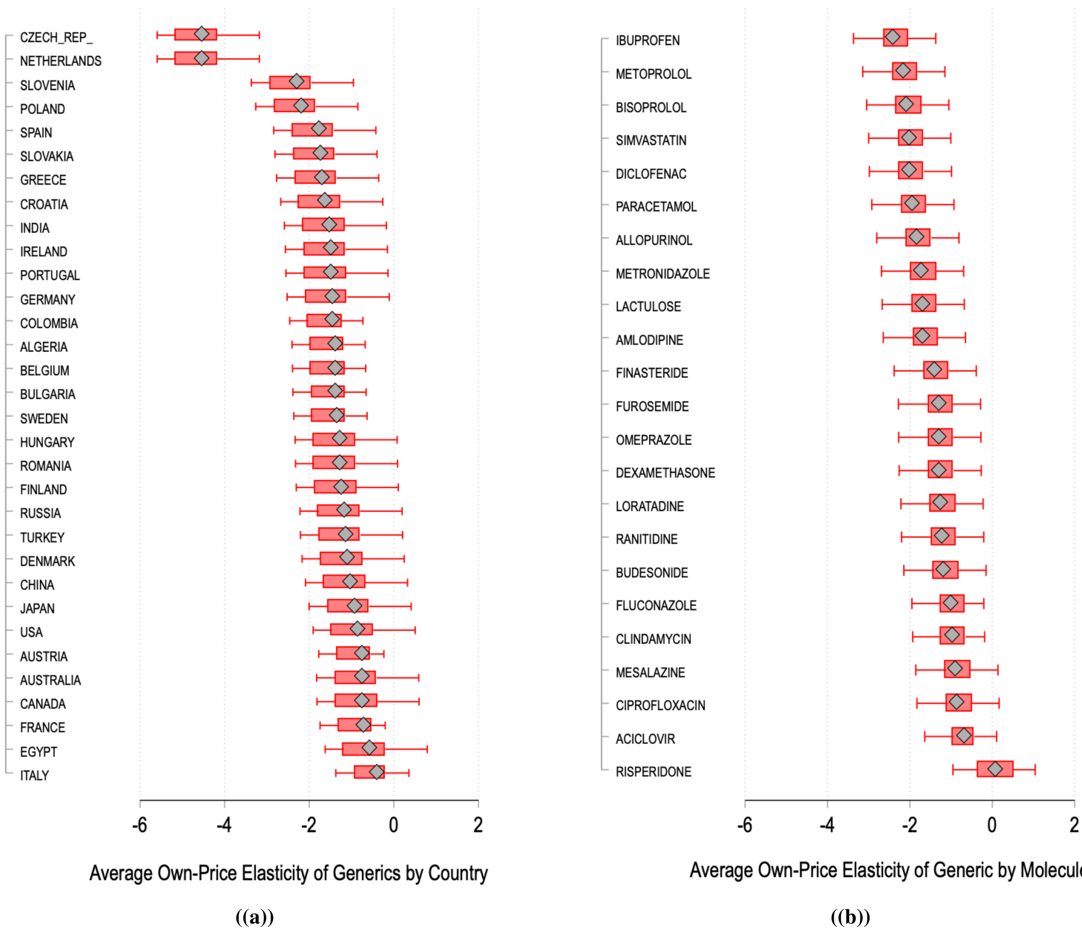
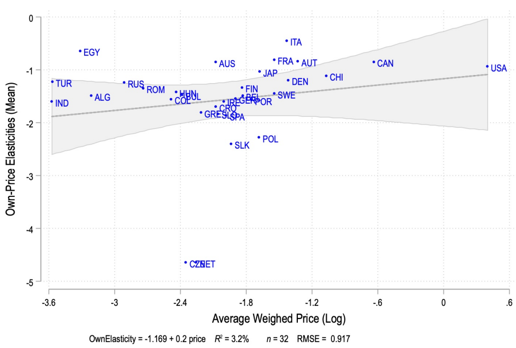


Figure 8. Price Elasticities by Country and Molecule for Generic Pharmaceuticals

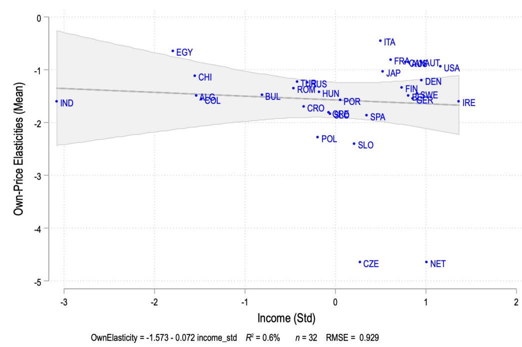
On the other hand, the most price-inelastic countries are expected to have a low share of generics and/or similar prices between generic and originator drugs. For instance, in Italy, the proportion of prescriptions filled with generics amounts to 19% (Wouters et al., 2017) and the price of generics is above the EU average. Similarly, in France, the price of generics amounts to two-thirds of the price of the originator counterpart (Wouters et al., 2017). Regarding Egypt, the market of generics is not developed mainly due to the lack of education of physicians or pharmacists who, often, do not know the generic counterpart of a branded medicine (Elhiny et al., 2021).

Figure 8(b) presents the molecule-wise average price elasticities of generic products. The ranking preserves the general trend of having molecules adopted in chronic diseases as being less responsive to price variations. Nevertheless, there are also significant changes —see Aciclovir, Ciprofloxacin, and Allopurinol— with respect to the elasticity estimates of the full range of products (Figure 3(b)). The most price-elastic molecules in our estimates —Ibuprofen, Metoprolol, and Bisoprolol— treat relatively common diseases and, therefore adopted by a larger share of the population. On the contrary, the most price-inelastic molecules —Risperidone, Aciclovir, and Ciprofloxacin— treat less common chronic diseases and are generally prescribed by a physician. Accordingly, those molecules have relatively smaller markets. As the literature underlines (Torres et al., 2007), generics' entry into small markets is typically more difficult. In addition, there is no clear repositioning of the market from originators to generics after originators are off-patent (Datta and Dave, 2017; Treur et al., 2009; Terrizzi and Meyerhoefer, 2020). A possible reason for such failures might be the highly observable side effects of generic counterparts of some originator drugs (Gallelli et al., 2013). Moreover, patients consuming such medicines may be reluctant to move to the generic counterpart when the originator version is covered by their insurance or because of the fear of losing the benefits given by the originator drug (Treur et al., 2009).

Overall, compared to the average price elasticities for the full range of pharmaceuticals (Figure 3), generic products are substantially more price-elastic, both at the country and molecule level. This result is intuitive because generic pharmaceuticals are off-patent and thus can have many substitutes within the country and globally, leading to higher price sensitivity.



((a))



((b))

Figure 9. Relationship between Price elasticity and (a) weighted price (b) Income as measured by GDP Per Capita (PPP) for generics. The y-axis displays the own-price elasticity of generics, averaged on 33 molecules and period 2008-2020, for each country. The x-axis of (a) displays the log price of generics, averaged on molecules and years, weighted by standard units. The x-axis of (b) displays the standardized GDP per capita averaged for the period 2008-2020.

Figure 9(a) presents the relation between price elasticity and the price of generics. An inverse relation is visible, yet is somewhat weaker compared to the full range of pharmaceutical products (Figure 5(a)). This evidence validates hypothesis H1. The latter might be explained in light of the regulations adopted by the majority of the countries in the sample to contain the variations in the price of generics (see Elhiny et al., 2021, Puig-Junoy, 2010, Simoens, 2012 among others). Furthermore, even though price variations were not artificially imposed, the price of generics tended to stabilize as an result of massive competition in the generics market. These observations are key to investigating the weaker inverse elasticity rule of generics compared to the overall range of drugs, since price variations are crucial to explaining price elasticities. When we look at HICs and MLICs separately, the inverse relation between price elasticities and prices is very strong for HICs (Figure E.3), while it is almost neutral for MLICs (Figure E.4).

The relationship between price elasticities of generics and GDP per capita income appears to be positive, as presented in Figure 9(b). The slightly positive association for the full range of countries turn out to be negative when we consider HIC (Figure E.3) and MLIC (Figure E.4) separately. However, each specification exhibits a low correlation between the two measures.

Overall, the analyses above suggest that cross-country Ramsey pricing of generic pharmaceuticals exists in HICs while it is questionable in MLICs. This pattern might be driven by the fact that the majority of HICs acknowledge that reference pricing strategies might lead to welfare losses, which could be fulfilled through differential pricing manoeuvres (Persson and Jönsson, 2016). Regarding MLICs, differential pricing appears to remain a difficult task due to the fact that, even if firms tried to set differential pricing, the end-patient price in low-income countries *"could be high due to price-gouging at the wholesale or retail levels due to lack of retail or wholesale competition" (Yadav, 2010b).*

4.3.4 A Robustness Check with Fixed Effects

In an effort to enforce our findings, we conducted a fixed effects analysis, the dependent variable being $\overline{log(P)_{i.}}$. The bar indicates that the variable has been averaged on time t[46] and the unit of observation is either the country or the molecule. Furthermore, elasticities were in absolute terms.

The equation takes the following form:

$$\overline{\log(P)}_{i\cdot} = \beta_0 \theta_0^{\text{own}} + \beta_1 X + FE + \varepsilon$$

and

$$\overline{\log(P)}_{i\cdot} = \beta_0 \theta_0^{\text{cross}} + \beta_1 X + FE + \varepsilon.$$

Thus, an inverse proportionality implies that the higher the absolute value of θ_0^{own} or θ_0^{cross} (i.e., the more own-price or cross-price elastic), the lower the price.

The primary reason for conducting two separate regressions is that own elasticities and cross elasticities are already recovered by controlling for cross elasticities and own elasticities, respectively, in the main analysis [47]. The goal is not only to focus on the average own and cross-elasticities as done in previous sections, but to examine the robustness of both own and cross-elasticities across different countries (i.e. by controlling for molecule fixed effects we examine the effect of own and cross elasticities within each molecule and across different countries) and across various molecules (i.e., by controlling for country fixed effects we examine the effect of own and cross elasticities within each country and across molecules). Accounting for these fixed effects is crucial because it eliminates the possible sources of unobserved heterogeneity. It also provides a solid robustness check since –as highlighted in Appendix H –the higher fixed effects by country belonged to those countries having a higher GDP (and higher average prices). This result is enforced when turning just to generics, thus validating our previous findings. Finally, it allows us to summarize our findings and to check if the source of variability across countries and across molecules is given by the cross or by the own elasticity.

[46] The logic is to eliminate the time dimension and to avoid incurring into unobserved time fixed effects.

[47] In the sense that we controlled for the own-elasticity when estimating the cross-elasticity effects, and vice versa. The latter is a consequence of the FWL theorem in our demand model described in the methodological section. By doing the above exercise, instead, we can separately identify and measure the effects of own-price changes and cross-price changes on the outcome of interest (average prices).

Tables 1 and 2 report the aforementioned analyses. Notice that the coefficients are reported in absolute values. Namely, in the regressions, the regressor is $|\hat{\varepsilon}|$. As an example when the estimated slope is negative: $\hat{\beta}_{\theta_0^{\text{own}}} = -0.5443$ (s.e. 0.061, $p < 0.01$) we reported its absolute value 0.5443 for compactness.

Table 1. Within-Between estimation (own elasticity)

	All Products		Generics	
	(1)	(2)	(3)	(4)
	Within	Between	Within	Between
$\beta_{\theta_0^{own}}$	0.4297*	0.0454	0.0601	0.5443***
	(0.2112)	(0.1023)	(0.1162)	(0.0610)
r2 (own)	0.1971	0.5871	0.2895	0.5601
N (own)	1319	1319	1089	1089
absvars	id_ctry	id_mol	id_ctry	id_mol
clustvar	id_ctry	id_mol	id_ctry	id_mol

Note: The dependent variable is $\overline{log(P)_{i.}}$, the bar indicating that the variable has been averaged on time t. The unit of observation i is either country or molecule. *id_ctry* refers to country fixed effects; *id_mol* refers to molecule fixed effects. $\beta_{\theta_0^{own}}$ is the coefficient of the effect of the own-elasticity on $\overline{log(P)_{i.}}$. Coefficients are in absolute values. *** significant at 1%; **significant at 5%; *significant at 10%.

Table 2. Within-Between estimation (cross elasticity)

	All Products		Generics	
	(1)	(2)	(3)	(4)
	Within	Between	Within	Between
$\beta_{\theta_0^{cross}}$	0.0299**	0.0099	0.0728***	0.0151*
	(0.0140)	(0.0090)	(0.0149)	(0.0128)
r2	0.159	0.0586	0.213	0.384
N	34848	34848	31947	30628
absvars	id_ctry	id_mol	id_ctry	id_mol
clustvar	id_ctry	id_mol	id_ctry	id_mol

Note: The dependent variable is $\overline{log(P)_{i.}}$, the bar indicating that the variable has been averaged on time t. The unit of observation i is either country or molecule. *id_ctry* refers to country fixed effects; *id_mol* refers to molecule fixed effects. $\beta_{\theta_0^{cross}}$ is the coefficient of the effect of the cross-elasticity on $\overline{log(P)_{i.}}$ Coefficients are in absolute values. *** significant at 1%; **significant at 5%; *significant at 10%.

The findings presented in Tables 1 and 2 highlight a significant impact of cross-elasticity, a factor previously underestimated in the literature. This effect is particularly evident across different countries and within the generics market, which is known for its competitive nature. Such evidence supports the notion of a Ramsey logic driven more by cross-elasticity than by own-elasticity, thereby strengthening the credibility of our earlier estimates where we accounted for cross-elasticities. A strong and significant coefficient of own-elasticity is reached across molecules and for generics. This latter evidence also confirms the presence of Ramsey pricing across molecules.

Moreover, the analysis of fixed effects coefficients, as presented in Appendix H, reveals that more developed countries exhibit higher, positive, and statistically significant coefficients. This indicates that, when controlling for elasticity, more developed countries tend to maintain higher average prices compared to less developed ones, with Algeria serving as the reference point. Notably, India, Russia, and Turkey show a negative and significant coefficient. These trends can be readily attributed to the greater income levels and technological opportunities prevalent in more developed countries.

To highlight the importance of conducting a causal analysis to identify the presence of Ramsey pricing among different groups of pharmaceuticals, it is useful to compare the findings of Tables 1 and 2 with the descriptive results from the previous sections. Specifically, when contrasting Figure 5(a) and Figure 9(a), we observed weaker evidence of the inverse relationship for generics than for all molecules. The explanation we provided was that generic drug prices tend to stabilize to a constant level due to regulations and competition[48]. However, in the regressions presented in Tables 1 and 2, we found that Ramsey pricing strategies are more commonly used for generics between molecules and within countries. At first glance, these results may seem contradictory without further explanation. The crucial factor is controlling for fixed effects (FE). The intuition is that, once we control for country-level fixed effects, we account for the possibility that regulations may flatten the price of generics. Simultaneously, competition between molecules is controlled through the "between effects.".

[48] When averaged over time and across molecules

In summary, the regression analysis reveals that the coefficient $\beta_{\theta_0^{cross}}$ [49] demonstrates significant statistical relevance for both within and across generic products. This finding indicates that Ramsey pricing strategies are more commonly applied within the generic market as opposed to the wider product market. The adoption of Ramsey pricing in highly competitive markets, along with the high substitutability inherent to generic products, can explain this prevalence. Importantly, the notable impact observed is primarily driven by cross-price elasticity, rather than own-price elasticity. This highlights the critical need to consider own-price elasticity in the estimation of these elasticities. Past research may have mistakenly attributed a positive effect to own-price elasticity, possibly due to challenges in accurately estimating cross-price elasticity. This perceived effect is likely distorted by the previously underestimated role of cross-price elasticity. Additionally, it's important to note the significant between effects of own elasticities in the case of generics, which stands in contrast to the scenario observed for all products.

Table 3 showcases a list of drug pairs that exhibit cross-elasticity in the 95th percentile across the majority of the countries surveyed, suggesting a significant degree of substitution or complementary usage among these medications. The calculation of cross-elasticities, based on their absolute values, was employed to confirm the strength of these relationships. For example, the combination of Mesalazine and Loratadine is highlighted for its remarkable cross-elasticity, reaching the 95th percentile in 25 countries, which points to their potential for concurrent use or as substitutes for one another.

Table 3. Molecule pairs whose cross elasticity appears in the 95^{th} quantile for the majority (more than half) of countries.

Molecule pairs in 95^{th} quantile	
Cross-Molecules	**# of countries 95^{th} quantile**
Allopurinol - Ciprofloxacin	18
Loratadine - Betamethasone	19
Amlodipine - Simvastatin	20
Acetylsalicylic Acid - Dexamethasone	20
Bisoprolol - Paracetamol	21
Carbamazepine - Omeprazole	21
Loratadine - Paracetamol	21
Loratadine - Salbutamol	23
Mesalazine - Loratadine	25

The observed phenomenon of drug pairings exhibiting significant cross-elasticity as per Table 3, can be attributed to several pharmacological considerations that underpin logical therapeutic strategies. For instance, Loratadine, an antihistamine, and Betamethasone, a corticosteroid, are commonly combined to manage allergic reactions, reflecting a standard approach to treating various allergic conditions. Similarly, Amlodipine and Simvastatin are often co-prescribed in cardiovascular risk reduction efforts, leveraging the complementary benefits of a calcium channel blocker and a statin. The co-administration of Acetylsalicylic Acid (Aspirin) and Dexamethasone is another example, where their anti-inflammatory properties are used synergistically in clinical practice. Bisoprolol and Paracetamol may be used together for pain management in patients with cardiovascular concerns, while Mesalazine and Loratadine could be prescribed concurrently for

[49] A negative coefficient (when controlling for molecule fixed effects) suggests that molecules with high substitutability with other molecules (high cross elasticity) across countries sustain lower average prices for molecule i. Hence, higher coefficients in absolute value demonstrate a higher adherence to Ramsey pricing.

individuals suffering from both inflammatory bowel disease and allergic symptoms. In scenarios requiring infection treatment in gout patients, Allopurinol and Ciprofloxacin might be administered together. Carbamazepine, known to cause gastrointestinal discomfort, is often sensibly paired with Omeprazole to mitigate such adverse effects. The combination of Loratadine and Paracetamol is frequently observed in treating cold and flu symptoms, providing both antihistaminic and analgesic/antipyretic benefits. Additionally, the co-administration of Loratadine and Salbutamol is rationalized in asthma management, particularly when allergic reactions are a contributing factor. These pairings, driven by patient-specific treatment protocols or broader prescription trends, underscore the influence of clinical guidelines on prescribing practices.

5 Conclusion

The present study provides the first empirical attempt to analyze the Ramsey pricing of pharmaceuticals by testing the validity of the inverse elasticity rule. It does so by (i) cautioning against using GDP per capita as a stand-in for price elasticity of demand under non-homothetic preferences and heterogeneous insurance, showing formally that income shifts not necessarily move own-price elasticities when drug baskets behave as near-necessities and insurance is uneven, (ii) establishing that, under Ramsey-Boiteux pricing, the deadweight-loss share of total welfare is generically larger in more elastic (MLIC) environments than in more inelastic (HIC) ones, thereby rationalizing why the inverse-elasticity rule should manifest more cleanly across HICs, and (iii) extending the Ramsey rule to interdependent demands and insurance-induced price distortions, yielding a closed-form markup system in which optimal prices depend on both own and cross-price elasticities, a "Standard Ramsey Component," and an Insurance Distortion Amplifier term.

These results motivate estimating, at a granular level, both own and cross elasticities—rather than treating drugs as isolated goods—because, in the presence of substitution and coverage, cross-elasticities can do “most of the work” in the optimal-pricing condition. Previous studies either theoretically analyzed Ramsey pricing (Danzon, 1997; Danzon, Towse, and Mestre-Ferrandiz, 2015a; Barros and Martinez-Giralt, 2008) or indirectly assessed its existence by looking at the relationship between pharmaceutical prices and income (Maskus, 2001; Sherer and Watal, 2002). However, Ramsey pricing can be accurately analyzed only by considering the association between price elasticities and prices, *i.e.* the inverse elasticity rule. To this end, we first estimate the price elasticity of demand. It is well known that previous demand analyses —that are within-country and at the insured patient level — considered pharmaceutical products as isolated entities, assuming no cross-price effects of substitutes and complements. Moreover, those studies suffered from omitted variables bias and endogeneity of prices. To overcome these methodological issues, we employ a recently developed machine learning method (DDML) to estimate price elasticities.

The results show that the own-price elasticity of pharmaceuticals ranges from -2.53 to -0.06, even positive for some molecules. Unlike the long-standing idea that pharmaceutical demand tends to be inelastic, the unit-specific estimates we perform demonstrate that pharmaceutical demand can be rather price-elastic when high-granular data and methods are used. Furthermore, we consider non-self price, *i.e.* price of substitutes and complements, in the demand model, yielding relatively elastic estimates compared to previous within-country coinsurance elasticity estimates, which neglected cross-price effects.

A cross-country assessment of the inverse elasticity rule shows that pharmaceutical prices vary inversely with the price elasticities, both in HIC and MLIC, indicating the existence of Ramsey pricing in cross-country pharmaceutical markets. However, within-country Ramsey pricing is questionable when we look at pharmaceutical markets at the molecule level. For cross-country generic pharmaceuticals, the inverse elasticity rule holds only in HIC, while within-country Ramsey pricing exists in both HIC and MLIC.

As with any other study, this study is not free of limitations. First of all, we accept the pharmaceutical prices given and are not interested in how pharmaceutical prices are set in markets. What we aim to do is to analyse whether ex-post prices inversely vary with price elasticities. However, it is well known that regulations, parallel trade, and reference pricing affect pharmaceutical prices and thus can affect price elasticities. Moreover, Ramsey principles require a substantial qualification that may not be achieved in real markets due to several barriers. First, pharmaceutical markets are heavily regulated, especially in the EU, where individual governments can enjoy monopsony power. In such an environment, demand elasticities may not be accurate (Maskus, 2001). Second, Ramsey pricing assumes well-segmented markets where price strategies can be implemented independently (Schmidt et al., 2001a). For instance, the price for a product in Italy should not be affected by the price of the same product in Germany. It is doubtful that this condition is met in the EU because of parallel trade and reference pricing as mentioned above (Towse et al., 2015). Third, the classical Ramsey problem applies to utilities when a sunk cost needs to be covered by a regulator that earns normal profit (Danzon, Towse, and Mestre-Ferrandiz, 2015a). However, pharmaceutical companies can earn more than normal profit thanks to patented products that provide monopolistic rents to manufacturers (Yadav, 2010a). Accordingly, future studies should address these issues to better frame Ramsey pricing of pharmaceuticals.

-

A Appendix for Theory

To simplify the analysis in what follows, we model entire categories–such as all branded drugs (B) and all generics (G)–as composite goods. Each category is treated as a single representative good, characterized by an average price, quantity, and income-sensitive demand behavior. This aggregation strategy is widely adopted in structural models (Höffler, 2006; Barros and Martinez-Giralt, 2008; Comin et al., 2021), and is supported by theoretical results showing that nested CES frameworks can capture within-category heterogeneity through substitution elasticities, taste parameters, and composite price indices (Redding and Weinstein, 2019). While this approach enhances tractability, it necessarily abstracts from molecular-level heterogeneity in preferences, substitution patterns, and pricing. For instance, some branded drugs may be life-saving essentials while others are elective, leading to potentially divergent income elasticities. Likewise, not all generics exhibit the same substitutability or pricing structure, and insurance coverage often varies by molecule rather than just by category. Nonetheless, the aggregation framework –through income sensitive taste parameters, substitution elasticities, and composite price indices–captures key within-category heterogeneity, thereby mitigating the aforementioned potential distortions (Redding and Weinstein, 2019; Comin et al., 2021). Another interesting feature of our framework is its scalability: it does not impose a restriction on the number of varieties in the economy. In principle, it can be extended to include multiple subcategories within the branded (B) and generic (G) classes –for example, $B_0, B_1, \ldots, B_{N_B}$ and $G_0, G_1, \ldots, G_{N_G}$. However, such an extension would require a more complex CES utility structure capable of capturing the full matrix of substitution elasticities across all subcategories. While analytically feasible, this generalization introduces significant dimensionality and is left for future research.

A.1 Consumers

In this section, we model consumer behavior by focusing on how individuals allocate their income across branded drugs, generic alternatives, and a numeraire good. Consumers are assumed to face relative prices and a budget constraint, and make consumption choices that maximize utility. To reflect realistic expenditure patterns and preference heterogeneity across income groups, we employ a non-homothetic CES framework inspired by Comin et al. (2021) and Hanoch (1975).

To simplify the analysis, we present the case where generic and branded drugs are substitutes (i.e. we will impose $\sigma_{BG} > 1$). This approach can be generalized to situations where no assumption is made about the elasticity of substitution between branded and generic drugs, as long as the elasticity is treated as an exogenously given parameter — a more credible assumption in contexts like ours where it can be empirically estimated. We force, however, the insurer to provide a fixed copayment rate δ since the analysis is focused on relatively standard drugs. In other words, we force a common coinsurance rate applying equally to B and G, so relative prices within the drugs block are unaffected. As a side note, we recognize that if δ differed by type, all formulas hold replacing P_k with $\tilde{P}_k$ and relative prices and shares would then change.

To derive the demand functions and expenditure shares, Comin et al. (2021) consider the consumer's utility maximization problem, where preferences are represented by a non-homothetic CES utility aggregator:

$$1 = \sum_{k \in \{B,G,y\}} (f_k(U_l))^{\frac{1}{\sigma}} C_{k,l}^{\frac{\sigma-1}{\sigma}},$$

where $k \in \{B, G, y\}$ indexes branded drugs (B), generic drugs (G), and a numeraire good (y); $f_k(U_l)$ is an income-dependent taste parameter for group $l \in \{L_R, L_P, H_R, H_P\}$ – that is, $f_k(U_l)$ and subsequent quantities are specified at the *group level*, representing the behavior of a *representative consumer* in each group. In order to make the problem treatable, we treat branded and generic as composite types. Each underlying product has a single posted price common across groups; we suppress the product index and denote the class-level price by P_k. Group heterogeneity enters only via coverage. For the moment, we retain the general form $f_k(U_l)$ rather than imposing a specific functional form to allow for flexibility in capturing heterogeneous, income-sensitive preferences across consumer groups. $\sigma > 1$ represents the constant elasticity of substitution. Assuming a constant elasticity of substitution may seem restrictive, however when thinking about it as the average elasticity of substitution between the goods in the economy, the latter becomes more reasonable. Notice that the CES block is defined separately for each subgroup l. Hence, σ is the elasticity of substitution only between goods $k \in \{B, G, y\}$ within a given l. Quantities with different l belong to different representative consumers and are not traded off in a single optimization. In particular, for instance, "B for L_R" and "B for L_P" are not substitutes.

In the non-homothetic CES (NH-CES) framework of Comin et al. (2021), imposing a single elasticity $\sigma > 1$ at the aggregate level makes all goods in the block gross substitutes (here: B with G, B with y, and G with y). This may be restrictive when drug spending should not substitute closely with other consumption. We therefore assume weak separability between pharmaceuticals and the numeraire and adopt a nested NH-CES.

To keep the exposition parsimonious and aligned with standard NH-CES derivations, we restrict attention to the drugs block only and model the allocation across $k \in \{B, G\}$. Since the final composition of the consumer's consumption bundle at income y is not central to our analysis, we restrict attention to a drugs-only NH–CES block exploiting weak-separability. For each group l, let $E_{D,l}$ denote the group-specific total drugs expenditure (taken as given in this within-drugs problem). Preferences over B and G are represented by a non-homothetic CES aggregator, normalized to one:

$$1 = \sum_{k \in \{B,G\}} f_k(U_l)^{1/\sigma} C_{k,l}^{(\sigma-1)/\sigma}, \qquad \sigma > 1,\ f_k(U_l) > 0. \tag{A.1}$$

The latter normalization is a choice of units ("one unit of drug utility"). It only fixes the scale of the utility index so that prices and the taste terms $f_k(U_l)$ have consistent units. Marginal rates of substitution and optimal choices are unaffected by that normalization.

A convenient Lagrangian for the static primal problem (drugs-only, aggregator normalized) is

$$\mathscr{L} = U_l + \rho \left(1 - \sum_{k \in \{B,G\}} \left[f_k(U_l) \right]^{\frac{1}{\sigma_{BG}}} C_{k,l}^{\frac{\sigma_{BG}-1}{\sigma_{BG}}} \right) + \lambda \left(E_{D,l} - \sum_{k \in \{B,G\}} \tilde{P}_{k,l} C_{k,l} \right),$$

where ρ and λ are multipliers. In order to simplify the notation, we redefine $\sigma \stackrel{\text{def}}{=} \sigma_{BG}$. We follow Comin et al. (2021) that treats $f_k(\cdot)$ as a function of the utility index U_l for expositional convenience. Notice that the utility U_l appears directly in the Lagrangian because the optimization problem is framed in dual form. U_l affects demands only via $f_k(U_l)$.

The first-order condition with respect to $C_{k,l}$ is

$$\frac{\partial \mathscr{L}}{\partial C_{k,l}} = -\rho \left(f_k(U_l) \right)^{\frac{1}{\sigma}} \frac{\sigma - 1}{\sigma} C_{k,l}^{\frac{\sigma-1}{\sigma}-1} - \lambda \tilde{P}_{k,l} = 0,$$

which rearranges to

$$\tilde{P}_{k,l} C_{k,l} = \frac{\rho}{\lambda} \left(\frac{1-\sigma}{\sigma} \right) \left(f_k(U_l) \right)^{\frac{1}{\sigma}} C_{k,l}^{\frac{\sigma-1}{\sigma}}.$$

Summing over all goods $k \in \{B, G\}$, we have that

$$E_{D,l} = \sum_{k \in \{B,G\}} \tilde{P}_{k,l} C_{k,l} = \frac{\rho}{\lambda} \left(\frac{1-\sigma}{\sigma} \right) \sum_{k \in \{B,G\}} (f_k(U_l))^{\frac{1}{\sigma}} C_{k,l}^{\frac{\sigma-1}{\sigma}}.$$

By the utility aggregator's definition (such that the sum is equal to 1, see above), this simplifies to

$$E_{D,l} = \frac{\rho}{\lambda} \left(\frac{1-\sigma}{\sigma} \right).$$

Define the expenditure share $s_{k,l}$ for each good k as $s_{k,l} = \frac{\tilde{P}_{k,l} C_{k,l}}{E_{D,l}}$. Substituting from above, we obtain

$$s_{k,l} = (f_k(U_l))^{\frac{1}{\sigma}} C_{k,l}^{\frac{\sigma-1}{\sigma}}.$$

Solving for $C_{k,l}$, we derive the demand function for each good:

$$C_{k,l} = \left(\frac{E_{D,l}}{\tilde{P}_{k,l}} \right)^{\sigma} f_k(U_l),$$

with expenditure shares given by

$$s_{k,l} = \left(\frac{\tilde{P}_{k,l}}{P_{D,l}} \right)^{1-\sigma} f_k(U_l).$$

Finally, substituting $C_{k,l} = \left(\frac{E_{D,l}}{\tilde{P}_{k,l}} \right)^{\sigma} f_k(U_l)$ back into the $E_{D,l}$ device yields

$$E_{D,l} = \sum_{k \in \{B,G\}} \tilde{P}_{k,l} C_{k,l} = \sum_{k \in \{B,G\}} \tilde{P}_{k,l} \left(\frac{E_{D,l}}{\tilde{P}_{k,l}} \right)^{\sigma} f_k(U_l),$$

which simplifies to

$$E_{D,l}^{1-\sigma} = \sum_{k \in \{B,G\}} \tilde{P}_{k,l}^{1-\sigma} f_k(U_l).$$

This equation characterizes $E_{D,l}$ as a function of prices $\tilde{P}_{k,l}$ and income-dependent preferences $f_k(U_l)$.

We can then rearrange $E_{D,l}$ as:

$$E_{D,l} = \left(\sum_{k \in \{B,G\}} \tilde{P}_{k,l}^{1-\sigma} f_k(U_l) \right)^{\frac{1}{1-\sigma}},$$

The elasticity of total drug expenditures for group l with respect to utility, denoted $\eta_{U_l}^{E_l}$, –where we have called $E_l := E_{D,l}$ to ease up the notation and for indicating more clearly that we are referring to an expenditure –is defined as:

$$\eta_{U_l}^{E_l} \equiv \frac{\partial E_l}{\partial U_l} \frac{U_l}{E_l}.$$

Taking the derivative of E_l with respect to U_l, we obtain:

$$\frac{\partial E_l}{\partial U_l} = \frac{1}{1-\sigma} E_l^{\sigma} \sum_{k \in \{B,G\}} \tilde{P}_{k,l}^{1-\sigma} \frac{\partial f_k(U_l)}{\partial U_l}.$$

Then, multiplying by $\frac{U_l}{E_l}$, we have:

$$\eta_{U_l}^{E_l} = \frac{1}{1-\sigma} \sum_{k \in \{B,G\}} s_{k,l} \eta_{U_l}^{f_k},$$

since $\tilde{P}_{k,l}^{1-\sigma} f_k(U_l) = E_l^{1-\sigma} s_{k,l}$ and E_l moves out of the sum over k being independent from k. $s_{k,l} = \frac{\tilde{P}_{k,l} C_{k,l}}{E_l}$ is the expenditure share for each good k, and $\eta_{U_l}^{f_k} = \frac{\partial f_k(U_l)}{\partial U_l} \frac{U_l}{f_k(U_l)}$ is the elasticity of the taste parameter $f_k(U_l)$ with respect to U_l.

Using the derived demand functions, the relative demand between any two goods $C_{k',l}$ and $C_{k'',l}$ (where $k', k'' \in \{B,G\}$ and $l \in \{L_R, L_P, H_R, H_P\}$) is given by:

$$\frac{C_{k',l}}{C_{k'',l}} = \left(\frac{\tilde{P}_{k'',l}}{\tilde{P}_{k',l}}\right)^{\sigma} \frac{f_{k'}(U_l)}{f_{k''}(U_l)}. \tag{A.2}$$

For the sake of completeness it is easy to show – using Eq. A.2 – that in this set up, combinations like, for instance, "B for L_R" and "B for L_P" are not substitutes. Indeed, with a uniform posted price P_B and fixed copay rates $\delta_{B,l}$, $\frac{C_{B,L_R}}{C_{B,L_P}} = \left(\frac{I_{L_R}}{I_{L_P}}\right)^{\sigma} \left(\frac{1-\delta_{B,L_P}}{1-\delta_{B,L_R}}\right)^{\sigma} \frac{f_B(U_{L_R})}{f_B(U_{L_P})}, \qquad \Rightarrow \qquad \frac{\partial \log\left(C_{B,L_R}/C_{B,L_P}\right)}{\partial \log P_B} = 0.$

The elasticity of substitution between goods $k^{'}$ and $k^{''}$, denoted $\eta_{\tilde{P}_{k',l}/\tilde{P}_{k'',l}}^{C_{k',l}/C_{k'',l}}$, measures the response of the relative demand $\frac{C_{k',l}}{C_{k'',l}}$ to changes in the relative price $\frac{\tilde{P}_{k',l}}{\tilde{P}_{k'',l}}$:

$$\eta_{\tilde{P}_{k',l}/\tilde{P}_{k'',l}}^{C_{k',l}/C_{k'',l}} = \frac{\partial \log(C_{k',l}/C_{k'',l})}{\partial \log(\tilde{P}_{k',l}/\tilde{P}_{k'',l})} = \sigma.$$

The elasticity of relative demand between goods (drugs specifically) $k^{'}$ and $k^{''}$ with respect to utility U_l, denoted $\eta_{U_l}^{C_{k',l}/C_{k'',l}}$, is given by:

$$\eta_{U_l}^{C_{k',l}/C_{k'',l}} = \frac{\partial \log(C_{k',l}/C_{k'',l})}{\partial \log U_l} = \frac{\partial \log f_{k'}(U_l)}{\partial \log U_l} - \frac{\partial \log f_{k''}(U_l)}{\partial \log U_l} = \eta_{U_l}^{f_{k'}} - \eta_{U_l}^{f_{k''}}.$$

The elasticity of demand for drug type $k^{'}$ with respect to total expenditure E_l is:

$$\eta_{E_l}^{C_{k',l}} := \frac{\partial \log C_{k',l}}{\partial \log E_l}.$$

Using $C_{k',l} = \left(E_l/\tilde{P}_{k',l}\right)^{\sigma} f_{k'}(U_l)$,

$$\eta_{E_l}^{C_{k',l}} = \frac{\partial \log C_{k,l}}{\partial \log E_l} = \sigma + \frac{\partial \log f_{k'}(U_l)}{\partial \log E_l} = \sigma + \frac{\eta_{U_l}^{f_{k'}}}{\eta_{U_l}^{E_l}},$$

where $\eta_{U_l}^{f_{k'}} := \frac{\partial \log f_{k'}(U_l)}{\partial \log U_l}$ and $\eta_{U_l}^{E_l} := \frac{\partial \log E_l}{\partial \log U_l}$. From the expenditure identity $E_l^{1-\sigma} = \sum_{j \in \{B,G\}} \tilde{P}_{j,l}^{1-\sigma} f_j(U_l)$,

$$\eta_{U_l}^{E_l} = \frac{1}{1-\sigma} \sum_{j \in \{B,G\}} s_{j,l}\, \eta_{U_l}^{f_j}, \qquad s_{j,l} := \frac{\tilde{P}_{j,l} C_{j,l}}{E_l}.$$

Thus, $\eta_{E_l}^{C_{k'},l}$ equals the substitution effect σ plus a non-homothetic correction term $\eta_{U_l}^{f_{k'}}/\eta_{U_l}^{E_l}$, which captures how tastes vary with utility relative to how total drug expenditure varies with utility.

We now turn to the income isoelastic case, which is particularly relevant in our context, as it allows drug demand to vary systematically with income—an important feature given the well-documented sensitivity of pharmaceutical consumption to income levels in MLICs. In this specification, each good's taste parameter depends on group-specific consumption, meaning that richer groups place relatively greater weight on goods with higher income elasticity, while poorer groups allocate relatively more to necessities. Namely, assuming a CES utility function where each good variety k has a taste parameter ω_k and a non-homothetic elasticity parameter ei_k, we express demand functions and elasticities as follows. Define $f_k(U_l) = f_k(C_l) = \omega_k C_l^{\text{ei}_k}$, where C_l represents aggregate consumption for group l. Then, the demand for good variety k and its expenditure share are given by:

$$C_{k,l} = \omega_k \left(\frac{E_l}{\tilde{P}_{k,l}}\right)^{\sigma} C_l^{\text{ei}_k} = \omega_k \left(\frac{\tilde{P}_{k,l}}{\tilde{P}_l}\right)^{-\sigma} C_l^{\text{ei}_k+\sigma},$$

$$s_{k,l} = \omega_k \left(\frac{\tilde{P}_{k,l}}{E_l}\right)^{1-\sigma} C_l^{\text{ei}_k} = \omega_k \left(\frac{\tilde{P}_{k,l}}{\tilde{P}_l}\right)^{1-\sigma} C_l^{\text{ei}_k-(1-\sigma)}.$$

The total expenditure function and the aggregate price index for group l are then:

$$E_l = \left(\sum_{k\in\{B,G\}} \omega_k \tilde{P}_{k,l}^{1-\sigma} C_l^{\text{ei}_k}\right)^{\frac{1}{1-\sigma}},$$

$$\tilde{P}_l = \frac{1}{C_l}\left(\sum_{k\in\{B,G\}} \omega_k \tilde{P}_{k,l}^{1-\sigma} C_l^{\text{ei}_k}\right)^{\frac{1}{1-\sigma}}.$$

The income elasticity of demand for each drug k is given by:

$$\eta_{C_l}^{C_{k,l}} = \sigma \eta_{C_l}^{E_l} + e_{ik}$$

where $\eta_{C_l}^{E_l} = \frac{\partial E_l}{\partial C_l}\frac{C_l}{E_l}$. It is easy to see that, $log(C_{k,l}) = log(\omega_k) + \sigma log(E_l) - \sigma log(\tilde{P}_{k,l}) + e_{ik} log(C_l)$, therefore, $\frac{log(C_{k,l})}{\partial C_l} = \underbrace{\frac{log(\omega_k)}{\partial C_l}}_{=0} + \sigma \underbrace{\frac{\partial log(E_l)}{\partial C_l}}_{:=\eta_{C_l}^{E_l}} - \sigma \underbrace{\frac{\partial log(\tilde{P}_{k,l})}{\partial C_l}}_{=0} + e_{ik} \underbrace{\frac{\partial log(C_l)}{\partial C_l}}_{=e_{ik}} = \sigma \eta_{C_l}^{E_l} + e_{ik}$.

Finally, the elasticity of total drug expenditure with respect to aggregate consumption is:

$$\eta_{C_l}^{E_l} = \frac{1}{1-\sigma}\sum_{k\in\{B,G\}} s_{k,l}\eta_{C_l}^{f_k} = \frac{1}{1-\sigma}\sum_{k\in\{B,G\}} s_{k,l}\text{ei}_k.$$

The income elasticity of demand for each drug i, $\eta_{E_l}^{C_{k,l}}$, is:

$$\eta_{E_l}^{C_{k,l}} = \sigma + (1-\sigma)\frac{\text{ei}_k}{\bar{\text{ei}}_l}, \tag{A.3}$$

where $\bar{\text{ei}}_l = \sum_{k\in\{B,G\}} s_{k,l}\text{ei}_k$ is the average income elasticity across drugs for group l. This expression implies that if $\text{ei}_k > \bar{\text{ei}}_l$, then good variety k is a luxury (i.e., it has an expenditure elasticity greater than 1); if $\text{ei}_k < \bar{\text{ei}}_l$, then it is a necessity.

A.2 Producer

The producer faces a fixed R&D cost, F, such that it is impossible to set prices equal to marginal cost. F is unavoidable and sunk in the short run. The producer supplies two types of goods, branded drugs (B) and generic drugs (G), to a population divided into two consumer groups: L_R, H_R (richer consumers) comprising a share r of the population, and H_P, L_P (poorer consumers) comprising the remaining $1-r$.

Let P_k be the price of a drug k in a country. For the sake of simplicity we assume that a uniform pricing regime is adopted. P_k is independent of the groups in l in the sense that, *a priori*, no price differentiation is made based on income [A.1]. More formally and at the risk of being repetitive, there are n distinct products indexed by $i \in \{1, \ldots, n\}$. Each product carries a type label $k \in \{B, G\}$ (branded or generic). The posted (administered) price P_i is *uniform across income groups* l; the *effective* price faced by group l is $\tilde{P}_{i,l} = \left(1-\delta_{i,l}\right) P_i$, where $\delta_{i,l} \in [0,1)$ is the group– and product–specific coverage rate. Let $q_{i,l}$ denote demand of product i by group l and $q_i := \sum_l q_{i,l}$ the total quantity.

In line with other models (Adida, 2021), the producer is aware of the consumers' demand and decision process. In particular the following is given:

i The inverse demand curve for good k for group l is derived as

$$\tilde{P}_{k,l} = E_{D,l} \left(\frac{f_k(U_l)}{C_{k,l}} \right)^{1/\sigma}, \qquad \sigma > 1. \tag{A.4}$$

where $E_{D,l} = \sum_{k\in\{B,G\}} \tilde{P}_{k,l} C_{k,l}$ represents total (drugs) expenditure for a consumer in group l, $f_k(U_l)$ is the income-dependent taste parameter, and $\sigma > 1$ is the elasticity of substitution In this simplified framework, all goods of type B and G are modeled as symmetric substitutes – i.e. a change in the price of any generic drug (i.e. the representative generic drug) affects the demand for every branded drug, and vice versa. Extending the model to allow for heterogeneous interdependencies (such as the possibility that certain pairs of drugs may be complements or independent) is left for future research, though it falls outside the scope of the present analysis due to added complexity in functional form and estimation.
Differentiating with respect to $C_{k,l}$ yields:

$$\frac{\partial \tilde{P}_{k,l}}{\partial C_{k,l}} = -\frac{1}{\sigma} E_{D,l} \, f_k(U_l)^{1/\sigma} C_{k,l}^{-1/\sigma - 1} = -\frac{1}{\sigma} \frac{\tilde{P}_{k,l}}{C_{k,l}} \; < \; 0,$$

since $I_l > 0$, $f_k(U_l) > 0$, $C_{k,l} > 0$, and $\sigma > 1$. Hence, the inverse demand curve is downward-sloping.

ii The expenditure function for each group l responds to prices:

$$E_{D,l} = \left(\sum_{k\in\{B,G\}} \tilde{P}_{k,l}^{1-\sigma} f_k(U_l) \right)^{\frac{1}{1-\sigma}},$$

Taking the derivative with respect to any price $\tilde{P}_{j,l}$ yields:

$$\frac{\partial E_{D,l}}{\partial \tilde{P}_{j,l}} = E_{D,l}^{\sigma} \tilde{P}_{j,l}^{-\sigma} f_j(U_l),$$

[A.1] However, the effective price paid by each group differs due to group-specific insurance coverage, with $\tilde{P}_{k,l} = (1-\delta_{k,l})P_k$, where $\delta_{k,l}$ reflects the proportion of the nominal price covered by insurance. The effective price P_k satisfies

$$\max_{P_k} \sum_l \left[(P_k - MC_k)) D(P_{k,l}) \right].$$

showing that $E_{D,l}$ adjusts when prices change. This is consistent with the CES structure, where relative prices influence expenditure allocation.

iii We keep a general formulation in what follows. However, in order to obtain a closed-form and to simplify calculations we may want to set $f_k(U_l) = \omega_k C_l^{ei_k}$, where $C_l > 0$ is aggregate consumption for group l, $\omega_k > 0$ is a constant, and ei_k governs non-homothetic preferences. It can be easily verified that $f_k(U_l)$ is , in this case, well-defined, continuous, and differentiable with respect to C_l. Notice that economically, $f_k(\cdot)$ is a group-specific taste weight that captures non-homotheticity. Higher f_k means group l places relatively more weight on good k at a given price vector. In line with standard NH–CES practice, we assume f_k depends on the group's income/utility (e.g., U_l or C_l) but not directly on posted prices. This is an important observation since – as we will see– FOCs depend on the price-derivative matrix $\partial q_{j,l}/\partial P_k$ and on the break-even constraint. Thus, as long as f_k does not depend on prices, the structure of those derivatives (and hence the markup system) is unchanged.

As we will better see in the next Section, on the cost side, technology is captured by a joint cost with a fixed R&D component F and product-level marginal costs c_k:

$$C(q) = F + \sum_k c_k q_k, \qquad MC_k = c_k, \qquad \sum_k (P_k - c_k) q_k = F.$$

The regulator chooses posted prices $P = \{P_k\}_k$ to maximize welfare subject to break-even:

$$\max_P \ W(P) \ \text{ s.t. } \sum_k (P_k - c_k) q_k(P) = F,$$

where W aggregates consumer surplus (integrated over $\tilde{P}_{k,l}$), producer surplus, and the fiscal cost of coverage $(1+\xi)\delta_{k,l} P_k q_{k,l}(P)$. The first-order conditions will yield a multiproduct Ramsey system.

A.3 Welfare Optimization and Ramsey Pricing

The literature acknowledges that Ramsey pricing applies in this context because the demand system, while dependent on both prices and income, features separable preferences where income effects ($f_k(U_l)$) adjust demand independently of price. As Wilson (1993) highlights, Ramsey principles extend to nonlinear contexts by incorporating effective elasticities, which in this case reflect both substitution and income effects. Thus, price-cost margins can be optimally adjusted to balance welfare and revenue constraints, consistent with the broader applicability of Ramsey pricing (Laffont and Tirole, 1993).

We assume the existence of a regulator that will define prices as to maximize social welfare. The welfare function aggregates consumer surplus, producer surplus, and public distortion costs across income groups and drug types, reflecting the trade-offs inherent in Ramsey pricing. Specifically, consumer surplus is derived from the inverse demand curves of each income group for branded and generic drugs, capturing their heterogeneous preferences and expenditures. Producer surplus ensures the recovery of production and R&D costs, while the public distortion cost penalizes subsidies or reimbursements funded through distortionary mechanisms Barros and Martinez-Giralt (2008). This structure of W allows for the explicit representation of heterogeneity in consumer preferences (via income-dependent taste parameters) and substitution effects between branded and generic drugs (via CES utility), ensuring that welfare maximization respects budget constraints and equity considerations. The welfare function is given by:

$$W = \sum_l \sum_{k \in \{B,G\}} \left(\int_{\tilde{P}_{k,l}}^{\infty} q_{k,l}(u;P)\, du \, - \, (1+\xi)\, \delta_{l,k}\, P_k\, q_{k,l}(P) \, + \, (P_k - c_k)\, q_{k,l}(P) \right) \, - \, F, \tag{A.5}$$

where u is the integration variable and the consumer surplus is written as a price–integral. Namely, following Barros and Martinez-Giralt (2008) we integrate quantity with respect to price from the effective price $\tilde{P}_{k,l}$ actually paid by group l (not the posted price P_k) up to the choke price, which is $+\infty$ under CES/isoelastic demand (since $q \to 0 \Rightarrow p \to \infty$); hence the integrand is the demand $q_{k,l}$. We furthermore have that:

- $l \in \{L_P, L_R, H_R, H_P\}$ represents income groups (low-income and high-income),
- $k \in \{B, G\}$ represents drug types (branded and generic),
- $\tilde{P}_{k,l}(C_{k,l})$ is the inverse demand for good k and group l. We highlight here that the stress is on effective prices according to the intuition of Barros and Martinez-Giralt (2008),
- $q_{k,l}$ is the demand of good k by group l. Total demand for each good k across all income groups is $q_k = \sum_{l \in \{L_P, L_R, H_R, H_P\}} q_{k,l}$, making the welfare function directly comparable to the setup in the Ramsey-Boiteux framework. This formulation retains the heterogeneity of preferences (via $f_k(U_l)$) and substitution effects (via σ), while maintaining the downward-sloping nature of the inverse demand curve,
- $\delta_{l,k}$ is the copayment rate,
- c_k is the marginal cost of producing good k,
- ξ is the distortion cost of public funds [A.2],
- F is the fixed R&D cost.

To derive the optimal pricing rule for each drug type k (branded B or generic G), we start with the first-order condition (FOC) from the welfare maximization problem. The welfare function is given by Eq. (A.5) and the cost-recovery constraint ensures that the producer surplus covers the fixed costs F:

$$\sum_{k \in \{B,G\}} (P_k - c_k) q_k = F,$$

where $q_k = \sum_l q_{k,l}$ is the total demand for drug k. An implicit assumption here is that market demand aggregates linearly. Furthermore, as aforementioned, unless stated otherwise, we assume a common σ across groups (one may also allow $\sigma = \sigma_l$).

The Lagrangian for this constrained maximization problem is:

$$\mathscr{L} = W + \lambda \left(\sum_{k \in \{B,G\}} (P_k - c_k) q_k - F \right),$$

where λ is the Lagrange multiplier associated with the cost-recovery constraint.

Differentiating the Lagrangian $\mathscr{L}$ with respect to P_k, the price of drug k, gives the FOC:

$$\frac{\partial \mathscr{L}}{\partial P_k} = \sum_l \Bigg[-(1-\delta_{l,k})\, q_{k,l} - (1+\xi)\, \delta_{l,k} \left(q_{k,l} + P_k \frac{\partial q_{k,l}}{\partial P_k} \right) - (1+\xi) \sum_{j \neq k} \delta_{l,j}\, P_j \frac{\partial q_{j,l}}{\partial P_k}$$
$$+ q_{k,l} + (P_k - c_k) \frac{\partial q_{k,l}}{\partial P_k} + \sum_{j \neq k} (P_j - c_j) \frac{\partial q_{j,l}}{\partial P_k} \Bigg] + \lambda \left[q_k + (P_k - c_k) \frac{\partial q_k}{\partial P_k} + \sum_{j \neq k} (P_j - c_j) \frac{\partial q_j}{\partial P_k} \right] = 0.$$

[A.2] *The effective price $\tilde{P}_{k,l}$ reflects what consumers actually pay and drives their demand behavior, while the term $(1+\xi)\delta_{l,k} P_k C_{k,l}$ captures the social cost of public subsidies, i.e. the cost of raising the money for copaymets, which causes distortions (e.g., via taxes).*

where the sum over k disappears since linearity let us take the derivatives inside the sums and since we are focusing on a specific good k. After combining the level terms it turns out to be

$$\frac{\partial \mathscr{L}}{\partial P_k} = \sum_l \left[-\xi\, \delta_{l,k}\, q_{k,l} - (1+\xi) \sum_j \delta_{l,j}\, P_j \frac{\partial q_{j,l}}{\partial P_k} + \sum_j (P_j - c_j) \frac{\partial q_{j,l}}{\partial P_k} \right] + \lambda \left[q_k + \sum_j (P_j - c_j) \frac{\partial q_j}{\partial P_k} \right] = 0.$$

where the first term is derived by differentiating the consumer surplus integral with respect to P_k, applying the Leibniz rule to account for the dependence of the lower limit $\tilde{P}_{k,l} = (1-\delta_{l,k})P_k$ on P_k. The sum taken over j includes good k unless otherwise specified. From CES utility function we know that each good's demand depends not only on its own price but also on the prices of other goods due to the underlying substitution effects. In such settings, deriving the FOCs for welfare maximization requires accounting for the full Jacobian matrix of the demand system—that is, all partial derivatives $\frac{\partial q_{k',l}}{\partial P_k}$ for all $k, k' \in \{B, G\}$. This is because changes in the price of one good affect the consumption of other goods, altering both consumer surplus and the feasibility of the cost-recovery constrain. Accounting for this ,we obtain the following FOCs:

$$\begin{aligned} 0 = \frac{\partial \mathscr{L}}{\partial P_k} = \sum_l & \left[-\xi\, \delta_{l,k}\, q_{k,l} - (1+\xi) \sum_j \delta_{l,j} \frac{P_j}{P_k} q_{j,l}\, \varepsilon_{kj,l} + \sum_j \frac{P_j - c_j}{P_k} q_{j,l}\, \varepsilon_{kj,l} \right] \\ & + \lambda \left[q_k + \sum_j \frac{P_j - c_j}{P_k} \sum_l q_{j,l}\, \varepsilon_{kj,l} \right], \end{aligned} \tag{A.6}$$

Then by separating the $j = k$ terms from the $j \neq k$ ones, we obtain the following:

$$\begin{aligned} 0 = \frac{\partial \mathscr{L}}{\partial P_k} = \sum_l & \Bigg[-\xi\, \delta_{l,k}\, q_{k,l} - (1+\xi)\, \delta_{l,k}\, q_{k,l}\, \varepsilon_{kk,l} - (1+\xi) \sum_{j \neq k} \delta_{l,j} \frac{P_j}{P_k} q_{j,l}\, \varepsilon_{kj,l} \\ & \qquad + \frac{P_k - c_k}{P_k} q_{k,l}\, \varepsilon_{kk,l} + \sum_{j \neq k} \frac{P_j - c_j}{P_k} q_{j,l}\, \varepsilon_{kj,l} \Bigg] \\ & + \lambda \left[q_k + \frac{P_k - c_k}{P_k} \sum_l q_{k,l}\, \varepsilon_{kk,l} + \sum_{j \neq k} \frac{P_j - c_j}{P_k} \sum_l q_{j,l}\, \varepsilon_{kj,l} \right], \end{aligned} \tag{A.7}$$

where $\varepsilon_{kj,l} = \frac{\partial q_{j,l}}{\partial P_k} \cdot \frac{P_k}{q_{j,l}}$ is the cross-price elasticity of demand (for $j \neq k$), and $\varepsilon_{kk,l} = \frac{\partial q_{k,l}}{\partial P_k} \cdot \frac{P_k}{q_{k,l}}$ is the own-price elasticity of demand. Define the following aggregates (all sums are over l):

$$A_k := \sum_l q_{k,l} \varepsilon_{kk,l}, \quad B_{kj} := \sum_l q_{j,l} \varepsilon_{kj,l}, \quad D_k := \sum_l \delta_{l,k} q_{k,l}, \quad \Delta_k := \sum_l \delta_{l,k} q_{k,l} \varepsilon_{kk,l}, \quad \Delta_{kj} := \sum_l \delta_{l,j} q_{j,l} \varepsilon_{kj,l}.$$

Let $\mu_i := \dfrac{P_i - c_i}{P_i}$ denote the markup on product i. A_k is the own–elasticity–weighted quantity for k; B_{kj} collects the cross responses of j to P_k; the Δ's are the same objects but insured-weighted via $\delta_{l,\cdot}$.

We can now collect the μ-terms on the left hand side. Using the definitions above, (A.7) can be rewritten as

$$(1+\lambda)\, \mu_k A_k + (1+\lambda) \sum_{j \neq k} \mu_j \frac{P_j}{P_k} B_{kj} = \xi D_k + (1+\xi) \Big[\Delta_k + \sum_{j \neq k} \frac{P_j}{P_k} \Delta_{kj} \Big] - \lambda q_k.$$

where $q_k := \sum_l q_{k,l}$.

Each bracket in (A.7) contributes either a level term ($-\xi\,\delta_{l,k} q_{k,l}$ and $-\lambda q_k$) or an elasticity term multiplied by a markup (producer surplus pieces) or by $(1+\xi)$ times a subsidy (fiscal distortion). The factor $(1+\lambda)$ comes from adding the same elasticity block inside and outside the constraint term.
Solving for the target mark-up μ_k (division by $(1+\lambda)A_k \neq 0$)

$$\mu_k = -\sum_{j\neq k} \mu_j \left(\frac{B_{kj}}{A_k}\frac{P_j}{P_k}\right) + \frac{\xi D_k + (1+\xi)\left[\Delta_k + \sum_{j\neq k}\frac{P_j}{P_k}\Delta_{kj}\right] - \lambda q_k}{(1+\lambda)A_k}$$

.

This is a linear system in the markups. The own object A_k sits in the denominator (Ramsey logic); cross-market couplings enter via B_{kj}.
Define

$$\bar{\varepsilon}_{kk} := \frac{A_k}{q_k}, \quad \bar{\varepsilon}_{jk} := \frac{B_{kj}}{q_j}, \quad \bar{\varepsilon}^{\delta}_{kk} := \frac{\Delta_k}{q_k}, \quad \bar{\varepsilon}^{\delta}_{jk} := \frac{\Delta_{kj}}{q_j}, \quad \delta_k := \frac{D_k}{q_k}.$$

These are natural q-weighted averages across groups l. With standard demand, $\bar{\varepsilon}_{kk} < 0$.

$$\boxed{\mu_k = -\sum_{j\neq k} \mu_j \left(\frac{q_j}{q_k}\frac{\bar{\varepsilon}_{jk}}{\bar{\varepsilon}_{kk}}\frac{P_j}{P_k}\right) + \frac{\xi\,\delta_k + (1+\xi)\bar{\varepsilon}^{\delta}_{kk} + (1+\xi)\sum_{j\neq k}\frac{q_j}{q_k}\frac{\bar{\varepsilon}^{\delta}_{jk}}{\bar{\varepsilon}_{kk}}\frac{P_j}{P_k} - \lambda}{(1+\lambda)\,\bar{\varepsilon}_{kk}}\,.} \tag{A.8}$$

The first sum is the cross–markup spillover; the fraction with $\bar{\varepsilon}_{kk}$ in the denominator is the Ramsey "core". The insured-weighted elasticities produce fiscal/insurance corrections. While the latter looks like a very complex formula, it can be easily related to the literature on Ramsey pricing. Namely, let's define the (directional) Höffler term from j to k and its policy-corrected counterpart (Höffler, 2006):

$$M^{(j\to k)}_{\text{Höffler}} := -\frac{q_j}{q_k}\frac{\bar{\varepsilon}_{jk}}{\bar{\varepsilon}_{kk}}\frac{P_j}{P_k}, \qquad M^{\delta,(j\to k)}_{\text{Höffler-corrected}} := \frac{1+\xi}{1+\lambda}\,\frac{q_j}{q_k}\frac{\bar{\varepsilon}^{\delta}_{jk}}{\bar{\varepsilon}_{kk}}\frac{P_j}{P_k}.$$

Also set

$$\text{IDA} := \frac{\xi\,\delta_k + (1+\xi)\bar{\varepsilon}^{\delta}_{kk}}{(1+\lambda)\,\bar{\varepsilon}_{kk}}, \qquad \text{SRC} := -\frac{\lambda}{(1+\lambda)\,\bar{\varepsilon}_{kk}}.$$

where SRC stands for Standard Ramsey Component and IDA for Insurance Distortion Amplifier. In particular, the IDA term is the part of the optimal markup that internalizes the fiscal externality created by insurance coverage. Another way to write the IDA is as follows

$$\text{IDA} = \frac{1+\xi}{1+\lambda}\,\theta_k - \frac{\xi\,\delta_k}{(1+\lambda)\,\eta_k},$$

where ξ is the cost of public funds, λ is the multiplier on the break-even constraint, $\theta_k \in [0,1]$ is the coverage-weighted share of marginal consumption (formally $\theta_k = \bar{\varepsilon}^{\delta}_{kk}/\bar{\varepsilon}_{kk}$), δ_k is the usual average coverage share, and $\eta_k := -\bar{\varepsilon}_{kk} > 0$ is the absolute own-price elasticity. It easy to see that the larger the fraction of insured marginal units (θ_k) and the higher the fiscal distortion (ξ), the higher the markup—tempered by $1+\lambda$. The second (negative) term is an elasticity correction: with broad average coverage (δ_k) but elastic demand (large η_k), a small price rise already cuts subsidized quantities, so less extra markup is needed. If there is no insurance ($\theta_k = \delta_k = 0$), IDA $= 0$; if the revenue constraint dominates ($\lambda \to \infty$), IDA fades; as

$\eta_k \to \infty$, IDA $\to \frac{1+\xi}{1+\lambda}\theta_k$. This is in line with the observations in Barros and Martinez-Giralt (2008). With the above definitions, (A.8) becomes

$$\boxed{\frac{P_k - c_k}{P_k} = \sum_{j\neq k} \mu_j M^{(j\to k)}_{\text{Höffler}} + \sum_{j\neq k} M^{\delta,(j\to k)}_{\text{Höffler-corrected}} + \text{IDA} + \text{SRC}\,.} \tag{A.9}$$

.

Proposition A.1 (Optimal Ramsey markups with insurance and interdependencies in demand): *Consider two drug types $k \in \{B, G\}$ and income groups l, with (i) separable preferences yielding inverse demand in effective prices $\tilde{P}_{k,l}$, (ii) interdependent demands, (iii) per-unit coverage $\delta_{l,k}$ and fiscal distortion ξ, and (iv) a regulator maximizing W as in Eq. (A.5) subject to the break-even constraint $\sum_k (P_k - c_k) q_k = F$ with multiplier λ, where $q_k = \sum_l q_{k,l}$.*
Let $\mu_i := \frac{P_i - c_i}{P_i}$, then the Ramsey rule applies as follows:

$$\frac{P_k - c_k}{P_k} = \sum_{j\neq k} \mu_j M^{(j\to k)}_{\textit{Höffler}} + \sum_{j\neq k} M^{\delta,(j\to k)}_{\textit{Höffler-corrected}} + \textit{IDA} + \textit{SRC}\,.$$

where

$$M^{(j\to k)}_{\textit{Höffler}} := -\frac{q_j}{q_k}\frac{\bar{\varepsilon}_{jk}}{\bar{\varepsilon}_{kk}}\frac{P_j}{P_k}, \qquad M^{\delta,(j\to k)}_{\textit{Höffler-corrected}} := \frac{1+\xi}{1+\lambda}\,\frac{q_j}{q_k}\frac{\bar{\varepsilon}^{\delta}_{jk}}{\bar{\varepsilon}_{kk}}\frac{P_j}{P_k}.$$

and

$$\textit{IDA} := \frac{\xi\,\delta_k + (1+\xi)\bar{\varepsilon}^{\delta}_{kk}}{(1+\lambda)\,\bar{\varepsilon}_{kk}}, \qquad \textit{SRC} := -\frac{\lambda}{(1+\lambda)\,\bar{\varepsilon}_{kk}}.$$

The whole analysis can be tailored and connected to the consumer and producer models studies above. Specifically, using NH–CES and under isoelastic inverse demand, the consumer–surplus integral collapses to

$$\text{CS}_{k,l} = \frac{\tilde{P}_{k,l}\,q_{k,l}}{\sigma - 1},$$

so welfare becomes a sum of simple price–quantity terms, with all own- and cross-price effects embedded in $s_{k,l}(P)$. This preserves the FOCs and the Ramsey markup system, because the full demand Jacobian $\partial q_{j,l}/\partial P_k$ arises from the dependence of the shares/price index on P.

A.4 Example with Uneven Insurance Coverage in MLICs

In what follows assumptions T.1-T.4 of the main text hold and the focus shifts on a representative MLIC, L. Additionally, for this example, we conjecture that in low-income country L insurance coverage may be uneven across income groups (van Hees et al., 2019; Osei Afriyie et al., 2022). Wealthier consumers (L_R) enroll in private insurance reducing effective prices similar to those in H and enabling a higher allocation in pharmaceutical expenses. On the contrary, poorer consumers (L_P), do either rely on poorly designed social insurance or do not have an insurance plan at all (van Hees et al., 2019). Accordingly, $\delta_{L_P,B} + \delta_{L_P,G} \simeq 0$, and L_P face full prices. Hence, a further assumption to (T.1-T.4) is that

T.5 In L, the L_R share of the population faces prices $\tilde{P}_k < P_k, \quad k \in \{B, G\}$ since $\delta_{L_R,B} + \delta_{L_R,G} >> 0$, whereas people in the L_P share face prices $\tilde{P}_k \simeq P_k, \quad k \in \{B, G\}$ since $\delta_{L_P,B} + \delta_{L_P,G} \simeq 0$

Suppose that assumptions (T.1-T.5) hold, then the income elasticity of demand for each type of good, $\eta_{E_l}^{C_{k,l}}$, is:

$$\eta_{E_l}^{C_{k,l}} = \sigma + (1-\sigma)\frac{\mathrm{ei}_k}{\bar{\mathrm{ei}}_l}, \tag{A.10}$$

where $\bar{\mathrm{ei}}_l = \sum_{k\in\{B,G\}} s_{k,l}\mathrm{ei}_k$ is the average income elasticity across drugs for group l with $s_{k,l}$ being the expenditure-share of good k for group l, i.e. the fraction of that group's total outlay I_l devoted to k.[A.3]. To clarify the distinction between ρ and s, we remind the reader that ρ_B and ρ_G represent *within–drug-basket* shares—that is, the fractions of pharmaceutical consumption that are branded and generic, respectively, such that $\rho_B + \rho_G = 1$. In contrast, $s_{k,l}$ denotes a *budget* share: the proportion of group l's total expenditure allocated to good k, with $s_{B,l} + s_{G,l} + s_{y,l} = 1$. In other words, ρ describes how consumers split their drug basket, while s describes how they split their entire wallet. Importantly, expression (A.10) implies that if $\mathrm{ei}_k > \bar{\mathrm{ei}}_l$, then good variety k is a luxury, while if $\mathrm{ei}_k < \bar{\mathrm{ei}}_l$, then it is a necessity.

For the poor group L_P, limited income I_{L_P} and high effective prices for pharmaceuticals due to OOP expenses ($\tilde{P}_{B,L_P} \approx P_B$ and $\tilde{P}_{G,L_P} \approx P_G$) lead–whenever the numeraire is a substitute to pharmaceutical expenditure – to a larger budget share allocated to the numeraire good y, which typically has a lower income elasticity. Consequently the share of income allocated to the numeraire, s_{y,L_P}, is relatively large, reducing the average income elasticity $\bar{\mathrm{ei}}_{L_P} = \sum_{k\in\{B,G\}} s_{k,L_P}\mathrm{ei}_k$. We hence have $\mathrm{ei}_B, \mathrm{ei}_G > \bar{\mathrm{ei}}_{L_P}$, classifying branded and generic drugs as luxuries for L_P with $\eta_E^{C_k} > 1$ for $k \in \{B,G\}$[A.4]

In contrast, for the rich group L_R, higher income I_{L_R} and insurance coverage reduce effective prices ($\tilde{P}_{B,L_R} < P_B$ and $\tilde{P}_{G,L_R} < P_G$), leading to a higher budget share allocated to pharmaceuticals relative to the numeraire good. This increases the shares of income allocated to branded and generics by L_R, s_{B,L_R} and s_{G,L_R}, in the calculation of $\bar{\mathrm{ei}}_{L_R} = \sum_{k\in\{B,G\}} s_{k,L_R}\mathrm{ei}_k$, making $\bar{\mathrm{ei}}_{L_R}$ relatively high. Thus, we may plausibly conclude that $\mathrm{ei}_B, \mathrm{ei}_G < \bar{\mathrm{ei}}_{L_R}$, branded and generic drugs are classified as necessities for L_R with $\eta_E^{C_k} < 1$ for $k \in \{B,G\}$. Thus,

T.6 Given assumptions (T.1–T.5) and the restriction that the monopolist sets uniform nominal prices for all consumers (with differences in effective prices arising from group-specific insurance coverage), pharmaceuticals are considered luxury goods for the poor group L_P, where income elasticities ($\eta_E^{C_k} > 1$) classify them as luxuries due to higher price sensitivity, and necessity goods for the rich group L_R, where insurance coverage reduces effective prices and results in lower income elasticities ($\eta_E^{C_k} < 1$).

Let's now suppose that a proportional drop in GDP in L, affecting both I_{L_P} and I_{L_R}, occurs. Such a proportional drop in GDP has asymmetric effects on the classification of pharmaceuticals as luxury or necessity goods for the poor (L_P) and rich (L_R) groups. For L_P, pharmaceuticals (B and G) are already classified as luxury goods, as their income elasticity of demand satisfies $\eta_{E_{L_P}}^{C_{k,L_P}} = \sigma + (1-\sigma)\frac{\mathrm{ei}_k}{\bar{\mathrm{ei}}_{L_P}} > 1$ due to $\mathrm{ei}_k > \bar{\mathrm{ei}}_{L_P}$, where $\bar{\mathrm{ei}}_{L_P} = \sum_{k\in\{B,G\}} s_{k,L_P}\mathrm{ei}_k$. A reduction in income I_{L_P} increases the budget share allocated to the numeraire good y, as $s_{k,L_P} = \left(\frac{\tilde{P}_{k,L_P}}{I_{L_P}}\right)^{1-\sigma} f_k(U_{L_P})$, and $\tilde{P}_{y,L_P} < \tilde{P}_{B,L_P}, \tilde{P}_{G,L_P}$ makes y more affordable relative to drugs. As s_{y,L_P} increases and the income elasticity for y (ei_y) is typically low, the average income elasticity $\bar{\mathrm{ei}}_{L_P}$ decreases, further reinforcing $\mathrm{ei}_B, \mathrm{ei}_G > \bar{\mathrm{ei}}_{L_P}$. This deepens the classification of branded and generic

[A.3]The derivation of Eq.A.10 is provided in Appendix A.

[A.4]The expenditure shares s_k are defined as $s_k = \left(\frac{\tilde{P}_k}{I_l}\right)^{1-\sigma} f_k(U)$, where $\tilde{P}_k$ is the effective price of good k for income group l, I_l is the income of the group, σ is the elasticity of substitution, and $f_k(U)$ reflects the utility-based preferences for good k.

drugs as luxuries for L_P, causing drug consumption C_k to drop disproportionately, as $\frac{\partial C_k}{\partial I_{L_P}} = \eta_E^{C_k} \cdot \frac{C_k}{I_{L_P}} > \frac{C_k}{I_{L_P}}$ [A.5]. For L_R, pharmaceuticals are classified as necessities, with $\eta_{E_{L_R}}^{C_{k,L_R}} = \sigma + (1-\sigma)\frac{\mathrm{ei}_k}{\bar{\mathrm{ei}}_{L_R}} < 1$ due to $\mathrm{ei}_k < \bar{\mathrm{ei}}_{L_R}$, where $\bar{\mathrm{ei}}_{L_R} = \sum_{k\in\{B,G\}} s_{k,L_R}\mathrm{ei}_k$. While I_{L_R} also decreases proportionally, insurance coverage ensures that effective prices remain low, stabilizing the allocation of budget shares s_{B,L_R} and s_{G,L_R} as long as $\tilde{P}_{B,L_R} < P_B$ and $\tilde{P}_{G,L_R} < P_G$. This modestly increases s_{y,L_R} but does not significantly affect $\bar{\mathrm{ei}}_{L_R}$, preserving the condition $\mathrm{ei}_B, \mathrm{ei}_G < \bar{\mathrm{ei}}_{L_R}$ and maintaining the classification of branded and generic drugs as necessities for L_R. As a consequence, the proportional GDP drop leads to reduced pharmaceutical consumption for L_P, while consumption for L_R remains stable.

Let us now assume that there is an external regulator setting prices according to the Ramsey rule and analyze how prices are affected by an exogenous shock to GDP. More precisely, the Ramsey pricing rule in Equation (A.9) of Appendix A applies. The latter is an interesting result per se. In this setting, the L_P group is more likely to be excluded from the consumption of pharmaceuticals even when Ramsey rule applies.

To see this notice that to ensure that the uninsured low-income group L_P can still purchase a positive bundle of branded and generic drugs after a proportional GDP shock, the two posted prices (P_B, P_G) must satisfy

$$\left(\frac{P_B}{I_{L_P}}\right)^{1-\sigma} f_B(U_{L_P}) + \left(\frac{P_G}{I_{L_P}}\right)^{1-\sigma} f_G(U_{L_P}) \leq 1, \qquad \sigma > 1.$$

since the drug–budget shares that come out of the CES demand must leave at least some non-negative income for the numeraire good y [A.6]. Because $1-\sigma < 0$, this inequality traces a downward-sloping *affordability frontier* in the (P_B, P_G)-plane. Holding the generic price fixed yields an upper bound for the branded price,

$$P_B \leq P_B^{\max}(P_G) := \left[I_{L_P}^{1-\sigma}\left\{1 - (P_G/I_{L_P})^{1-\sigma} f_G\right\}/f_B\right]^{\frac{1}{1-\sigma}},$$

and, symmetrically, fixing P_B generates the ceiling $P_G^{\max}(P_B)$. A fall in income shifts the entire frontier inward, so at least one posted price must fall (or effective prices must be lowered via insurance) to keep L_P

[A.5] This occurs since $\eta_E^{C_k} := \frac{\partial C_k}{\partial I_{L_P}} \cdot \frac{I_{L_P}}{C_k}$, from which $\frac{\partial C_k}{\partial I_{L_P}} = \eta_E^{C_k} \cdot \frac{C_k}{I_{L_P}}$. Since drug consumption is a luxury for L_P, $\eta_E^{C_k} > 1$ for $k \in \{B, G\}$, $\eta_E^{C_k} \cdot \frac{C_k}{I_{L_P}} > \frac{C_k}{I_{L_P}}$. The term $\frac{C_k}{I_{L_P}}$ represents the average consumption per unit of income for good k. If the income I_{L_P} decreases, the proportional drop in C_k would match $\frac{C_k}{I_{L_P}}$ only if the income elasticity $\eta_{I_{L_P}}^{C_k} = 1$, implying proportional responsiveness of consumption to income. However, if $\eta_{I_{L_P}}^{C_k} > 1$, the response of C_k to a change in income is more than proportional. From $\frac{\partial C_k}{\partial I_{L_P}} = \eta_{I_{L_P}}^{C_k} \cdot \frac{C_k}{I_{L_P}}$, we see that $\frac{\partial C_k}{\partial I_{L_P}} > \frac{C_k}{I_{L_P}}$ whenever $\eta_{I_{L_P}}^{C_k} > 1$, meaning that a decrease in I_{L_P} results in a larger percentage drop in C_k than would occur with a proportional relationship. For example, if I_{L_P} decreases by 10% and $\eta_{I_{L_P}}^{C_k} = 1.5$, the consumption C_k decreases by 15%, which is larger than the proportional decline. Thus, C_k decreases disproportionately with respect to I_{L_P} when $\eta_{I_{L_P}}^{C_k} > 1$.

[A.6] In Appendix A we derive CES demands and expenditure shares for each drug $k \in \{B, G\}$:

$$C_{k,\ell} = \left(\frac{I_\ell}{\tilde{P}_{k,\ell}}\right)^{\sigma} f_k(U_\ell), \qquad s_{k,\ell} \equiv \frac{\tilde{P}_{k,\ell} C_{k,\ell}}{I_\ell} = \left(\frac{\tilde{P}_{k,\ell}}{I_\ell}\right)^{1-\sigma} f_k(U_\ell).$$

These follow from the Lagrangian with the non-homothetic CES aggregator and the budget constraint. For the uninsured poor group LP, $\tilde{P}_{k,LP} = P_k$. Requiring that the consumer can still purchase some numeraire y (i.e., $s_{y,LP} \geq 0$) implies that the drug shares cannot exhaust the budget: $s_{B,LP} + s_{G,LP} \leq 1$. Substituting the share expressions yields

$$\left(\frac{P_B}{I_{LP}}\right)^{1-\sigma} f_B(U_{LP}) + \left(\frac{P_G}{I_{LP}}\right)^{1-\sigma} f_G(U_{LP}) \leq 1, \qquad \sigma > 1,$$

which is the "affordability frontier" already mentioned previously.

inside its budget set while preserving the empirical ordering $P_B > P_G$.

The observations above let us define a feasibility and an inclusion regions. Namely, let the feasibility region be defined as

$$\mathscr{F} := \Big\{(P_B, P_G) : P_B \geq c_B, P_G \geq c_G\Big\}$$

Since our aim is to test if L_P is excluded from the market after a GDP negative shock in prices, we need to define what we mean by market exclusion. With the latter, in particular, we refer to the fact that L_P is still able to buy a non-negligible share of drugs, i.e. drugs receive at least a small budget share $\underline{s} \in (0,1)$.

The inclusion region for L_P is thus the set under the frontier

$$\mathscr{A}_{L_P}(\underline{s}) := \left\{(P_B, P_G) : \left(\frac{P_B}{I_{L_P}}\right)^{1-\sigma} f_B(U_{L_P}) + \left(\frac{P_G}{I_{L_P}}\right)^{1-\sigma} f_G(U_{L_P}) \ \geq \ \underline{s}\right\}.$$

where we notice that prices are effective since no insurance is applied to L_P by assumption and total drug share can't exceed 1 (budget constraint). Since $1-\sigma < 0$, the left-hand side is decreasing in each price, so $\mathscr{A}_{L_P}(\underline{s})$ places upper bounds on posted prices (for L_P posted and effective prices coincide). If $\mathscr{A}_{L_P}(\underline{s}) \cap \mathscr{F} \neq \varnothing$, L_P can be included with $P_k \geq c_k$; if $\mathscr{A}_{L_P}(\underline{s}) \cap \mathscr{F} = \varnothing$, posted-price inclusion is infeasible and inclusion requires lowering L_P's effective prices via insurance or subsidies while maintaining $P_k \geq c_k$.

Because $1-\sigma < 0$, $\mathscr{A}_{L_P}(\cdot)$ is decreasing in each price, among all feasible $(P_B, P_G) \in \mathscr{F}$, the largest value of $\mathscr{A}_{L_P}$ occurs at the lowest feasible prices, i.e. at the boundary point (c_B, c_G). Hence, there exists at least one feasible price pair that includes L_P (i.e. $\mathscr{A}_{L_P}(\underline{s}) \cap \mathscr{F} \neq \varnothing$) iff the boundary point itself satisfies the inclusion inequality. Thus, if the inclusion inequality at the minimal feasible prices ($P_B = c_B$ and $P_G = c_G$), it fails everywhere in $\mathscr{F}$. If instead it holds there, the intersection is nonempty. A convenient test, therefore, evaluates inclusion at cost:

$$\left(\frac{c_B}{I_{L_P}}\right)^{1-\sigma} f_B(U_{L_P}) + \left(\frac{c_G}{I_{L_P}}\right)^{1-\sigma} f_G(U_{L_P}) \ \underset{\text{exclude}}{\overset{\text{include}}{\gtrless}} \ \underline{s},$$

so that inclusion at (or above) cost is feasible iff the inequality holds with "$\geq$".

Consider now the proportional GDP decline for the poor group, $I'_{L_P} = \theta I_{L_P}$ with $0 < \theta < 1$. Feasible posted-price inclusion (at or above cost) is characterized by the inclusion-at-cost test evaluated at (c_B, c_G):

$$\left(\frac{c_B}{I'_{L_P}}\right)^{1-\sigma} f_B(U'_{L_P}) \ + \ \left(\frac{c_G}{I'_{L_P}}\right)^{1-\sigma} f_G(U'_{L_P}) \ \underset{\text{exclude}}{\overset{\text{include}}{\gtrless}} \ \underline{s}, \qquad \sigma > 1.$$

Two forces push the left-hand side downward after the shock: (i) lower income I'_{L_P}, which tightens the affordability term because $1-\sigma < 0$, and (ii) non-homotheticity, which typically lowers the taste shifters $f_k(U'_{L_P})$ as utility falls. Assumption T.6 explains this direction –pharmaceuticals behave as luxuries for L_P. Exclusion of L_P arises only if the inequality fails, i.e. when the left-hand side drops below the concrete participation threshold $\underline{s}$. In short, T.6 provides the direction (the expression tends to fall after a negative income shock), while the inequality provides the decision: inclusion remains feasible at posted prices $P_k \geq c_k$ if and only if the test holds with "$\geq$"; otherwise, posted-price inclusion is infeasible and would require lowering L_P's effective prices via insurance or subsidies. The Ramsey mechanism does not influence the inequality. When we focus on the uninsured group L_P ($\delta_{l,k} = 0$), indeed, two things happen simultaneously: (i) the insurance–distortion amplifier IDA becomes negligible, and (ii) the Hoffler–correction term $M^{\delta}_{\text{Hoffler-corrected}}$ becomes small. Intuitively, the correction is meant to offset distortions created by subsidised demand interactions across drugs. If a group receives no subsidy, it doesn't generate that kind of distortion, so the planner has almost nothing to correct—hence the term collapses toward zero. Thus the Ramsey mechanism won't generate a special corrective discount for L_P.

In short, in an economy where insurance is concentrated among the rich, a proportional GDP decline lowers the poor group's affordability. If, at the new income, even pricing at marginal cost fails the inclusion condition, then the poor group cannot be kept in via posted prices alone; Ramsey pricing does not supply a corrective discount for the uninsured and may select higher prices, reinforcing the tendency toward exclusion. Thus the shock favors exclusion but does not make it automatic unless the inclusion-at-cost inequality fails.

By contrast, L_R (the richer, insured group) is more profitable to target because their demand is less elastic [A.7] thanks to the mitigating effect of insurance, which reduces their effective price $\tilde{P}_{k,l}$ thus relaxing their inclusion constraint. Formally, for a small participation threshold $\underline{s} \in (0,1)$ the L_R inclusion set is

$$\mathscr{A}_{L_R}(\underline{s}) := \Big\{(P_B, P_G) : \big(\tfrac{(1-\delta_{B,L_R})P_B}{I_{L_R}}\big)^{1-\sigma} f_B(U_{L_R}) + \big(\tfrac{(1-\delta_{G,L_R})P_G}{I_{L_R}}\big)^{1-\sigma} f_G(U_{L_R}) \geq \underline{s}\Big\},$$

Given $I_{L_R} > I_{L_P}$ and $\delta_{k,L_R} > 0$ (with $\delta_{k,L_P} \simeq 0$), we typically have $\mathscr{A}_{L_R}(\underline{s})$ much larger than $\mathscr{A}_{L_P}(\underline{s})$. Hence, there often exist uniform posted prices (P_B, P_G) that both satisfy the no-loss constraints $(P_k \geq c_k)$ and lie in $\mathscr{A}_{L_R}(\underline{s})$ but not in $\mathscr{A}_{L_P}(\underline{s})$. Participation of L_R at such prices is therefore secured by affordability (lower effective prices and higher income), not by regulatory favoritism; L_P may self-exclude because its inclusion set does not intersect the feasible set at those prices.

This phenomenon is widely discussed in the pharmaceutical-economics literature though–to the best of our knowledge–rarely formalized within a unified Ramsey/NH–CES framework. Several studies document that the poorest populations often lack affordability unless prices approach competitive production cost, especially where insurance coverage is limited and inequality is high (Moon et al., 2011; Yadav, 2010b), so large price cuts may leave their effective access essentially unchanged. This observation is consistent with settings in which budget constraints dominate purchase decisions.

It is natural to ask whether this condition may be mitigated by Ramsey-pricing policies. Some theoretical settings, however, show that negative Ramsey Lerner indexes (prices below marginal cost) may arise not in an effort of the planner to accomodate the share of uninsured patients, rather due to the shape of specific demand structures with particular cross-price interactions with the outside good (Bertoletti, 2018). In our context, therefore, the uninsured poor group L_P faces an *affordability* constraint (captured by the inclusion-at-cost condition) rather than a Ramsey mechanism that pushes $P < c$. By contrast, for the richer, insured group L_R, higher income and lower *effective* prices via insurance relax the same inclusion constraint; uniform posted prices that satisfy $P_k \geq c_k$ are thus more likely to be affordable for L_R than for L_P, without implying any general guarantee on markups for L_R.

Accordingly, we propose the following hypothesis:

Hypothesis 3 (H3): *As a consequence of a negative GDP shock, if prices are set according to the Ramsey rule, even though general price elasticities may increase, we do not expect significant changes in prices.*

B Semenova et al. (2021)'s Setting and Major Results.

The aim of this section is to provide some basic understanding of Semenova et al. (2021) for the sake of the preset work.

[A.7] The demand elasticity for L_P and L_R is given by $\eta^{C_{k,l}}_{\tilde{P}_{k,l}} = -\sigma + (1-\sigma)\eta^{f}_{U_l}$ where the income-dependent preference elasticity $\eta^{f}_{U_l}$ for L_P is small, making their demand more price-sensitive. It is high for L_R due to the mitigating effect of insurance.

B.1 Identification with All Exogenous Covariates

We start with a general form equation:

$$Y_{it} := \beta_0(X_{it}, P_{it}) + e_0(X_{it}) + \xi_i^E + U_{it} \quad E[U_{it}|P_{it}, X_{it}, \phi_{it}] = 0 \tag{B.1}$$

where ϕ_{it} is a filtration that consists of predetermined variables before t, i.e. $\phi_{it} = \{(X_{it'}, P_{it'}, Y_{it'})_{t'=1}^{t-1}\}$. X_{it} may include lags of P_{it} and Y_{it}. as well as known functions of $M_i := \{M_{it}\}_{t=1}^{T}$ representing a collection of fixed variables. For the moment, the variables are assumed to be exogenous. In this specification $\beta_0(.)$ represents the CATE for changing from policy p_1 to policy p_0 since $\beta_0(X_{it}, p_1) - \beta_0(X_{it}, p_0) = E[Y_{it}(p_1)|X_{it}] - E[Y_{it}(p_0)|X_{it}]$. In Semenova et al., 2021, $\beta_0(X_{it}, P_{it})$ is assumed to be well approximated by a linear combination of X_{it} and P_{it} such that

$$\beta_0(X_{it}, P_{it}) = D_{it}'\beta_0 \quad D_{it} := D(X_{it}, P_{it})$$

An example may be $D_{it} := d_{i0}(X_{it}) + V_{it}$, $E[V_{it}|X_{it}, P_{it}] = 0$. Hence substituting in Eq.(B.1) we obtain that

$$Y_{it} = D_{it}'\beta_0 + e_0(X_{it}) + \xi_i^E + U_{it}$$

Since $E[D_{it}|X_{it}, \Phi_{it}] := d_{0i}(X_{it}) = d_0(X_{it}, \xi_i)$, $\xi = (\xi_1, \xi_2, \ldots, \xi_N)$[B.1], the unit specific treatment reduced form can be written as

$$\begin{aligned} E[Y_{it}|X_{it}, \Phi_{it}] :=& E[D_{it}|X_{it}, \Phi_{it}] + e_0(X_{it}) + \xi_i^E = \\ &= d_{i0}(X_{it})'\beta_0 + e_0(X_{it}) + \xi_i^E = \\ &= l_{i0}(X_{it}) \end{aligned}$$

Hence, the treatment and outcome residuals are, by definition $V_{it} = D_{it} - d_{i0}(X_{it})$, $\tilde{Y}_{it} = Y_{it} - l_{i0}(X_{it})$. Remembering the definitions of Y_{it} and $E[D_{it}|X_{it}, \Phi_{it}]$ and substituting them in the expression of $\tilde{Y}_{it}$ we obtain that

$$\tilde{Y}_{it} = V_{it}'\beta_0 + U_{it} \quad E[U_{it}|X_{it}, \Phi_{it}, V_{it}] = 0 \tag{B.2}$$

Notice that the latter equation identifies β_0 as the coefficient of the best linear approximation of $\tilde{Y}_{it}$ on V_{it} since it represents an OLS coefficient (Semenova et al., 2021). For dealing with the unit fixed effects, Semenova et al. (2021) adopt a FE approach assuming that the fixed-effects ξ^E are weakly sparse. This allows the use of regularization techniques to consistently estimate them and eliminating them. In order to illustrate that the weak sparsity assumption is not restrictive, Semenova et al. (2021) modeled the unit fixed effects in Mundlak style[B.2]. The latter amounts to model

$$e_0(X_{it}) + \xi_i^E = \overline{X}_{it}'\delta_0^X + \underbrace{\overline{M}_i'\delta_{M0} + \xi_i^E}_{a_i^E}$$

where the key is that while x_i^E is restricted to be weakly sparse, no restrictions are imposed to a_i^E. Hence weak sparsity assumption on ξ_i^E is not binding. The same logic can be applied to the unit fixed effects present in the treatment equation.

The estimation and inference strategy assumes that there is high-dimensional sparsity in the sense that $d = dim(D_{it}) = dim(\beta_0) >> NT$ but only a number $s << NT$ of them has non-zero effects, i.e. $||\beta_0||_0 = s$. In this case Semenova et al. (2021) build the following algorithm:

[B.1] Where ξ_i are fixed vectors of unit specific fixed effects.

[B.2] That is allowing for the inclusion of time averages of the covariates in the regression equation to account for unobserved heterogeneity coming from unit fixed effects.

1. Estimate $\tilde{Y}_{it}$ and V_{it} using machine-learning (ML) with a specific form of cross-fitting when dealing with panel data called NLO [B.3] by the authors. The problem with panel data is that traditional cross-fitting techniques do not guarantee a split into independent groups (regardless of whether the split is done on the time or unit dimension). It is hence mandatory to device a cross fitting strategy that deals with weakly dependent data.
2. Estimate the CATE function via Lasso penalized regression of estimated $\tilde{Y}_{it}$ (i.e. $\hat{\tilde{Y}}_{it}$) and V_{it} (i.e. $\hat{V}_{it}$).
3. Perform Gaussian inference using Debiased Lasso on classes of linear functionals of β_0 of the CATE function $D'_{it}\beta_0$.

In the first step of the algorithm, the residuals $\tilde{Y}it$ and V_{it} are estimated via NLO cross-fitting. Specifically for each k, estimates of the reduced form of the residuals are computed using quasi-complementary blocks, such that the following estimates are reached:

$$\hat{\tilde{Y}}_{it} := Y_{it} - \hat{l}_{ik}(X_{it}), \;\; \hat{V}_{it} := D_{it} - \hat{\delta}_{ik}(X_{it})$$

In the second step, CATE functions are estimated via Orthogonal LASSO under Neyman Orthogonality. Namely, let $\lambda_B = C_\beta \sqrt{log(d)/NT}$[B.4] and C_β a penalty parameter. Then define

$$\hat{\beta}_L := argmin_{\beta \in \mathscr{R}^d} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\hat{\tilde{Y}}_{it} - \hat{V}_{it}\beta)^2 + \lambda_\beta \sum_{j=1}^{d} |\beta_j|$$

which is shown to reach almost oracle rate of convergence for the CATE function.
The third step consists finally in describing a debiased inference method on parameters of CATE function. Specifically, since the L1-shrinkage induces the well-known bias of penalized regression it is not possible to run inference directly on $\hat{\beta}_L$ based on standard Gaussian approximations. Thus, it is necessary to construct another estimator –the debiased LASSO estimator– that allows for asymptotically normal confidence intervals:

$$\hat{\beta}_{DL} := \hat{\beta}_L + \hat{\Omega}^{CLIME} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{V}_{it}(\hat{\tilde{Y}}_{it} - \hat{V}_{it}\hat{\beta}_L)$$

where $\hat{\Omega}^{CLIME}$ is a symmetric approximation of the inverse $\hat{\Omega}$ of the sample covariance matrix of the residuals: $\hat{Q} := \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{V}_{it}\hat{V}'_{it}$. One of the main results of Semenova et al. (2021) is that $\sqrt{NT}(\hat{\beta}_{DL} - \hat{\beta}_0)$ is approximately $N(0,\Sigma)$ which, along with the convergence rate of orthogonal LASSO serves as a proof of identification for the case with only exogenous covariates.

C Reduced form of Eq.(17) and Eq.(18)

In what follows we will re-write Eq.(17) and Eq.(18) according to the simplified notation introduced in the main text (Section 4.2.2).
To make the operator steps precise, we (i) work within a molecule and stack products so that s_t and p_t

[B.3] In short, NLO builds quasi-complementary blocks. Say we have a partition $[1,\ldots,T]$ into adjacent blocks $\{\mathscr{M}\}_{k=1}^{K}$. the authors then define $\mathscr{N}(k)$ as k and its immediate neighbors. Hence the quasi-complement of $\mathscr{M}_k$, $\mathscr{M}_k^{QC} = \{\mathscr{M}_1, \mathscr{M}_2, \ldots, \mathscr{M}_K\} \setminus \{\mathscr{M}_j : l \in \mathscr{N}(k)\}$, where "\" denotes the exception sign. The corresponding blocks are $\mathscr{B}_k := \{W_{.,t} : t \in \mathscr{M}_k\}$ and $\mathscr{B}_k^{QC} := \{W_{.,t} : t \in \mathscr{M}_k^{QC}\}$ where $W_{.,t}$ is the data vector at all units at time t.

[B.4] See Semenova et al. (2021) for details are $n \times 1$ vectors, (ii) use the lag operator L with $L^{\ell} x_t = x_{t-\ell}$, (iii) define degree–48 lag polynomials with no contemporaneous term $\alpha_S(L) := \sum_{\ell=1}^{48} \alpha_{S,\ell} L^{\ell}$ and $\alpha_P(L) := \sum_{\ell=1}^{48} \alpha_{P,\ell} L^{\ell}$, and (iv) let $D(K_t)$ be the $n \times n$ diagonal matrix with the observed vector K_t on the diagonal; writing b_0 for the contemporaneous price coefficient and b_K for the coefficient on the interaction $K_t \circ p_t$, the stacked sales equation is

$$s_t \;=\; b_0\, p_t \;+\; b_K\, D(K_t)\, p_t \;+\; \alpha_P(L)\, p_t \;+\; \alpha_S(L)\, s_t \;+\; U_t,$$

so that, after collecting terms in s_t and premultiplying by the inverse operator (which exists under the usual stability condition on $\alpha_S(L)$),

$$\begin{aligned} (I-\alpha_S(L))\, s_t \;&=\; \big[b_0 I + b_K D(K_t) + \alpha_P(L)\big]\, p_t \;+\; U_t \quad \Longrightarrow \\ s_t \;&=\; \underbrace{(I-\alpha_S(L))^{-1}\big[b_0 I + b_K D(K_t) + \alpha_P(L)\big]}_{:=\Upsilon_0(L,t)}\, p_t \;+\; \underbrace{(I-\alpha_S(L))^{-1}}_{:=\Upsilon_U(L)}\, U_t. \end{aligned}$$

For prices, start from the 48–lag reduced form $p_t = \varphi(L)\, p_t + V_t$ with $\varphi(L) := \sum_{\ell=1}^{48} \varphi_{\ell} L^{\ell}$ (no contemporaneous term); then $p_t = (I-\varphi(L))^{-1} V_t =: \Upsilon_1(L)\, V_t$. Substituting $p_t = \Upsilon_1(L) V_t$ into the sales equation yields the reduced form stated in the main text with the correct error filtering:

$$s_t \;=\; \Upsilon_0(L,t)\, \Upsilon_1(L)\, V_t \;+\; \Upsilon_U(L)\, U_t,$$

that is, the "system" operator $\Upsilon_0(L,t)$ multiplies p_t (and hence V_t through Υ_1), while the disturbance is transformed only by $(I-\alpha_S(L))^{-1}$. Briefly: $\alpha_S(L)$ and $\alpha_P(L)$ are degree–48 lag polynomials; $(I-\alpha_S(L))^{-1}$ filters both the regressors block and the error, and the price process is solved by its own inverse operator before substitution.

D Within-Country Analysis

Figure D.1 provides within-country analysis of the relation between the price elasticities of molecules and molecule prices, while it also displays a cross-country comparison. Each subgraph shows the prediction for the price elasticities of molecules available in the country in question from a linear regression of price elasticities on average prices (log) at the molecule level. The results exhibit some degree of variation. Algeria and Turkey have a somewhat strong positive slope, indicating an inverse relation between price elasticities and price. Netherlands, Russia, Slovenia and the USA show a rather neutral relation between price elasticities and prices. Austria, Australia, Hungary, Italy, Portugal, and Slovakia have a clear negative slope, meaning that pharmaceutical prices at the molecule level do not inversely vary with price elasticities. Overall, Figure D.1 provides moderate evidence on the non-existence of within-country Ramsey pricing at the molecule level in many national pharmaceutical markets. Country-wise differences in pharmaceutical markets can be addressed by appropriate consideration of various determinants including national health systems, insurance coverage, availability of generic products, reference pricing, and parallel trade.

We will here derive also the theoretical results in the main text referring to the within country analysis. In particular we will compute the exact derivatives from the linearized Ramsey price equation (6).
Write

$$p_i \;=\; MC_i\Big[\,1 - DR_i - CI_i\Big], \qquad DR_i := \frac{\lambda}{1+\lambda}\cdot\frac{1}{ep_{ii}}, \qquad CI_i := \sum_{j\neq i} M_j\, \frac{ep_{ji}}{ep_{ii}}\;\frac{q_j}{q_i}\;\frac{p_j}{p_i}.$$

Fix i and consider a cross–elasticity ep_{ki} with $k \neq i$. Differentiating p_i with respect to ep_{ki} and using the product/chain rules yields

$$\frac{\partial p_i}{\partial ep_{ki}} = -MC_i\left(\underbrace{\frac{\partial DR_i}{\partial ep_{ki}}}_{=0} + \frac{\partial CI_i}{\partial ep_{ki}} + \frac{\partial CI_i}{\partial p_i}\cdot\frac{\partial p_i}{\partial ep_{ki}}\right).$$

Hence, collecting all the $\frac{\partial p_i}{\partial ep_{ki}}$ terms on the left hand side, we have:

$$\frac{\partial p_i}{\partial ep_{ki}}\left[1 + MC_i\,\frac{\partial CI_i}{\partial p_i}\right] = -MC_i\,\frac{\partial CI_i}{\partial ep_{ki}}.$$

The two partial derivatives inside are

$$\frac{\partial CI_i}{\partial ep_{ki}} = M_k\,\frac{1}{ep_{ii}}\,\frac{q_k}{q_i}\,\frac{p_k}{p_i}, \qquad \frac{\partial CI_i}{\partial p_i} = -\sum_{j\neq i} M_j\,\frac{ep_{ji}}{ep_{ii}}\,\frac{q_j}{q_i}\,\frac{p_j}{p_i^2}.$$

Substituting and solving for $\partial p_i/\partial ep_{ki}$ gives the *exact* linearized derivative:

$$\boxed{\frac{\partial p_i}{\partial ep_{ki}} = \frac{-MC_i\,M_k\,\frac{1}{ep_{ii}}\,\frac{q_k}{q_i}\,\frac{p_k}{p_i}}{1 - MC_i\,\sum_{j\neq i} M_j\,\frac{ep_{ji}}{ep_{ii}}\,\frac{q_j}{q_i}\,\frac{p_j}{p_i^2}}\;.} \tag{D.1}$$

An analogous expression holds for any ep_{ji} with $j \neq i$:

$$\boxed{\frac{\partial p_i}{\partial ep_{ji}} = \frac{-MC_i\,M_j\,\frac{1}{ep_{ii}}\,\frac{q_j}{q_i}\,\frac{p_j}{p_i}}{1 - MC_i\,\sum_{\ell\neq i} M_\ell\,\frac{ep_{\ell i}}{ep_{ii}}\,\frac{q_\ell}{q_i}\,\frac{p_\ell}{p_i^2}}\;.} \tag{D.2}$$

These formulas are exact *conditional on the linearized price equation* $p_i = MC_i[1-(DR_i + CI_i)]$ and capture the implicit feedback through the p_i-dependence of CI_i via the term p_j/p_i. For substitutes ($ep_{ji} > 0$ for all relevant j), the numerator in (D.1) is positive because $ep_{ii} < 0$, and the denominator exceeds zero for small interaction terms, implying $\partial p_i/\partial ep_{ki} > 0$.

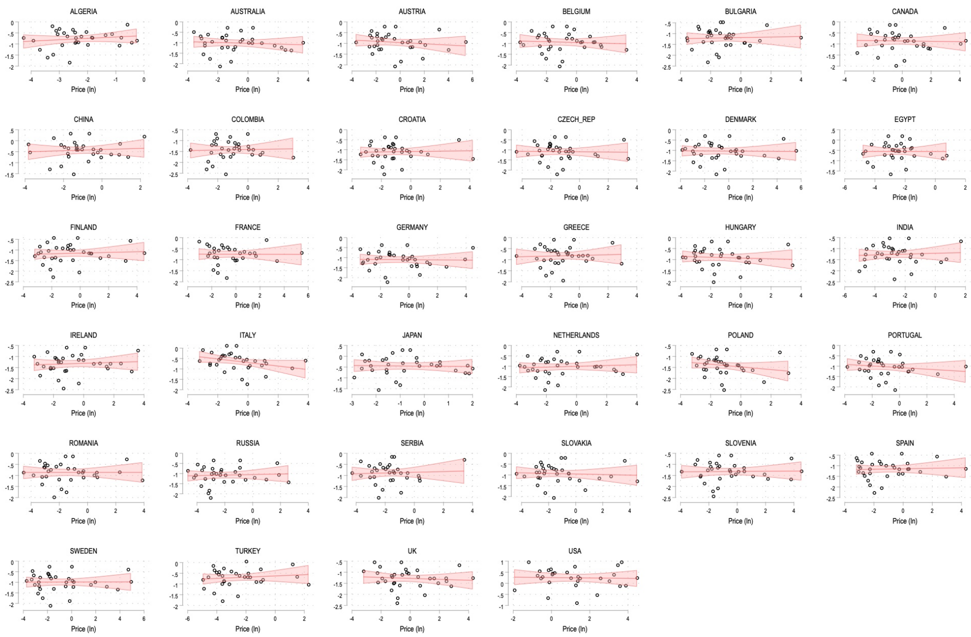


Figure D.1. Relationship Between Price Elasticities and Price of Molecules by Country

Figure D.2 restricts the within-country analysis to generics. Figure D.2, in particular, shows that the well-known inverse elasticity rule holds in national pharmaceutical markets; thus, confirming the presence of a Ramsey pricing within-country for generic products. Unlike the full range of pharmaceutical products (Figure D.1), each sub-graph exhibits a clear positive slope, indicating a strong negative relationship between price elasticity and the price of generics in each national market.

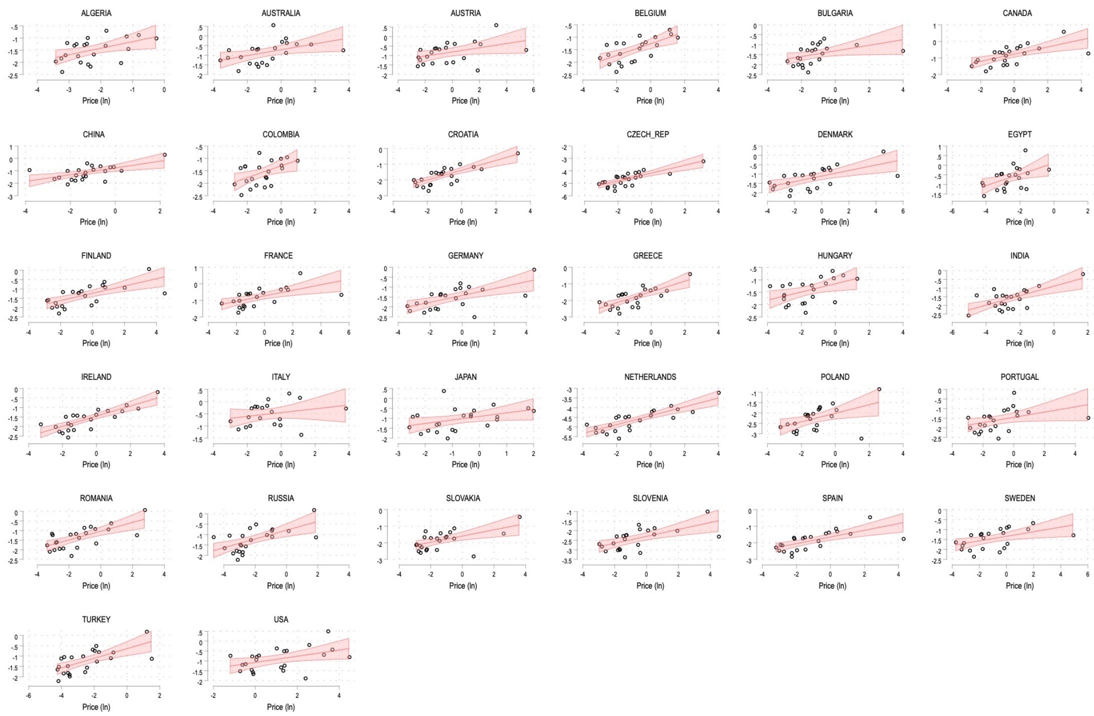


Figure D.2. Price Elasticities by Molecule-Country Pairs for Generic Pharmaceuticals

D.1 Generic vs branded within country comparison

Let's compare how much *within-generics* cross-price elasticities move generic prices versus how much *within-brands* cross elasticities move branded prices.
We know from above that for a focal product i and a rival $\ell \neq i$:

$$\frac{\partial p_i}{\partial \mathrm{ep}_{\ell i}} = \frac{-MC_i M_\ell \dfrac{1}{\mathrm{ep}_{ii}} \dfrac{q_\ell}{q_i} \dfrac{p_\ell}{p_i}}{1 - MC_i \displaystyle\sum_{m \neq i} M_m \dfrac{\mathrm{ep}_{mi}}{\mathrm{ep}_{ii}} \dfrac{q_m}{q_i} \dfrac{p_m}{p_i^2}} \;\; =: \;\; \frac{-MC_i M_\ell \dfrac{1}{\mathrm{ep}_{ii}} \dfrac{q_\ell}{q_i} \dfrac{p_\ell}{p_i}}{D_i}. \tag{D.3}$$

Assume substitutes: $\mathrm{ep}_{ii} < 0$ and $\mathrm{ep}_{\ell i} > 0$. Then $D_i > 1$ and, after putting the absolute values,

$$\frac{\partial p_i}{\partial \mathrm{ep}_{\ell i}} = \frac{MC_i}{|\mathrm{ep}_{ii}|} \frac{M_\ell}{D_i} \frac{q_\ell}{q_i} \frac{p_\ell}{p_i}. \tag{D.4}$$

For convenience (since many terms simplify), let us now define a unit-free normalized sensitivity as follows:

$$S_i^{(\ell \to i)} := \frac{|\mathrm{ep}_{ii}|}{MC_i} \frac{q_i}{q_\ell} \frac{p_i}{p_\ell} \frac{\partial p_i}{\partial \mathrm{ep}_{\ell i}}. \tag{D.5}$$

Plug now (D.4) into (D.5):

$$\begin{aligned} S_i^{(\ell\to i)} &= \left(\frac{|\mathrm{ep}_{ii}|}{MC_i}\right)\left(\frac{q_i}{q_\ell}\right)\left(\frac{p_i}{p_\ell}\right)\left[\frac{MC_i}{|\mathrm{ep}_{ii}|}\,\frac{M_\ell}{D_i}\,\frac{q_\ell}{q_i}\,\frac{p_\ell}{p_i}\right] \\ &= \underbrace{\frac{|\mathrm{ep}_{ii}|}{MC_i}\cdot\frac{MC_i}{|\mathrm{ep}_{ii}|}}_{=1}\,\underbrace{\frac{q_i}{q_\ell}\cdot\frac{q_\ell}{q_i}}_{=1}\,\underbrace{\frac{p_i}{p_\ell}\cdot\frac{p_\ell}{p_i}}_{=1}\cdot\frac{M_\ell}{D_i}=\frac{M_\ell}{D_i}. \end{aligned} \tag{D.6}$$

Thus, conveniently, the normalized effect of a rival ℓ on the focal i depends only on the rival's markup M_ℓ and the focal-good feedback factor D_i.
We are now ready to aggregate within category (branded-generic). For a focal generic $i = G$ (with rivals $k \in \mathscr{G}'$) and a focal brand $i = B$ (with rivals $j \in \mathscr{B}'$):

$$\mathscr{S}_G := \sum_{k\in\mathscr{G}'} S_G^{(k\to G)} = \sum_{k\in\mathscr{G}'}\frac{M_k}{D_G} = \frac{\sum_{k\in\mathscr{G}'} M_k}{D_G}, \qquad \mathscr{S}_B := \sum_{j\in\mathscr{B}'} S_B^{(j\to B)} = \frac{\sum_{j\in\mathscr{B}'} M_j}{D_B}. \tag{D.7}$$

the idea is that $\mathscr{S}_G$ ($\mathscr{S}_B$) describes how strongly other generics (branded) –not the focal one– can move the generic (branded)'s price through cross-substitution, after stripping out size/units effects. It is therefore easy to see that generics' cross-elasticities matter more for generic prices than brands' cross-elasticities matter for brand prices iff

$$\mathscr{S}_G > \mathscr{S}_B \iff \frac{\sum_{k\in\mathscr{G}'} M_k}{D_G} > \frac{\sum_{j\in\mathscr{B}'} M_j}{D_B} \iff \frac{\sum_{k\in\mathscr{G}'} M_k}{\sum_{j\in\mathscr{B}'} M_j} > \frac{D_G}{D_B}. \tag{D.8}$$

where D_i can be recovered from (D.3) as:

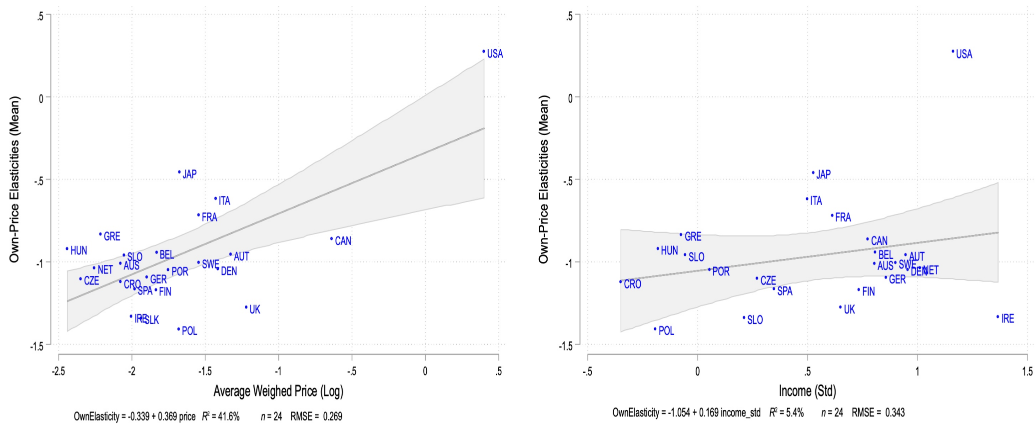
$$D_i = 1 - MC_i \sum_{m\neq i} M_m \frac{\mathrm{ep}_{mi}}{\mathrm{ep}_{ii}}\,\frac{q_m}{q_i}\,\frac{p_m}{p_i^2} = 1 + MC_i \sum_{m\neq i} M_m \frac{|\mathrm{ep}_{mi}|}{|\mathrm{ep}_{ii}|}\,\frac{q_m}{q_i}\,\frac{p_m}{p_i^2} \;>\; 1. \tag{D.9}$$

$$=: 1 + MC_i K_i, \qquad K_i := \sum_{m\neq i} M_m \frac{|\mathrm{ep}_{mi}|}{|\mathrm{ep}_{ii}|}\,\frac{q_m}{q_i}\,\frac{p_m}{p_i^2}. \tag{D.10}$$

The bigger D_i is, the harder it is for a change in elasticities to move product i's price (because D_i is at the denominator of $\mathscr{S}_i$). Higher MC_i and stronger substitution into i (larger K_i) make D_i larger, which dampens the price response.

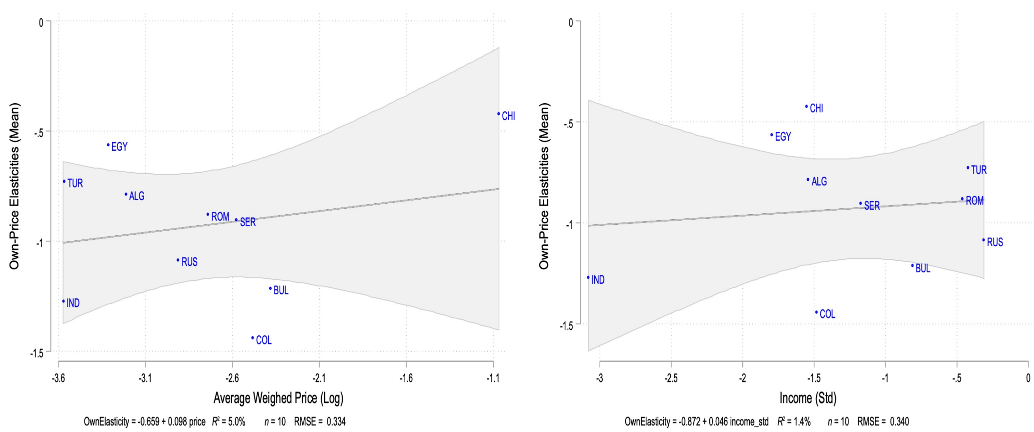
E Analysis for High Income Countries and Middle-Low Income Countries

This section reproduces the analysis of subsection 4.3.2 and 4.3.3 by focusing separately on HIC and MLIC.



((a)) ((b))

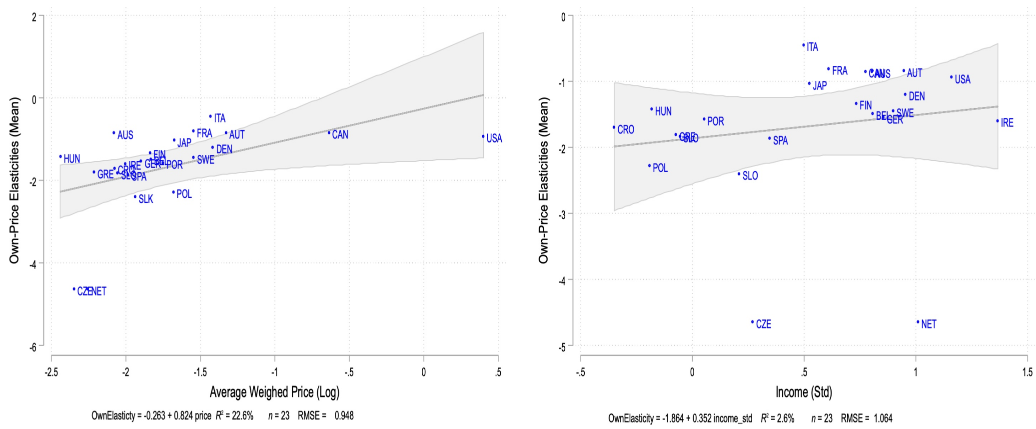
Figure E.1. Relationship Between Price Elasticities and Price (a) and GDP Per Capita (PPP) Income (b) for High Income Countries (HIC)



((a)) ((b))

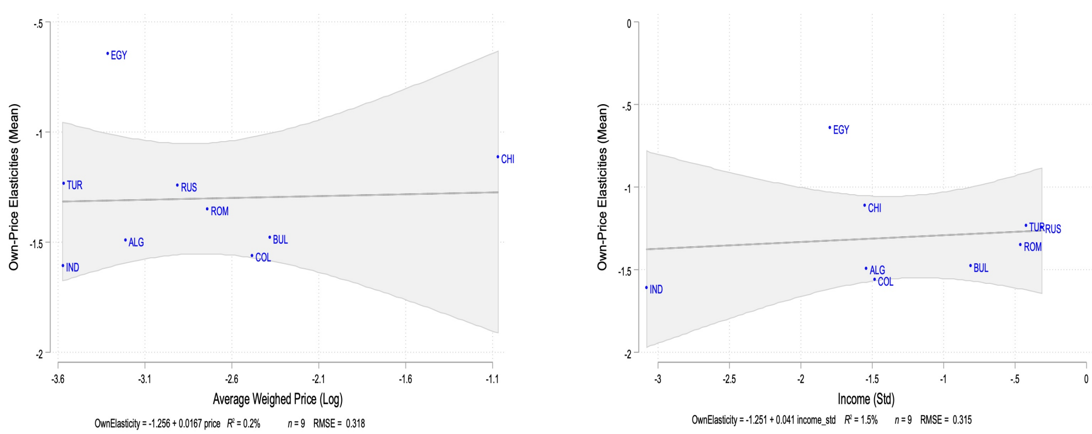
Figure E.2. Relationship Between Price Elasticities and Price (a) and GDP Per Capita (PPP) Income (b) for Middle and Low Income Countries (MLIC)

E.1 Generics



((a)) ((b))

Figure E.3. Generic Pharmaceuticals: Relationship Between Price Elasticities and Price (a) and GDP Per Capita (PPP) Income (b) for High Income Countries (HIC)

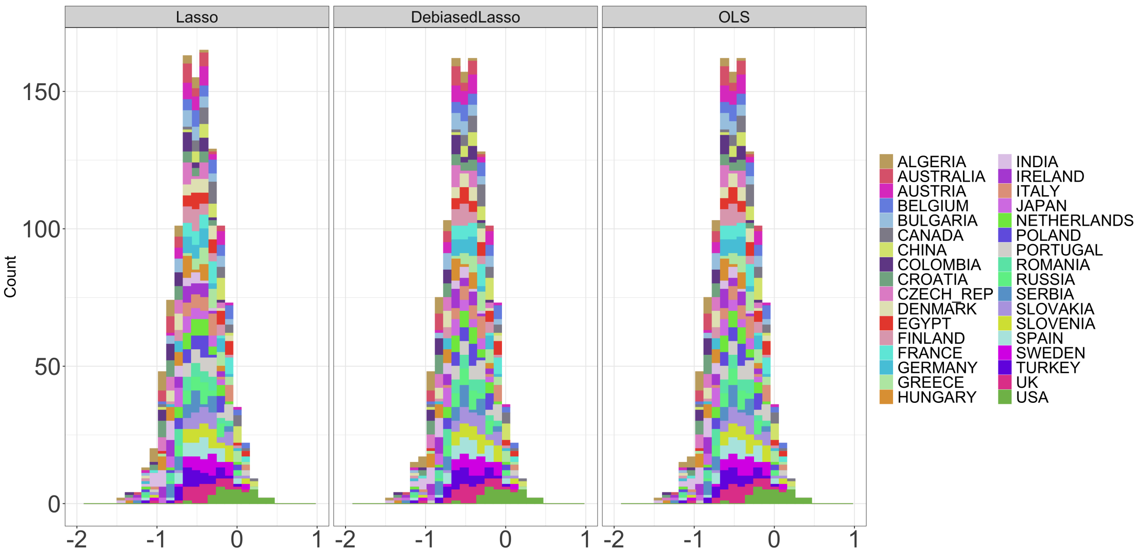


((a)) ((b))

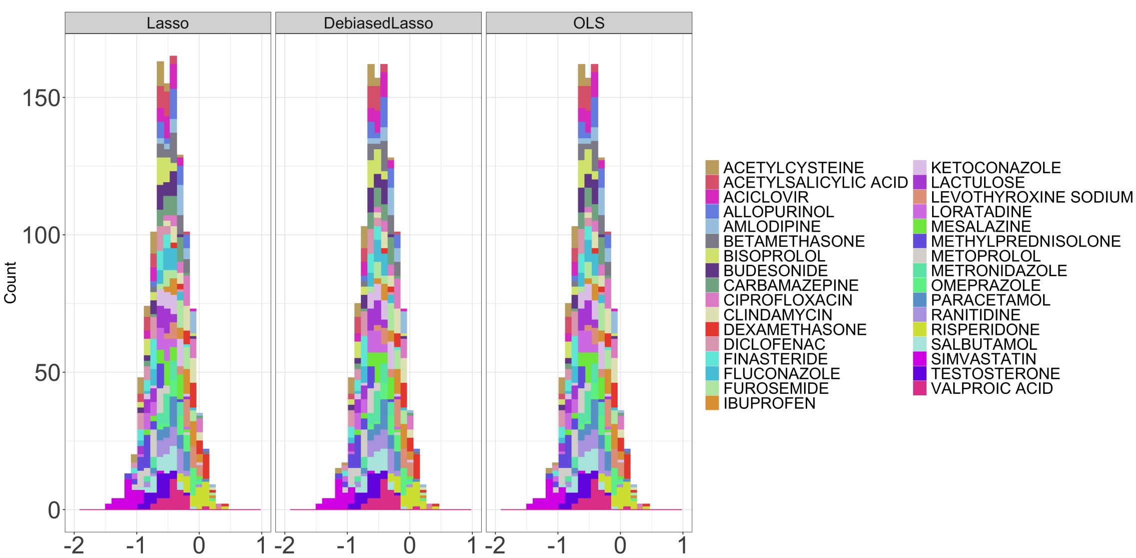
Figure E.4. Generic Pharmaceuticals: Relationship Between Price Elasticities and Price (a) and GDP Per Capita (PPP) Income (b) for Middle and Low Income Countries (MLIC)

F Different Second Stage Models

As a robustness check, we analyse if the price elasticity estimates are robust to different second-stage model choices. Figure F.1 displays the results.



((a))



((b))

Figure F.1. Different Second Stage Models

Following Semenova et al. (2021), we estimate heterogeneous elasticities with three distinct methods: cross-validated Lasso (first panel), Debiased Lasso (middle panel) and OLS (last panel). Each point on the histograms represents one panel unit, *i.e.* country-molecule pair. Each bar is a collection of estimated price elasticities for country-molecule pairs. Consequently, pairs located in the same vertical bin have the same or very close estimated elasticities. Y-axis indicates how many times a particular elasticity value is estimated. The total number of estimated price elasticities is 1122 ($33x34$), which is the number of units, *i.e.* molecule-country pairs.

Figures F.1(a) and F.1(b) differ between each other due to the grouping level. Figure F.1 colours elasticity estimates at the country level, while Figure F.1 colours at the molecule level to display the heterogeneity of elasticity estimates in both levels. The results show that OLS and Debiased Lasso methods produce the same results. Cross-validated Lasso produces slightly different results, while the overall picture does not change. The results are pretty robust to different model choices.

G Placebo Test

We introduce a placebo test to examine the effectiveness of the DML algorithm employed to estimate the price elasticities of demand. Specifically, we randomize the input data and re-estimate own and cross price elasticities to show that the algorithm does not produce spurious results.

As the first step, we construct the placebo data set by including molecules serving different needs (unrelated ATCs) and countries that are relatively unrelated in terms of complexity pattern (Albeaik et al., 2017)(*e.g.* Germany and Ecuador). As the second step, we perform randomization to re-assign the standard units of different molecules in different countries. Starting from the original database, we make a permutation of the standard units at the molecule and country levels. In this way, we associate the standard units of molecule G in country c to the standard units of molecule j in country k. The randomization step ensures the absence of any relationship between the movements of the prices. Therefore, it is reasonable to expect mostly insignificant elasticity estimates. Finally, we apply the DDML algorithm to the placebo data set. Figure G.1 displays the results. To facilitate the visualization, we report the matrix of molecules' elasticities ($\mathbf{E} \in \mathbb{R}^{m \times m}$) in the form of heat map. By construction, the diagonal elements $e_{i,i}$ of $\mathbf{E}$ represent own-price elasticities, whereas non-diagonal elements $e_{i,j}, i \neq j$ represent the cross price elasticity of i with respect to j. More formally:

$$E_{i,j} := \begin{cases} \textit{Own-elasticity}, & \text{if } i = j, \\ \textit{Cross-elasticity}, & \text{otherwise}. \end{cases}$$

Zero values have been set to black (NaN). The results show that the 96% of the elasticity estimates are not significantly different from zero when countries with no relevant trade relationship have interacted with each other (both on the main diagonal, *i.e.* own elasticities, and outside-of-diagonal, *i.e.* cross elasticities).

The placebo test has been repeated for several country pairs, yielding very similar results. For the sake of brevity, we only report the results for Germany and Ecuador.

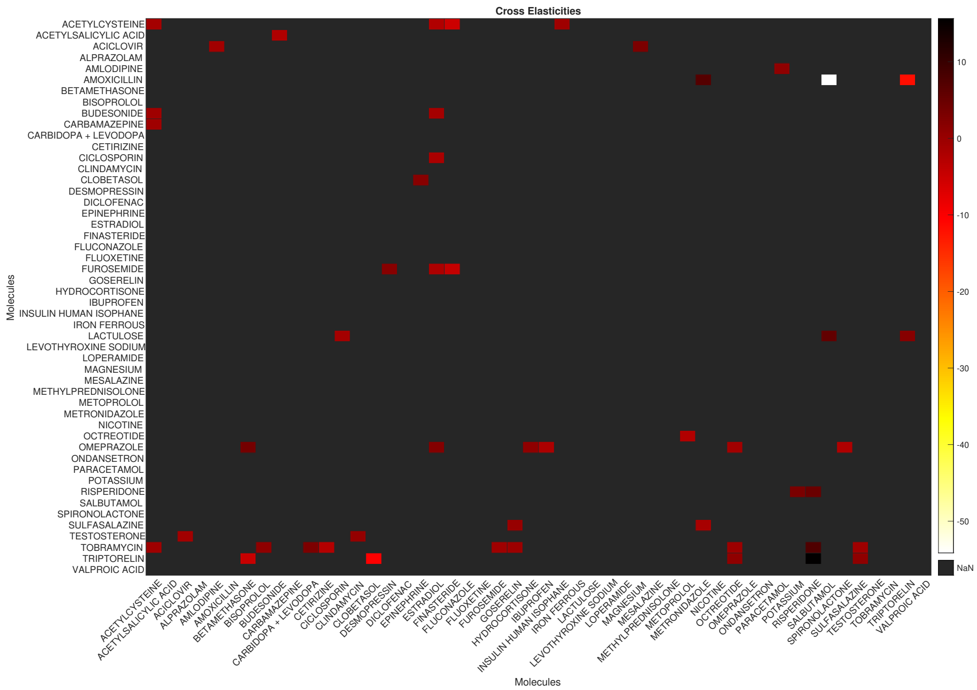


Figure G.1. Heat Map of Placebo Price Elasticities Between the Molecules of Ecuador and Germany

H Fixed Effects for Table 1

Table H.1. Within-Between estimation (own elasticity)

	All Products		Generics	
	(1)	(2)	(3)	(4)
	Within	Between	Within	Between
ALGERIA (comparison category)	.	.	.	.
	(.)	(.)	(.)	(.)
AUSTRALIA	1.1100***	.	0.3848	.
	(0.3191)	.	(0.2955)	.
AUSTRIA	1.3030***	.	0.7831***	.
	(0.3191)	.	(0.2787)	.
BELGIUM	1.1109***	.	0.9434***	.
	(0.3191)	.	(0.3435)	.
BULGARIA	0.8084**	.	0.0174	.
	(0.3191)	.	(0.2954)	.
CANADA	1.1784***	.	1.3486***	.
	(0.3191)	.	(0.2953)	.
CHINA	0.7809**	.	0.0611	.
	(0.3191)	.	(0.3473)	.
COLOMBIA	0.7575**	.	0.3871	.
	(0.3191)	.	(0.3515)	.
CROATIA	0.9322***	.	1.0023***	.
	(0.3191)	.	(0.3434)	.
DENMARK	0.8647***	.	0.4899	.
	(0.3191)	.	(0.3436)	.
EGYPT	-0.5014	.	-0.2894	.
	(0.3191)	.	(0.2785)	.
FINLAND	1.0425***	.	1.0209***	.
	(0.3191)	.	(0.2700)	.
FRANCE	1.1682***	.	0.5043*	.
	(0.3191)	.	(0.2999)	.
GERMANY	0.7846**	.	0.5879**	.
	(0.3191)	.	(0.2965)	.
GREECE	0.8644***	.	1.1776***	.
	(0.3191)	.	(0.3515)	.
HUNGARY	0.5413*	.	0.9074***	.
	(0.3191)	.	(0.2966)	.
INDIA	-0.8518***	.	-1.1618***	.
	(0.3194)	.	(0.2779)	.
IRELAND	1.3452***	.	1.3819***	.
	(0.3192)	.	(0.3439)	.
ITALY	1.1091***	.	0.8840***	.
	(0.3191)	.	(0.2953)	.
JAPAN	1.3514***	.	-0.0013	.
	(0.3191)	.	(0.2791)	.
NETHERLANDS	0.7016**	.	0.4720*	.
	(0.3191)	.	(0.2411)	.
POLAND	0.5686*	.	0.2470	.
	(0.3191)	.	(0.3434)	.
PORTUGAL	0.8222**	.	0.8296**	.
	(0.3191)	.	(0.3434)	.
ROMANIA	0.5216	.	0.2730	.
	(0.3191)	.	(0.3332)	.
RUSSIA	-0.3024	.	-0.5534**	.
	(0.3191)	.	(0.2699)	.
SERBIA	0.5226	.	0.5841*	.
	(0.3191)	.	(0.3473)	.
SLOVAKIA	0.8087**	.	0.4380	.
	(0.3191)	.	(0.3473)	.
SLOVENIA	1.0896***	.	1.0800***	.
	(0.3191)	.	(0.3515)	.
SPAIN	0.5525*	.	0.4592	.
	(0.3191)	.	(0.3473)	.
SWEDEN	0.6392**	.	1.2756***	.
	(0.3191)	.	(0.2953)	.
TURKEY	-0.7319**	.	-0.9302***	.
	(0.3191)	.	(0.2975)	.
UK	0.5725*	.	0.4611*	.
	(0.3191)	.	(0.2143)	.
USA	1.4691***	.	0.5812***	.
	(0.3192)	.	(0.1434)	.

Country Fixed Effects for Table 1: It is observed, as anticipated, that countries with a high GDP exert the most significant effects, exhibiting an average price higher than that of other countries. This observation aligns with our previous descriptive findings

Online Appendix

A Technical details for the case $DR + CI < 1$

A.1 Computations for case (ii)

Case 1: Generic Drug (i is generic, $k \in \mathscr{G}'$). We consider the effect of the cross-price elasticity ep_{ki} on the price p_i of a generic drug i. The relevant term in the pricing formula is

$$T_k = M_k \cdot \frac{\mathrm{ep}_{ki}}{\mathrm{ep}_{ii}} \cdot \frac{q_k}{q_i} \cdot \frac{p_k}{p_i}$$

which appears inside the summation. Differentiating the price equation with respect to ep_{ki}, we obtain

$$\frac{\partial p_i}{\partial \mathrm{ep}_{ki}} = -MC_i \cdot \frac{\partial}{\partial \mathrm{ep}_{ki}} \left(M_k \cdot \frac{\mathrm{ep}_{ki}}{\mathrm{ep}_{ii}} \cdot \frac{q_k}{q_i} \cdot \frac{p_k}{p_i} \right)$$

$$= -MC_i \cdot M_k \cdot \frac{1}{\mathrm{ep}_{ii}} \cdot \frac{q_k}{q_i} \cdot \frac{p_k}{p_i} + MC_i \cdot M_k \cdot \frac{\mathrm{ep}_{ki}}{\mathrm{ep}_{ii}} \cdot \frac{q_k}{q_i} \cdot \frac{p_k}{p_i^2} \cdot \frac{\partial p_i}{\partial \mathrm{ep}_{ki}}$$

Solving for the derivative yields:

$$\frac{\partial p_i}{\partial \mathrm{ep}_{ki}} = \frac{-MC_i \cdot M_k \cdot \frac{1}{\mathrm{ep}_{ii}} \cdot \frac{q_k}{q_i} \cdot \frac{p_k}{p_i}}{1 - MC_i \cdot M_k \cdot \frac{\mathrm{ep}_{ki}}{\mathrm{ep}_{ii}} \cdot \frac{q_k}{q_i} \cdot \frac{p_k}{p_i^2}}$$

Case 2: Branded Drug (i is branded, $j \in \mathscr{B}, j \neq i$). We now consider the effect of the cross-price elasticity ep_{ji} on the price p_i of a branded drug i. The relevant term is

$$T_j = M_j \cdot \frac{\mathrm{ep}_{ji}}{\mathrm{ep}_{ii}} \cdot \frac{q_j}{q_i} \cdot \frac{p_j}{p_i}$$

Differentiating the price expression with respect to ep_{ji} gives:

$$\frac{\partial p_i}{\partial \mathrm{ep}_{ji}} = -MC_i \cdot \frac{\partial}{\partial \mathrm{ep}_{ji}} \left(M_j \cdot \frac{\mathrm{ep}_{ji}}{\mathrm{ep}_{ii}} \cdot \frac{q_j}{q_i} \cdot \frac{p_j}{p_i} \right)$$

$$= -MC_i \cdot M_j \cdot \frac{1}{\mathrm{ep}_{ii}} \cdot \frac{q_j}{q_i} \cdot \frac{p_j}{p_i} + MC_i \cdot M_j \cdot \frac{\mathrm{ep}_{ji}}{\mathrm{ep}_{ii}} \cdot \frac{q_j}{q_i} \cdot \frac{p_j}{p_i^2} \cdot \frac{\partial p_i}{\partial \mathrm{ep}_{ji}}$$

Solving for the derivative, we find:

$$\frac{\partial p_i}{\partial \mathrm{ep}_{ji}} = \frac{-MC_i \cdot M_j \cdot \frac{1}{\mathrm{ep}_{ii}} \cdot \frac{q_j}{q_i} \cdot \frac{p_j}{p_i}}{1 - MC_i \cdot M_j \cdot \frac{\mathrm{ep}_{ji}}{\mathrm{ep}_{ii}} \cdot \frac{q_j}{q_i} \cdot \frac{p_j}{p_i^2}}$$

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge (2023). “When should you adjust standard errors for clustering?” In: *The Quarterly Journal of Economics* 138.1, pp. 1–35.

Adebayo, Esther F, Olalekan A Uthman, Charles S Wiysonge, Erin A Stern, Kim T Lamont, and John E Ataguba (2015). “A systematic review of factors that affect uptake of community-based health insurance in low-income and middle-income countries”. In: *BMC health services research* 15, pp. 1–13.

Adida, Elodie (2021). “Outcome-based pricing for new pharmaceuticals via rebates”. In: *Management Science* 67.2, pp. 892–913.

Albeaik, Saleh, Mary Kaltenberg, Mansour Alsaleh, and Cesar A Hidalgo (2017). “Improving the economic complexity index”. In: *arXiv preprint arXiv:1707.05826*.

Armstrong, Mark and John Vickers (2018). “Multiproduct pricing made simple”. In: *Journal of Political Economy* 126.4, pp. 1444–1471.

Bajari, P., D. Nekipelov, S. P. Ryan, and M. Yang (Feb. 2015). *Demand Estimation with Machine Learning and Model Combination*. Working Paper 20955. National Bureau of Economic Research. URL: `http://www.nber.org/papers/w20955`.

Baltagi, Badi H and Francesco Moscone (2010). “Health care expenditure and income in the OECD reconsidered: Evidence from panel data”. In: *Economic modelling* 27.4, pp. 804–811.

Barros, P. P. and X. Martinez-Giralt (2008). “On international cost-sharing of pharmaceutical R&D”. In: *International Journal of Health Care Finance and Economics* 8.4, pp. 301–312.

Baumol, W. J. and D. F. Bradford (1970). “Optimal Departures From Marginal Cost Pricing”. In: *The American Economic Review* 60.3, pp. 265–283. ISSN: 00028282. URL: `http://www.jstor.org/stable/1817977`.

Belloni, A. and V. Chernozhukov (2013). “Least squares after model selection in high-dimensional sparse models”. In: *Bernoulli* 19.2, pp. 521–547. URL: `https://doi.org/10.3150/11-BEJ410`.

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.

Berry, S., J. Levinsohn, and A. Pakes (1995). “Automobile Prices in Market Equilibrium”. In: *Econometrica* 63.4, pp. 841–890. URL: `http://www.jstor.org/stable/2171802`.

Bertoletti, Paolo (2018). “A note on Ramsey pricing and the structure of preferences”. In: *Journal of Mathematical Economics* 76, pp. 45–51.

Bigerna, S. and C. A. Bollino (2016). “Ramsey prices in the Italian electricity market”. In: *Energy Policy* 88, pp. 603–612. ISSN: 0301-4215. DOI: `https://doi.org/10.1016/j.enpol.2015.06.037`.

Björnerstedt, Jonas and Frank Verboven (2012). “Does Merger Simulation Work? A’Natural Experiment’in the Swedish Analgesics Market”. In.

Blankart, Katharina E and Sotiris Vandoros (2024). “Explaining why increases in generic use outpace decreases in brand name medicine use in multisource markets and the role of regulation”. In: *Plos one* 19.5, e0301716.

Blazquez-Fernandez, Carla, David Cantarero, and Patricio Perez (2014). “Disentangling the heterogeneous income elasticity and dynamics of health expenditure”. In: *Applied Economics* 46.16, pp. 1839–1854.

Borgheini, Giuseppe (2003). “The bioequivalence and therapeutic efficacy of generic versus brand-name psychoative drugs”. In: *Clinical therapeutics* 25.6, pp. 1578–1592.

Brot-Goldberg, Z. C., A. Chandra, B. R. Handel, and J. T. Kolstad (2017). “What does a Deductible Do? The Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics”. In: *The Quarterly*

Journal of Economics 132.3, pp. 1261–1318. ISSN: 0033-5533. URL: https://doi.org/10.1093/qje/qjx013.
Buyukyazici, Duygu (2022). "Essays in Applied Economics". PhD thesis. IMT School for Advanced Studies Lucca. URL: http://e-theses.imtlucca.it/id/eprint/359.
Cameron, Alexandra, Margaret Ewen, Dennis Ross-Degnan, Douglas Ball, and Richard Laing (2009). "Medicine prices, availability, and affordability in 36 developing and middle-income countries: a secondary analysis". In: *The lancet* 373.9659, pp. 240–249.
Chandra, A., J. Gruber, and R. McKnight (2014). "The impact of patient cost-sharing on low-income populations: Evidence from Massachusetts". In: *Journal of Health Economics* 33, pp. 57–66. ISSN: 0167-6296. URL: https://www.sciencedirect.com/science/article/pii/S0167629613001409.
Chaudhuri, Sudip (2012). "Multinationals and monopolies: Pharmaceutical industry in India after TRIPS". In: *Economic and Political Weekly*, pp. 46–54.
Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018a). "Double/debiased machine learning for treatment and structural parameters". In: *The Econometrics Journal* 21.1, pp. 1–68. URL: https://doi.org/10.1111/ectj.12097.
Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins (2018b). *Double/debiased machine learning for treatment and structural parameters*.
Chevalier, Judith A, Anil K Kashyap, and Peter E Rossi (2003). "Why don't prices rise during periods of peak demand? Evidence from scanner data". In: *American Economic Review* 93.1, pp. 15–37.
Comin, Diego, Danial Lashkari, and Martí Mestieri (2021). "Structural change with long-run income and price effects". In: *Econometrica* 89.1, pp. 311–374.
Conrad, Ryan and Randall Lutter (2019). "Generic competition and drug prices: new evidence linking greater generic competition and lower generic drug prices". In: *US Food & Drug Administration*.
Costa-Font, Joan, Caroline Rudisill, and Stefanie Tan (2014). "Brand loyalty, patients and limited generic medicines uptake". In: *Health policy* 116.2-3, pp. 224–233.
Cutler, David M and Keith M Marzilli Ericson (2010). "Cost-effectiveness analysis in markets with high fixed costs". In: *Pharmacoeconomics* 28, pp. 867–875.
Danzon, M. P. (1998). "The Economics of Parallel Trade". In: *Pharmaeconomics* 13.3, pp. 293–304.
— (2019). "Drug Pricing and Value in Oncology". In: *Regulatory and Economic Aspects in Oncology* 213.
Danzon, M. P. and A. Towse (2003a). "Differential Pricing for Pharmaceuticals: Reconciling Access, RD and Patents". In: *International Journal of Health Care Finance and Economics* 3, pp. 183–205.
Danzon, M. P., A. Towse, and J. Mestre-Ferrandiz (2015a). "Value-Based Differential Pricing: Efficient Prices for Drugs in a Global Context". In: *Health Economics* 24.3, pp. 294–301.
Danzon, P. M. (1997). "Price Discrimination for Pharmaceuticals: Welfare Effects in the US and the EU". In: *International Journal of the Economics of Business* 4.3.
— (2000). "Making sense of drug prices". In: *Regulation : the Cato review of business and government* 21.1.
— (2018). "Differential Pricing of Pharmaceuticals: Theory, Evidence and Emerging Issues". In: *Pharmacoeconomics* 36.12, pp. 1395–1405.
Danzon, Patricia, Adrian Towse, and Jorge Mestre-Ferrandiz (2015b). "Value-based differential pricing: Efficient prices for drugs in a global context". In: *Health economics* 24.3, pp. 294–301.
Danzon, Patricia M, Andrew W Mulcahy, and Adrian K Towse (2015). "Pharmaceutical Pricing in Emerging Markets: Effects of Income, Competition, and Procurement". In: *Health economics* 24.2, pp. 238–252.
Danzon, Patricia M and Adrian Towse (2003b). "Differential pricing for pharmaceuticals: reconciling access, R&D and patents". In: *International journal of health care finance and economics* 3, pp. 183–205.

Datta, Anusua and Dhaval Dave (2017). "Effects of physician-directed pharmaceutical promotion on prescription behaviors: longitudinal evidence". In: *Health economics* 26.4, pp. 450–468.

Desai, Rishi J, Ameet Sarpatwari, Sara Dejene, Nazleen F Khan, Joyce Lii, James R Rogers, Sarah K Dutcher, Saeid Raofi, Justin Bohn, John G Connolly, et al. (2019). "Comparative effectiveness of generic and brand-name medication use: A database study of US health insurance claims". In: *PLoS medicine* 16.3, e1002763.

Desmarais, Julie Eve, Linda Beauclair, and Howard C Margolese (2011). "Switching from brand-name to generic psychotropic medications: A literature review". In: *CNS neuroscience & therapeutics* 17.6, pp. 750–760.

Duggan, Mark and Fiona Scott Morton (2010). "The effect of Medicare Part D on pharmaceutical prices and utilization". In: *American Economic Review* 100.1, pp. 590–607.

Dunne, Suzanne, Bill Shannon, Ailish Hannigan, Colum Dunne, and Walter Cullen (2014). "Physician and pharmacist perceptions of generic medicines: What they think and how they differ". In: *Health policy* 116.2-3, pp. 214–223.

Dunne, Suzanne S and Colum P Dunne (2015). "What do people really think of generic medicines? A systematic review and critical appraisal of literature on stakeholder perceptions of generic drugs". In: *BMC medicine* 13, pp. 1–27.

Dusetzina, Stacie B, Juliette Cubanski, Leonce Nshuti, Sarah True, Jack Hoadley, Drew Roberts, and Tricia Neuman (2020). "Medicare Part D Plans Rarely Cover Brand-Name Drugs When Generics Are Available: Study examines Medicare Part D coverage to quantify how often brand-name drugs are designated preferred over generics." In: *Health Affairs* 39.8, pp. 1326–1333.

Dusetzina, Stacie B, Shelley Jazowski, Ashley Cole, and Joehl Nguyen (2019). "Sending the wrong price signal: why do some brand-name drugs cost Medicare beneficiaries less than generics?" In: *Health Affairs* 38.7, pp. 1188–1194.

Einav, L., A. Finkelstein, and M. Polyakova (2018). "Private Provision of Social Insurance: Drug-Specific Price Elasticities and Cost Sharing in Medicare Part D". In: *American Economic Journal: Economic Policy* 10.3, pp. 122–53. URL: `https://www.aeaweb.org/articles?id=10.1257/pol.20160355`.

Elhiny, Rehab, Amr S Amin, Hatem A Sarhan, and Mohamed M Abdel-Latif (2021). "Generic prescribing: the Egyptian semblance". In: *Journal of advanced Biomedical and Pharmaceutical Sciences* 4.2, pp. 87–94.

Ellis, R. P., B. Martins, and W. Zhu (2017). "Health care demand elasticities by type of service". In: *Journal of Health Economics* 55, pp. 232–243.

Ford, G. S., W. J. Baumol, and J. G. Sidak (1993). "Toward Competition in Local Telephony". In: *Southern Economic Journal* 62, p. 1109.

Frank, Richard G, Thomas G McGuire, and Ian Nason (2021). "The evolution of supply and demand in markets for generic drugs". In: *The Milbank Quarterly* 99.3, pp. 828–852.

Frech, Harry Edward, Mark V Pauly, William S Comanor, Joseph R Martinez, et al. (2023). *Pharmaceutical pricing and R&D as a global public good*. Tech. rep. National Bureau of Economic Research.

Gallelli, Luca, Caterina Palleria, Antonio De Vuono, Laura Mumoli, Piero Vasapollo, Brunella Piro, and Emilio Russo (2013). "Safety and efficacy of generic drugs with respect to brand formulation". In: *Journal of Pharmacology and Pharmacotherapeutics* 4.1_suppl, S110–S114.

Gatwood, Justin, Teresa B Gibson, Michael E Chernew, Amanda M Farr, Emily Vogtmann, and A Mark Fendrick (2014). "Price elasticity and medication use: cost sharing across multiple clinical conditions". In: *Journal of Managed Care Pharmacy* 20.11, pp. 1102–1107.

Gilman, B. H. and J. Kautter (2008). "Impact of Multitiered Copayments on the Use and Cost of Prescription Drugs among Medicare Beneficiaries". In: *Health Research and Educational Trust* 43.3, pp. 478–495.

Goldman, D. P., G. F. Joyce, J. J. Escarce, J. E. Pace, M. D. Solomon, M. Laouri, P. B. Lansman, and S. M. Teutsch (2004). “Pharmacy Benefits and the Use of Drugs by the Chronically Ill”. In: *JAMA The Journal of the American Medical Association* 291.19, pp. 2344–2350.

Guell, Robert C and Marvin Fischbaum (1995). “Toward allocative efficiency in the prescription drug industry”. In: *The Milbank Quarterly*, pp. 213–230.

— (1997). “Estimating allocative inefficiency in the prescription drug industry”. In: *Applied Economics Letters* 4.7, pp. 419–423.

Guru, Sakshi, Nitin Padole, Pankaj Dhapke, Nilakshi Dhoble, and Jagdish Baheti (2025). “A Review: A Comparative Study of Branded and Generic Antibiotics”. In: *Asian Journal of Research in Pharmaceutical Sciences* 15.1, pp. 99–104.

Halme, Merja, Kari Linden, and Kimmo Kääriä (2009). “Patients’ preferences for generic and branded over-the-counter medicines”. In: *The Patient: Patient-Centered Outcomes Research* 2.4, pp. 243–255.

Hamada, Yuhei, Masashi Uchida, Sayaka Arai, Kaori Yamazaki, Mariko Takeda, Kenichi Arai, Takako Nakamura, Takaaki Suzuki, and Itsuko Ishii (2020). “Analysis of patients’ request to switch from a generic drug to the original drug in external prescriptions”. In: *Journal of pharmaceutical health care and sciences* 6, pp. 1–9.

Hanoch, Giora (1975). “Production and demand models with direct or indirect implicit additivity”. In: *Econometrica: Journal of the Econometric Society*, pp. 395–419.

Harris, Rebecca Arden (2022). “Price elasticity of demand for new and recurring opioid prescriptions”. In: *medRxiv*.

Hernández-Izquierdo, Cristina, Beatriz González López-Valcárcel, Stephen Morris, Mariya Melnychuk, and Ignacio Abásolo Alessón (2019). “The effect of a change in co-payment on prescription drug demand in a National Health System: The case of 15 drug families by price elasticity of demand”. In: *PloS one* 14.3, e0213403.

Höffler, Felix (2006). “Monopoly prices versus Ramsey-Boiteux prices: are they “Similar”, and: does it matter?” In: *Journal of Industry, Competition and Trade* 6, pp. 27–43.

Hooley, Brady, Doris Osei Afriyie, Günther Fink, and Fabrizio Tediosi (2022). “Health insurance coverage in low-income and middle-income countries: progress made to date and related changes in private and public health expenditure”. In: *BMJ Global Health* 7.5, e008722.

Huber, Martin (2021). *Causal Analysis*.

Jack, W. and J. O. Lanjouw (2005). “Financing Pharmaceutical Innovation: How Much Should Poor Countries Contribute?” In: *The World Bank Economic Review* 19.1, pp. 45–67.

Jarusriwanna, Atthakorn, Saran Malisorn, Sirikarn Tananoo, Kwanchanok Areewong, Supachok Rasamimongkol, and Artit Laoruengthana (2024). “Efficacy and safety of generic alendronate for osteoporosis treatment”. In: *Orthopedic Research and Reviews*, pp. 85–91.

Kalo, Z., L. Annemans, and L. P. Garrison (2013). “Differential pricing of new pharmaceuticals in lower income European countries”. In: *Expert Review of Pharmacoeconomics & Outcomes Research* 13.6. PMID: 24219049, pp. 735–741. DOI: 10.1586/14737167.2013.847367. URL: https://doi.org/10.1586/14737167.2013.847367.

Laffont, Jean-Jacques and Jean Tirole (2001). *Competition in telecommunications*. MIT press.

Landsman, P. B., W. Yu, X. Liu, S. M. Teutsch, and M. L. Berger (2005). “Impact of 3-tier pharmacy benefit design and increased consumer cost-sharing on drug utilization”. In: *The American Journal of Managed Care* 11.10, pp. 621–628.

Li, S. and D. Chollet (2006). “Price and Income Elasticity of the Demand for Health Insurance and Health Care Services:A Critical Review of the Literature”. In: URL: https://www.mathematica.org/

publications/price-and-income-elasticity-of-the-demand-for-health-insurance-and-health-care-services.

Lichtenberg, F. (2011). “Pharmaceutical Companies’ Variation Of Drug Prices Within And Among Countries Can Improve Long-Term Social Well-Being”. In: *Health Affairs* 30.8, pp. 1539–1544.

Liu, Keli, Jelena Markovic, and Robert Tibshirani (2018). “More powerful post-selection inference, with application to the lasso”. In: *arXiv preprint arXiv:1801.09037*.

Lőrinczy, Markéta (2013). “Impact of the crisis on the pharmaceutical market in the Czech Republic and Hungary”. In: *Forum Scientiae Oeconomia*. Vol. 1. 2, pp. 21–33.

Mackintosh, Maureen, Geoffrey Banda, Paula Tibandebage, and Watu Wamae (2015). *Making medicines in Africa: The political economy of industrializing for local health*. Springer Nature.

Maleug, D. A. and M. Schwartz (1994). “Parallel imports, demand dispersion, and international price discrimination”. In: *Journal of International Economics* 34, pp. 167–195.

Manning, W. G., J. P. Newhouse, N. Duan, E. B. Keeler, and Ar. Leibowitz (1987). “Health Insurance and the Demand for Health Care”. In: *American Economic Review* 77.3, pp. 251–277.

Martin-Cejas, R. R. (1997). “Airport pricing systems in europe and an application of Ramsey pricing to Spanish airports”. In: *Transportation Research Part E: Logistics and Transportation Review* 33.4, pp. 321–327. ISSN: 1366-5545. DOI: https://doi.org/10.1016/S1366-5545(97)00032-X.

Martínez Giralt, Xavier and Pedro Pita Barros (2006). “On insurance and the cost-sharing of pharmaceutical R&D”. In.

Maskus, K. E. (2001). “Parallel Imports in Pharmaceuticals Implications for Competition and Prices in Developing Countries”. In.

Melnick, G. A., J. Zwanziger, A. Bamezai, and R. Pattison (1992). “The effects of market structure and bargaining position on hospital prices”. In: *Journal of Health Economics* 11.3, pp. 217–233. ISSN: 0167-6296. DOI: https://doi.org/10.1016/0167-6296(92)90001-H.

Moon, Suerie, Elodie Jambert, Michelle Childs, and Tido von Schoen-Angerer (2011). “A win-win solution?: A critical analysis of tiered pricing to improve access to medicines in developing countries”. In: *Globalization and Health* 7, pp. 1–11.

Morel, C. M., A. McGuire, and E. Mossialos (2011). “The Level Of Income Appears To Have No Consistent Bearing On Pharmaceutical Prices Across Countries”. In: *Health Affairs* 30.8, pp. 1545–1552.

Newey, Whitney K and James L Powell (2003). “Instrumental variable estimation of nonparametric models”. In: *Econometrica* 71.5, pp. 1565–1578.

Osei Afriyie, Doris, Blerina Krasniq, Brady Hooley, Fabrizio Tediosi, and Günther Fink (2022). “Equity in health insurance schemes enrollment in low and middle-income countries: A systematic review and meta-analysis”. In: *International journal for equity in health* 21.1, p. 21.

Outterson, K. (2005). “Pharmaceutical Arbitrage: Balancing Access and Innovation in International Prescription Drug Markets”. In: *Yale Journal of Health Policy, Law & Ethics* 193.

Panchal, Sanjeev Kumar, Bilal Mustafa Khan, and S Ramesh (2012). “Importance of ‘brand loyalty, brand awareness and perceived quality parameters’ in building brand equity in the Indian pharmaceutical industry”. In: *Journal of Medical Marketing* 12.2, pp. 81–92.

Persson, Ulf and Bengt Jönsson (2016). “The end of the international reference pricing system?” In: *Applied health economics and health policy* 14.1, pp. 1–8.

Petrou, P. and S. Vandoros (2016). “Pharmaceutical price comparisons across the European Union and relative affordability in Cyprus”. In: *Health Policy and Technology* 5.4, pp. 350–356.

Premanath, Manjunath and Praveen Kulkarni (2024). “Generic Drugs or Branded Generics, Which One You Prefer to Prescribe?” In: *APIK Journal of Internal Medicine* 12.3, pp. 195–196.

Puig-Junoy, Jaume (2010). "Impact of European pharmaceutical price regulation on generic price competition". In: *Pharmacoeconomics* 28.8, pp. 649–663.
Ramsey, F. P. (1927). "A Contribution to the Theory of Taxation". In: *The Economic Journal* 37.145, pp. 47–61. ISSN: 00130133, 14680297. URL: http://www.jstor.org/stable/2222721.
Redding, Stephen J and David E Weinstein (2019). "Aggregation and the gravity equation". In: *AEA Papers and Proceedings*. Vol. 109. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, pp. 450–455.
Reisinger, M., L. Sauri, and H. Zenger (2019). "Post-TRIPS Options for Access to Patented Medicines in Developing Nations". In: *Journal of Health Economics* 66, pp. 163–179.
Sanchez-Borras, M., C. Nash, P. Abrantes, and A. Lopez-Pita (2010). "Rail access charges and the competitiveness of high speed trains". In: *Transport Policy* 17.2, pp. 102–109. ISSN: 0967-070X. DOI: https://doi.org/10.1016/j.tranpol.2009.12.001.
Schmidt, F., M. Egler, and R. G. Geursen (2001a). "Ramsey Pricing as an Explanation and Justification for Price Differences in the European Pharmaceutical Market". In: *Drugs Made in Germany* 44.3.
Schmidt, Franziska, Martin Egler, and Robert G Geursen (2001b). "Ramsey Pricing as an Explanation and Justification for Price Differences in the European Pharmaceutical Market". In: *Drugs made in Germany* 44.3, pp. 61–66.
Semenova, V., M. Goldman, V. Chernozhukov, and M. Taddy (2021). *Estimation and Inference on Heterogeneous Treatment Effects in High-Dimensional Dynamic Panels*. arXiv: 1712.09988 [stat.ML].
Sherer, F. M. and J. Watal (2002). "Post-TRIPS Options for Access to Patented Medicines in Developing Nations". In: *Journal of International Economic Law* 5.4, pp. 913–939.
Simoens, Steven (2012). "A review of generic medicine pricing in Europe". In: *GaBI Journal* 1.1, pp. 8–12.
Soni, Aparna (2019). "Health insurance, price changes, and the demand for pain relief drugs: evidence from Medicare Part D". In: *Kelley School of Business Research Paper* 19-4.
Steinman, Michael A, Mary-Margaret Chren, and C Seth Landefeld (2007). "What's in a name? Use of brand versus generic drug names in United States outpatient practice". In: *Journal of general internal medicine* 22, pp. 645–648.
Syrgkanis, Vasilis, Greg Lewis, Miruna Oprescu, Maggie Hei, Keith Battocchi, Eleanor Dillon, Jing Pan, Yifeng Wu, Paul Lo, Huigang Chen, et al. (2021). "Causal inference and machine learning in practice with econml and causalml: Industrial use cases at microsoft, tripadvisor, uber". In: *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, pp. 4072–4073.
Taylor, Jonathan and Robert Tibshirani (2018). "Post-selection inference for-penalized likelihood models". In: *Canadian Journal of Statistics* 46.1, pp. 41–61.
Terrizzi, Sabrina and Chad Meyerhoefer (2020). "Estimates of the price elasticity of switching between branded and generic drugs". In: *Contemporary Economic Policy* 38.1, pp. 94–108.
Torres, Iván Moreno, Jaume Puig, Joan-Ramon Borrell-Arqué, et al. (2007). *Generic entry into a regulated pharmaceutical market*. Tech. rep. Department of Economics and Business, Universitat Pompeu Fabra.
Towse, A., A. Cole, and B. Zamora (2018). "The Debate on Indication-Based Pricing in the U.S. and Five Major European Countries". In: *OHE Consulting Report*.
Towse, A., M. Pistollato, J. Mestre-Ferrandiz, Z. Khan, S. Kaura, and L. Garrison (2015). "European Union Pharmaceutical Markets: A Case for Differential Pricing?" In: *International Journal of the Economics of Business* 22.2, pp. 263–275.
Treur, Maarten, Bart Heeg, Hans-Jürgen Möller, Annette Schmeding, and Ben van Hout (2009). "A pharmacoeconomic analysis of patients with schizophrenia switching to generic risperidone involving a possible compliance loss". In: *BMC Health Services Research* 9.1, pp. 1–9.

Van Hees, Suzanne GM, Timothy O'Fallon, Miranda Hofker, Marleen Dekker, Sarah Polack, Lena Morgon Banks, and Ernst JAM Spaan (2019). "Leaving no one behind? Social inclusion of health insurance in low-and middle-income countries: a systematic review". In: *International journal for equity in health* 18, pp. 1–19.

Wagstaff, Adam (2002). "Poverty and health sector inequalities". In: *Bulletin of the world health organization* 80, pp. 97–105.

Waning, B., W. Kaplan, A. C. King, D. A. Lawrence, H. G. Leufkens, and M. P. Fox (2009a). "Global strategies to reduce the price of antiretroviral medicines: Evidence from transactional databases". In: *Bulletin of the World Health Organisation* 87, pp. 520–528.

— (2009b). "Global strategies to reduce the price of antiretroviral medicines: evidence from transactional databases". In: *Bulletin of the World Health Organization* 87.7, pp. 520–528.

Wilson, Robert B (1993). *Nonlinear pricing*. Oxford University Press, USA.

Wouters, Olivier J, Panos G Kanavos, and Martin McKee (2017). "Comparing generic drug markets in Europe and the United States: prices, volumes, and spending". In: *The Milbank Quarterly* 95.3, pp. 554–601.

Wu, Anpeng, Kun Kuang, Ruoxuan Xiong, and Fei Wu (2022). "Instrumental variables in causal inference and machine learning: A survey". In: *arXiv preprint arXiv:2212.05778*.

Yadav, P. (2010a). "Differential pricing for pharmaceuticals". In: *UK Department for International Development*. URL: `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/67672/diff-pcing-pharma.pdf`.

Yadav, Prashant (2010b). "Differential pricing for pharmaceuticals". In: *DFID: London*.

Yeung, Kai, Anirban Basu, Ryan N Hansen, and Sean D Sullivan (2018). "Price elasticities of pharmaceuticals in a value based-formulary setting". In: *Health economics* 27.11, pp. 1788–1804.