

# On the Effect of Parallel Trade on Manufacturers' and Retailers' Profits in the Pharmaceutical Sector

Pierre Dubois\*

Morten Sæthre†

Toulouse School of Economics

Norwegian School of Economics

December 2017‡

## Abstract

Differences in regulated pharmaceutical prices within the European Economic Area create arbitrage opportunities that pharmacy retailers can use through parallel imports. For prescription drugs under patent, such provision decisions affect the sharing of profits among an innovating pharmaceutical company, retailers, and parallel traders. We develop a structural model of demand and supply in which retailers can choose the set of goods to sell to consumers, thus foreclosing the consumers' access to some less-profitable drugs, which allows retailers to bargain and obtain lower wholesale prices with the manufacturer and parallel trader. With detailed transaction data, we identify a demand model with unobserved choice sets using supply-side conditions for optimal assortment decisions of pharmacies. Estimating our model, we find that retailer incentives play a significant role in fostering parallel trade penetration. Our counterfactual simulations show that parallel imports of drugs allows retailers to gain profits at the expense of the manufacturer, whereas parallel traders also gain but earn more modest profits. Finally, a policy preventing pharmacies from foreclosing the manufacturer's product is demonstrated to partially shift profits from pharmacists to both the parallel trader and the manufacturer, and a reduction in the regulated retail price favors the manufacturer even more.

JEL codes: I11, L22

Key words: Parallel trade, pharmaceuticals, vertical contracts, demand estimation, foreclosure.

---

\*Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France. e-mail: pierre.dubois@tse-fr.eu

†Norwegian School of Economics, Bergen, Norway. e-mail: morten.saethre@nhh.no

‡We are grateful to Matthew Backus, Christian Bontemps, Kurt Brekke, Estelle Cantillon, Liran Einav, Kate Ho, J.F. Houde, Bruno Jullien, Robin Lee, Charles Murry, Aviv Nevo, Peter Newbery, Volker Nocke, Ariel Pakes, Patrick Rey, John Rust, Tobias Salz, Nicolas Schutz, Howard Smith, and Amanda Starc for useful comments and to many participants in seminars in Bergen, Bocconi, ECARES Brussels, Columbia, Cornell, CREST Paris, Harvard/MIT, Helsinki, Mannheim, McGill, Northwestern, Paris, Penn State, Royal Holloway, Toulouse, Toronto, Wharton, Yale (Cowles) and at many conferences.

# 1 Introduction

Within the European Economic Area, arbitrage trade of pharmaceutical drugs across countries is fully legal. Cross-country differences in national price regulations have maintained substantive differences in prices. This has led to an increase in parallel trade<sup>1</sup>, estimated at 5.5 billion euros in 2012, with highly heterogeneous national market shares that can be up to 25% in some countries. The cross-country price differences can be as large as 300% and are due to regulatory caps or strict government rules in price setting. Differences in price regulation depend on the aggressiveness of each member state's authorities in negotiating with manufacturers (Kyle, 2007). Not surprisingly, these price differences result in parallel imports of pharmaceuticals by high-price countries from low-price countries. Drugs are often bought in Southern European countries, such as Greece, Portugal and Spain, and resold in Northern European countries, such as the UK, Netherlands, Norway and Sweden (Kyle, 2011). At the same time, there are large variations in parallel import penetration across otherwise similar countries. In many European countries, consumers are covered by national health insurance and will often face substantially lower costs than the full price. Combined with price regulation, this often leads to very small or no price differences between direct and parallel imported drugs. There are large cross-country differences in the share of parallel import sales, which according to the findings of Kanavos et al. (2004) and Kanavos and Vandoros (2010) seem to have a clear link to regulation governing margins at the pharmacy and domestic supply level. A prime example is Germany, where pharmacies are subject to regulations fixing their margins and, to a lesser extent, supplied parallel trade prior to the national insurance authorities introducing a minimum quota of 5%. Meanwhile, British pharmacies have no direct caps on their margins and have large shares of parallel imports (Kanavos et al., 2004). Even though price studies have not found evidence of any significant reduction in price dispersion across EU countries (Kanavos et al., 2004; Kyle, 2011), Ganslandt and Maskus (2004) report that parallel imports might have led to a reduction in drug prices on the order of 12–19% for drug segments subject to parallel imports entry in Sweden. As the entry of parallel traders needs to be performed through pharmaceutical retailers, retailers' incentives are potentially decisive in determining the extent of parallel trade due to retailers' role as intermediaries in the supply chain. The strategic role of profit-maximizing pharmacies, both towards drug manufacturers providing directly imported drugs<sup>2</sup> and parallel traders providing parallel imports, can thus be important in the organization of the pharmaceutical sector. The parallel import version of a drug directly sold in a given country is the version of that drug marketed in other countries by the same pharmaceutical company. Direct and parallel

---

<sup>1</sup>These operations are executed by firms specializing in parallel trade and require necessary logistical capacity and facilities suitable for repackaging of drugs. Repackaging is required for drugs for which the imported package and accompanying information sheets is in another language than the language of the destination country.

<sup>2</sup>Directly imported drugs are drugs supplied by the manufacturers or their marketing agencies.

imports are potentially differentiated (by appearance and packaging) from the consumer point of view but are essentially the same products and likely to be highly substitutable. Parallel imports create an upstream provision alternative to the manufacturer for pharmacists in their the wholesale negotiations. This may have significant implications for the distribution of surplus in the market. In the case of prescription drugs under patent, the monopoly rights of the manufacturer are supposed to give them the possibility to extract the willingness to pay of consumers when setting prices either directly to the market or when negotiating with governments the consumer price level. Past research has shown that innovation is indeed elastic to such reward (Acemoglu and Linn (2004), Dubois et al. (2015)). However, if intermediaries such as pharmacies or parallel traders manage to extract a large share of the monopoly rent of manufacturers, the innovation incentive may be inefficiently reduced. It is thus important to study how the organization of retailing and parallel trade affects profit sharing.

**Contribution** To study the sales of parallel imported pharmaceutical drugs, we develop a structural model of demand and supply with intermediaries such as pharmacists. Specifically, we address the question of how incentives of retail pharmacies facilitate sales of parallel imports amid regulatory policies towards parallel trade and price setting of pharmaceuticals. Our model can explain how parallel imports can capture substantial market shares, even though the savings afforded to consumers might be negligible or even non-existent. The mechanism of our model is that a retailer selling at regulated prices might wish to restrict the supply of less-profitable products to increase purchases of more-profitable products. This might reduce the total demand for the retailer, as restrictions in the choice set decrease the expected utility of visiting the retailer, particularly when (some) consumers have a preference for a restricted product. The problem of the retailer is thus a trade-off between foreclosing access to lower margins products and staying attractive to consumers. There is reason to suspect that some consumers would prefer the direct import variety, even though parallel and direct imports are the same drugs produced by the same company. Indeed, pharmacists must inform patients when a drug is parallel imported. The packaging will usually display the brand name of the parallel importer, and the product can differ in visual appearance and inactive ingredients. An example is tablets for which the directly imported variety is round and white, whereas the parallel import comes from a country in which they are octagonal and red. This type of differentiation in appearance and specification across countries has been linked to attempts to reduce the scope for parallel trade (Kyle, 2009, 2011). In this sense, as we would suspect consumers to be either indifferent between direct and parallel import or skeptical about parallel import and because the prices they pay are usually the same, it seems necessary to consider the incentives of the retail side of the market to explain the penetration of parallel imports.

We estimate our model using a very rich data set regarding the Norwegian pharmaceutical market, for which we are able to observe detailed demand data and also pharmacy margins. In particular, we observe all purchases by individual consumers over time, the pharmacy chain at which a given purchase happened, and whether the specific drug dispensed was imported through the original manufacturer (direct import) or by parallel traders. We also have data regarding the retail price the pharmacy charges for all dispensed sales, in addition to data about the wholesale prices paid by the pharmacy chain to the upstream firms for each specific drug package. Thus, we observe the gross margin obtained by the chain on all products, which affects retailers' incentives to dispense parallel imports.

As the choice set of consumers changes across pharmacies and is not observable to the econometrician, we develop an estimation method based on observed transactions with unobserved choice sets. This method nests the Nash equilibrium in pharmacies' strategic choice sets within the construction of choice probabilities. Maintaining the assumption of existence of a Nash equilibrium between the pharmacy chains, we can identify the demand model due to exogenous variation in pharmacy margin for parallel and direct imports that leads to varying choice sets in equilibrium. Our nested fixed-point algorithm could be applied to other settings in which retailer incentives to propose an assortment of products can be characterized by an equilibrium condition. We find that inclusion of retailer incentives in our model plays an important role in explaining consumer choices. Then, we model the wholesale price setting between pharmacy chains and the manufacturer or the parallel traders as a simultaneous Nash bargaining problem. We identify the bargaining weights of each party using the Nash-in-Nash equilibrium equations for wholesale price determinations and exogenous shocks on prices in source countries together with exchange rate shocks that affect the opportunity value of parallel imports versus direct imports. In the case of the main statin market during the period of 2004-2007, our GMM estimation results for bargaining parameters show that the manufacturer has higher bargaining power than parallel trader with respect to pharmacy chains but that pharmacy chains enjoy significant bargaining ability. We then use the estimated bargaining model to simulate three counterfactual situations related to *i*) the possibility for pharmacy chains to use parallel imports, *ii*) their ability to use foreclosure strategies and *iii*) the level of the retail price cap imposed by the government.

Our counterfactual simulations imply that even though, on average, consumers prefer directly imported products, parallel imports allow the retail pharmacy chains to capture a much larger share of industry profits than would otherwise be the case, particularly at the expense of the manufacturer. On the Atorvastatin market (patented and marketed by Pfizer under trade name Lipitor during 2004-2007), the manufacturer's profit would increase by at least 24% if there was no parallel trade, whereas pharmacy chains would lose between 21% and 93%, depending on the chain. The shift in profits to retailers is driven by two mechanisms:

*i*) the creation of price competition between the upstream firms from chains' ability to shift sales as a response to differences in profitability and *ii*) the outside option conferred to a chain from being able to sell parallel imported drugs when bargaining over wholesale prices with the direct importer. We also perform a counterfactual simulation in which pharmacies are forced to always propose both versions of drugs to customers. In this case, the banning of foreclosure allows the manufacturer to gain and makes the pharmacies' profits go down, though not by a large amount. This result shows that differentiation of drugs and the concentration of pharmacy chains with the possibility to purchase parallel imports at much cheaper price still allows them to capture a large part of profits. Finally, we perform counterfactuals in which the retail price cap is lowered by 20%. The results demonstrate that most of the reduction is borne by pharmacy chains and also parallel traders because the margin of negotiation is much reduced by the lower difference between prices in source countries and the maximum allowed retail price. The manufacturer loses very little profit, whereas the total government expenses associated with this market are reduced by 20%.

**Related literature** Many industries rely on a downstream retailing sector to market goods. Not only do vertical relationships do affect price competition among substitutes and differentiated goods, but retailers – as intermediaries between manufacturers and consumers – can also affect competition by engaging in other strategic actions affecting final consumers' demand. Such strategic actions can include choices regarding the assortment of goods (Draganska et al. (2009)). The equilibrium results of such structures can be analyzed using game theoretic models and estimated through structural models. Typical sectors in which retailers' behavior have attracted attention of economists are Internet platforms and the food retailing industry with large supermarket chains. Pharmacy retailing has been less studied, although the growth in health care expenses among developed countries raises questions about how to design policies to contain spending on pharmaceutical drugs while ensuring or improving access to innovation for patients. In Europe, most countries regulate the prices of prescription drugs, although other aspects of competitive behavior, such as strategic choice of entry across different markets, matter substantially (Danzon and Chao, 2000; Danzon et al., 2005; Maini and Pammolli, 2017). How pharmacists choose the assortment of drugs, proposing a parallel import, direct import or both, is similar to strategically choosing to stock out or foreclose access to some versions of drugs. Previous literature has provided reduced-form evidence for this type of response to markup differences in prescription drug markets. In a simpler setting in which physicians can prescribe and dispense drugs, Izuka (2013) shows that Japanese physicians respond to markup differentials between originator and generics. In the Norwegian market for off-patent drugs, Brekke and Straume (2013) find a strong relationship between market share and differences in pharmacy margins for branded and generic drugs. Crawford et al. (2017b)

show similar foreclosure strategies in distribution of TV channels. Such a strategy can also be profitable in other industries, but it should especially be the case in tightly regulated markets in which price setting is constrained, as is common in many European countries for pharmaceuticals.<sup>3</sup> Our demand estimation with unobserved choice sets is also related to the literature regarding consideration sets or unobserved stock outs. In a seminal paper, Goeree (2008) uses advertising to identify the likely variation in consideration sets using aggregate demand data. Crawford et al. (2017a) use sufficient statistics on consideration sets to estimate a discrete choice model with unobserved choice sets using individual-level transaction data. We use the retail pharmacists' incentives to manipulate choice sets to identify our demand model. In a different context, Gaynor et al. (2016) estimate a demand model that explicitly captures choice constraints imposed on patients by physicians. Our identification relies on the observation of individual choices and modeling pharmacists' strategic choices. Conlon and Mortimer (2013) uses the fact that they observe periodical stock-outs of products in vending machines to estimate a demand model with varying choice sets. Our supply-side vertical relationship model is related to the empirical IO literature using Nash-in-Nash bargaining equilibrium. Grennan (2013) uses a model of bargaining on prices of medical devices between hospitals and upstream suppliers. Gowrisankaran et al. (2015) model bargaining between managed care organization and hospitals in the US. Ho and Lee (2017) also use bargaining to model the negotiated provider prices. Finally, some literature has studied the impact of parallel trade on pharmaceuticals in Europe. Using a structural model of demand estimated with data on the German market for oral anti-diabetic drugs, Duso et al. (2014) evaluate the welfare impact of parallel imports. Their estimates imply that parallel imports have reduced the prices of on-patent drugs by 11% but that the impact on consumer surplus is modest. The effect of parallel imports on drugs prices therefore depends crucially on country specific regulation of the pharmacies. In contrast to the approach of Duso et al. (2014), we explicitly model both the vertical relationship between manufacturers and pharmacy retail chains and the strategic role of retailer incentives in the development of parallel imports market shares. Using data from Norway, Brekke and Straume (2015) study the interaction between price cap regulation and parallel imports across a large number of drugs. They find reduced-form evidence that original manufacturers might benefit from lower price ceilings when there is competition from parallel trade. They also use a Nash bargaining model of contracting between pharmacy chains and upstream manufacturers to motivate their empirical analysis, and, although the channels they highlight are slightly different from ours, we also find a similar effect in our last counterfactual on a single, large drug. Novel features of our paper include the strategic decisions by the retailers regarding the drugs proposed to consumers, the structural estimation of the bargaining model and the analysis of counterfactual policies and incentive configurations.

---

<sup>3</sup>For details about pharmaceutical market regulation in different countries, see, e.g., Kanavos et al. (2008).

**Structure of the paper** In Section 2, we present the market and data. We present the structural model of demand and supply in section 3. In Section 4, we describe the empirical specification and identification of our model and present the estimation results. In Section 5, we present the results from our counterfactual simulations, while Section 6 concludes.

## 2 The Norwegian Pharmaceutical Market and Parallel Imports

### 2.1 Overview and Regulation

The supply side of the market for prescription drugs consists mainly of three large pharmacy retail chains, which are vertically integrated with each of their upstream wholesalers. The three largest chains, Apotek 1, Boots and Vitus, cover 85 % of all pharmacies, and public hospital pharmacies (6 %), a smaller retail chain (5 %), and independent pharmacies (4 %) comprise the rest.

The Norwegian market for drugs is subject to a wide array of regulations. The Norwegian Medicines Agency, a governmental organization under the Ministry of Health and Care Services, is the main regulatory body for drug affairs, in charge of marketing authorization, drug classification, vigilance, price regulation, reimbursement regulation, and providing information about drugs to both prescribers and the public.

With the exception of over-the-counter drugs, all drugs sold on the Norwegian market are subject to a price cap, which is set by the Norwegian Medicines Agency. As a general rule, this price cap is set as the average of the three lowest among market prices in a fixed group of European comparison countries, consisting of Sweden, Finland, Denmark, Germany, the United Kingdom, the Netherlands, Austria, Belgium and Ireland. The price caps should normally not change more than once every year. Reconsideration of the price caps is initiated by the Norwegian Medicines Agency, and selection is based on sales volume over the past 12 months. The price caps are set according to the active ingredient in the drug and amount of active ingredient (dosage). Per unit price caps (with the unit defined by Defined Daily Dose (DDD) for drugs for which this quantity is defined) should generally be equal within the category of a given dosage for a given active ingredient, although the Norwegian Medicines Agency requires large package sizes to have a lower per-unit price in some cases.

In cases in which the patient has a long-term ailment, defined as requiring treatment for at least three months, and the drug under question has been judged to have sufficient effect compared to the costs, government reimbursement is available. The prescribing physician is responsible for deciding whether the patient satisfies the criteria for treatment length, whereas the Norwegian Medicines Agency determines if a drug satisfies the cost-efficiency criteria for reimbursement. When patients get reimbursed, they face a co-payment

of 38 % of the total price, capped at 520 NOK in 2013 (approximately 65 EUR) per three months. The co-payments for drugs and health care spending are capped at 2040 NOK yearly in 2013 (approximately 260 EUR). For drugs that are subject to patent, the government will reimburse the full cost of the drug to the patient, net of co-payments. When the drug is off-patent and generic drugs have entered the market, the reimbursement rate will generally be reduced below the price ceiling, but there are almost no parallel imports of off-patent drugs.

## 2.2 Parallel Trade

Parallel traders have to obtain a license for selling drugs in Norway from the Norwegian Medicines Agency, unless they already have obtained a license for sales in the European Economic Area through the centralized European Union procedure. Parallel traders sell to one or more of the three large wholesalers, as only full-line wholesalers are allowed by law to supply pharmacies with drugs.<sup>4</sup> A license will be for a specific drug package imported from a specific country, with the exception of licenses granted through the European Union procedure. In Figure 2.1, we see the number of licenses granted by the Norwegian Medicines Agency by source country, which is in line with the countries that are usually reported as major source countries for parallel imported drugs.

In our dataset, which contains information about prescription filings at pharmacies in Norway for the period of 2004–2007, we can identify whether each sold product is directly imported or parallel imported. Parallel trade happens most prominently in the under-patent period and makes up a negligible share for drugs where generics are present. The average share of DDD of parallel import in the ATC codes (Anatomical Therapeutic Chemical classification system) after generic entry is 3%, whereas it is roughly 27% among all ATC codes and periods without generics<sup>5</sup>. All of the following analysis is performed on a subsample of the 50 most important active ingredients for which parallel trade occurs in our sample period, thus excluding active ingredients for which parallel imported products obtain a very low share of sales in their segment. Monetary units are reported in nominal NOK ( $\approx$  0.12 EUR / 0.16 USD in the period) unless noted otherwise.

Figure 2.2 shows the parallel import share of sales within each pharmacy chain. It is interesting to note that there is large variation both between chains and over time. When analyzing the retail prices of parallel imported and directly imported versions in each chain, it appears that the price ceiling is binding for both

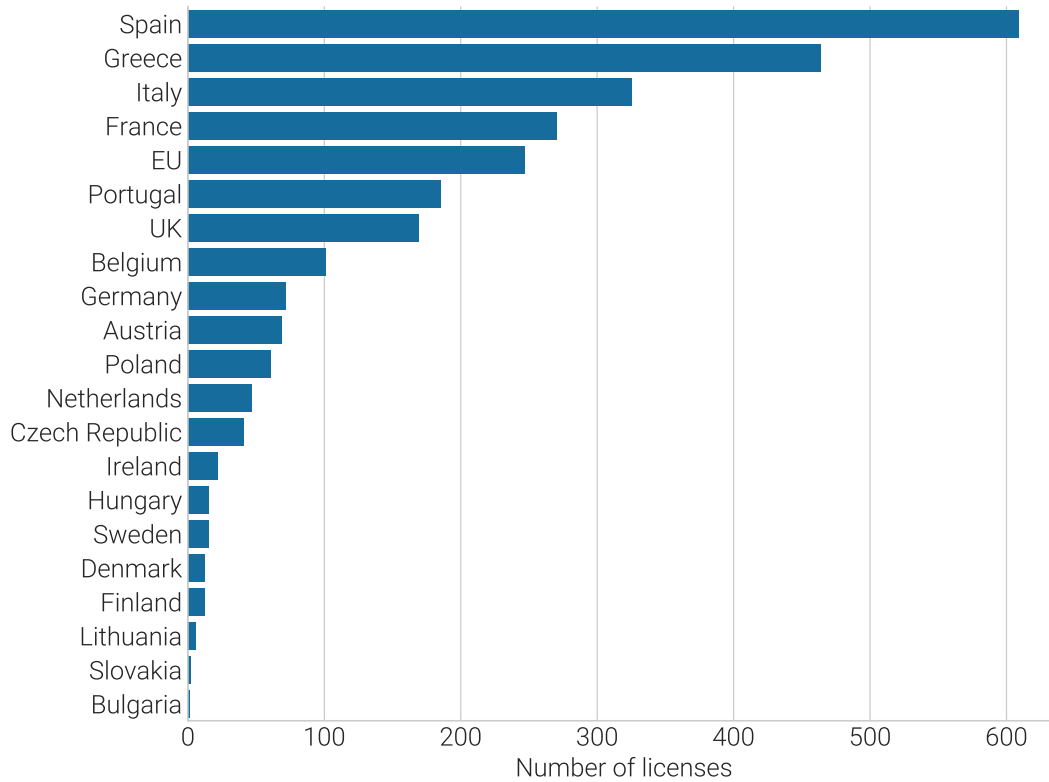
---

<sup>4</sup>Whether this matters is an open question, as the market is almost fully vertically integrated at the wholesaler-pharmacy level, although the full-line supply regulation could be an explanation for the concentration and vertical integration observed in the market.

<sup>5</sup>In the sample, 11 out of 530 unique parallel imported products are generics, featured in 4 out of 109 ATC codes with parallel import. They are responsible for less than 1% of the observations of parallel imports and about 0.15% of parallel import DDD. Parallel import of generics is thus a minor phenomena in the Norwegian market in this period.



Figure 2.1: *Granted licenses in Norway for parallel import by country*



Source: *The Norwegian Medicines Agency.*

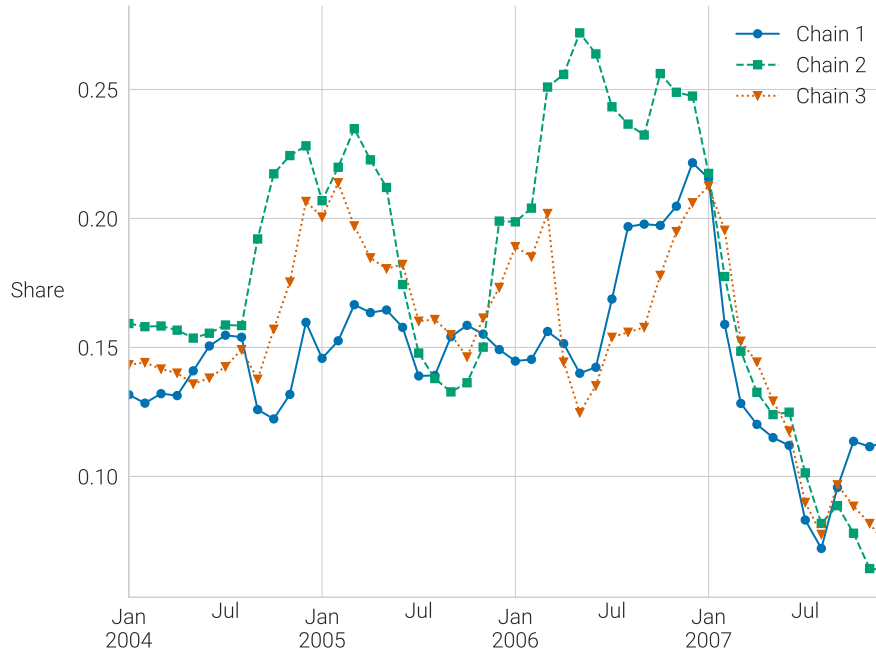
categories for all active ingredients, dosages and package sizes<sup>6</sup>. Thus, there is no retail price difference between parallel and directly imported versions of the same molecule, and the price is equal to the price cap. Note that the price cap is the reimbursement price, up to the common regulated co-payment described above.

We also compare the margin that each pharmacy chain obtains on comparable products, i.e., within categories defined by active ingredient, dosage and package size. Here, the pharmacy chain margin is defined as the sales price in the pharmacy net of the price the pharmacy chain's integrated wholesaler pays to the supplier for obtaining the drug, where the supplier is either a marketing agency of the manufacturer, in the case of direct imports, or the parallel trading firm. These (gross) margin differences are shown in Figure 2.3. The average difference varies between 4% and 16% over the 4 years of data across the 3 chains.

The seeming correlation between margin differences and the parallel import share of sales in Figure 2.2 is confirmed by a significant chain-month level positive correlation between the parallel import shares and the margin difference between parallel and direct imports. This cannot be given a causal interpretation by

<sup>6</sup>The package size is defined as DDD per package, which for tablets with the same active ingredient and dosage would be equivalent to tablets per package.

Figure 2.2: *Parallel import share of sales in DDD by chain*



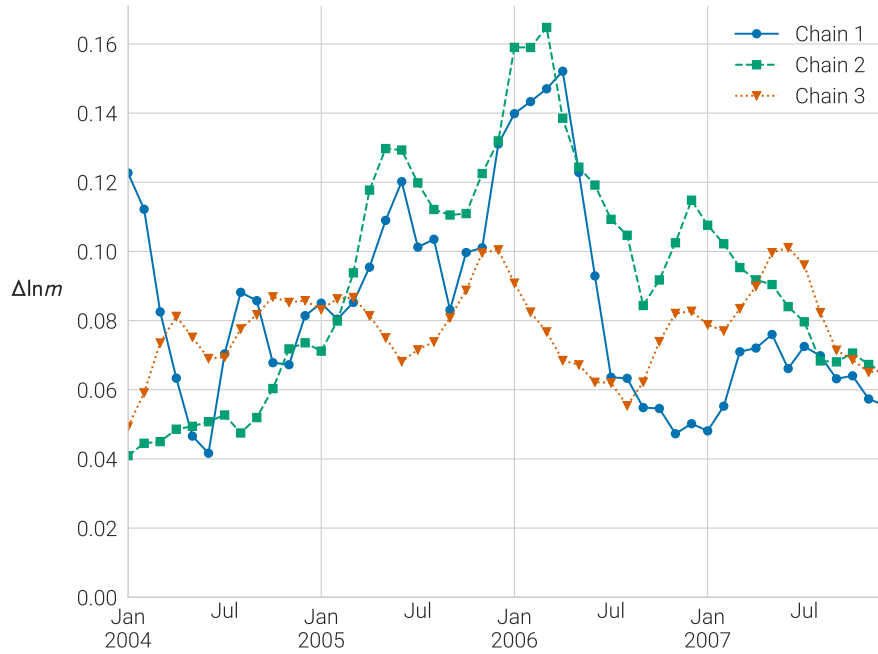
Note: *Only molecules featuring sales of parallel imports over the sample period are included.*

itself, but it is a first indication of pharmacy incentives mattering for the composition of drugs dispensed to consumers.

The correlation is, however, far from perfect and tells us that the margin difference is not the only factor driving the variation that we see in the evolution of parallel import sales. We would expect there to be strategic interaction between the importing firms and the pharmacy chains, which will make the simple estimate biased if the benchmark is a causal estimate of the effect of margin difference on product sales and uninformative when used as input for a model evaluating the market in terms of market power and policy evaluation.

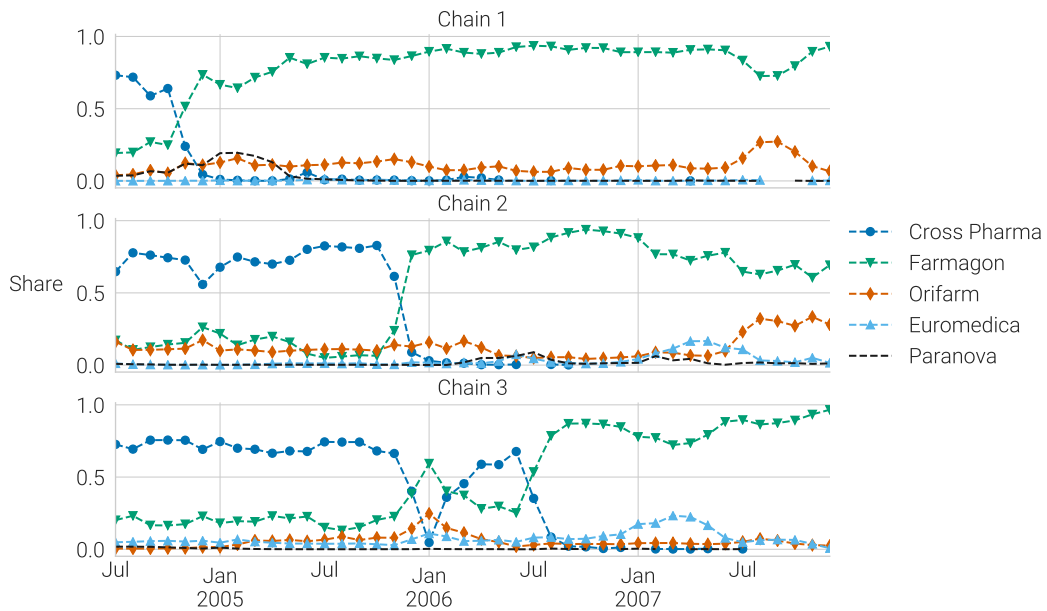
In the Norwegian market in this period, there are five companies specializing in parallel trade with any noticeable activity, namely Cross Pharma, Euromedica, Farmagon, Orifarm and Paranova. The share of parallel import sales within each pharmacy chain for each of these companies is shown in Figure 2.4. There is some variation both between pharmacies and over time in terms of the relative presence for these companies. Considering the active ingredient level, each pharmacy chain deals with one parallel importer at a given time, although the identity of the parallel importer varies across chains for the same drug and also changes over time for a drug within a chain (not shown here).

Figure 2.3: *Difference in product margin between direct and parallel imports*



Note: Margin differences are reported in NOK per DDD and correspond to the margin of parallel import minus the margin of direct import. Differences are calculated for packages that are of the same ATC code, with the same amount of active ingredient and of comparable size.

Figure 2.4: *Composition of parallel importers*



Note: Share of parallel import sales in DDD within pharmacy chain for each parallel importer.

### 3 A Structural Model of Demand and the Supply Chain

To explain the variation in the data, we aim at creating an estimable model that can be used for counterfactuals in terms of policy implications and cross-country differences.

#### 3.1 Consumer Behavior and Demand for Parallel Trade Products

We assume that the consumer has an exogenous need for a drug with a particular active ingredient and dosage. Thus, we abstract from the issue of therapeutic choice decided by prescribers, which, as we show below, does not seem to be affected by the availability of parallel traded versions of active ingredients or pharmacy margin differences and can thus be considered exogenous to the fundamental mechanisms of our model.

The consumer chooses which pharmacy chain  $c$  to visit and – once in the pharmacy – makes a choice among the available products in the pharmacy. When the consumer chooses a pharmacy  $c$ , he does not know if parallel imported (PI) or directly imported (DI) versions of the drug will be available, although we assume that the consumer is aware of the expected availability. Because pharmacies potentially have higher margins on drugs that the consumer do not strictly prefer, it may be optimal not to propose the lower-margin drug with certainty to induce consumers to buy the other option. However, as described previously, the parallel imported drug can be noticeably different from the directly imported variety.<sup>7</sup> It may thus be optimal to propose consumers' preferred drug with a non-zero probability to attract them. This phenomenon is confirmed by casual observation, and the fact that pharmacists do consider this policy of non-permanent availability is acknowledged in discussions with pharmacists.<sup>8</sup> We thus assume that consumers know the probabilities of availability chosen by the pharmacy chains.<sup>9</sup>

For a given active ingredient, the choice set at pharmacies can be  $\{PI\}$ ,  $\{DI\}$  or  $B \equiv \{DI, PI\}$ . We let the origin of the drug be indexed by  $k \in \{0, 1\}$  where 0 denotes  $PI$  and 1 denotes  $DI$ .

We denote by  $\theta_{ct}^0$  and  $\theta_{ct}^1$  the probabilities that the choice sets are  $\{PI\}$  or  $\{DI\}$ , respectively, and thus  $1 - \theta_{ct}^0 - \theta_{ct}^1$  is the probability that the choice set is  $B = \{DI, PI\}$ . We assume that the utility of consumer  $i$

---

<sup>7</sup>In addition, pharmacists are required to inform the consumer that the drug is parallel imported.

<sup>8</sup>According to an industry source with whom we spoke, most customers do not object to substitution to parallel trade, although some consumers are concerned and can even insist on obtaining the directly imported version. According to our source, the chain sets the standard policy for deliveries of stocks to the pharmacies, which the pharmacist can then alter or make additional orders for non-standard selections ex post. The experience was that for several drugs, the standard policy included few or none of the directly imported versions, such that the pharmacy easily could end up only having the parallel traded version available. Since obtaining an additional order would take at least one day, there was a worry that customers insisting on the directly imported version would rather go to a competing pharmacy.

<sup>9</sup>In Appendix 7.2, we present an alternative demand model in which consumers know ex ante the products available at each pharmacy chain and choose the product to purchase and the pharmacy to visit from amongst the different pharmacies according to their preference for each version and for each pharmacy chain. The methodology and results of the structural estimation are then similar.

is given by

$$u_{ikct} = V_{ikct} + \varepsilon_{ict} + \lambda_c \epsilon_{ikct}$$

where  $V_{ikct}$  is the mean utility consumer  $i$  obtains from choosing the drug of origin  $k$  in pharmacy chain  $c$  in market  $t$  and  $\varepsilon_{ict}$  and  $\epsilon_{ikct}$  are chain-specific and product-specific sequentially observed shocks, respectively. We assume that they are distributed independently across drugs and chains according to a Gumbel distribution. Our choice model is, however, not a nested logit but rather a model with two extreme value distributed shocks observed sequentially by the decision maker<sup>10</sup>.

Thus, within the pharmacy, the probability that consumer  $i$  chooses  $k \in \{0, 1\}$  conditional on choice of pharmacy chain  $c$  when both products are available in the pharmacy is given by

$$s_{ikt|c,B} = \frac{e^{V_{ikct}/\lambda_c}}{e^{V_{i0ct}/\lambda_c} + e^{V_{i1ct}/\lambda_c}} = \frac{1}{1 + e^{V_{ikct}/\lambda_c - V_{ik'ct}/\lambda_c}} \text{ with } k' = 1 - k$$

because  $\epsilon_{ikct}$  is i.i.d. extreme value distributed.

Assuming that the consumer always prefers any available drug to no drug, the choice probability of product  $k$  conditional on the choice of pharmacy  $c$  is

$$\underbrace{s_{ikt|c}}_{\substack{\text{choice probability of } k \\ \text{conditional on} \\ \text{going to chain } c}} = \underbrace{\theta_{ct}^k}_{\substack{\text{probability that} \\ \text{only } k \text{ is} \\ \text{available at } c}} + \underbrace{(1 - \theta_{ct}^0 - \theta_{ct}^1)}_{\substack{\text{probability that} \\ \text{both versions} \\ \text{are available at } c}} \underbrace{s_{ikt|c,B}}_{\substack{\text{choice probability of } k \\ \text{given both versions} \\ \text{are available at } c}},$$

that is, the probability of drug  $k$  being the only available plus the probability that drug  $k$  is chosen when both are available times the probability that both are available.

We assume that the consumer's shock  $\epsilon_{ikct}$  is only known within the chosen pharmacy chain, such that the consumer chooses a chain using the expected utility of choosing a pharmacy by taking expectations with respect to the possible choice sets and with respect to the shock  $\epsilon_{ikct}$ . The consumer utility of visiting pharmacy  $c$  is then  $I_{ict} + \varepsilon_{ict}$ , where

$$I_{ict} \equiv \sum_{k \in \{0,1\}} \underbrace{\theta_{ct}^k}_{\text{prob. only } k \text{ available}} \underbrace{V_{ikct}}_{\text{utility of } k} + \underbrace{(1 - \theta_{ct}^0 - \theta_{ct}^1)}_{\text{prob. both versions}} \underbrace{E_{\epsilon_{ikct}} \left[ \max_{k \in \{0,1\}} (V_{ikct} + \lambda_c \epsilon_{ikct}) \right]}_{\text{expected utility preferred version}}$$

<sup>10</sup>Cardell (1997) shows that there exists a distribution of random variable  $\varepsilon_{ict}$  such that with  $\epsilon_{ikct}$  extreme value i.i.d., the random variable  $\varepsilon_{ict} + \lambda_c \epsilon_{ikct}$  is also an extreme value and gives rise to a nested logit model.

with the log-sum formula for the inclusive value in case the choice set contains both products<sup>11</sup>:

$$E_{\epsilon_{ikct}} \left[ \max_{k \in \{0,1\}} (V_{ikct} + \lambda_c \epsilon_{ikct}) \right] = \lambda_c \ln \left( \sum_{k \in \{0,1\}} e^{V_{ikct}/\lambda_c} \right)$$

which is always greater than  $\max(V_{i0ct}, V_{i1ct})$ .

Then, as  $\epsilon_{ikct}$  is extreme value Gumbel-distributed independently across chains, patient  $i$  chooses chain  $c$  with probability

$$s_{ict} = \frac{e^{I_{ict}}}{\sum_{\bar{c}} e^{I_{i\bar{c}t}}}.$$

Denoting by  $F(\cdot|\beta)$  the cumulative distribution function of consumer preferences  $\mathbf{V}_{it} \equiv (V_{i01t}, \dots, V_{i0Ct}, V_{i11t}, \dots, V_{i1Ct})$  conditional on the parameter vector  $\beta$ , we can write the aggregate choice probability or market share of drug  $k$  sold by  $c$  in period  $t$  as

$$s_{kct} = \int s_{ikct} dF(\mathbf{V}_{it}|\beta) = \int s_{ict} s_{ikt|c} dF(\mathbf{V}_{it}|\beta), \quad (3.1)$$

and the aggregate market share of drug  $k$  within the pharmacy chain  $c$  as

$$s_{kt|c} = \int s_{ikt|c} dF(\mathbf{V}_{it}|\beta) = \theta_{ct}^k + (1 - \theta_{ct}^0 - \theta_{ct}^1) \int s_{ikt|c,B} dF(\mathbf{V}_{it}|\beta).$$

### 3.2 Pharmacy Chains Behavior

Let us now turn to the behavior of the pharmacy chains. The profits of chain  $c$  normalized by market size in time  $t$  are

$$\pi_{ct} = \sum_{k \in \{0,1\}} (p_{kct} - w_{kct}) s_{kct},$$

where  $p_{kct}$  is the retail price and  $w_{kct}$  the wholesale price of drug  $k$  in pharmacy  $c$  at  $t$ . As retail prices are regulated with a price ceiling that applies to both the direct and parallel import versions of a drug, pharmacies can choose the set of products they prefer to sell but cannot fixed prices higher than the price ceiling such that it must be the case that  $p_{kct} \leq \bar{p}_t$ . However, as for almost all under-patent drugs (including the one used in the structural model estimation), the retail prices happen to always be equal to the price ceiling, we treat the price ceiling chosen by the regulator as binding ( $p_{kct} = \bar{p}_t$ ). We show in appendix 7.3 that it may be constrained-optimal for the pharmacy to set both prices of parallel and direct imports at the price ceiling.

---

<sup>11</sup>Note that we omit the means of all Gumbel-distributed random utility terms,  $\epsilon_{ijkct}$ , in the following. It is equal to the Euler-Mascheroni constant for all terms involving expectations of random utility terms and will thus not affect choices.

We also take the wholesale prices as given when considering the optimal choice of  $\theta$ 's and return to their determination when discussing the behavior of manufacturers in the next section. We now denote by  $m_{kct} \equiv \bar{p}_t - w_{kct}$  the product price-cost margin, where  $w_{kct}$  allows wholesale price discrimination across pharmacy chains.

We implicitly assume that both margins are positive, such that pharmacy chains accept both procurement channels. Necessary first-order conditions for an interior solution for the  $\theta$ 's are

$$0 = \frac{\partial \pi_{ct}}{\partial \theta_{ct}^0} = \frac{\partial \pi_{ct}}{\partial \theta_{ct}^1}. \quad (3.2)$$

For  $\theta_{ct}^0$ , this is

$$\begin{aligned} 0 &= \sum_k m_{kct} \frac{\partial s_{kct}}{\partial \theta_{ct}^0} = \sum_k m_{kct} \int \frac{\partial}{\partial \theta_{ct}^0} [s_{ict} s_{ikt|c}] dF(\mathbf{V}_{it}|\boldsymbol{\beta}) \\ &= \int \sum_k m_{kct} \left[ \underbrace{\frac{\partial s_{ikt|c}}{\partial \theta_{ct}^0}}_{\substack{\text{change in} \\ \text{probability to} \\ \text{choose } k \text{ in chain } c}} \underbrace{s_{ict}}_{\substack{\text{probability} \\ \text{to choose} \\ \text{chain } c}} + \underbrace{s_{ikt|c}}_{\substack{\text{probability} \\ \text{to choose } k \\ \text{in chain } c}} \underbrace{\frac{\partial s_{ict}}{\partial \theta_{ct}^0}}_{\substack{\text{change in} \\ \text{probability to} \\ \text{choose chain } c}} \right] dF(\mathbf{V}_{it}|\boldsymbol{\beta}), \end{aligned}$$

which shows that an increase in  $\theta_{ct}^0$  has two effects. The first term shows the increase in profit through higher sales of the more-profitable good 0 at the expense of sales of the less-profitable good 1 because good 0 is more often the only option for the consumer; the second term shows the profit loss from a loss in market share due to chain  $c$ 's less-attractive foreclosure policy of the other good from the consumer's point of view.

As

$$\frac{\partial s_{ikt|c}}{\partial \theta_{ct}^{k'}} = 1_{\{k=k'\}} - s_{ikt|c,B} \quad \text{and} \quad \frac{\partial s_{ict}}{\partial \theta_{ct}^{k'}} = \left[ V_{ik'ct} - \lambda_c \ln \left( \sum_k e^{V_{ikct}/\lambda_c} \right) \right] s_{ict} (1 - s_{ict}) \leq 0,$$

using the fact that

$$\frac{\partial s_{ik't|c}}{\partial \theta_{ct}^0} s_{ict} + s_{ik't|c} \frac{\partial s_{ict}}{\partial \theta_{ct}^0} = (1_{\{k'=0\}} - s_{ik't|c,B}) s_{ict} + s_{ik't|c} \left[ V_{i0ct} - \lambda_c \ln \left( \sum_k e^{V_{ikct}/\lambda_c} \right) \right] (1 - s_{ict}) s_{ict},$$

we obtain that the first-order condition for optimal  $\theta_{ct}^0$  implies

$$\frac{m_{0ct}}{m_{1ct}} = \frac{\int s_{i1t|c,B} s_{ict} + s_{i1t|c} \left[ \lambda_c \ln \left( \sum_k e^{V_{ikct}/\lambda_c} \right) - V_{i0ct} \right] (1 - s_{ict}) s_{ict} dF(\mathbf{V}_{it})}{\int s_{i1t|c,B} s_{ict} - s_{i0t|c} \left[ \lambda_c \ln \left( \sum_k e^{V_{ikct}/\lambda_c} \right) - V_{i0ct} \right] (1 - s_{ict}) s_{ict} dF(\mathbf{V}_{it})} \quad (3.3)$$

because  $1 - s_{i0t|c,B} = s_{i1t|c,B}$  and  $1 - s_{i0t|c} = s_{i1t|c}$ .

Similarly, the first-order condition with respect to  $\theta_{ct}^1$  (for an interior solution) can be written

$$\frac{m_{1ct}}{m_{0ct}} = \frac{\int s_{i0t|c,B} s_{ict} + s_{0t|c} \left[ \lambda_c \ln \left( \sum_k e^{V_{ikct}/\lambda_c} \right) - V_{i1ct} \right] (1 - s_{ict}) s_{ict} dF(\mathbf{V}_{it}|\boldsymbol{\beta})}{\int s_{i0t|c,B} s_{ict} - s_{1t|c} \left[ \lambda_c \ln \left( \sum_k e^{V_{ikct}/\lambda_c} \right) - V_{i1ct} \right] (1 - s_{ict}) s_{ict} dF(\mathbf{V}_{it}|\boldsymbol{\beta})}. \quad (3.4)$$

We can see that only one of the first-order conditions will be satisfied. Indeed, as  $1 - s_{i0t|c} = s_{i1t|c}$ ,

$$\begin{aligned} & s_{i1t|c,B} s_{ict} + s_{i1t|c} \left[ \lambda_c \ln \left( \sum_k e^{V_{ikct}/\lambda_c} \right) - V_{i0ct} \right] (1 - s_{ict}) s_{ict} \\ = & s_{i1t|c,B} s_{ict} - s_{i0t|c} \left[ \lambda_c \ln \left( \sum_k e^{V_{ikct}/\lambda_c} \right) - V_{i0ct} \right] (1 - s_{ict}) s_{ict} + \left[ \lambda_c \ln \left( \sum_k e^{V_{ikct}/\lambda_c} \right) - V_{i0ct} \right] (1 - s_{ict}) s_{ict} \\ > & s_{i1t|c,B} s_{ict} - s_{i0t|c} \left[ \lambda_c \ln \left( \sum_k e^{V_{ikct}/\lambda_c} \right) - V_{i0ct} \right] (1 - s_{ict}) s_{ict}, \end{aligned}$$

and similarly,

$$s_{i0t|c,B} s_{ict} + s_{0t|c} \left[ \lambda_c \ln \left( \sum_k e^{V_{ikct}/\lambda_c} \right) - V_{i1ct} \right] (1 - s_{ict}) s_{ict} > s_{i0t|c,B} s_{ict} - s_{1t|c} \left[ \lambda_c \ln \left( \sum_k e^{V_{ikct}/\lambda_c} \right) - V_{i1ct} \right] (1 - s_{ict}) s_{ict}$$

Thus, equation (3.3) cannot be true if  $m_{1ct} > m_{0ct}$ , and equation (3.4) cannot be true if  $m_{1ct} < m_{0ct}$ .

In the case in which  $m_{1ct} < m_{0ct}$ , there is no interior solution for  $\theta_{ct}^1$ , and thus we will have  $\theta_{ct}^1 = 0$ , meaning that the pharmacy chain never proposes the drug with the lowest margin alone. Then,  $\theta_{ct}^0$  is a solution of equation (3.3). The intuitive explanation is that when the chain increases the probability of only having the lower margin product available, profits are hurt both due to the opportunity cost of consumers who would otherwise have bought the high margin product when both were available and the loss of market share due to offering less variety on average.

For simplicity, in the following, we assume that parallel imports (good 0) is the high-margin product for all chains (which is the case in our data, as we will show later). Thus, we can set the probability of proposing direct imports alone,  $\theta_{ct}^1$ , to zero for all  $c$  in the following exposition and define

$$\theta_{ct} \equiv 1 - \theta_{ct}^0,$$

i.e., the probability that both goods are available in pharmacy chain  $c$ .

We can now express the expected inclusive value as

$$I_{ict} = (1 - \theta_{ct}) V_{i0ct} + \theta_{ct} \lambda_c \ln \left( \sum_k e^{V_{ikct}/\lambda_c} \right) = V_{i0ct} + \theta_{ct} \lambda_c \delta_{ict},$$



where

$$\delta_{ict} \equiv \ln \left( 1 + e^{\Delta V_{ict}/\lambda_c} \right) \quad \text{and} \quad \Delta V_{ict} = V_{i1ct} - V_{i0ct}$$

Thus,  $\delta_{ict}$  is the incremental expected utility from having both drugs available to choose from, as opposed to parallel import alone. Furthermore, let

$$\rho_{ict} \equiv s_{i1t|c,B} = \frac{1}{1 + e^{V_{i1ct}/\lambda_c - V_{i0ct}/\lambda_c}},$$

that is, the probability that consumer  $i$  chooses the directly imported variety in chain  $c$  at  $t$  when both are available. It will be helpful to note that  $\delta_{ict} = -\ln(1 - \rho_{ict})$ , which has the natural interpretation that individual  $i$ 's incremental utility from having both goods available is increasing in the probability that she will choose the direct imported variety when both are available. Then, the individual choice probabilities are

$$s_{i1ct}(\boldsymbol{\theta}_t) = \frac{e^{V_{i0ct} + \theta_{ct} \lambda_c \delta_{ict}}}{\sum_{\bar{c}} e^{V_{i0\bar{c}t} + \theta_{\bar{c}t} \lambda_{\bar{c}} \delta_{i\bar{c}t}}} \theta_{ct} \frac{e^{V_{i1ct}/\lambda_c}}{e^{V_{i0ct}/\lambda_c} + e^{V_{i1ct}/\lambda_c}}$$

and

$$s_{i0ct}(\boldsymbol{\theta}_t) = \frac{e^{V_{i0ct} + \theta_{ct} \lambda_c \delta_{ict}}}{\sum_{\bar{c}} e^{V_{i0\bar{c}t} + \theta_{\bar{c}t} \lambda_{\bar{c}} \delta_{i\bar{c}t}}} \left( 1 - \theta_{ct} \frac{e^{V_{i1ct}/\lambda_c}}{e^{V_{i0ct}/\lambda_c} + e^{V_{i1ct}/\lambda_c}} \right)$$

where  $\boldsymbol{\theta}_t \equiv (\theta_{0t}, \dots, \theta_{Ct})'$  is the vector of the probabilities that both goods are available at each chain.

Integrating over the distribution of preferences, we obtain the market share of each product as

$$s_{kct}(\boldsymbol{\theta}_t) = \int s_{ikct}(\boldsymbol{\theta}_t) dF(\mathbf{V}_{it}|\boldsymbol{\beta}).$$

The profit maximization problem for each chain  $c$  at  $t$  now implies the following optimality condition:

$$\frac{\partial \pi_{ct}}{\partial \theta_{ct}}(\boldsymbol{\theta}_t) \begin{cases} \leq 0 & \text{if } \theta_{ct} = 0, \\ = 0 & \text{if } 0 < \theta_{ct} < 1, \\ \geq 0 & \text{if } \theta_{ct} = 1. \end{cases} \quad (3.5)$$

The derivative of profits with respect to the probability that both goods are available is

$$\frac{\partial \pi_{ct}}{\partial \theta_{ct}} = m_{0ct} \frac{\partial s_{0ct}}{\partial \theta_{ct}} + m_{1ct} \frac{\partial s_{1ct}}{\partial \theta_{ct}}, \quad (3.6)$$

where the derivatives of shares with respect to  $\theta_{ct}$  are

$$\begin{aligned}\frac{\partial s_{0ct}}{\partial \theta_{ct}} &= \int (-\rho_{ict}s_{ict} + (1 - \theta_{ct}\rho_{ict})\lambda_c\delta_{ict}s_{ict}(1 - s_{ict}))dF(V_{it}|\boldsymbol{\beta}), \text{ and} \\ \frac{\partial s_{1ct}}{\partial \theta_{ct}} &= \int (\rho_{ict}s_{ict} + \theta_{ct}\rho_{ict}\lambda_c\delta_{ict}s_{ict}(1 - s_{ict}))dF(V_{it}|\boldsymbol{\beta}).\end{aligned}$$

From these expressions, we see that there are basically two effects from increasing the probability that both products are available. To give a better sense of how the model works, we first discuss these effects from the point of view of an individual  $i$ . The first effect is a change in the conditional choice probability of the product—that is, the choice probability given that the individual has chosen pharmacy chain  $c$ —weighted by the probability  $s_{ict}$  that chain  $c$  is chosen by individual  $i$  in the first place. This is *negative* for parallel imports, as it reduces the number of times for which it is the only product available, whereas it is *positive* for the direct import, as it increases the number of times for which it is part of the choice set. The second effect is a change in the probability of choosing chain  $c$ , weighted by individual  $i$ 's conditional probability of choosing the product. This effect is *positive* for both products, since the incremental expected utility of having both drugs available,  $\delta_{ict}$ , is positive for all individuals; i.e., more individuals will choose chain  $c$  when the variety is greater. The *aggregate* effect then depends on the distribution of individual tastes in the population. As an example, let us consider a *decrease* in  $\theta_{ct}$  to induce more consumers to buy the parallel imported variety. This will have a larger impact on the relative shares of the goods within pharmacy chain  $c$  when consumers have a strong preference for the directly imported variety on average, and even more so if this correlates positively with the probability of choosing chain  $c$  in the population. However, if people on average have a strong preference for the directly imported variety, the incremental utility  $\delta_{ict}$  will tend to be large, thus implying a stronger substitution away from chain  $c$ . This negative aggregate effect will be weaker if people have strong preferences for a specific pharmacy, such that  $s_{ict}$  tends to be either very high or very low, and also if there is a positive correlation between the taste for direct imports and chain  $c$ . From this, we can see that the distribution of tastes in the population will be central in the decision of pharmacy chains on how to foreclose the lower margin product.

From equation (3.6), together with the previous discussion, it is apparent that an increase in  $m_{0ct}$ —the margin on parallel imports—will lead to a decrease in  $\theta_{ct}$ , as  $\frac{\partial s_{0ct}}{\partial \theta_{ct}} < 0$ , whereas the opposite holds for an increase in  $m_{1ct}$ . Also note that only the relative margin matters for the decision of the pharmacy chain, although the relative margin will depend on both the wholesale prices and the price ceiling.

We assume that each pharmacy chain  $c$  sets  $\theta_{ct}$  to maximize its profits conditional on the wholesale prices it faces, while taking the choices of all other pharmacy chains as given. The equilibrium in each market  $t$  will

then be given by the vector  $\boldsymbol{\theta}_t^*(\mathbf{w}_{0t}, \mathbf{w}_{1t})$ , which consists of the elements  $\theta_{ct}^*(\mathbf{w}_{0t}, \mathbf{w}_{1t})$ , where the vectors of equilibrium  $\theta$ 's are functions of the wholesale prices of direct and parallel import in the market ( $\mathbf{w}_{1t}$  and  $\mathbf{w}_{0t}$  respectively) and implicitly of the exogenously given retail price ceiling  $\bar{p}_t$ , such that equation (3.5) is satisfied simultaneously for all pharmacy chains at  $t$ .

### 3.3 Upstream Manufacturer and Importers

We now model the behavior of the manufacturer and parallel importers. We assume that upstream firms and pharmacy chains bargain over wholesale prices, leading to the *Nash-in-Nash bargaining* model, which was first proposed by Horn and Wolinsky (1988). As documented by Brekke and Straume (2015), the prohibition against side-payments in contracts between manufacturers and wholesalers in the Norwegian pharmaceutical market explains why only linear pricing transaction are observed. We thus consider bargaining over a piece-rate price between upstream firms and pharmacy chains.

#### 3.3.1 Manufacturer Behavior

The total sales of the manufacturer of a drug in a given market (country) come from two channels: the direct import channel of its product (good 1) to all chains  $c$  and the parallel imports of the same patented active ingredient (good 0) by all chains  $c$ . Here, we hypothesize a fully rational manufacturer, internalizing the sales in a given market induced by parallel trade with other countries.

Thus, using the simpler notation  $\boldsymbol{\theta}_t^*$  for  $\boldsymbol{\theta}_t^*(\mathbf{w}_{0t}, \mathbf{w}_{1t})$ , the profits of the manufacturer are given by

$$\Pi_t(\mathbf{w}_{1t}, \boldsymbol{\theta}_t^*) = \sum_c \underbrace{(w_{1ct} - c_t) s_{1ct}(\boldsymbol{\theta}_t^*)}_{\text{Profit to Manufacturer of Direct Imports profit in chain } c} + \underbrace{(p_{1ct}^I - c_t) s_{0ct}(\boldsymbol{\theta}_t^*)}_{\text{Profit to Manufacturer of Parallel Import in chain } c \text{ at wholesale source price } p_{1ct}^I},$$

where  $c_t$  is the marginal cost of production,  $p_{1ct}^I$  is the manufacturer price in the source country of the parallel importer supplying chain  $c$ , and, as before,  $w_{1ct}$  is the wholesale prices charged for direct imported drugs to chain  $c$  at time  $t$ .

We assume that in each pairwise negotiation with the pharmacy chains, the manufacturer and pharmacy chain  $c$  set wholesale prices to maximize the Nash-product

$$(\Pi_t - \Pi_{-c,t})^{b_{1c}} (\pi_{ct} - \pi_{-1,ct})^{1-b_{1c}}, \quad (3.7)$$

where  $b_{1c}$  is the bargaining weight of the manufacturer when negotiating with chain  $c$ ,  $\Pi_{-c,t}$  is the manufacturer's profit in absence of an agreement with pharmacy chain  $c$ , and  $\pi_{-1,ct}$  is likewise pharmacy chain  $c$ 's profit in absence of an agreement with the manufacturer. We make the assumptions that all contracts remain the same if another negotiation fails and that each bargaining pair observes the wholesale prices of parallel imports to each pharmacy chain  $\mathbf{w}_{0t} = (w_{01t}, w_{02t}, \dots, w_{0Ct})$ . These assumptions are commonplace in the literature estimating structural bargaining models (see e.g., Gowrisankaran et al. (2015), Crawford and Yurukoglu (2012) and Ho and Lee (2017)). The first-order condition for a solution to equation (3.7) is

$$b_{1c} \frac{\partial \Pi_t / \partial w_{1ct}}{\Pi_t - \Pi_{-c,t}} + (1 - b_{1c}) \frac{\partial \pi_{ct} / \partial w_{1ct}}{\pi_{ct} - \pi_{-1,ct}} = 0. \quad (3.8)$$

In maximizing the Nash product, there will be an effect on the manufacturer's profit from how changes in wholesale prices affect the equilibrium  $\boldsymbol{\theta}_t^*(\mathbf{w}_{0t}, \mathbf{w}_{1t})$  in the next stage of the game.

Note that in the case where the manufacturer has all the bargaining power, that is,  $b_{1c} = 1$ , equation (3.8) reduces to the first-order condition for an optimal take-it-or-leave-it contract on  $w_{1ct}$  for the manufacturer, whereas in the case of  $b_{1c} = 0$ , it can be rewritten as the condition for an optimal contract proposed by the chain.

Denote the net value of agreement for the manufacturer and chain  $c$  as  $\Delta_c \Pi_t \equiv \Pi_t - \Pi_{-c,t}$  and  $\Delta_1 \pi_{ct} \equiv \pi_{ct} - \pi_{-1,ct}$ , respectively. The derivative of the manufacturer's profits with respect to the wholesale price is

$$\begin{aligned} \frac{\partial \Pi_t(\mathbf{w}_{1t}, \boldsymbol{\theta}_t^*(\mathbf{w}_{0t}, \mathbf{w}_{1t}))}{\partial w_{1ct}} &= s_{1ct}(\boldsymbol{\theta}_t^*) + \sum_{\bar{c}} \left[ (w_{1\bar{c}t} - c_t) \frac{\partial s_{1\bar{c}t}(\boldsymbol{\theta}_t^*(\mathbf{w}_{0t}, \mathbf{w}_{1t}))}{\partial w_{1ct}} + (p_{1\bar{c}t}^I - c_t) \frac{\partial s_{0\bar{c}t}(\boldsymbol{\theta}_t^*(\mathbf{w}_{0t}, \mathbf{w}_{1t}))}{\partial w_{1ct}} \right] \\ &= s_{1ct}(\boldsymbol{\theta}_t^*) + \sum_{\bar{c}} \left[ w_{1\bar{c}t} \frac{\partial s_{1\bar{c}t}(\boldsymbol{\theta}_t^*(\mathbf{w}_{0t}, \mathbf{w}_{1t}))}{\partial w_{1ct}} + p_{1\bar{c}t}^I \frac{\partial s_{0\bar{c}t}(\boldsymbol{\theta}_t^*(\mathbf{w}_{0t}, \mathbf{w}_{1t}))}{\partial w_{1ct}} \right] \end{aligned}$$

where we use the fact that aggregate demand is fixed and thus  $\sum_{\bar{c}} (\frac{\partial s_{1\bar{c}t}}{\partial w_{1ct}} + \frac{\partial s_{0\bar{c}t}}{\partial w_{1ct}}) = 0$ .

This first-order condition shows that the marginal cost of production  $c_t$  is not identified because the total market size is exogenously given for a prescription drug with a regulated retail price ( $\bar{p}_t$ ). The marginal cost  $c_t$  does not affect wholesale prices, except by imposing implicit bounds conditions for non-negative profits of manufacturers that we assume are satisfied for all drugs present on the market.

Similarly, the derivative of chain  $c$ 's profits with respect to the wholesale price  $w_{1ct}$  is

$$\frac{\partial \pi_{ct}(w_{0ct}, w_{1ct}, \boldsymbol{\theta}_t^*(\mathbf{w}_{0t}, \mathbf{w}_{1t}))}{\partial w_{1ct}} = -s_{1ct}(\boldsymbol{\theta}_t^*) + (\bar{p}_t - w_{1ct}) \frac{\partial s_{1ct}(\boldsymbol{\theta}_t^*(\mathbf{w}_{0t}, \mathbf{w}_{1t}))}{\partial w_{1ct}} + (\bar{p}_t - w_{0ct}) \frac{\partial s_{0ct}(\boldsymbol{\theta}_t^*(\mathbf{w}_{0t}, \mathbf{w}_{1t}))}{\partial w_{1ct}}$$

In the two expressions above, the derivatives of market shares with respect to wholesale prices will depend on the derivatives of market shares with respect to equilibrium  $\theta$ 's and the derivatives of equilibrium  $\theta$ 's

with respect to wholesale prices, which can be obtained using the optimal behavior of pharmacies, as detailed below.

Using vector notations for market shares  $\mathbf{s}_{0t} = (s_{01t}, \dots, s_{0Ct})$  and  $\mathbf{s}_{1t} = (s_{11t}, \dots, s_{1Ct})$ , we can then rewrite equation (3.8) governing the solution to the bargaining between the manufacturer and chain  $c$  as

$$s_{1ct} + \mathbf{w}'_{1t} \frac{\partial \mathbf{s}_{1t}}{\partial w_{1ct}} + \mathbf{p}'_{1t} \frac{\partial \mathbf{s}_{0t}}{\partial w_{1ct}} = \frac{1 - b_{1c}}{b_{1c}} \frac{\Delta_c \Pi_t}{\Delta_1 \pi_{ct}} \left( s_{1ct} - m_{1ct} \frac{\partial s_{1ct}}{\partial w_{1ct}} - m_{0ct} \frac{\partial s_{0ct}}{\partial w_{1ct}} \right) \quad (3.9)$$

which shows that the manufacturer considers the change in all the shares in the market through the change in equilibrium  $\boldsymbol{\theta}_t^*$ , whereas the pharmacy chain only considers the change in their own shares. The expression in parentheses on the right hand side is the (negative of) loss in profits to chain  $c$  from a change in the direct import wholesale price, which will depend on how much is lost in direct import sale from the marginal change in equilibrium  $\boldsymbol{\theta}_t^*$  and how much is gained in parallel import sale. The larger the relative bargaining power of the chain,  $\frac{1-b_{1c}}{b_{1c}}$ , and the larger the net value of agreement for the manufacturer relative to that of the chain,  $\Delta_c \Pi_t / \Delta_1 \pi_{ct}$ , the larger weight will be given to the change in profits for the pharmacy chain in determining the wholesale price.

Letting  $s_{j\tilde{c}t \setminus 1c}$  denote the share of chain  $\tilde{c}$ 's product  $j$  in  $t$  when direct imports are not available at chain  $c$ , we can express the net value for the manufacturer, suppressing arguments  $\boldsymbol{\theta}_t^*$ , as

$$\begin{aligned} \Delta_c \Pi_t &= \sum_{\tilde{c}} [(w_{1\tilde{c}t} - c_t) s_{1\tilde{c}t} + (p_{1\tilde{c}t}^I - c_t) s_{0\tilde{c}t}] \\ &\quad - \sum_{\tilde{c}} [(w_{1\tilde{c}t} - c_t) s_{1\tilde{c}t \setminus 1c} + (p_{1\tilde{c}t}^I - c_t) s_{0\tilde{c}t \setminus 1c}] \\ &= \sum_{\tilde{c}} (w_{1\tilde{c}t} \Delta_{1c} s_{1\tilde{c}t} + p_{1\tilde{c}t}^I \Delta_{1c} s_{0\tilde{c}t}), \end{aligned}$$

because  $s_{j\tilde{c}t \setminus 1c} = 0$ , and defining  $\Delta_{1c} s_{j\tilde{c}t} \equiv s_{j\tilde{c}t} - s_{j\tilde{c}t \setminus 1c}$ , that is, the difference in share of product  $j$  in chain  $\tilde{c}$  between the case of agreement and disagreement in the negotiations between the manufacturer and chain  $c$ . Since aggregate demand is constant, such that market shares sum to one both in the case of agreement and disagreement, the cost of production is immaterial to the change in manufacturer profit. Moreover, because of price regulation, we have that the manufacturer takes as given the price obtained on sales in source countries for parallel imports to chain  $c$  ( $p_{1ct}^I$ ).

Similarly, the net value for the chain is

$$\Delta_1 \pi_{ct} = (\bar{p}_t - w_{1ct}) s_{1ct} + (\bar{p}_t - w_{0ct}) \Delta_{1c} s_{0ct},$$

When the shape of demand is identified, it is possible to calculate the differences in shares,  $\Delta_{1c} s_{j\bar{c}t}$ , using the estimated demand system.

Note that the derivatives of market shares with respect to wholesale price follow from the chain rule and the implicit function theorem governing the change in equilibrium  $\theta_t^*$  when wholesale prices change due to pharmacists' optimal behavior. Details about how to obtain the derivatives of  $\theta_t^*(\mathbf{w}_{0t}, \mathbf{w}_{1t})$  with respect to wholesale prices are given in Appendix 7.4.

Once the demand shape is identified, together with the optimal behavior of pharmacy chains, the system (3.9) has one equation per molecule-pharmacy chain-period, with in principle one unknown parameter  $b_{1c}$ . The system also depends on the exogenous wholesale price of drugs earned by the manufacturer in the foreign country  $p_{1ct}^I$  but not the manufacturer marginal cost  $c_t$  because of the fixed market size (implying that  $\sum_c (s_{0ct} + s_{1ct}) = 1$ ). If  $p_{1ct}^I$  is known, the system of equations (3.9) allows us to identify the bargaining weight of each pharmacy chain.

When the condition in equation (3.9) holds for all  $c$ , we have a Nash-in-Nash solution for the bargaining between the manufacturer and each of the pharmacy chain. The full Nash-in-Nash solution is obtained when we also consider the conditions for bargaining between the parallel importer and each of the pharmacy chains, as described below.

### 3.3.2 Parallel Importers Behavior

We now consider the parallel importer's profits from its total sales of a drug in the importing market. This profit is given by

$$\Pi_t^{PI} = \sum_c (w_{0ct} - p_{0ct}^I) s_{0ct}(\theta_t^*),$$

where  $w_{0ct}$  is the wholesale price paid for parallel imported drugs by chain  $c$  at time  $t$  and  $p_{0ct}^I$  is the price that the importer has to pay for the drug in the source country, which we allow to vary across chains  $c$  for full generality because each chain may require different source countries. The wholesale price that parallel importers obtain from the pharmacy chains must be in the interval  $[p_{0ct}^I, \bar{p}_t]$  because parallel importers can only make a profit if the imported drug price in the source country is less than the maximum retail price, i.e., if  $p_{0ct}^I < \bar{p}_t$ .

We assume that the parallel importer bargains over the wholesale price with each pharmacy chain  $c$ , where they take as given the negotiated wholesale prices of originator products to each pharmacy chain  $\mathbf{w}_{1t} = (w_{11t}, w_{12t}, \dots, w_{1Ct})$ . When bargaining over the wholesale prices charged to the chains,  $\mathbf{w}_{0t}$ , the parallel importer will also take into account how changes in these prices will affect the equilibrium  $\theta_t^*(\mathbf{w}_{0t}, \mathbf{w}_{1t})$ . Similarly to equation (3.8), the first-order conditions for the solution to the Nash bargaining between each

pharmacy chain  $c$  and the parallel importer is

$$b_{0c} \frac{\partial \Pi_t^{PI} / \partial w_{0ct}}{\Pi_t^{PI} - \Pi_{-c,t}^{PI}} + (1 - b_{0c}) \frac{\partial \pi_{ct} / \partial w_{0ct}}{\pi_{ct} - \pi_{-0,ct}} = 0, \quad (3.10)$$

which can be rewritten, following the approach in previous section 3.3.1 using vector notations for prices and market shares stacked over the chains  $c$ , as

$$s_{0ct} + (\mathbf{w}_{0t} - \mathbf{p}_{0t}^I)' \frac{\partial \mathbf{s}_{0t}}{\partial w_{0ct}} = \frac{1 - b_{0c}}{b_{0c}} \frac{\Delta_c \Pi_t^{PI}}{\Delta_0 \pi_{ct}} \left( s_{0ct} - m_{1ct} \frac{\partial s_{1ct}}{\partial w_{0ct}} - m_{0ct} \frac{\partial s_{0ct}}{\partial w_{0ct}} \right) \quad (3.11)$$

where the left-hand side is the derivative of parallel importer profits with respect to the wholesale price  $w_{0ct}$  and we denoted  $\Delta_c \Pi_t^{PI} = \Pi_t^{PI} - \Pi_{-c,t}^{PI}$  with  $\Pi_{-c,t}^{PI} = \sum_{\tilde{c} \neq c} (w_{0\tilde{c}t} - p_{0\tilde{c}t}^I) s_{0\tilde{c}t \setminus 0c}$  and

$$\Delta_0 \pi_{ct} = \pi_{ct} - \pi_{-0,ct} = (\bar{p}_t - w_{1ct}) \Delta_{0c} s_{1ct} + (\bar{p}_t - w_{0ct}) s_{0ct}$$

where as defined previously  $\Delta_{0c} s_{1ct}$  corresponds to the market share of the direct imports at chain  $c$  when there are no parallel trade version at chain  $c$ .

Again, since wholesale prices are observed, one can use these optimality conditions to identify the parallel importers bargaining parameters  $b_{0c}$ , provided we observe or can model the prices at which imports are paid from the source country  $p_{0ct}^I$ .

## 4 Data, Identification and Empirical Results

### 4.1 Data and Descriptive Statistics

We estimate our model on the Norwegian market for Atorvastatin, which is a member of the statins drug class that is used to lower blood cholesterol. It is marketed by Pfizer under the trade name *Lipitor*. The patent expired towards the end of 2011, and the drug is thus under patent for the whole period from 2004 to 2007 covered by our data. The drug comes in four distinct strengths in the Norwegian market: tablets with 10, 20, 40 and 80 milligrams of the active ingredient. The prescription determines which of these strengths the consumer can obtain at the pharmacy, and the pharmacy can freely propose the directly imported or parallel imported alternatives. Atorvastatin was used by roughly 140,000 individuals in 2004 and 2005, but the number of users dropped to approximately 100,000 in 2006 and 85,000 in 2007.<sup>12</sup> The explanation for this can largely be attributed to a change in the regulation of statin prescriptions introduced in June 2005. The

<sup>12</sup>The population of Norway was roughly 4.6 million in this period.

regulation required that Simvastatin was to be prescribed for all new cases requiring statin treatment, whereas present users were to be put on treatment with Simvastatin within a year, unless medical considerations dictated otherwise.<sup>13</sup> The motivation for the regulation was to reduce expenditure for the Norwegian National Insurance Administration.

We combine data from several sources: transaction data from the *Norwegian Directorate of Health* covering all purchases of reimbursable drugs by individuals in Norway; wholesale registry data from the *Norwegian Institute of Public Health* containing monthly wholesale prices of drug wholesalers in Norway; data regarding price regulation, substitutability and parallel marketing licenses from the *Norwegian Medicines Agency*; and data about aggregate wholesale prices in several countries from *IMS Health*. We thus have data concerning all purchases of Atorvastatin in Norway for the period of 2004–2007, which amounts to approximately 1.4 million transactions. The transactions are performed by approximately 170,000 individuals, where a pseudo-ID for each individual allows us to track individual choices over time. The demographic information on individuals is otherwise limited to age and gender. For each transaction, we know the price charged for the drug by the pharmacy chain, the co-payment paid, the specific pharmacy at which the transaction happened, the number of packages bought, and the specific drug package.<sup>14</sup>

The supply side of the market for prescription drugs mainly consists of three large pharmacy retail chains that are vertically integrated with their upstream wholesalers (where the wholesaler is unique to each chain). The three largest chains, Apotek 1, Boots and Vitus, cover 85 % of all pharmacies, whereas public hospital pharmacies (6 %), a smaller retail chain (5 %), and independent pharmacies (4 %) comprise the rest.<sup>15</sup>

Table 4.1 shows the yearly size of the Atorvastatin market in Norway in millions of *Defined Daily Doses* (DDD), segmented by the amount of active ingredient.<sup>16</sup> We have also calculated the parallel import share of DDD within each segment. We see that for 10 and 20 mg, parallel imports were not present before 2007. For 40 and 80 mg, parallel imports often cover a substantial share of the market, constituting approximately 90% of the 80 mg segment in the period of 2004–2006. The reason for the differences in parallel import shares is likely a combination of differences in parallel export opportunities, differences in profitability across parallel import locations and differences in the relative price in the source country and Norway. We do not include the entry decision of parallel imports in each market in our structural model, but these entries can

---

<sup>13</sup>More details about this regulatory change can be found in Sakshaug et al. (2007).

<sup>14</sup>An example of a specific drug package is *Lipitor* with 40 mg of the active ingredient, containing 98 tablets, and imported by Farmagon from France.

<sup>15</sup>The shares are calculated from our own data and checked against data obtained from the Norwegian Medicines Agency. The numbers correspond exactly to official statistics reported by the Norwegian Medicines Agency and the Norwegian Pharmacy Association.

<sup>16</sup>Our definition of the market includes direct purchases in pharmacies by individuals exclusively. Although there might be some usage of Atorvastatin in hospitals—for instance, as part of statin treatment after heart attacks—the numbers in our data are virtually identical to official statistics regarding drug utilization in Norway for aggregate usage of Atorvastatin, which makes us conclude that this usage represents a negligible share of sales.



Table 4.1: *Market size in million DDD, share of parallel imports, price to consumers (Price) and wholesale prices (Wholesale) in NOK/DDD*

		2004	2005	2006	2007
10 mg	DDD (mill.)	16.36	15.10	9.13	4.61
	Share parallel	0.00	0.00	0.00	0.13
	Price	8.78	8.84	8.39	8.43
	Wholesale direct	6.21	6.20	5.86	5.86
	Wholesale parallel	-	-	-	4.42
20 mg	DDD (mill.)	34.15	34.99	22.14	12.07
	Share parallel	0.00	0.00	0.00	0.10
	Price	6.62	6.67	6.37	6.42
	Wholesale direct	4.74	4.74	4.52	4.53
	Wholesale parallel	-	-	-	3.15
40 mg	DDD (mill.)	23.78	31.22	26.42	29.32
	Share parallel	0.79	0.48	0.07	0.17
	Price	4.16	4.21	3.82	3.90
	Wholesale direct	3.00	3.01	2.71	2.76
	Wholesale parallel	2.91	2.93	2.87	2.03
80 mg	DDD (mill.)	12.03	20.12	27.38	35.69
	Share parallel	0.93	0.86	0.96	0.63
	Price	2.15	2.23	1.98	1.97
	Wholesale direct	1.55	1.60	1.40	1.39
	Wholesale parallel	1.52	1.50	1.38	1.35

Note: DDD stands for Defined Daily Dose.

be rationalized by exogenous wholesale earnings variations in source countries, e.g., regulatory changes in source countries and exchange rate shocks. Our model explains parallel imports market shares for markets in which they are present, which are the ones that we use in our estimation regarding upstream manufacturer and importer behavior. There is also substantial variation across some of these years. The market size for 10 and 20 mg decreases substantially over the sample period, whereas it stays at roughly the same level for 40 mg and increases substantially for 80 mg. It seems likely that the large changes in the number of consumers underlying these figures will have an impact on the distribution of preferences in the market. We will allow the average taste for each available drug to change across segments and time, something we return to when discussing the specification of the consumer choice model. The price to consumers reflects the regulatory price ceiling set by the Norwegian Medicines Agency, as all packages—both parallel and direct imports—are consistently priced at the price ceiling. From the wholesale prices, we see that the aggregate margin is larger for parallel imports in almost all cases, except for 40 mg in 2006, though this is related to a reduction in the price ceiling early this year, after which parallel importers withdraw from the market (see Figure 4.1 below).

The relative margin only appears to be reversed here, due to the average being taken over the full year for direct import but only for part of the year for parallel imports.

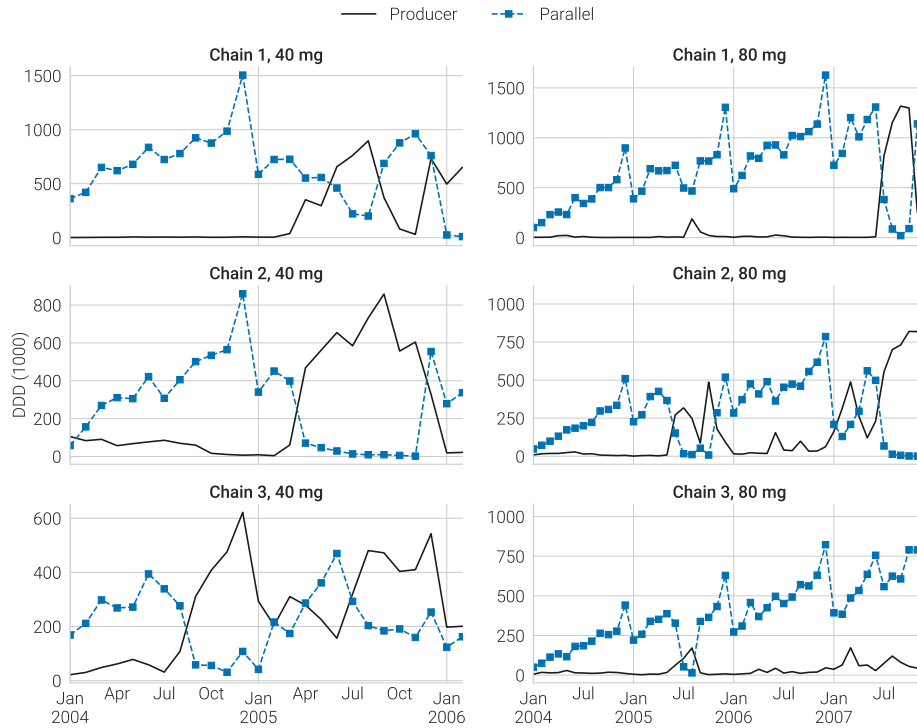
Table 4.2: *Drug packages and parallel import licensing*

Dose	Company	#Tablets	Source country	License Year
10 mg	Pfizer	100	-	-
	Farmagon	100	Czech Rep.	2006
	Orifarm	100	Poland	2006
20 mg	Pfizer	100	-	-
	Farmagon	100	Czech Rep.	2006
	Orifarm	100	Poland	2006
40 mg	Pfizer	100	-	-
	Farmagon	98	UK, France	2002, 2004
	Farmagon	100	Poland, Czech Rep.	2004, 2006
	Orifarm	98	UK	2002
80 mg	Pfizer	100	-	-
	Farmagon	98	UK, France	2002, 2004
	Farmagon	100	Czech Rep.	2006

Table 4.2 presents information about the specific packages sold in the Norwegian market for Atorvastatin in the sample period. The data do not contain the expiration date of the drug, but the lifetime is typically very long for this type of drug (a couple of years), and anecdotal evidence indicates that it does not differ between parallel and direct imports. The active upstream firms are Pfizer, Farmagon and Orifarm; Pfizer holds the patent and is responsible for the direct imports, whereas Farmagon and Orifarm are parallel importers. The parallel importers have licenses to import from the United Kingdom, France, Czech Republic and Poland, where most of the licenses for the Eastern European countries were acquired in 2006. The underlying sales data shows that the packages imported from Eastern Europe were only sold in 2007. Where several source countries are listed, the packages imported from the different countries are given the same identifier in the national drug classification system, which means that they are identical in all respects. The parallel import process is such that the drugs will be repackaged by the parallel importer to be in accordance with nation specific guidelines on package labels, language and warnings. In several of the cases, the parallel importers have license to import the package from two countries, but inspection of the data shows that parallel imports of Lipitor are exclusively from the Eastern European countries for 10 and 20 mg, where parallel imports enter in the second half of 2007, whereas for 40 and 80 mg, parallel imports were from the Western European countries until 2007, when there was a switch to Eastern European imports after a large drop in parallel imports in 2006. In Figure 4.1, we show monthly sales of parallel imports and the manufacturer (Pfizer) in

thousands of DDD for each segment (amount of active ingredient) and pharmacy chain. These graphs show the important variation over time, products and chains of the parallel imports or direct imports sales.

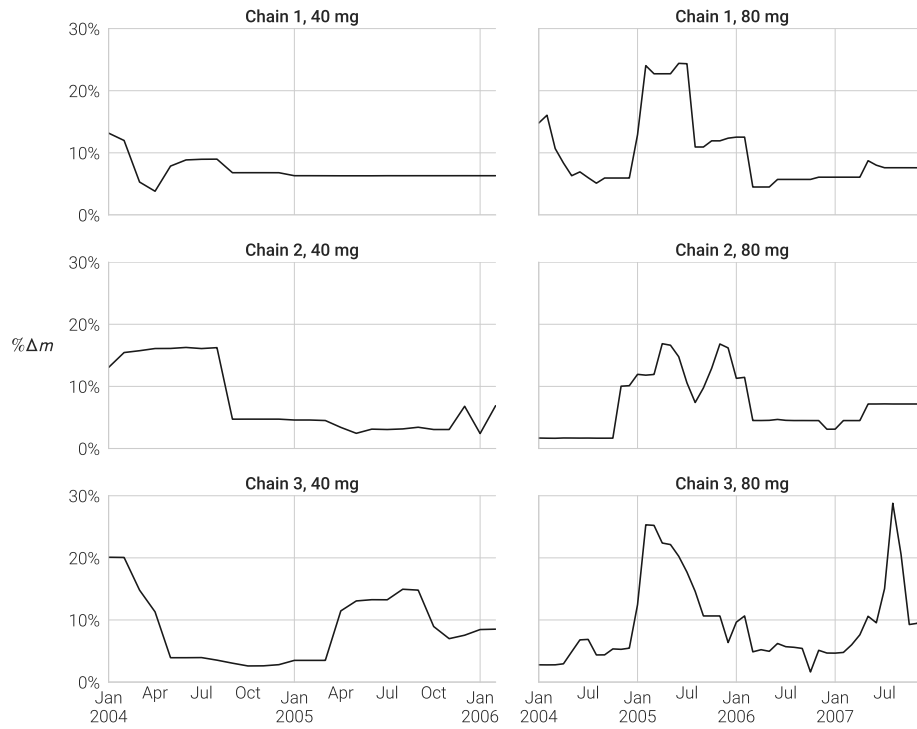
Figure 4.1: *Monthly sales in 1000 DDD of DI and PI for each chain and dosage*



In Figure 4.2, we show the percentage margin difference between parallel and direct imports separately for each segment and pharmacy chain. As the consumer price is entirely decided by the price cap, which is binding for both the direct and parallel imported varieties, the retail pharmacist margin difference between parallel and direct imports is exactly equal to the wholesale prices difference of parallel and direct imports. As margins of parallel imports are consistently higher than direct imports, wholesale prices are consistently lower than the direct import wholesale price.

As Figure 4.1 shows, for the 40 mg version, sales of parallel imports are largely more important than are sales of direct imports, and the former grow over time during 2004 across the three chains. However, they go down strongly after that period for chains 1 and 2, as the margin advantage of parallel imports decreases simultaneously in 2005 for both chains, as seen in Figure 4.2. For chain 3, the parallel imports sales decrease earlier in the second part of 2004, when margins of parallel imports decrease relative to margins of direct imports (see Figure 4.2), but during 2005, unlike in chains 1 and 2, parallel imports sales increase again in chain 3 and exhibit at the same time a growing margin compared to direct imports. For the 80 mg version, parallel imports dominate sales over direct imports, except at the end of 2007 for chain 2 and temporarily

Figure 4.2: Margin difference between PI and DI in percentage of DI wholesale price



for chain 1. These figures show that the form of total sales of Lipitor varies importantly over time between parallel and direct imports.

## 4.2 Reduced-Form Evidence

To further investigate the descriptive evidence of correlation between pharmacy margins and sales of different versions of the same drug, we perform a set of reduced-form regressions showing that sales of parallel imports do react to the pharmaceutical chain margins. The results in Table 4.3 demonstrate that sales of parallel imports do react to pharmacy chain margins. As  $s_{kct}$  stands for the market share of drug version  $k$  in pharmacy chain  $c$  within market  $t$  (where  $\sum_{c \in \{1, \dots, C\}, k \in \{0, 1\}} s_{kct} = 1$ ), we regress the log relative margin of direct imports ( $\ln(s_{1ct}/(s_{0ct} + s_{1ct}))$ ) on the pharmacy  $c$  margins for each version  $k$  equal to the retail price minus the wholesale price ( $m_{kct} \equiv p_{kct} - w_{kct}$ ).

Table 4.3: *Reduced-form evidence for a relationship between parallel imports and pharmacy margins on Lipitor*

Dependent Variable $\left(\ln \frac{s_{1ct}}{s_{0ct} + s_{1ct}}\right)$	(OLS) (1)	(OLS) (2)	(2SLS) (3)	(OLS) (4)	(OLS) (5)	(2SLS) (6)
Direct imports margin $m_{1ct}$	2.125*** (0.523)	1.911*** (0.558)	2.428*** (0.597)			
Parallel imports margin $m_{0ct}$	-0.119*** (0.033)	-0.247*** (0.059)	-0.572*** (0.118)			
Margin difference ( $m_{1ct} - m_{0ct}$ )				0.175*** (0.043)	0.329*** (0.057)	0.717*** (0.097)
Chain 1	-0.412*** (0.085)	-0.415*** (0.089)	-0.390*** (0.089)	-0.485*** (0.085)	-0.474*** (0.089)	-0.453*** (0.093)
Chain 2	0.135 (0.072)	0.087 (0.073)	-0.026 (0.091)	0.112 (0.076)	0.055 (0.078)	-0.081 (0.089)
Year*Month fixed effects	No	Yes	Yes	No	Yes	Yes
N	574	574	574	574	574	574

Notes: Standard errors in parenthesis are clustered at the market level. Chain 3 is the omitted dummy. Constant not shown. In the case of Two Stage Least Squares (2SLS) estimates, the instruments are the wholesale prices in France, Germany, Italy, Spain, the United Kingdom, and Greece and the exchange rates between NOK and the euro, Swiss Franc and US dollar.

The regressions in Columns (1) through (6) include chain-fixed effects, time-fixed effects and the margins and are estimated using either OLS or Two-Stage Least Squares. In the case of Two Stage Least Squares, we instrument the margins or the margin difference with the average quarterly wholesale prices of Lipitor in France, Germany, Italy, Spain, and the UK, in addition to the NOK exchange rates with the US dollar, the euro, and the Swiss Franc. These regressions show that the larger the parallel import margin is and the lower the direct import margin is, the larger the sales of parallel imports. Costa-Font (2016) finds a similar effect using data from the Netherlands by regressing the market share of parallel imports of statins on price differences in source countries and other distance variables, showing that they are driven by cross-country differences in margins. We thus have clear evidence that strategic behavior of pharmacies allows them to sell more of the drugs for which they have a higher margin. The trade-off mechanism exhibited in our model is consistent with these findings, predicting that pharmacies will sell even more of the high-margin version of

the drug when the margin difference increases. In Appendix 7.7, we show additional reduced-form regressions based on 9 molecules (including Atorvastatin) for which there is substantial parallel imports over the 4 years of data and for which we have data concerning wholesale prices in other European countries that are used as instrumental variables. The results show similar effects of margins on the share of direct imports within the chain. Similarly, Brekke and Straume (2013) shows that in the case of off-patent drugs (for which there are almost no parallel imports), the shares of generics versus the originator brand are also related to pharmacy-chain margins even controlling for (consumer) price differences.

### 4.3 Econometric Identification and Estimation

Our structural model of demand and supply can be estimated using data regarding consumer choices between parallel trade and directly imported versions of a drug and data about the pharmacy retail chain margins or wholesale prices. In Section 3.1, we developed a consumer discrete choice model in which consumers choose between pharmacy chains and directly versus parallel imported drugs. Our random utility model resembles a classic random coefficients logit model, although with the difference that random utilities depend on unobserved, strategic choices of the firms: pharmacies' strategic choices of the distribution of assortment of parallel trade versus direct imported drugs. To address this issue, we simultaneously estimate preference parameters and the assortment set probabilities of pharmacy chains using pharmacy chains' profit maximization conditions explained in section 3.2 together with the demand model. We first show how to identify and estimate this model of demand and pharmacy chain behavior. In a second step, we use the estimated parameters and choice set probabilities to identify the source country cost for parallel imports, source country earnings for the manufacturer, and the bargaining parameters using the vertical chain bargaining model developed in section 3.3.

#### 4.3.1 Demand Identification with Consumer and Pharmacy Chain Behaviors

From the discrete choice demand model described in section 3.1, the individual choice probability for consumer  $i$  choosing version  $j \in \{0, 1\}$  at pharmacy chain  $c$  and period  $t$  is given by

$$s_{ijct}(\boldsymbol{\theta}_t) = s_{ict}s_{ijt|c} = \frac{e^{V_{i0ct} + \theta_{ct}\lambda_c\delta_{ict}}}{\sum_{\tilde{c}} e^{V_{i0\tilde{c}t} + \theta_{\tilde{c}t}\lambda_{\tilde{c}}\delta_{i\tilde{c}t}}} \left( 1_{\{j=0\}} + (-1)^{1_{\{j=0\}}} \theta_{ct} \frac{e^{V_{i1ct}/\lambda_c}}{e^{V_{i0ct}/\lambda_c} + e^{V_{i1ct}/\lambda_c}} \right).$$

We specify individual  $i$ 's utility from product version  $j$  bought at pharmacy chain  $c$  in market  $t$  as

$$V_{ijct} = \alpha_{jct} + \nu_{ijct}$$

where  $\alpha_{jct}$  is the average utility in market  $t$  for product  $j$  at chain  $c$ , common to all individuals, and  $\nu_{ijct}$  is the individual deviation from the mean utility for that good, capturing heterogeneity in consumers' tastes. Just as there is typically significant heterogeneity in preferences for generics related to education (Bronnenberg et al., 2015), a similar source of unobserved heterogeneity is possible for parallel imports. In our setting, unobserved heterogeneity in the consumers' distances to stores, for example, could be important, as could other chain-specific variation in preferences. Since the common mean effects  $\alpha_{jct}$  vary freely across version-chain-market, they can capture unobserved market effects for each product in addition to mean-chain effects.

To allow a flexible distribution of preferences, we model  $\nu_{ijct}$  as a mixture of normal distributions. In practice, we specify  $\nu_{ijct}$  such that

$$\nu_{ijct} = \underbrace{\delta_j^{g_i} + \sigma_j^{g_i} \nu_i^j}_{\text{Drug version specific taste}} + \underbrace{\delta_c^{g_i} + \sigma_c^{g_i} \nu_i^c}_{\text{Pharmacy specific taste}} \quad (4.1)$$

where  $\nu_i^k$  is individual  $i$ 's taste characteristics for characteristic  $k$ , which is either the product version  $j$  or a specific chain  $c$ . We allow unobserved latent groups, where  $g_i \in \mathcal{G}$  denotes the latent group of  $i$  and  $\mathcal{G}$  is the set of groups in the population. We assume that  $\nu_i^k$  obeys a standard normal distribution in the population, such that  $\sigma_k^{g_i}$  measures the scale of individual heterogeneity in taste for  $k$  for an individual in group  $g_i$  and  $\delta_k^{g_i}$  is the mean deviation in taste for  $k$  for individuals in this group. The group of individual  $i$  is unobserved and is thus treated as a latent class during estimation. After some initial estimates and tests with a growing number of classes, we allow four latent classes, where one is arbitrarily chosen as the base group,  $g = 0$  with  $\delta_j^0 = \delta_c^0 = 0$ .<sup>17</sup> Each group  $g$  has a population share  $\tau_g$ , assumed to be the same across markets, which is introduced as a parameter to be estimated in the likelihood.

Denoting by  $\beta = (\delta_j^g, \sigma_j^g, \delta_c^g, \sigma_c^g, \lambda_1, \dots, \lambda_C, \tau_1, \dots, \tau_G)$  the full vector of parameters governing heterogeneous preferences, for some given mean preference parameters  $\alpha_{0ct}$ ,  $\alpha_{1ct}$  and  $\theta_{ct}$ , one can estimate  $\beta$  via the maximum likelihood. The likelihood of individual  $i$ 's choice sequence is given by

$$L_i(\beta; \alpha_{0ct}, \alpha_{1ct}, \theta_{ct}) = \sum_{g \in \mathcal{G}} \tau_g \int \left( \prod_{p \in \mathcal{P}_i} s_{ij(p)c(p)t(p)}(\nu_i) \right) dF(\nu_i | \beta), \quad (4.2)$$

where  $\mathcal{P}_i$  is the set of purchase events in which consumer  $i$  is involved,  $j(p)$  and  $c(p)$  denote consumer  $i$ 's choice of product and chain under purchase event  $p$ , and  $t(p)$  is the market in which purchase event  $p$  happens. Thus,  $s_{i,j(p),c(p),t(p)}(\nu_i)$  is the individual  $i$  choice probability conditional on his unobserved heterogeneity  $\nu_i \equiv (\nu_i^j, \nu_i^c)$  and  $F(\nu_i | \beta)$  is the cumulative distribution function of  $\nu_i$ .

<sup>17</sup>This normalization is necessary for identification, since the  $\alpha_{jct}$  average utility parameters will pin down the baseline mean utility of version and chain across the unobserved groups.

The mean parameters  $\alpha_{0ct}$ ,  $\alpha_{1ct}$  and  $\theta_{ct}$  are identified by adding the pharmacy chain optimality equilibrium conditions for  $\theta_{ct}$  and a condition for equality between predicted and observed chain-product market shares for all  $c, t$ :

$$\theta_{ct}^* = \arg \max_{0 \leq \theta_{ct} \leq 1} \pi_{ct}(m_{0ct}, m_{1ct}, \theta_{ct}^*, \alpha_{0ct}, \alpha_{1ct}) \quad (4.3)$$

$$\hat{s}_{jct} = s_{jct}(\theta_{ct}, \alpha_{0ct}, \alpha_{1ct}, \beta) \quad (4.4)$$

where  $\hat{s}_{jct}$  is the observed market share of product  $j$  in chain  $c$ . Note that pharmacy chain profits depend on observed margins  $m_{0ct}$  and  $m_{1ct}$  as

$$\pi_{ct}(m_{0ct}, m_{1ct}, \theta_{ct}, \alpha_{0ct}, \alpha_{1ct}, \beta) = m_{0ct}s_{0ct}(\theta_{ct}, \alpha_{0ct}, \alpha_{1ct}, \beta) + m_{1ct}s_{1ct}(\theta_{ct}, \alpha_{0ct}, \alpha_{1ct}, \beta)$$

with predicted shares from our model given by

$$s_{jct}(\theta_{ct}, \alpha_{0ct}, \alpha_{1ct}, \beta) = \sum_i s_{ijct} = \sum_i \sum_{g \in \mathcal{G}} \tau_g \int s_{ijct}(\mathbf{v}_i) dF(\mathbf{v}_i | \beta)$$

Under the assumption of existence of a Nash equilibrium between the three chains, the pharmacy chains' necessary incentives equation (4.3) can be described by the first-order condition given in equation (3.5) such that for all  $c$  and  $t$ ,<sup>18</sup>

$$\left. \frac{\partial \pi_{ct}}{\partial \theta_{ct}} \right|_{\theta_{ct} = \theta_{ct}^*} \begin{cases} = 0 & \text{if } 0 < \theta_{ct}^* < 1 \\ \leq 0 & \text{if } \theta_{ct}^* = 0 \\ \geq 0 & \text{if } \theta_{ct}^* = 1 \end{cases} \quad (4.5)$$

These optimal choices of  $\theta_{ct}$  mean that they can be expressed as functions of the vector of margins or wholesale prices,  $\theta_{ct}^*(\mathbf{m}_{0t}, \mathbf{m}_{1t})$  or  $\theta_{ct}^*(\mathbf{w}_{0t}, \mathbf{w}_{1t})$ , in addition to being functions of the mean utility parameters  $\alpha_{jct}$  and the vector  $\beta$ . The identification of the demand model is given by the properties of the likelihood (4.2), but if one does not want to rely on its functional form, the identification relies on the assumption that margins  $(\mathbf{m}_{0t}, \mathbf{m}_{1t})$  (or equivalently wholesale prices  $(\mathbf{w}_{0t}, \mathbf{w}_{1t})$ ) vary independently of preferences  $(\alpha_{0t}, \alpha_{1t})$ . We could allow the heterogeneity of preferences (4.1) to be time varying – which we do not do for simplicity given the already large time flexibility introduced but the mean preferences  $(\alpha_{0t}, \alpha_{1t})$  – provided that we also assume that the variability of margins  $(\mathbf{m}_{0t}, \mathbf{m}_{1t})$  is independent of the varying heterogeneity of preferences.

<sup>18</sup>We can also remark that  $s_{i1t|c}(\theta_{ct}) = \theta_{ct} \rho_{ict} \leq \theta_{ct}$ , implying that the *observable* relative market share of direct imports within the chain  $s_{i1t|c}$  is a lower bound on  $\theta_{ct}$ , which means that we can search for the optimal  $\theta_{ct}$  in the interval  $[\hat{s}_{i1t|c}, 1]$ .



Then, observing individual choice variation across choice occasions gives us a lot of identifying power with respect to mean preferences  $\alpha_{0t}, \alpha_{1t}$ , as individuals have fixed heterogeneity of preferences.

Even if there are many parameters since we have  $(\alpha_{0ct}, \alpha_{1ct}, \theta_{ct})$  for each chain-market combination, utilizing the fact that these parameters are common across consumers within each chain-market, they can be solved for by a simpler root-finding algorithm, conditional on the parameter vector  $\beta$ . The intuition is that within each market  $t$ , these parameters can be set such that observed market shares are equal to predicted aggregate shares and such that the equilibrium conditions for optimal chain behavior hold.

The nested fixed-point algorithm we use is as follows:

**Inner loop for given preference parameters  $\beta$ :**

The inner loop of our estimation algorithm intends to find the mean preference parameters  $\alpha_{jct}$  and the choice set parameters  $\theta_{ct}$  that satisfy the Nash equilibrium necessary conditions across pharmacy chains and the equality condition between observed and simulated market shares. Existence will be guaranteed under some sufficient conditions detailed below.

For a given vector  $(\theta_t, \beta)$ , we know from Berry (1994) and Berry et al. (1995) that one can solve for all  $\alpha_{0ct}, \alpha_{1ct}$  such that (4.4) is true for all  $j, c$ . This means that we can uniquely define  $\alpha_{0ct}(\theta_t, \beta), \alpha_{1ct}(\theta_t, \beta)$  that are continuous in all  $\theta_{ct}$ .

For any  $\alpha_{0ct}, \alpha_{1ct}$  we assume that there exists<sup>19</sup> a Nash equilibrium in  $\theta_t$  of (4.3). As for each pharmacy chain  $c$ , the profit function  $\pi_{ct}$  is continuous in all  $\theta_{ct}$ , the best response of each chain is well defined, and we only require best response functions to cross. We will assume this is the case, which can be verified empirically. Thus, we can define  $\theta_{ct}(\alpha_{0t}, \alpha_{1t}, \beta) \in [0, 1]$  that solves the maximization (4.3) and are continuous in all  $\alpha_{0ct}, \alpha_{1ct}$  because  $\pi_{ct}(\theta_t, \alpha_{0ct}, \alpha_{1ct}, \beta)$  is continuous in all  $\theta_{ct}$  that belong to  $[0, 1]$ .

Then, assuming that the image of  $[0, 1]^C$  by  $\theta_t(\alpha_{0t}(\cdot, \beta), \alpha_{1t}(\cdot, \beta), \beta)$  is  $[0, 1]^C$ , we can use Brouwer's fixed point theorem and obtain that there is a vector  $\theta_t$  that is solution of

$$\theta_t(\alpha_{0t}(\theta_t, \beta), \alpha_{1t}(\theta_t, \beta), \beta) = \theta_t \tag{4.6}$$

This proves that there is a vector  $(\alpha_{0t}(\hat{s}_t, \mathbf{m}_{0t}, \mathbf{m}_{1t}, \beta), \alpha_{1t}(\hat{s}_t, \mathbf{m}_{0t}, \mathbf{m}_{1t}, \beta), \theta_t(\hat{s}_t, \mathbf{m}_{0t}, \mathbf{m}_{1t}, \beta))$  solution of (4.3) and (4.4). At this step, we can search for the possibility of multiple solutions over the support of  $\theta$  which has the advantage to be bounded below and above<sup>20</sup>.

**Outer loop: maximizing the likelihood in  $\beta$ :**

<sup>19</sup>We do not need to assume unicity, and we will look for possible multiple equilibria.

<sup>20</sup>We provide details about our numerical procedure corresponding to the inner loop algorithm in Appendix Section 7.5.

We then maximize the following likelihood function in  $\beta$

$$L_i(\beta; \hat{\mathbf{s}}_t, \mathbf{m}_{0t}, \mathbf{m}_{1t}) = L_i(\beta; \alpha_{0t}(\hat{\mathbf{s}}_t, \mathbf{m}_{0t}, \mathbf{m}_{1t}, \beta), \alpha_{1t}(\hat{\mathbf{s}}_t, \mathbf{m}_{0t}, \mathbf{m}_{1t}, \beta), \theta_t(\hat{\mathbf{s}}_t, \mathbf{m}_{0t}, \mathbf{m}_{1t}, \beta)) \quad (4.7)$$

The estimation routine then becomes a nested fixed point algorithm, where we solve for the parameters  $\alpha_{0t}(\beta)$ ,  $\alpha_{1t}(\beta)$  and  $\theta_t(\beta)$  conditional on the current value of  $\beta$  in the inner loop, while searching for the parameter vector  $\beta$  that maximizes the log likelihood in the outer loop.

Finally, we note that the corner solutions of  $\theta_{ct}(\beta) = 1$  allow some independent variation of the likelihood in parameters  $\beta$  not coming from the changes in  $\theta_{ct}$  driven by  $\beta$  when  $\theta_{ct}$  is interior. This intuitively allows us to separately identify the effect of preferences from the effect of choice sets. Intuitively,  $\theta_{ct}$  will be equal to one when the margins for each version of the drug are sufficiently similar given the region of preference parameters  $\beta$ , and the individual choices will vary only because of preferences.

### 4.3.2 Identifying Bargaining in the Supply-Side Model

We now use the vertical structure competition game developed in section 3.3 to identify the supply-side parameters of the model. The objective is to identify all the bargaining parameters  $b_{0c}$  and  $b_{1c}$  respectively for the parallel importer and the manufacturer negotiation with each pharmacy chain  $c$ .

The optimality conditions (3.9) and (3.11) of the bargaining game between the manufacturer or the parallel importer and pharmacy chains relate demand and bargaining parameters to the marginal source country opportunity costs of drugs for the parallel importer ( $\mathbf{p}_{0t}^I$ ) and the manufacturer ( $\mathbf{p}_{1t}^I$ ). We note that all  $p_{0ct}^I$  and  $p_{1ct}^I$  can be different because of the costs related to packaging and extra logistics when importing from source countries and the pricing between the manufacturer, the source-country wholesaler and the parallel importer. We assume that parallel importers' costs ( $\mathbf{p}_{0t}^I = (p_{01t}^I, \dots, p_{0Ct}^I)$ ) and the source countries' wholesale prices of the manufacturer ( $\mathbf{p}_{1t}^I = (p_{11t}^I, \dots, p_{1Ct}^I)$ ) are functions of observables  $X_t$ , such as the wholesale prices in the source countries, company-fixed effects for the manufacturer or parallel importer, and interactions with source country prices<sup>21</sup>. With  $\mathbf{p}_{0t}^I$  and  $\mathbf{p}_{1t}^I$  from the optimal bargaining equations (3.8) and (3.10), stacked in the vector  $\mathbf{p}_t^I = (\mathbf{p}_{0t}^I, \mathbf{p}_{1t}^I)$  for each market  $t$ , we specify

$$p_t^I(\mathbf{b}) = X_t\eta + \epsilon_t,$$

where  $\mathbf{b}$  is the vector of bargaining parameters  $\mathbf{b} = (b_{00}, \dots, b_{0C}, b_{10}, \dots, b_{1C})$ .

---

<sup>21</sup>We use the wholesale prices of the source countries France and the UK but also those in Germany, Italy, Spain, Turkey, France, the UK and the US, which will be informative about the price at which parallel traders acquire the drugs and what the manufacturer earns on parallel trade.

Then, we assume that we observe instrumental variables  $Z_t$  such that  $E[\epsilon_t|Z_t] = 0$  and then identify the parameter vector  $(\eta, \mathbf{b})$  using the moment condition  $E[\epsilon(\eta, \mathbf{b})|Z] = 0$ . Our instrumental variables  $Z_t$  include variables exchange rate shocks between the Norwegian Crown and the US dollar and Euro, in addition to the price ceiling  $\bar{p}_t$ , indicators for pharmacy chain identity, and interactions. The specific moment conditions that we use are the sample analogs of  $E[Z'\epsilon(\eta, \mathbf{b})] = 0$ , such that our GMM estimator is

$$(\hat{\eta}, \hat{\mathbf{b}}) = \arg \min_{\eta, \mathbf{b}} \epsilon(\eta, \mathbf{b})' ZW Z' \epsilon(\eta, \mathbf{b}), \quad (4.8)$$

where  $W$  is a weighting matrix for the moments.

The intuition for identifying the bargaining parameters in light of the instrument set is that pharmacy chain identity should be informative about the overall bargaining strength of the chain while being plausibly uncorrelated with unobserved determinants of costs related to parallel trade. We thus preclude the possibility that sorting of parallel importers across pharmacy chains is related to the costs of parallel trade.<sup>22</sup> In addition, the price ceiling affects sales revenues for a given product, with a potentially differential impact on the total value of agreement in the different pharmacy chains. The price ceiling can impact the relative net value of agreement between the upstream firm and pharmacy chain due to differences in the response of demand and other chains' strategies ( $\theta_{ct}$ ) in the event of a disagreement. Thus, the interactions between pharmacy chain indicators and the price ceiling can help identify the bargaining parameters because the equilibrium effect of changes in net values of agreement is dependent on the bargaining parameters.

The necessary assumption for the price ceiling—and, thus, the interactions with pharmacy chain indicators—to be valid instruments is that the price ceiling is uncorrelated with  $\epsilon_t$ , conditional on the wholesale prices in other countries included in  $X_t$ . It is possible that the price ceiling—being a function of prices in several other countries, as described in Section 2.1—is correlated with the unobserved determinants of parallel trade costs. However, the UK is the only source country in our sample that is also in the reference countries for regulatory price ceilings, we believe this to be less of a concern and perform robustness checks with respect to this. Most prices in countries in  $X_t$  should help capture general movements in trade costs, exchange rates and relative prices between different locations.

## 4.4 Empirical Results

As our data contain a very large amount of choices, we draw a random sample of 50,000 individuals from the full sample of approximately 170,000 for estimating the individual choice model. We also restrict our

---

<sup>22</sup>The costs here are interpretable as both the total costs of parallel traders, e.g., procurement and handling, sales value in the source country and differences in import costs between Norway and the source country.

attention to the markets for the 40 and 80 mg versions, as parallel imports only entered late in our period of analysis for the 10 and 20 mg strengths, so we do not have sufficient data for a careful estimation. The maximum simulated likelihood estimates of the demand model are presented in Tables 4.4, 4.5 and 4.6. Table 4.4 presents the parameters, common across all unobserved groups and the parameters for the unobserved groups with differing values from the baseline  $g = 0$ , whose mean utility parameters are normalized to zero.

From the estimates of parameters governing preferences according to unobserved, discrete groups in the population in Table 4.4, there are two striking features. The first is that the statistical and economic significance of these parameters imply that the specification is appropriate, compared to a more usual mixed parameters logit specification with a single distribution for each coefficient.<sup>23</sup> The second is a pattern in which each group has a stronger relative preference for each of the pharmacy chains. This pattern seems reasonable, as one would suspect that many unobserved factors – such as travel distance or chain-store preference – would contribute to exactly such a pattern. Finally, if we consider the preference for parallel import versus direct import by adding all drug version specific effects for parallel and direct import, we find that on average, there is a preference for direct imports, but with quite a lot of heterogeneity. Finally, all  $\lambda$  parameters are in the (0,1) interval, as should be the case, and are precisely estimated.

Table 4.5 presents the distribution of the estimates of the chain-market specific choice set probabilities  $\theta$  that are also estimated within the maximum likelihood model according to restrictions from chains' profit-maximizing behavior informed by the observed wholesale prices. These estimates show that  $\theta$  varies across markets and chains and are on average between 0.58 and 0.82 for the 40 mg market and between 0.39 and 0.67 for the 80 mg one. The estimates also show that there are many corner solutions for which  $\theta$  is equal to one, meaning that both parallel imports and direct imports are always proposed by that chain in a given market (dosage-month combination). The median and 25% and 75% quantiles show that for some years, more than half of market-chains have  $\theta_{ct} = 1$ . We also report the mean across chains and markets of the estimated standard errors of  $\theta_{ct}$ , which show that they are precisely estimated. The standard errors of each  $\theta_{ct}$  are obtained using their censored normal asymptotic distributions, as described in Appendix 7.6.

Table 4.6 indicates that chain 1 performs significant foreclosure of direct imports in 2004, whereas chain 2 never does, and chain 3 does moderately for the market for the 40 mg dosage. In 2005, the picture is similar for the 40 mg market, with a bit less foreclosure of direct imports by chain 1, but on the 80 mg market, chains 2 and 3 start performing some foreclosure. In 2006, chain 2 starts performing foreclosure on the 40 mg market but still does not do so on the 80 mg market. Chain 1 continues performing quite substantial foreclosure in 2006 and 2007, whereas chain 3 performs less in 2006 but a bit more in 2007.

---

<sup>23</sup>The specification is a finite mixture of normal distributions. The economic significance is based on comparisons of behavioral implications under a simpler distributional specification not reported here.

Table 4.4: *Parameter estimates for consumer choice model with supply constraints*

Latent groups	$g = 0$	$g = 1$	$g = 2$	$g = 3$
$\tau_g$	0.07 –	0.26 (0.00)	0.29 (0.00)	0.38 (0.00)
$\eta_g$	0.00 –	1.37 (0.07)	1.46 (0.07)	1.75 (0.07)
Drug version specific taste ( $\delta_j^g + \sigma_j^g \nu_i^j$ )				
$\delta_0^g$	0.00 –	0.53 (0.04)	–0.35 (0.04)	–0.39 (0.04)
$\sigma_0^g$	0.22 (0.13)	0.02 (0.83)	0.98 (0.01)	0.81 (0.02)
Chain specific taste ( $\delta_c^g + \sigma_c^g \nu_i^c$ )				
$\delta_2^g$	0.00 –	4.09 (0.03)	1.94 (0.05)	–4.30 (0.11)
$\delta_2^g$	0.00 –	–0.97 (0.12)	6.46 (0.07)	–3.80 (0.10)
$\sigma_2^g$	3.01 (0.13)	6.50 (0.23)	7.96 (0.10)	2.67 (0.13)
$\sigma_3^g$	2.75 (0.16)	3.27 (0.13)	3.59 (0.07)	2.52 (0.13)
$\lambda_1, \lambda_2, \lambda_3$	0.32, 0.54, 0.54 (0.01), (0.01), (0.01)			
$\ln \mathcal{L}(\hat{\beta})$	–168,093			
$N$	50,000			

Note: one observation is a choice sequence of transactions by an individual. Standard errors in parentheses. The drug version specific taste is for parallel imports, and the reference is for direct imports. All  $\alpha_{0ct}, \alpha_{1ct}$  jointly estimated are not shown, whereas  $\theta_{0ct}, \theta_{1ct}$  are presented in Table 4.5.

Table 4.5: *Foreclosure parameter estimates  $\theta_{ct}$*

Strength	40 mg			80 mg			
	2004	2005	2006	2004	2005	2006	2007
Mean	0.588	0.828	0.613	0.679	0.394	0.621	0.62
25% percentile	0.010	0.835	0.015	0.042	0.016	0.029	0.20
Median	1.000	1.000	1.000	1.000	0.064	1.000	1.00
75% percentile	1.000	1.000	1.000	1.000	1.000	1.000	1.00
Mean std. err.	(0.159)	(0.182)	(0.113)	(0.240)	(0.118)	(0.166)	(0.193)

Note: the last row shows the mean across markets of the estimated standard errors of  $\theta_{ct}$ .

The results of the estimation of the bargaining parameters following the estimation method presented in Section 4.3.2 are presented in Table 4.7. The bargaining parameters are the bargaining weights of the

Table 4.6: *Foreclosure parameter estimates  $\theta_{ct}$* 

Strength		40mg			80mg			
Chain	Year	2004	2005	2006	2004	2005	2006	2007
1	Mean	0.008	0.522	0.404	0.037	0.041	0.016	0.358
	25% percentile	0.006	0.081	0.004	0.007	0.007	0.006	0.005
	Median	0.007	0.632	0.012	0.020	0.010	0.011	0.106
	75% percentile	0.010	0.880	1.000	0.035	0.020	0.024	0.858
	Mean std. err.	(0.180)	(0.346)	(0.105)	(0.263)	(0.055)	(0.148)	(0.203)
2	Mean	1.000	1.000	0.437	1.000	0.839	1.000	1.000
	25% percentile	1.000	1.000	0.018	1.000	0.979	1.000	1.000
	Median	1.000	1.000	0.152	1.000	1.000	1.000	1.000
	75% percentile	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Mean std. err.	(0.234)	(0.052)	(0.182)	(0.111)	(0.143)	(0.041)	(0.042)
3	Mean	0.756	0.962	1.000	1.000	0.302	0.847	0.502
	25% percentile	0.438	1.000	1.000	1.000	0.035	1.000	0.208
	Median	1.000	1.000	1.000	1.000	0.055	1.000	0.308
	75% percentile	1.000	1.000	1.000	1.000	0.414	1.000	1.000
	Mean std. err.	(0.062)	(0.147)	(0.052)	(0.347)	(0.155)	(0.310)	(0.333)

Note: the last row of each chain specific panel estimates lists the mean across markets of the estimated standard errors of  $\theta_{ct}$ .

upstream firms, that is, the manufacturer for direct imports and the parallel importers for parallel imports. Remark that the constraint that bargaining parameters should be between 0 and 1 is not imposed in our estimation. The GMM estimates of equation (4.8) are obtained using the Lipitor wholesale price in the UK and the Czech Republic converted to NOK per DDD, both interacted with the indicator for parallel imports as explanatory variables  $X_t$  and some instrumental variables  $Z_t$  including  $X_t$  and some excluded instruments such as indicators for chain identity and upstream firm type (parallel trader versus manufacturer), exchange rates NOK/USD, NOK/EUR, NOK/CZK, interactions of exchange rates with indicator for parallel trade and the inclusive value of the upstream firm interacted with upstream firm type. The inclusive value of upstream firm is the (average) log-sum of exponential utility for each upstream firm in the market. This instrumental variable comes out of the demand model and measures of how “valuable” the firm’s presence is to consumers in the market. Gowrisankaran et al. (2015) use this type of instrument, namely a “predicted willingness-to-pay for the hospital” when estimating the bargaining weights between hospitals and Managed Care Organizations in the US. In our case, these inclusive values measure the willingness to pay of customers for parallel imports or direct imports; they are estimated first using the consumer choice model and can explain why the manufacturer or parallel importer may be able to negotiate better wholesale prices with the pharmacy chain, thus serving to identify bargaining weights, as in Gowrisankaran et al. (2015).

From these estimates, we can see that perhaps unsurprisingly, the parallel importers wield a smaller bargaining weight, on average, compared to the originator (the manufacturer). Pharmacy retailers, which are concentrated in Norway, constitute an important gatekeeper for parallel trade companies who want to export to Norway.

Table 4.7: *Bargaining parameter estimates (GMM)*

	Manufacturer	Parallel Importer
Pharmacy Chain 1	0.95 (0.02)	0.50 (0.38)
Pharmacy Chain 2	0.55 (0.12)	0.26 (0.41)
Pharmacy Chain 3	0.67 (0.14)	0.32 (0.36)

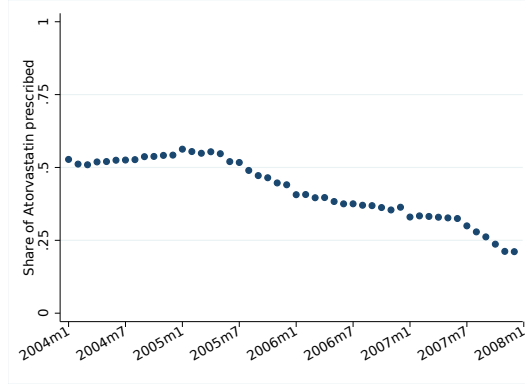
Note: estimates obtained using GMM equation (4.8), where  $X_t$  includes the UK and the Czech Republic Lipitor wholesale price interacted with indicator for parallel imports and excluded instruments in  $Z_t$  are exchange rates NOK/USD, NOK/EUR, NOK/CZK interacted with indicator for parallel trade and the inclusive value of the upstream firm interacted with upstream firm type.

## 4.5 Prescription behavior and parallel trade

One worry for the identification of our model is that doctors will change their prescription behavior if pharmacies induce consumers to consume parallel traded Lipitor more frequently. An example of what we have in mind is that consumers might oppose getting parallel traded drugs, thereby making their doctor prescribe them other types of statins for which there does not exist parallel traded alternatives. Over the sample period, there was an increase in the share of statin prescriptions going to Simvastatin due to new guidelines for statin prescriptions from the Norwegian Medicines Agency. This increase coincided with a similar decrease in the share of statin prescriptions going to Atorvastatin (the molecule contained in Lipitor), as shown in Figure 4.3. We regard this decrease as a function of the change in policy for statin prescriptions induced by the government, who implemented a lower price cap on Simvastatin than Atorvastatin, and not necessarily related to the preferences of consumers or doctors for directly imported versus parallel trade drugs.

We want to investigate the potential endogeneity issues arising from doctors responding to pharmacies strategies for selling parallel traded Lipitor by changing what statin they prescribe. Using data on the prescription behavior of individual doctors, we can look at the share of statin prescriptions going to Atorvastatin, together with the behavior of the pharmacies to which each doctor’s patients are exposed. This is feasible due

Figure 4.3: *Physicians' prescription of Atorvastatin as share of total statin prescription*



to availability of information linking the doctor to the prescription used by a patient for each transaction at each given pharmacy. Since we do not directly observe the behavior of pharmacies, we use information about the availability of parallel imports (assuming that if a pharmacy did not sell any parallel imports during a month, it means it was not available) and the ratio of margin for parallel and direct imports at a given pharmacy chain. The availability gives a sense of whether the doctor's patient potentially faced foreclosure of direct imports, whereas the margin can be thought of as a reduced-form measure of the pharmacy's decision to foreclose direct imports. To operationalize this, we calculate the weighted sum of availability and margin ratio in each chain for each doctor, where the measure is weighted by the share of the doctor's patients patronizing the different chains. More precisely, for doctor  $d$  in month  $t$

$$available_{dt} = \frac{1}{N_{dt}} \sum_{i=1}^{N_{dt}} \mathbf{1}_{\{parallel_{it}\}},$$

where  $N_{dt}$  is the number of patients for doctor  $d$  in month  $t$ , and  $\mathbf{1}_{\{parallel_{it}\}}$  is an indicator for whether patient  $i$  went to a pharmacy offering parallel traded Lipitor in month  $t$ . Similarly,

$$ratio_{dt} = \frac{1}{N_{dt}} \sum_{i=1}^{N_{dt}} \frac{m_{0c(i)t}}{m_{1c(i)t}},$$

where  $\frac{m_{0c(i)t}}{m_{1c(i)t}}$  is the ratio of margins for parallel (0) and direct (1) imported Lipitor at the pharmacy chain  $c(i)$  visited by patient  $i$  in month  $t$ . Overall, doctors prescribe Lipitor in 43% of the cases where a statin was prescribed, whereas parallel trade is available for 25% of the patients. The number of unique doctors in our sample is 14,051, who are observed for a maximum of 48 months between January 1, 2004 and December 31, 2007.



Table 4.8: *Effects of margins and availability of parallel imports on Atorvastatin prescription*

	(1)	(2)	(3)	(4)	(5)	(6)
$ratio_{dt} * available_{dt}$	-0.052*** (0.003)	-0.000 (0.003)	0.003 (0.003)	-0.036*** (0.002)	0.005** (0.002)	0.007*** (0.002)
$ratio_{dt}$	-0.018** (0.006)	-0.010 (0.006)	-0.013* (0.006)	-0.047*** (0.004)	-0.022*** (0.003)	-0.023*** (0.003)
Time trend		Yes			Yes	
Time FE			Yes			Yes
Physician FE				Yes	Yes	Yes
$N$	258,281	258,281	258,281	258,281	258,281	258,281
$R^2$	0.01	0.11	0.13	0.02	0.18	0.20

Note: OLS regression. Standard errors clustered by doctor. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .  
The dependent variable is the share of Atorvastatin prescribed by physician.

In Table 4.8, we present the results of OLS regressions of Atorvastatin's share of statin prescriptions on weighted margin ratios and parallel trade availability. The observation unit is a doctor-month. Column (1) shows a large negative coefficient on margin and availability, although this is driven by the overall downward trend in Atorvastatin prescriptions, together with a tendency for both the margin ratio and the availability of parallel trade to increase over time. This is confirmed by the coefficient on margin ratio going to a quite precisely estimated zero in Columns (2) and (3), where we add a linear time trend and time-fixed effects, respectively. When we add doctor-fixed effects together with a time trend or time-fixed effects in Columns (5) and (6), we obtain a positive coefficient on the margin ratio and a negative coefficient on availability, both of which are statistically significant. However, considering the size of the coefficients, none of them are economically significant. The coefficient on the margin ratio tells us that the effect of an increase of roughly two standard deviations (the standard deviation of that variable being 0.54), the Atorvastatin share of statin prescriptions will increase by roughly one half percentage point. Similarly for availability, an increase in availability from none to full would yield a decrease in Atorvastatin prescriptions by 2.2 percentage points. Considering that the average availability is 25%, this result implies that very large changes in pharmacies behavior is related to relatively small changes in the prescription behavior of doctors in our sample. We thus conclude that we should not be concerned by a potential identification problem due to doctors changing molecule prescriptions in response to pharmacies incentives to sell parallel traded Lipitor more frequently.

## 5 Counterfactual Simulations

Using our estimated model, we now study several counterfactual policies. The first counterfactual of interest investigates the role of parallel trade on market equilibrium, firms' profits and consumer welfare. Comparing

the current situation with the counterfactual equilibrium obtained absent parallel trade, we can better understand how parallel trade affects market outcomes. Then, given the findings on the negative impact of parallel trade on the sharing of profits between manufacturer and intermediaries in the pharmaceutical industry, we consider a possible regulation of pharmacies which would restrict their possibility to foreclose the choice of direct imports to consumers and capture a larger share of profits. Finally, we implement a counterfactual in which we also decrease the retail price of Lipitor by 20% and observe the new equilibrium wholesale prices and profits.

## 5.1 The Impact of Parallel Trade

Using our structural model, we can simulate a counterfactual situation in which parallel imports are unavailable, for instance, as the result of a ban. Pharmacy chains would then propose only the directly imported version of drugs and substitutions from parallel imports to the directly imported version would increase the demand for direct imports. As observed retail prices are equal to the regulated price ceilings even when parallel imports are present, retail prices will necessarily be equal to the regulated price ceilings when only direct imports are allowed. It is thus easy to identify the effect on demand of banning parallel trade since consumers will simply choose their preferred pharmacy chain.

In such a case, a consumer  $i$  chooses chain  $c$  with a counterfactual probability that is equal to the choice probability of the directly imported drug in chain  $c$ ,  $s_{i1ct_{noPI}}$ , and the aggregate counterfactual market share of chain  $c$  is

$$s_{1ct_{noPI}} = \int s_{i1ct_{noPI}} dF(\boldsymbol{\nu}_i|\boldsymbol{\beta}) = \int \frac{e^{V_{i1ct}}}{\sum_{\tilde{c}} e^{V_{i1\tilde{c}t}}} dF(\boldsymbol{\nu}_i|\boldsymbol{\beta}) \quad (5.1)$$

Once the counterfactual demand is known, we need to determine the counterfactual wholesale prices to compute profits. We assume that the same bargaining game is played, except that parallel importers are absent. In this bargaining, the pharmacy chains profits are zero in case of disagreement because they cannot sell parallel imports, and in case of agreement, their profits only depend on direct imports with  $\pi_{ct_{noPI}} = (\bar{p}_t - w_{1ct})s_{1ct_{noPI}}$ , where  $s_{1ct_{noPI}}$  is the demand for direct imports in chain  $c$  in the absence of parallel imports, as defined in (5.1). For the manufacturer, profits without parallel imports are given by  $\Pi_{t_{noPI}}(\mathbf{w}_{1t}) = \sum_c (w_{1ct} - c_t)s_{1ct_{noPI}}$  when agreeing with all chains.

Then, to determine the manufacturer profit in case of disagreement, we need to make an assumption about the shape of demand in case of bargaining disagreement between two parties. In our benchmark model, the pharmacy chain has the option of selling parallel imports, and thus consumers can shift to parallel imports within the chain or substitute away to other chains. In this counterfactual, without the possibility of offering parallel imports, if the manufacturer disagrees with a chain, consumers can only substitute direct imports

of other chains, thus allowing the manufacturer to obtain direct imports margins on all units demanded in every chain. Then, the manufacturer has the incentive to set all wholesale prices  $w_{1ct}$  equal to the price when bargaining with a chain the manufacturer would lose by agreeing a lower wholesale price in that chain. Then, as there is no loss to disagree with each chain because  $\Pi_{t_{noPI}} - \Pi_{-c,t_{noPI}} = 0$ , the wholesale prices  $w_{1ct}$  can equal the retail price ceiling  $\bar{p}_t$ , in which case the pharmacy chains profits are zero and the manufacturer profits are maximal. In this case, we obtain that the full market revenue is captured by the manufacturer, and pharmacy chains obtain zero profits.

We consider this as a first possible scenario that can be thought of as a upper bound on what the manufacturer can obtain. In this counterfactual case, computing the profit changes is very simple because pharmacy chains obtain zero profits and the manufacturer has a revenue equal to the retail price cap times the demand.

However, there is another possible scenario concerning the demand substitutions across chains when both parallel and direct imports are absent, which can be considered as the other extreme case. If instead, when there are no parallel imports, counterfactual demands in other chains  $\tilde{c} \neq c$  do not change when chain  $c$  disagrees with the manufacturer, meaning that customers' demand of direct imports in each chain would be completely lost in case of disagreement with the chain. Then, the Nash bargaining between the manufacturer and any chain  $c$  is the solution of

$$\max_{w_{1ct}} \{((w_{1ct} - c_t)s_{1ct_{noPI}})^{b_{1c}}((\bar{p}_t - w_{1ct})s_{1ct_{noPI}})^{1-b_{1c}}\}$$

which leads to the following simple determination of the wholesale price as

$$w_{1ct_{noPI}} = b_{1c}\bar{p}_t + (1 - b_{1c})c_t \quad (5.2)$$

However, whereas the estimation of our model allows identifying the bargaining parameter, it does not lead to identifying the marginal cost of production  $c_t$  using only the equilibrium wholesale price equations (3.9) and (3.11). We thus use some additional restrictions to identify a lower bound on the marginal cost by assuming that beforehand the simultaneous negotiation with the manufacturer and parallel importers, the pharmacy chains can decide to not use parallel imports and simply negotiate with the direct importer (manufacturer), in which case they would obtain the profits  $\pi_{ct_{noPI}}$  obtained in this counterfactual. We thus impose that for all chains  $c$  the counterfactual profits without parallel imports are lower or equal than the observed equilibrium profits, that is,

$$\pi_{ct_{noPI}} \equiv (\bar{p}_t - w_{1ct_{noPI}})s_{1ct_{noPI}} \leq \pi_{ct}^*$$

where  $\pi_{ct}^*$  is the observed equilibrium profit of chain  $c$ . Thus, we have the following lower bound for the counterfactual wholesale price at chain  $c$ :

$$w_{1ct_{noPI}} \geq \bar{p}_t - \frac{\pi_{ct}^*}{s_{1ct_{noPI}}}$$

These inequalities and equation (5.2) lead to the following lower bound for the marginal cost of production:

$$c_t \geq \bar{p}_t - \min_{c \in \{1, \dots, C\}} \left\{ \frac{1}{1 - b_{1c}} \frac{\pi_{ct}^*}{s_{1ct_{noPI}}} \right\}$$

Using this lower bound, we can obtain a lower bound on the wholesale price (5.2).

Then, a natural upper bound on the marginal cost is the minimum of all observed wholesale prices. As wholesale prices in source countries are typically less than the Norwegian direct and parallel imports wholesale prices, we can use the minimum of wholesale prices of source countries imports denoted  $p_{1ct}^I$  (as in section 3.3.1) for the ones at chain  $c$ . Thus, we obtain the following lower and upper bounds on marginal costs as

$$\bar{p}_t - \min_{c \in \{1, \dots, C\}} \left\{ \frac{1}{1 - b_{1c}} \frac{\pi_{ct}^*}{s_{1ct_{noPI}}} \right\} \leq c_t \leq \min_{c \in \{1, \dots, C\}} \{p_{1ct}^I\} \quad (5.3)$$

We can then obtain an upper and lower bound on the counterfactual wholesale prices, manufacturer profit and pharmacy chains profit using minimum and maximum values of counterfactuals over the set of marginal costs in the obtained interval of (5.3). Table 7.3 in Appendix 7.8 presents the estimates of these bounds on marginal costs together with a few other descriptive statistics.

Table 5.1 presents the counterfactual changes in quantities ( $\Delta q$ ), wholesale prices of direct imports ( $\Delta w_1$ ) and profits ( $\Delta \pi$  and  $\Delta \Pi$ ) from the observed equilibrium to the counterfactual case in which we remove parallel trade. As we can see, the demand changes are point identified using simply (5.1) and do not depend on the scenario of the demand shape in case of disagreement in bargaining. However, the bargaining outcome in wholesale prices and thus in profits depends on those scenarios. In the case of scenario, we obtain trivially that whereas wholesale prices and profits are not point identified, we can identify bounds as explained above, and Table 5.1 demonstrates that the bounds are still quite informative. First, the quantity sales in case of removing parallel imports are such that of course parallel imports disappear and aggregate demand is redistributed between direct imports across the three chains so that chain 1 sells more Lipitor than before the ban while chains 2 and 3 sell less as there is less substitution towards direct imports than the initial parallel imports sold by these chains. Then, even if changes in wholesale prices and profits are not set identified, the results demonstrate that the change in profits would favor the upstream manufacturer and penalize pharmacy

chains that would not be able to use intra-brand competition between parallel trade and direct imports to extract part of manufacturers profits. The total revenue or total profit (total cost of producing aggregate demand does not change) of the manufacturer would increase between 15.56 and 20.29 millions NOK per year, the parallel trader would disappear (their 1.13 millions profit would disappear) and pharmacy chains would lose significantly, with chain 1 losing much more than the others because accepting a much higher wholesale price.

Given our assumption of no substitution away from a chain of their sales of direct imports if not selling direct imports while parallel imports are absent, this can be considered as a lower bound on what the pharmacies would lose and what the manufacturer would gain. As mentioned above, in the opposite case of full substitution away to other chains when both parallel and direct imports are absent from a chain, the pharmacy chains would lose 100% of their profits and the manufacturer would gain even more, with 27.07 millions of NOK per year instead of between 15.56 and 20.29.

In any case, this counterfactual shows that banning parallel imports would benefit the manufacturer and reduce substantially the profit of retail chains. However, depending on the demand substitutions away from a chain when it disagrees with the manufacturer while it has no parallel imports to offer, the change in profits of the manufacturer can be almost half (potentially as low as 15.56 millions NOK) that in the case in which there is full substitution across chains (27.07 millions NOK). This result shows that the bargaining power of pharmacy chains can help them obtain important profits if demand does not substitute perfectly across chains even when there are no parallel imports.

Table 5.1: *Impact of removing parallel imports*

	$\Delta q_0$	$\Delta q_1$	$\Delta w_1$	$\Delta \pi$
Pharmacy Chain 1	-12.56 -100%	14.67 536%	[0.71, 0.75] [35%, 37%]	[-12.21, -11.59] [-93%, -89%]
Pharmacy Chain 2	-5.08 -100%	4.03 108%	[0.04, 0.37] [2%, 18%]	[-3.83, -1.58] [-51%, -21%]
Pharmacy Chain 3	-6.27 -100%	5.20 245%	[0.25, 0.50] [12%, 24%]	[-4.91, -3.05] [-68%, -42%]
$\Delta \Pi$				
Manufacturer		23.91 278%	[0.33, 0.54] [17%, 27%]	[15.56, 20.29] [24%, 31%]
Parallel Importer	-23.91 -100%			-1.13 -100%

Note: numbers in brackets indicate the lower and upper bounds when the outcome is not point identified. Quantities are in millions of DDD per year. Prices are in NOK, and profits are in millions of NOK per year.

## 5.2 The Impact of Direct Imports Foreclosure by Pharmacy Chains

We now consider a different policy in which, e.g., some regulation would prevent pharmacies from foreclosing access to directly imported versions of drugs to consumers. Under such a policy, parallel imports are allowed and used by pharmacy chains, but pharmacies are not allowed to propose only parallel imports to consumers. With our estimates of the bargaining model, parallel importer costs and manufacturer opportunity prices in source countries, it is also possible to assess the impact of the pharmacy chains' strategic foreclosure of parallel imports done by optimally choosing the probability with which a drug will be proposed to the consumer. We consider the case in which each pharmacy chain  $c$  has to propose both versions, meaning that it sets  $\theta_{ct} = 1$ . Inspection of pharmacies' offerings to consumers would easily allow implementing such a regulation. Our estimates show that among the chain-market combinations featuring parallel imports, the estimated  $\theta$  varies significantly between zero and one but is less than one on average for 45% of markets-chains, meaning that the consumer will face a restricted choice set in those instances. The quantitative effect of setting  $\theta$  equal to one on the pharmacy chain demand will depend on the preferences of the consumers. Moreover, when the pharmacy chains are required to always propose both varieties, it will also have an effect on the bargained wholesale prices between the upstream firms—the direct and parallel importers—and the pharmacy chains. This implies that the wholesale prices in general will increase, since there is no longer an incentive for the upstream firms to reduce wholesale prices to increase sales.

This counterfactual situation is simulated using the same bargaining model as in Section 3.3 with the estimated bargaining parameters but in which the counterfactual demand model is obtained by imposing pharmacies to set  $\theta_{ct} = 1$ . For these calculations, we take not only consumer preferences but also the prices in the source country for the parallel importer and the wholesale price in the source country obtained by the manufacturer as given. We then solve for demand and the bargaining outcomes. Solving for the new wholesale price equilibrium is obtained by solving the Nash bargaining model between the manufacturer or parallel importer and the pharmacy chains.

To be more precise, in this case where foreclosure is prevented, the counterfactual individual choice probabilities  $s_{ikct}$  are given by  $s_{ikct} = s_{ict}s_{ikt|c}$ , where the choice probability of version  $k$  of the drug conditional on pharmacy chain  $c$  is as before:

$$s_{ikt|c} = s_{ikt|c,B} = \frac{e^{V_{ikct}/\lambda_c}}{e^{V_{i0ct}/\lambda_c} + e^{V_{i1ct}/\lambda_c}} \text{ with } k' = 1 - k$$

However, now, the expected consumer utility of visiting pharmacy  $c$  is

$$I_{ict} \equiv E_{\epsilon_{ikct}} \left[ \max_{k \in \{0,1\}} (V_{ikct} + \lambda_c \epsilon_{ikct}) \right] = \lambda_c \ln \left( \sum_{k \in \{0,1\}} e^{V_{ikct}/\lambda_c} \right)$$

such that

$$s_{ict} = \frac{e^{I_{ict}}}{\sum_{\bar{c}} e^{I_{i\bar{c}t}}} = \frac{\left( \sum_{k \in \{0,1\}} e^{V_{ikct}/\lambda_c} \right)^{\lambda_c}}{\sum_{\bar{c}} \left( \sum_{k \in \{0,1\}} e^{V_{ik\bar{c}t}/\lambda_{\bar{c}}} \right)^{\lambda_{\bar{c}}}}$$

which gives the following individual choice probability:

$$s_{ikct} = \frac{\left( \sum_{k \in \{0,1\}} e^{V_{ikct}/\lambda_c} \right)^{\lambda_c - 1} e^{V_{ikct}/\lambda_c}}{\sum_{\bar{c}} \left( \sum_{k \in \{0,1\}} e^{V_{ik\bar{c}t}/\lambda_{\bar{c}}} \right)^{\lambda_{\bar{c}}}}$$

Then, the manufacturer-retailer and parallel importer-retailer Nash bargaining will yield a new wholesale prices equilibrium that satisfies the set of equations for all  $c \in 1, \dots, C$ :

$$\Delta_1 \pi_{ct} = \frac{1 - b_{1c}}{b_{1c}} \Delta_c \Pi_t \quad \text{and} \quad \Delta_0 \pi_{ct} = \frac{1 - b_{0c}}{b_{0c}} \Delta_c \Pi_t^{PI}$$

where

$$\begin{aligned} \Delta_c \Pi_t &= \sum_{\bar{c}} (w_{1\bar{c}t} \Delta_{1c} s_{1\bar{c}t} + p_{1\bar{c}t}^I \Delta_{1c} s_{0\bar{c}t}) \\ \Delta_1 \pi_{ct} &= (\bar{p}_t - w_{1ct}) s_{1ct} + (\bar{p}_t - w_{0ct}) \Delta_{1c} s_{0ct} \\ \Delta_0 \pi_{ct} &= \pi_{ct} - \pi_{-0,ct} = (\bar{p}_t - w_{1ct}) \Delta_{0c} s_{1ct} + (\bar{p}_t - w_{0ct}) s_{0ct} \end{aligned}$$

In Table 5.2 we present the changes from the current situation in terms of quantities ( $\Delta q$ ), wholesale prices ( $\Delta w$ ), profits of the pharmacy chains ( $\Delta \pi$ ) and profits of the upstream firm ( $\Delta \Pi$ ). We see that such a regulatory change would have an impact on sales of parallel imports and direct imports and on profits. First, preventing partial foreclosure of direct imports would raise total direct import sales by 10.29 millions DDD per year (reducing sales of Parallel Imports by the same amount). The largest part of this substitution would occur at chain 1. Then, as wholesale prices would increase, the three pharmacy chains lose profits, with losses from 5% at chain 2 to 10% at chain 1. We see that the manufacturer would gain from such a change, with an overall increase of revenue of 0.97 millions NOK per year. This increase occurs because there is no longer an element of competition for the upstream firm when bargaining over wholesale prices with the chains, such that the manufacturer wins both because of an increase in wholesale price of direct imports and a substitution between parallel imports and direct imports. However, the total sales of parallel imports in

Norway would still be very substantial, as they represent on average more than 70% of sales quantities to start with. Parallel importers would earn because the increases in wholesale prices more than compensate for the decreases in quantities sold. Pharmacy chains would lose more than the manufacturer earns because parallel importers would also gain with such policy. This experiment thus shows that even if it is possible to reduce the part of profits obtained by pharmacy chains at the expense of the manufacturer, it would then imply that some profits would also be shifted from pharmacy chains to the parallel trade companies.

Table 5.2: *Impact of preventing parallel imports foreclosure ( $\theta_{ct} = 1$ )*

	$\Delta q_0$	$\Delta q_1$	$\Delta w_0$	$\Delta w_1$	$\Delta \pi$
Pharmacy Chain 1	-8.75 -70%	9.38 343%	0.09 5%	0.07 4%	-1.26 -10%
Pharmacy Chain 2	-0.23 -5%	-0.13 -4%	0.01 0%	0.01 1%	-0.38 -5%
Pharmacy Chain 3	-1.30 -21%	1.04 49%	0.02 1%	0.02 1%	-0.53 -7%
					$\Delta \Pi$
Manufacturer		10.29 120%		0.03 2%	0.97 1%
Parallel Importer	-10.29 -43%		0.04 2%		0.36 32%

Note: quantities are in millions of DDD per year. Prices are in NOK and profits are in millions of NOK per year.

### 5.3 Decrease in the Price Ceiling and Preventing Foreclosure

We now perform a third counterfactual that consists of preventing foreclosure but also reducing the retail prices by 20%. In this counterfactual, as new wholesale prices may decrease substantially and parallel importers profits are low, it is likely that participation constraints of positive profits of parallel importers may bind, leading to exit of parallel imports from some markets. We thus check for positive profits for the new counterfactual Nash-in-Nash equilibrium and search for the counterfactual equilibrium in which no firm has non-positive profits.

From Table 5.3, we can see that such retail price decrease that leads to a 20% decrease in total drug expenses by the government on this market has a much lower effect on the manufacturer but a large effect on reducing rents of pharmacy retailers. In fact, the 20% retail price decrease leads to a wholesale price decrease of direct imports of only 1% in chain 1, 8% in chain 2 and 4% in chain 3. The sales of direct imports increase substantially in chain 1, and thus the total profits of the manufacturer decrease by only 3% that is on average



Table 5.3: *Impact of preventing parallel imports foreclosure and reducing the price ceiling by 20%*

	$\Delta q_0$	$\Delta q_1$	$\Delta w_0$	$\Delta w_1$	$\Delta \pi$
Pharmacy Chain 1	-8.76 -70%	9.42 344%	-0.04 -2%	-0.03 -1%	-8.88 -68%
Pharmacy Chain 2	-0.27 -5%	-0.08 -2%	-0.44 -22%	-0.17 -8%	-3.35 -45%
Pharmacy Chain 3	-1.45 -23%	1.15 54%	-0.17 -9%	-0.08 -4%	-4.45 -62%
					$\Delta \Pi$
Manufacturer		10.49 122%		-0.09 -4%	-1.92 -3%
Parallel Importer	-10.49 -44%		-0.22 -11%		-0.47 -42%

Note: quantities are in millions of DDD per year. Prices are in NOK, and profits are in millions of NOK per year.

1.92 millions NOK per year, whereas total expenses decrease by 19.47 million NOK per year (20% of the total expenses on these markets).

With other retail price reduction amounts, the effects are qualitatively similar, but with a 10% retail price reduction, for example, the effect on the manufacturer profit is even smaller and negligible, whereas most of the reduction in expenses is then attributed to a reduction in pharmacy chains profits. Table 7.8 in Appendix 7.8 shows the changes in profits for different retail price reductions from 10 to 30%.

As mentioned earlier, when performing these counterfactuals, we also need to check that the price reduction still allows parallel importers to remain in the market. It is possible that when the retail price is too small, some parallel importers exit the market because their source cost corresponding to some wholesale price in a source country is too high compared to the maximum price allowed in Norway. When checking the participation constraints, indeed, some parallel traders' non-negative profits are binding, and they exit some markets. In the case of a 20% price reduction, there are approximately 10% of chains-months-parallel importer combinations for which there is exit, that is, 19 chain-market exits of parallel traders. Of course, when some parallel trader stops dealing with a chain in a given market, it reduces the competition between chains on that market and marginally benefits the manufacturer, also, as the retail price regulation constraint make parallel imports exit from some chains and markets. Table 7.8 in Appendix 7.8 also reports the number of chain-markets exits according to the price reduction.

## 6 Conclusion

In this paper, we investigate the incentives of pharmacy chains in selling parallel traded drugs. Our estimates demonstrate that foreclosure of directly imported drugs is plausibly used by pharmacy chains to increase profits and bargaining position relative to the manufacturer. This behavior is driven by parallel importers generally giving the pharmacy chains lower wholesale prices than the manufacturer. We find that the possibility of foreclosure of direct imports by pharmacy chains is at the expense of the manufacturer. In this market, where prices are constrained by regulation, being able to distort offers between the varieties, pharmacy chains effectively introduce competition between the upstream suppliers. In our counterfactual simulations, we also find that a lower retail price may not be very detrimental to the manufacturer as it can squeeze the possibility of pharmacies to extract some of the rent using parallel imports.

The specific random foreclosure mechanism that we highlight—in which pharmacies can distort availability of drugs for which they have differing margins—has not been formalized in the previous literature, although pharmacists’ incentives have been mentioned as a plausible factor impacting sales of drugs for which substitution at the pharmacy level is available (see, e.g., Caves et al., 1991). The incentives to distort availability seems particularly important in many European countries, where price regulation is prevalent.

Furthermore, we show how we can identify a consumer demand model when choice sets are unobserved to the econometrician but modeling of the retailer incentives to choose the optimal set of product varieties to propose allows recovering all preferences parameters. In our case, this is achieved by using rich data regarding retailers’ (pharmacies) margins, in which the incentives of the retailer to foreclose partially the access of the less profitable products is clear, even though it might reduce the retailer’s attractiveness to consumers. In other settings, the method developed in this paper may be useful for studying the foreclosure and strategic behavior of retailers or intermediaries who can affect choice sets strategically.

Finally, this paper shows that we should also consider the vertical relationships and market structure of pharmacy retailing in the debate on the impact of parallel trade on long-run welfare. In fact, parallel trade can be considered as a threat to third-degree price discrimination and might result in a manufacturer only serving high-demand markets (Malueg and Schwartz, 1994). Danzon et al. (2005) already shows that launch delays are correlated with price regulation. However, Grossman and Lai (2008) have shown that when regulators respond optimally to the presence of parallel trade, i.e., determining price regulation to trade off static and dynamic efficiency, international intellectual property rights exhaustion might lead to more innovation than if parallel trade is not possible. This hinges on the regulators in each country being able to fully incorporate the effect the price ceilings they set have on innovation and on being able to politically trade off price levels and innovation in an optimal manner. This paper shows that it may be important to consider the pharmacy

retailing structure and regulation when setting optimal price levels because unregulated pharmacy chains may manage to extract a large part of the margin reward to innovators. We leave for future research the study of optimal price regulation across countries when parallel trade and strategic pharmacies interact with the pharmaceutical industry manufacturers.

## References

- ACEMOGLU, D. AND J. LINN (2004): “Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry,” *Quarterly Journal of Economics*, 119, 1049–1090.
- BERRY, S. (1994): “Estimating discrete-choice models of product differentiation,” *The RAND Journal of Economics*, 25, 242–262.
- BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): “Automobile prices in market equilibrium,” *Econometrica: Journal of the Econometric Society*, 63, 841–890.
- BREKKE, KURT R., T. H. H. AND O. R. STRAUME (2013): “Margins and market shares: Pharmacy incentives for generic substitution,” *European Economic Review*, 61, 116–131.
- (2015): “Price Regulation and Parallel Imports of Pharmaceuticals,” *Journal of Public Economics*, 129, 92–105.
- BRONNENBERG, B. J., J.-P. DUBÉ, M. GENTZKOW, AND J. SHAPIRO (2015): “Do Pharmacists Buy Bayer? Sophisticated Shoppers and the Brand Premium,” *Quarterly Journal of Economics*, 130, 1669 – 1726.
- CARDELL, S. N. (1997): “Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity,” *Econometric Theory*, 13, 185–213.
- CAVES, R., M. WHINSTON, M. HURWITZ, A. PAKES, AND P. TEMIN (1991): “Patent expiration, entry, and competition in the US pharmaceutical industry,” *Brookings papers on economic activity. Microeconomics*, 1991, 1–66.
- CONLON, C. AND J. H. MORTIMER (2013): “Demand estimation under incomplete product availability,” *American Economic Journal-Microeconomics*, 5, 1–30.
- COSTA-FONT, J. (2016): “Is medicines parallel trade regulatory arbitrage?” *International Journal of Health Care Finance and Economics*, 16, 321 – 336.
- CRAWFORD, G., R. GRIFFITH, AND A. IARIA (2017a): “Preference Estimation with Unobserved Choice Set Heterogeneity Using Sufficient Sets,” *CEP Discussion Paper 11675*.
- CRAWFORD, G., R. LEE, M. WHINSTON, AND A. YURUKOGLU (2017b): “The Welfare Effects of Vertical Integration in Multichannel Television Markets,” *NBER Working Paper 21832*.

- CRAWFORD, G. S. AND A. YURUKOGLU (2012): “The Welfare Effects of Bundling in Multichannel Television Markets,” *American Economic Review*, 102, 643–85.
- DANZON, P. AND L. CHAO (2000): “Cross-National Price Differences for Pharmaceuticals: How Large and Why?” *Journal of Health Economics*, 19, 159–195.
- DANZON, P., Y. R. WANG, AND L. WANG (2005): “The Impact of Price Regulation on the Launch Delay of New Drugs – Evidence from Twenty-Five Major Markets in the 1990s,” *Health Economics*, 14, 269–292.
- DRAGANSKA, M., M. MAZZEO, AND K. SEIM (2009): “Beyond plain vanilla: Modeling joint product assortment and pricing decisions,” *Quantitative Marketing and Economics*, 7, 105–146.
- DUBOIS, P., O. DE MOUZON, F. SCOTT-MORTON, AND P. SEABRIGHT (2015): “Market size and pharmaceutical innovation,” *Rand Journal of Economics*, 46, 844–871.
- DUSO, T., A. HERR, AND M. SUPPLIET (2014): “The Welfare Impact of Parallel Imports: A Structural Approach Applied to the German Market for Oral Anti-diabetics,” *Health Economics*, 23, 1036–1057.
- GANSLANDT, M. AND K. E. MASKUS (2004): “Parallel imports and the pricing of pharmaceutical products: evidence from the European Union,” *Journal of Health Economics*, 23, 1035 – 1057.
- GAYNOR, M., C. PROPPER, AND S. SEILER (2016): “Free to Choose? Reform, Choice, and Consideration Sets in the English National Health Service,” *American Economic Review*, 106, 3521–57.
- GOEREE, M. S. (2008): “Limited Information and Advertising in the U.S. Personal Computer Industry,” *Econometrica*, 76, 1017–1074.
- GOWRISANKARAN, G., A. NEVO, AND R. TOWN (2015): “Mergers When Prices are Negotiated: Evidence from the Hospital Industry,” *American Economic Review*, 105, 172–203.
- GRENNAN, M. (2013): “Price Discrimination and Bargaining: Empirical Evidence from Medical Devices,” *American Economic Review*, 103, 145–177.
- GROSSMAN, G. M. AND E. L.-C. LAI (2008): “Parallel Imports and Price Controls,” *The RAND Journal of Economics*, 39, 378 – 402.
- HO, K. AND R. LEE (2017): “Insurer Competition in Health Care Markets,” *Econometrica*, 85, 379–417.
- HORN, H. AND A. WOLINSKY (1988): “Bilateral Monopolies and Incentives for Merger,” *The RAND Journal of Economics*, 19, 408 – 419.

- IZUKA, T. (2013): “Physician Agency and Adoption of Generic Pharmaceuticals,” *American Economic Review*, 102, 2826–2858.
- JUDD, K. L. (1998): *Numerical Methods in Economics*, vol. 1 of *MIT Press Books*, The MIT Press.
- KANAVOS, P., J. COSTA-FONT, S. MERKUR, AND M. GEMMILL (2004): *The economic impact of pharmaceutical parallel trade in European Union member states: A stakeholder analysis*, London School of Economics and Political Science London.
- KANAVOS, P., J. COSTA-FONT, AND E. SEELEY (2008): “Competition in off-patent drug markets: Issues, regulation and evidence,” *Economic Policy*, 23, 499–544.
- KANAVOS, P. AND S. VANDOROS (2010): “Competition in prescription drug markets: is parallel trade the answer?” *Managerial and Decision Economics*, 31, 325–338.
- KYLE, M. (2009): *Parallel Trade in Pharmaceuticals: Firm Responses and Competition Policy*, Chapter 13 in *International Antitrust Law & Policy: Fordham Competition Law 2009*, edited by Barry Hawk (Juris Publishing, New York).
- KYLE, M. K. (2007): “Pharmaceutical Price Controls and Entry Strategies,” *The Review of Economics and Statistics*, 89, 88–99.
- (2011): “Strategic responses to parallel trade,” *The BE Journal of Economic Analysis & Policy*, 11, 1–32.
- MAINI, L. AND F. PAMMOLLI (2017): “Reference Pricing as a Deterrent to Entry: Evidence from the European Pharmaceutical Market,” *mimeo*.
- MALUEG, D. A. AND M. SCHWARTZ (1994): “Parallel imports, demand dispersion, and international price discrimination,” *Journal of International Economics*, 37, 167 – 195.
- SAKSHAUG, S., K. FURU, Ø. KARLSTAD, M. RØNNING, AND S. SKURTVEIT (2007): “Switching statins in Norway after new reimbursement policy – a nationwide prescription study,” *British Journal of Clinical Pharmacology*, 64, 476–481.

## 7 Appendix

### 7.1 Parallel trade products



Figure 7.1: *Example of parallel trade and direct imported products (outside and inside)*

### 7.2 Alternative demand model

In this alternative model, the consumer makes a choice over which pharmacy chain  $c$  to visit and which product to purchase depending on the available choices in the pharmacies. For a given active ingredient, the choice set at pharmacies can be  $\{PI\}$ ,  $\{DI\}$  or  $B \equiv \{DI, PI\}$ . As we don't observe the choice set, we denote  $\theta_{ct}^0$  and  $\theta_{ct}^1$  the probabilities that the choice sets are  $\{PI\}$  or  $\{DI\}$  respectively and thus  $1 - \theta_{ct}^0 - \theta_{ct}^1$  the probability that the choice set is  $B = \{DI, PI\}$ . We assume that the utility of consumer  $i$  is given by

$$u_{ikct} = V_{ikct} + \epsilon_{ikct}$$

where  $V_{ikct}$  is the mean utility consumer  $i$  obtains from choosing the drug of origin  $k$  in pharmacy chain  $c$  in market  $t$ , and  $\epsilon_{ikct}$  is an idiosyncratic i.i.d. random utility component, that we assume distributed independently across drugs and chains according to a Gumbel distribution.

Then, without loss of generality, denoting  $k' = 1 - k$ , we can write the choice probability that consumer  $i$  purchases version  $k$  in pharmacy 1 (other probabilities can be written in a similar way):

$$\begin{aligned}
s_{ik1t} (\theta_{1t}^0, \theta_{2t}^0, \theta_{3t}^0, \theta_{1t}^1, \theta_{2t}^1, \theta_{3t}^1) &= \theta_{1t}^k \theta_{2t}^k \theta_{3t}^k \frac{e^{V_{ik1t}}}{\sum_{c=1}^3 e^{V_{ikct}}} \\
&+ \theta_{1t}^k \theta_{2t}^k \theta_{3t}^{k'} \frac{e^{V_{ik1t}}}{e^{V_{ik'3t}} + \sum_{c=1}^2 e^{V_{ikct}}} \\
&+ \theta_{1t}^k \theta_{2t}^k (1 - \theta_{3t}^k - \theta_{3t}^{k'}) \frac{e^{V_{ik1t}}}{e^{V_{ik'3t}} + \sum_{c=1}^3 e^{V_{ikct}}} \\
&+ \theta_{1t}^k \theta_{3t}^k \theta_{2t}^{k'} \frac{e^{V_{ik1t}}}{e^{V_{ik'2t}} + \sum_{c=1,3} e^{V_{ikct}}} \\
&+ \theta_{1t}^k \theta_{3t}^k (1 - \theta_{2t}^k - \theta_{2t}^{k'}) \frac{e^{V_{ik1t}}}{e^{V_{ik'2t}} + \sum_{c=1}^3 e^{V_{ikct}}} \\
&+ (1 - \theta_{1t}^k - \theta_{1t}^{k'}) \theta_{2t}^k \theta_{3t}^k \frac{e^{V_{ik1t}}}{e^{V_{ik'1t}} + \sum_{c=1}^3 e^{V_{ikct}}} \\
&+ (1 - \theta_{1t}^k - \theta_{1t}^{k'}) \theta_{2t}^k \theta_{3t}^{k'} \frac{e^{V_{ik1t}}}{e^{V_{ik'1t}} + e^{V_{ik'3t}} + \sum_{c=1}^2 e^{V_{ikct}}} \\
&+ (1 - \theta_{1t}^k - \theta_{1t}^{k'}) \theta_{2t}^k (1 - \theta_{3t}^k - \theta_{3t}^{k'}) \frac{e^{V_{ik1t}}}{\sum_{c=1}^3 e^{V_{ikct}} + e^{V_{ik'1t}} + e^{V_{ik'3t}}} \\
&+ (1 - \theta_{1t}^k - \theta_{1t}^{k'}) \theta_{3t}^k \theta_{2t}^{k'} \frac{e^{V_{ik1t}}}{e^{V_{ik'1t}} + e^{V_{ik'2t}} + \sum_{c=1,3} e^{V_{ikct}}} \\
&+ (1 - \theta_{1t}^k - \theta_{1t}^{k'}) \theta_{3t}^k (1 - \theta_{2t}^k - \theta_{2t}^{k'}) \frac{e^{V_{ik1t}}}{e^{V_{ik'1t}} + e^{V_{ik'2t}} + \sum_{c=1}^3 e^{V_{ikct}}} \\
&+ (1 - \theta_{1t}^k - \theta_{1t}^{k'}) (1 - \theta_{2t}^k - \theta_{2t}^{k'}) (1 - \theta_{3t}^k - \theta_{3t}^{k'}) \frac{e^{V_{ik1t}}}{\sum_{c=1}^3 e^{V_{ikct}} + \sum_{c=1}^3 e^{V_{ik'ct}}}
\end{aligned}$$

With  $F(\cdot)$  the cumulative distribution function of consumer preferences  $\mathbf{V}_{it} \equiv (V_{i01t}, \dots, V_{iC1t}, V_{i11t}, \dots, V_{i1Ct})$ , the aggregate choice probability or market share of drug  $k$  sold by  $c$  at  $t$  is

$$s_{kct} (\theta_{1t}^0, \theta_{2t}^0, \theta_{3t}^0, \theta_{1t}^1, \theta_{2t}^1, \theta_{3t}^1) = \int s_{ikct} (\theta_{1t}^0, \theta_{2t}^0, \theta_{3t}^0, \theta_{1t}^1, \theta_{2t}^1, \theta_{3t}^1) dF(\mathbf{V}_{it})$$

### 7.3 Pharmacy retail pricing with price ceiling

Here we show that a pharmacy chain offering two goods,  $PI$  ( $j = 0$ ) and  $DI$  ( $j = 1$ ), subject to a common price ceiling  $\bar{p}$  will sometimes choose to price both goods at the price ceiling, even if consumers have a preference for one of the two. Let's assume that consumers have a preference for  $DI$ , such that  $PI$  will be bought to a lower extent if prices and availability are equal. It can be shown that the chosen prices will both sometimes be at the price ceiling and that the extent of pharmaceutical coverage and "tightness" of the price ceiling will make this even more likely.



Let the demand for each good  $j$  at pharmacy  $c$  be given by  $q_{jc}(p_{0c}, p_{1c}, p_{0-c}, p_{1-c})$ , where  $p_{jc}$  is the price paid by the consumer for good  $j$  in pharmacy  $c$ . The price set by the firm,  $r_{jc}$  is related to the price paid by the consumer through  $p_{jc} = \tau r_{jc}$ , where  $0 \leq \tau \leq 1$  is the co-payment rate. The profits of pharmacy chain  $c$  is given by

$$\pi_c = q_{0c}(r_{0c} - w_{0c}) + q_{1c}(r_{1c} - w_{1c}),$$

where  $w_{jc}$  is the pharmacy chain's wholesale price for good  $j$ . In a Nash equilibrium, given prices in other chains, the pharmacy chain solves the problem:

$$\max_{r_{0c}, r_{1c}} \pi_c \quad \text{s.t.} \quad r_{0c}, r_{1c} \leq \bar{p},$$

with the corresponding complementary slackness conditions for each  $j \in \{0, 1\}$

$$q_{jc} + \tau \frac{\partial q_{0c}}{\partial p_{jc}}(r_{0c} - w_{0c}) + \tau \frac{\partial q_{1c}}{\partial p_{jc}}(r_{1c} - w_{1c}) \geq 0, \quad r_{jc} \leq \bar{p}.$$

Assume that the price ceiling is sufficiently low to bind for good 1 ( $r_{1c} = \bar{p}$ ), which is the one which consumers value the most and will command the highest price in the absence of the price ceiling. To see that the pharmacy could find it optimal to price at the ceiling also for the other product, note that the unconstrained price for good 0 in this case would be

$$r_{0c}^* = w_{0c} + \frac{q_{0c}}{-\tau \frac{\partial q_{0c}}{\partial p_{0c}}} + \frac{\frac{\partial q_{1c}}{\partial p_{0c}}}{-\frac{\partial q_{0c}}{\partial p_{0c}}}(\bar{p} - w_{1c}).$$

It is straightforward to see that  $r_{0c}^*$  could exceed  $\bar{p}$  if the price ceiling is tight enough. From the second term, we see that the lower the co-payment rate  $\tau$ , the less responsive consumers are to any change in the retail price  $p_{0c}$ , thus increasing the optimal unconstrained price  $r_{0c}^*$ . A lower price ceiling will tend to reduce  $r_{0c}^*$  through the reduced sales of 0, since the price of good 1 becomes lower, and through reducing the profit margin on good 1, which lowers the value of the diverted sales to good 1 with an increase in the price of good 0, but unless the price of good 0 responds too much to a change in the price of good 1 (i.e., the slope of the "reaction function" is too large), it will be possible for both goods to be constrained by the price ceiling simultaneously.

## 7.4 Details on derivatives of $\theta_t^*(\mathbf{w}_{0t}, \mathbf{w}_{1t})$

In order to obtain how wholesale prices affect the equilibrium  $\theta_t^*(\mathbf{w}_{0t}, \mathbf{w}_{1t})$ , let  $\mathbf{F}_{\theta,t}$  denote the vector of derivatives of pharmacy chain profit with respect to direct import availability at time  $t$ , i.e.

$$\mathbf{F}_{\theta,t} \equiv \left( \frac{\partial \pi_{1t}}{\partial \theta_{1t}}, \frac{\partial \pi_{2t}}{\partial \theta_{2t}}, \dots, \frac{\partial \pi_{Ct}}{\partial \theta_{Ct}} \right)'$$

Implicit differentiation of the system of first order conditions  $\mathbf{F}_{\theta,t} = \mathbf{0}$  yields

$$\frac{\partial \mathbf{F}_{\theta,t}}{\partial \theta_t'} \Big|_{\theta_t = \theta_t^*} d\theta_t + \frac{\partial \mathbf{F}_{\theta,t}}{\partial \mathbf{w}'_{1t}} \Big|_{\theta_t = \theta_t^*} d\mathbf{w}_{1t} = \mathbf{0}.$$

The Jacobian of  $\theta_t^*(\mathbf{w}_{0t}, \mathbf{w}_{1t})$  with respect to  $\mathbf{w}_{1t}$  is then

$$\frac{\partial \theta_t^*}{\partial \mathbf{w}'_{1t}} = - \left( \frac{\partial \mathbf{F}_{\theta,t}}{\partial \theta_t'} \Big|_{\theta_t = \theta_t^*} \right)^{-1} \frac{\partial \mathbf{F}_{\theta,t}}{\partial \mathbf{w}'_{1t}} \Big|_{\theta_t = \theta_t^*}.$$

Of course, if some elements of  $\theta_t^*$  is not interior, the corresponding elements of  $\frac{\partial \theta_t^*}{\partial \mathbf{w}'_{1t}}$  will be zero.

Recalling that  $\frac{\partial \pi_{ct}}{\partial \theta_{ct}} = (\bar{p}_t - w_{0ct}) \frac{\partial s_{0ct}}{\partial \theta_{ct}}(\theta_t) + (\bar{p}_t - w_{1ct}) \frac{\partial s_{1ct}}{\partial \theta_{ct}}(\theta_t)$ , we have that

$$\frac{\partial \mathbf{F}_{\theta,t}}{\partial \mathbf{w}'_{1t}} = - \begin{pmatrix} \frac{\partial s_{11t}}{\partial \theta_{1t}} & 0 & \dots & 0 \\ 0 & \frac{\partial s_{12t}}{\partial \theta_{2t}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\partial s_{1Ct}}{\partial \theta_{Ct}} \end{pmatrix},$$

while

$$\frac{\partial \mathbf{F}_{\theta,t}}{\partial \theta_t'} = \begin{pmatrix} \sum_k m_{k1t} \frac{\partial^2 s_{k1t}}{\partial \theta_{1t}^2} & \sum_k m_{k1t} \frac{\partial^2 s_{k1t}}{\partial \theta_{1t} \partial \theta_{2t}} & \dots & \sum_k m_{k1t} \frac{\partial^2 s_{k1t}}{\partial \theta_{1t} \partial \theta_{Ct}} \\ \sum_k m_{k2t} \frac{\partial^2 s_{k2t}}{\partial \theta_{2t} \partial \theta_{1t}} & \sum_k m_{k2t} \frac{\partial^2 s_{k2t}}{\partial \theta_{2t}^2} & \dots & \sum_k m_{k2t} \frac{\partial^2 s_{k2t}}{\partial \theta_{2t} \partial \theta_{Ct}} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_k m_{kCt} \frac{\partial^2 s_{kCt}}{\partial \theta_{Ct} \partial \theta_{1t}} & \sum_k m_{kCt} \frac{\partial^2 s_{kCt}}{\partial \theta_{Ct} \partial \theta_{2t}} & \dots & \sum_k m_{kCt} \frac{\partial^2 s_{kCt}}{\partial \theta_{Ct}^2} \end{pmatrix}$$

Then, all the derivatives of market shares with respect to  $w_{1ct}$  in equation (3.9) can be obtained from elements of the stacked vector  $\frac{\partial s_{kt}}{\partial \mathbf{w}'_{1t}}$  for  $k = 0$  or  $1$  and which satisfies

$$\begin{aligned} \frac{\partial s_{kt}}{\partial \mathbf{w}'_{1t}} &= \left( \frac{\partial s_{kt}}{\partial \theta_t'} \Big|_{\theta_t = \theta_t^*} \right) \frac{\partial \theta_t^*}{\partial \mathbf{w}'_{1t}} \\ &= \left( \frac{\partial s_{kt}}{\partial \theta_t'} \Big|_{\theta_t = \theta_t^*} \right) \left( - \frac{\partial \mathbf{F}_{\theta,t}}{\partial \theta_t'} \Big|_{\theta_t = \theta_t^*} \right)^{-1} \frac{\partial \mathbf{F}_{\theta,t}}{\partial \mathbf{w}'_{1t}} \Big|_{\theta_t = \theta_t^*}, \end{aligned}$$

which shows that the change in a given market share,  $s_{kct}$ , caused by the change in a given wholesale price,  $w_{1\tilde{c}t}$ , will depend on the change in the full vector of  $\theta$ 's following from the change in the Nash equilibrium in the competition between chains.

## 7.5 Inner Loop Algorithm of Demand Estimation

In each period  $t$ , given the other chains choices for  $\theta_{\tilde{c}t}$  ( $\tilde{c} \neq c$ ), each pharmacy chain  $c$  solves the constrained maximization problem:

$$\max_{\theta_{ct}} \pi_{ct} \quad \text{s.t.} \quad 0 \leq \theta_{ct} \leq 1$$

Letting  $\mu_{ct}^L$  and  $\mu_{ct}^H$  denote the multipliers associated with the lower and upper bound on  $\theta_{ct}$  respectively, the necessary conditions for maximization of the corresponding Lagrangian are

$$\begin{aligned} \frac{\partial \pi_{ct}}{\partial \theta_{ct}} + \mu_{ct}^L - \mu_{ct}^H &= 0 \\ \mu_{ct}^L \geq 0, \quad \mu_{ct}^L \theta_{ct} &= 0, \quad \theta_{ct} \geq 0 \\ \mu_{ct}^H \geq 0, \quad \mu_{ct}^H (1 - \theta_{ct}) &= 0, \quad \theta_{ct} \leq 1 \end{aligned}$$

The equilibrium in each period  $t$  is given by the solution to these equations for each chain  $c$ . This equilibrium can be redefined as the solution to the following constrained minimization problem:

$$\begin{aligned} \min_{\{\theta_{ct}, \mu_{ct}^L, \mu_{ct}^H\}_{c \in \{1, \dots, C\}}} \sum_{c \in \{1, \dots, C\}} (\mu_{ct}^L \theta_{ct} + \mu_{ct}^H (1 - \theta_{ct})) \\ \text{s.t.} \\ \frac{\partial \pi_{ct}}{\partial \theta_{ct}} + \mu_{ct}^L - \mu_{ct}^H = 0 \quad \forall c \in \{1, \dots, C\}, \\ 0 \leq \theta_{ct} \leq 1, \quad \mu_{ct}^L \geq 0, \quad \mu_{ct}^H \geq 0 \quad \forall c \in \{1, \dots, C\}. \end{aligned}$$

The objective function in this minimization problem is the sum of the complementary slackness condition corresponding to the bounds on  $\theta$  for each chain, and will thus be zero at the solution, while the constraints ensure that the solution is a Nash-equilibrium.

The full problem in the inner loop of the estimation also includes fitting the mean utility parameters  $\alpha_{jct}$  for each product  $j$  at each chain  $c$  in each period  $t$ . These mean utility parameters are set such that observed shares  $\hat{s}_{jct}$  are equal to predicted shares  $s_{jct}(\boldsymbol{\theta}_t, \boldsymbol{\alpha}_t, \boldsymbol{\beta})$ , where  $\boldsymbol{\theta}_t$  is the vector of  $\theta_{ct}$  for all chains  $c$ , and  $\boldsymbol{\alpha}_t$

is the vector of mean utility parameters for all products in period  $t$ . We can write this restriction as

$$\mathbf{s}_t(\boldsymbol{\theta}_t, \boldsymbol{\alpha}_t, \boldsymbol{\beta}) = \hat{\mathbf{s}}_t,$$

where  $\mathbf{s}_t(\boldsymbol{\theta}_t, \boldsymbol{\alpha}_t, \boldsymbol{\beta})$  is the vector of predicted market shares from the model, and  $\hat{\mathbf{s}}_t$  is the vector of observed market shares. These conditions can then be included as constraints in the minimization problem characterizing the market equilibrium in period  $t$ .

The full constrained minimization problem can then be written (in vector notation)

$$\begin{aligned} \min_{\boldsymbol{\alpha}_t, \boldsymbol{\theta}_t, \boldsymbol{\mu}_t^L, \boldsymbol{\mu}_t^H} \quad & \boldsymbol{\mu}_t^L \cdot \boldsymbol{\theta}_t + \boldsymbol{\mu}_t^H \cdot (\mathbf{1} - \boldsymbol{\theta}_t) \\ \text{s.t.} \quad & \\ & \mathbf{s}_t(\boldsymbol{\theta}_t, \boldsymbol{\alpha}_t, \boldsymbol{\beta}) = \hat{\mathbf{s}}_t, \\ & \frac{\partial \pi_t}{\partial \boldsymbol{\theta}_t} + \boldsymbol{\mu}_t^L - \boldsymbol{\mu}_t^H = \mathbf{0}, \\ & \mathbf{0} \leq \boldsymbol{\theta}_t \leq \mathbf{1}, \quad \boldsymbol{\mu}_t^L \geq \mathbf{0}, \quad \boldsymbol{\mu}_t^H \geq \mathbf{0}. \end{aligned}$$

Informally, we can think about the market share constraint as particularly informative about the mean utility parameters  $\boldsymbol{\alpha}_t$ , while the constraints corresponding to first order conditions for profit maximization are particularly informative about  $\boldsymbol{\theta}_t$ , though in practice they will jointly inform all parameters. Note that observed margins only enter each chain's profit maximization problem (i.e., it does not have a direct effect on demand), which serves as an exclusion restriction in our model.

The solution to this constrained minimization program is computed for each market  $t$  using a sequential quadratic method.<sup>24</sup> In our empirical estimates, we find a solution satisfying all constraints with the minimized value equal to 0 for all markets. As further checks on the solutions of the inner loop, we perform several additional tests (at the parameter vector estimated in the outer loop). One is a check of the second order conditions of the firms' maximization problems, to verify that the  $\theta$ 's constitute a maximum. Another is a check of whether a firm would profit by unilaterally deviating by setting  $\theta$  to one of the corners (if  $\theta$  is already at a corner, only the other is tested), as a test of the Nash equilibrium. Also, we perform two tests for multiple equilibria. First, we recalculate the solution to the system of equations given above for the cases where we fix a firm's  $\theta$  at each of the corners (for each firm separately), thus removing the constraint corresponding to this firm's profit maximization problem. We then check whether any of the solutions satisfies the full set of

---

<sup>24</sup>See, e.g., (Judd, 1998, ch. 4.7). Specifically, we use sequential least squares programming (SLSQP), as implemented in the optimization routines of the Python package SciPy.

equations. Second, we solve the system of equations for many different starting values, checking whether we obtain non-unique solutions.

## 7.6 Asymptotic Disctribution of $\theta_{ct}$ Estimates

In the case of interior solution to the Nash equilibrium in  $\theta_{ct}$ ,  $\theta_{ct}$  must satisfy first order condition (3.5) given other  $\theta_{ct}$ . Let's denote  $\theta_{ct}^u(\beta)$  the solution to the first order condition whether it belongs to the  $[0,1]$  interval or not. Then we know that the solution of the Nash equilibrium is  $\theta_{ct}(\beta) = \theta_{ct}^u(\beta) \mathbf{1}_{\{\theta_{ct}^u(\beta) \in (0,1)\}} + \mathbf{1}_{\{\theta_{ct}^u(\beta) \geq 1\}}$ . Using the Delta method we can first find the asymptotic law of  $\theta_{ct}^u(\beta)$ . We need the gradient of  $\theta_{ct}^u(\beta)$  with respect to  $\beta$ . Fully differentiating the first order condition determining  $\theta_{ct}^u(\beta)$ , we obtain for all  $c$ :

$$\sum_{c'} \left( m_{0ct} \frac{\partial^2 s_{0ct}}{\partial \theta_{ct}^u \partial \theta_{c't}^u} + m_{1ct} \frac{\partial^2 s_{1ct}}{\partial \theta_{ct}^u \partial \theta_{c't}^u} \right) \frac{\partial \theta_{c't}^u(\beta)}{\partial \beta} + m_{0ct} \frac{\partial^2 s_{0ct}}{\partial \theta_{ct}^u \partial \beta} + m_{1ct} \frac{\partial^2 s_{1ct}}{\partial \theta_{ct}^u \partial \beta} = 0$$

where

$$\frac{\partial^2 s_{0ct}}{\partial \theta_{ct}^u \partial \theta_{c't}^u} = \int -\rho_{ict} \frac{\partial s_{ict}}{\partial \theta_{c't}^u} - \mathbf{1}_{\{c=c'\}} \rho_{ict} \lambda_c \delta_{ict} s_{ict} (1 - s_{ict}) + (1 - \theta_{ct}^u \rho_{ict}) \lambda_c \delta_{ict} [1 - 2s_{ict}] \frac{\partial s_{ict}}{\partial \theta_{c't}^u} dF(V_{it}|\beta),$$

$$\frac{\partial^2 s_{1ct}}{\partial \theta_{ct}^u \partial \theta_{c't}^u} = \int \rho_{ict} \frac{\partial s_{ict}}{\partial \theta_{c't}^u} + \mathbf{1}_{\{c=c'\}} \rho_{ict} \lambda_c \delta_{ict} s_{ict} (1 - s_{ict}) + \theta_{ct}^u \rho_{ict} \lambda_c \delta_{ict} [1 - 2s_{ict}] \frac{\partial s_{ict}}{\partial \theta_{c't}^u} dF(V_{it}|\beta)$$

with

$$\frac{\partial s_{ict}}{\partial \theta_{c't}^u} = \lambda_{c'} \delta_{ic't} (\mathbf{1}_{\{c=c'\}} - s_{ict}) s_{ic't}$$

and  $\frac{\partial^2 s_{1ct}}{\partial \theta_{ct}^u \partial \beta}$  and  $\frac{\partial^2 s_{0ct}}{\partial \theta_{ct}^u \partial \beta}$  come from taking derivatives with respect to  $\beta$  of:

$$\begin{aligned} \frac{\partial s_{0ct}}{\partial \theta_{ct}^u} &= \int (-\rho_{ict} s_{ict} + (1 - \theta_{ct}^u \rho_{ict}) \lambda_c \delta_{ict} s_{ict} (1 - s_{ict})) dF(V_{it}|\beta), \text{ and} \\ \frac{\partial s_{1ct}}{\partial \theta_{ct}^u} &= \int (\rho_{ict} s_{ict} + \theta_{ct}^u \rho_{ict} \lambda_c \delta_{ict} s_{ict} (1 - s_{ict})) dF(V_{it}|\beta). \end{aligned}$$

Then we know that  $\hat{\theta}_{ct}^u(\beta) \leftrightarrow N\left(\theta_{ct}^u(\hat{\beta}), \text{var}\left(\hat{\theta}_{ct}^u(\beta)\right)\right)$  with

$$\text{var}\left(\hat{\theta}_{ct}^u(\hat{\beta})\right) = \left[ \frac{\partial \theta_{ct}^u(\hat{\beta})}{\partial \beta} \right]' \text{var}(\hat{\beta}) \left[ \frac{\partial \theta_{ct}^u(\hat{\beta})}{\partial \beta} \right]$$

where  $\left[ \frac{\partial \theta_{ct}^u(\beta)}{\partial \beta} \right]$  is the Jacobian matrix of  $\theta_{ct}^u(\beta) = (\theta_{1t}^u(\beta), \dots, \theta_{Ct}^u(\beta))'$  with respect to the vector of parameters  $\beta$ , and  $\text{var}(\hat{\beta})$  is the variance-covariance matrix of  $\hat{\beta}$ .

As  $\theta_{ct}(\beta) = \theta_{ct}^u(\beta) \mathbf{1}_{\{\theta_{ct}^u(\beta) \in (0,1)\}} + \mathbf{1}_{\{\theta_{ct}^u(\beta) \geq 1\}}$  we obtain directly the asymptotic law of  $\theta_{ct}(\beta)$  using the one of  $\theta_{ct}^u(\beta)$ .  $\theta_{ct}$  is censored normally distributed. With  $\phi$  the  $N(0,1)$  c.d.f., the c.d.f. of  $\hat{\theta}_{ct}(\hat{\beta})$  is  $P(\hat{\theta}_{ct}(\hat{\beta}) \leq a) = \phi\left(\frac{a - \hat{\theta}_{ct}^u(\hat{\beta})}{\text{var}(\hat{\theta}_{ct}^u(\hat{\beta}))}\right) \mathbf{1}_{\{a \in (0,1)\}} + \mathbf{1}_{\{a \geq 1\}}$  which allows construct the confidence interval of  $\theta_{ct}(\beta)$ .

## 7.7 Additional reduced forms regressions

Table 7.1: *Reduced form evidence of Parallel Imports relationship with Pharmacy margins*

Dependent Variable $\left(\ln \frac{s_{1ct}}{s_{0ct} + s_{1ct}}\right)$	(OLS) (1)	(OLS) (2)	(2SLS) (3)	(OLS) (4)	(OLS) (5)	(2SLS) (6)
Direct imports margin $m_{1ct}$	0.001*** (0.000)	0.010*** (0.002)	0.017*** (0.002)			
Parallel imports margin $m_{0ct}$	-0.014*** (0.002)	-0.014*** (0.002)	-0.025*** (0.005)			
Margin difference $(m_{1ct} - m_{0ct})$				0.002*** (0.000)	0.013*** (0.002)	0.031*** (0.004)
Chain 1	-0.295*** (0.043)	-0.310*** (0.041)	-0.369*** (0.041)	-0.253*** (0.043)	-0.312*** (0.041)	-0.408*** (0.042)
Chain 2	0.001 (0.031)	-0.012 (0.029)	-0.060 (0.041)	0.047 (0.031)	-0.012 (0.029)	-0.096* (0.041)
Market fixed effects	No	Yes	Yes	No	Yes	Yes
N	2440	2440	2342	2440	2440	2342

Notes: Markets are defined as ATC code level 5 (molecule)-strength-month. Standard errors in parenthesis are clustered at the market level. Chain 3 is the reference dummy. In the case of Two Stage Least Squares estimates (2SLS), instruments are the wholesale price in France, Italy, Poland, Spain, United Kingdom, interacted with pharmacy chain dummies.

## 7.8 Additional Tables

Table 7.2: *Impact of preventing parallel imports foreclosure and reducing the price ceiling*

$\Delta p$		-10%	-15%	-20%	-25%	-30%
$\Delta\pi$						
	Chain 1	-4.95 -38%	-6.93 -53%	-8.88 -68%	-10.81 -83%	-12.66 -97%
	Chain 2	-2.42 -32%	-3.23 -43%	-3.35 -45%	-4.15 -56%	-4.83 -65%
	Chain 3	-2.52 -35%	-3.49 -48%	-4.45 -62%	-5.30 -74%	-6.15 -85%
$\Delta\Pi$						
	Manufacturer	-0.11 -0%	-0.67 -1%	-1.92 -3%	-2.72 -4%	-3.65 -6%
	Parallel	-0.07 -6%	-0.27 -24%	-0.47 -42%	-0.68 -60%	-0.88 -78%
	Number of chain-market exits	2	9	19	26	37

Notes: The profits changes are in millions of NOK per year. Percentage are indicated below absolute changes. There are 77 markets (strength-month combinations) and thus 231 chain-market observations.

Table 7.3: *Average across Chains and Months of Marginal Cost and Margins*

		Market	
		40mg	80mg
Marginal cost	$c$	[1.73, 2.76]	[0.81, 1.39]
Retail price	$p$	4.19	2.09
<i>Direct Imports</i>			
Pharmacy margin on DI	$p - w_1$	1.18	0.59
Manufacturer margin on DI	$w_1 - c$	[0.25, 1.27]	[0.10, 0.69]
<i>Parallel Imports</i>			
Manufacturer margin of PI in source country	$p_1^I - c$	[0, 1.03]	[0, 0.59]
Pharmacy margin on PI	$p - w_0$	1.27	0.69
Parallel importer margin on PI	$w_0 - p_0^I$	0.09	0.05

Notes: Intervals when marginal cost or margin is only set identified. Average in NOK per DDD.