



<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper published in *Journal of Health Economics*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Granlund, D., Bergman, M. (2018)

Price competition in pharmaceuticals: evidence from 1303 Swedish markets

Journal of Health Economics, 61: 1-12

<https://doi.org/10.1016/j.jhealeco.2018.06.009>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

© 2018. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
<http://creativecommons.org/licenses/by-nc-nd/4.0/>

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-150413>

Price competition in pharmaceuticals – evidence from 1303 Swedish markets¹

David Granlund* and Mats A. Bergman†

Accepted manuscript, July 2018. The final version is published in *Journal of Health Economics*, and available at <https://doi.org/10.1016/j.jhealeco.2018.06.009>

Abstract: We study the short- and long-term price effects of the number of competing firms, using panel-data on 1303 distinct pharmaceutical markets for 78 months within a reference-price system. We use actual transaction prices in an institutional setting with little scope for non-price competition and where simultaneity problems can be addressed effectively. In the long term, the price of generics is found to decrease by 81% when the number of firms selling generics with the same strength, form and similar package size is increased from 1 to 10. Nearly only competition at this fine-grained level matters; the effect of firms selling other products with the same active substance, but with different package size, form, or strength, is only a tenths as large. Half of the price reductions take place immediately and 70% within three months. Also, prices of originals are found to react to competition, but far less and much slower.

Keywords: Price competition; dynamic; adjustment; pharmaceutical industry; generic drugs; brand-name drugs; reference price; generic substitution.

JEL codes: D40, I13, L13, L65

¹ We would like to thank participants at the 2017 EARIE Conference in Maastricht, Chandra Kiran, two anonymous reviewers, and seminar participants at University of Minho, Södertörn University, and at Umeå University for helpful comments and suggestions. A research grant from the Swedish Competition Authority (project number 395/2015) is gratefully acknowledged. We are also grateful to IMS Sweden and the Swedish Agency for Growth Policy Analysis for supplying the data used in this article. Declarations of interest: none.

* Umeå University, SE-901 87 Umeå, Sweden; david.granlund@umu.se (corresponding author).

† Södertörn University - Stockholm, SE-141 89 Huddinge, Sweden;

1. Introduction

An important economic question is how the number of sellers affects prices. Many studies have attempted to determine this but very few of them are able to distinguish between short- and long-term effects. Weiss (1989) summarizes the results of studies in the old industrial-economics tradition. Mazzeo (2002), Davis (2005) and Singh and Zhu (2008) are more recent examples.

Within the context of a reference-price system for off-patent prescription drugs, where the reference price is set equal to the lowest price, i.e., so-called internal reference pricing, we find a strong price effect of the number of competing sellers. The setting, with clear and explicit market rules, allows us to make causal interpretations and also to estimate dynamic pricing models, hence enabling estimates of the speed of the price response.

We relate to a substantial empirical literature on pharmaceutical pricing, that addresses the effect of the number of generic firms on prices. Estimates by Caves et al. (1991), Frank and Salkever (1997), and Wiggins and Maness (2004), who all use US data, suggest that increasing the number of actual generic suppliers from 1 to 10 reduces prices of generics by about 50%. Reiffen and Ward (2005) estimate the effect to be slightly smaller, but Regan (2008), also using US data, and Brekke et al. (2011), who use Norwegian data, find no significant negative effects.

Danzon and Chau (2000) estimated that increasing the number of product per molecule from 1 to 10 was associated with a price reduction of 69% in the US and somewhat less in Canada, UK, and Germany, but they found no significant effects for France, Italy and Japan. That Danzon and Chau found so large associations for some countries, despite being unable to address endogeneity concerns, might be because they used cross-sectional data and hence estimated long-run associations. Bernt and Aitiken (2001) report data suggesting that the effect can be even larger in the US. They report that, for a sample of top-selling generic molecules that were still in the market 25 months after the initial entry, the average generic price had then fallen by about 94%, while the average number of generic firms had increased to 12.

Previous results regarding the effect on prices of originals are mixed. Frank and Salkever and Regan found that prices of originals increased in response to generic entry, while Caves et al., Wiggins and Maness, Saha et al. (2006), who also use US data, and Stargardt (2011), who uses German data, found negative price effects of more generic competition. Based on US data, Ching (2010a, b) reports mixed results; that some brand-name prices increase and a few decrease, as the number of generics becomes higher.

A central aim of this article is to examine how the number of firms in the market affects the prices of individual pharmaceutical products in a setting with well-defined markets and few non-price competitive actions available to the firms. Advertising directed towards consumers, for example, is banned by law for

prescription pharmaceuticals in Sweden and the physical and financial conditions for delivery and payment are fixed by the market regulator.

Our study is also relevant for the literature on reference pricing. Following the introduction of the first reference-price system in Germany in 1989, a large number of such systems have been introduced, including the 1993 Swedish system, subsequently reformed in 2002. Reference pricing aims to control costs indirectly, by making demand more elastic, rather than through direct price regulation (Brekke et al., 2007). A more elastic demand is achieved by requiring patients to pay the difference between the price of the subscribed product and the reference price for the cluster to which it belongs. The reference price can be, e.g., the lowest, second-lowest or average price within the cluster or a certain fraction of the original's price (internal reference pricing) or the average price of the same product in a group of countries (external reference pricing).

Virtually all studies confirm that the introduction of reference pricing results in lower prices in the short run,² but less is known about their long-run effects (Galizzi et al., 2011) and about how the design of the system influences its success (Kaiser and Mendez, 2015). Concerns have been raised that reference pricing will only have a transitory effect. Our research design does not allow us to evaluate reference pricing per se, but we do observe strong and sustained effects of competition within such a system. To some extent this may be due to the narrowly defined reference clusters and to some extent due to the reference price being set in a monthly bidding contest that confers substantial benefits to the low bidder.

Unlike the pharmaceutical-pricing studies mentioned above, we estimate dynamic models, allowing us to study the speed of adjustment and distinguishing between short- and long-term effects. Several mechanisms make it likely that the short-term effects are smaller than the long-term effects. For example, for an incumbent firm it might be easier to achieve a collusive equilibrium by initially maintaining the pre-entrance price so as to allow entering firms to adjust their prices, rather than reducing the price at entry and then attempt to achieve a coordinated price increase. Another reason is that, when a firm exits, each remaining firm gains by being the last to increase its price. Companies may also have limited abilities to predict what the new equilibrium price will be, which makes them adjust gradually to the new equilibrium. Lastly, for originals, market-specific rules can result in slow adjustment. In the Swedish pharmaceutical market there is a dynamic price-cap that may prevent a product that is already the most expensive among substitutes to increase its price if it wants to remain within the reimbursement system. Hence, for originals a price-cut that in retrospect is found to be too large cannot always be reversed.

By studying the speed of adjustment, we relate to the large experimental (and theoretical) literature on whether and how fast equilibrium is reached in one-sided and two-sided auctions (see e.g. Smith, 1962; Plott and George, 1992; List, 2003; Crocket et al., 2011). The Swedish generics market offers a large

² For surveys, see Galizzi et al. (2011), Puig-Junoy (2010), and Dylst et al. (2011).

number of recurring high-value auctions and provides an opportunity to learn about the behavior of professional bidders. Knowing the speed of adjustment when the number of firms changes is also important when forecasting expenditures for budget purposes and when evaluating reforms in the market.

Applying a dynamic model to monthly data and using the fact that the rules require firms to submit their price bids two months in advance allow us to identify the causal effects of the number of firms. The reason is that the monthly data and the bidding rules effectively solve the simultaneity problem that often troubles price–concentration studies – under the assumption that firms cannot predict future price shocks when submitting their bids. That the simultaneity problem is solved this way enables us to estimate the effect of competition, using indicator variables for the number of firms. To our knowledge, this has previously been done only by Reiffen and Ward (2005) and Regan (2008) using a few hundred observations. We find that the effect of additional firms is large, even then the initial number of firms is already large.

We use a dataset provided by IMS Sweden that covers all off-patent prescription pharmaceuticals sold in the Swedish reimbursement system at Swedish pharmacies from January 2006 through June 2012. The dataset contains a total of 168,188 observations of prices and total national sales. One advantage with the data is that the prices are actual transaction prices, not list prices, as Swedish law forbids pharmaceutical firms to give pharmacies discounts or rebates for pharmaceuticals with generic alternatives.³ Another is that the observations are at the product level⁴, meaning that the composition effects caused by, e.g., changes in the distribution over package size will not bias the results. The observations are related to 4 730 different products in 1 303 exchange groups. The exchange groups consist of products with the same combination of active substance, form of administration, strength, and packet size. At pharmacies, consumers can choose among products (brands) within the exchange group of the prescribed product and are incentivized to choose the lowest-priced product.

Comparing exchange groups within substances a given month, the data reveals that the price per defined daily dose is more than twice as large in the exchange groups with the lowest number of firms compared to the one with the most firms. From a policy perspective it is important to study to what extent this reflects a causal effect of the number of firms, since this can determine if it is profitable to, e.g., reduce the administrative fees in order to increase the number of active firms in small exchange groups.

This paper relates to Bergman, Granlund and Rudholm (2017) which used part of the data used in this study to investigate how changing the market share for the lowest bidder affects the cost per defined dose.

³ The last few years, some pharmaceutical firms have given the county councils, which employ most physicians in Sweden, chargebacks for some new on-patent drugs, but this does not affect the off-patent drugs under our study-period.

⁴ A product is defined as a unique combination of substance, form of administration, strength and package size, sold by a specific firm.

That study also analyzed the effect of the number of firms, but instead of having the price of individual products as the dependent variable, the dependent variable was cost per defined dose measured at the exchange-group level. This means that the effect estimated in Bergman, Granlund and Rudholm (2017) is a weighted average of the effect on generic prices and on original prices plus an effect that goes through changing products' market shares. Thus, this paper contributes by studying the effects on individual prices and by doing this separately for generics and originals. Unlike Bergman, Granlund and Rudholm, who implicitly assumed firms to have naïve expectations about the number of competitors, we also allow for rational expectations and find evidence consistent with this.

The results show that in the long term generic prices fall by 81% and original prices by 29% when the number of firms selling generics in the exchange group increases from 1 to 10. Flexible-form estimations reveal that the effect of the number of firms on prices is well described by constant elasticities; for example, the percentage effect on generic prices of going from 6 to 9 firms is nearly equally large as that of going from 2 to 3 firms.

For generics, we find a fast adjustment to a changed number of competitors. About half of the long-term effect on prices of a change of the number of firms occurs immediately and 70% of the adjustment takes place within three months. For originals, the corresponding figures are only about 10% and 20%, respectively. The slower adjustment for originals might in part be explained by the dynamic price-cap discussed above.

Most previous studies have estimated competition at the substance level, but our detailed data allows us to distinguish between competition at different levels. We find that the effect of additional firms selling the same substance in other exchange groups is close to zero. Thus, one reason as to why our estimates are larger than those of previous studies can be that we measure competition at a more fine-grained level. Also the effect of therapeutic competitors is small, but the results indicate that generic prices are reduced slightly when the number of therapeutic alternatives with generic versions increases. Still, the results clearly indicate that competition within the exchange group is the most important type of competition.

The article is organized as follows: Section 2 presents the Swedish pharmaceutical market and Section 3 discusses the data. In Section 4 the empirical method is discussed and results are presented in Section 5. Section 6 concludes the paper, while robustness checks are presented and discussed in an appendix.

2. The generics market

During the study period, a government-funded benefits scheme covered approximately 75% of the cost of prescription drugs for Swedish patients and, on the margin, patients with high costs paid nothing (National Board on Health and Welfare, 2013). Pharmaceutical firms were (and still are) free to set their own prices, but in order to be included in the pharmaceutical benefits scheme, the price must be approved by the Dental and Pharmaceutical Benefits Agency (DPBA). In 1993, reference pricing was introduced in Sweden, with the reference price set equal to 110% of the lowest price within the relevant cluster of products, the exchange group.⁵ For products with a price higher than the reference price, the consumer is required to pay the full difference between the price and the reference price, in addition to a copayment determined by the reference price and the consumer's purchase history. In 2002, generic substitution in combination with a national market for generics was introduced and, following the reform, the reference price is set equal to the lowest price.

Firms wanting their product to be included in the pharmaceutical benefit scheme must compete in the national market by submitting their price bids for month t to DPBA already in month $t-2$. Firms bid in prices that are uniform across Sweden and include transport to the pharmacies. Prices not exceeding the highest price within the exchange group the previous month are always approved by the DPBA. The fifth workday of month $t-1$, DPBA announces all purchase prices and the retail pharmacy prices, which are set with a simple algorithm that to the purchase price adds a margin that is continuously increasing in the pharmacy purchase price. Note that when the firms submit their bids in month $t-2$, the prices that will apply in month $t-1$ have already been announced. Consequently, the number of active firms in that period is also known.

Since the introduction of generic substitution, pharmacy personnel are required to inform consumers if cheaper substitute products are available in Sweden and in that way steer purchases to the lowest-price alternative. This obligation is waived, however, if the physician indicated on the prescription that no substitution should be allowed for medical reasons or if the pharmacist has reason to believe that the patient would be adversely affected, e.g., because the low-cost alternative has a package that is difficult to open for some patients. If consumers oppose substitution or choose to switch to another substitute than the cheapest available in Sweden, the entire incremental cost will be charged to them.

Pharmacies are by law required to provide the pharmaceutical a consumer demands within 24 hours, unless special circumstances motivate longer time (Ministry of Health and Social Affairs, 2009a). This gives pharmacies an incentive to keep the cheapest alternative in stock. To allow pharmacies to clear excess inventory, pharmacies are allowed to sell the product that was the cheapest in month $t-1$ during the first 15 days of month t without additional cost to the consumer (Dental and Pharmaceutical Benefits Agency, 2009). In the data used in this study, physicians opposed substitution for 3.4% of the packages,

⁵ Within the cluster, or exchange group, packet size is allowed to vary slightly; for example, substitution will be made from a 30-pill package to a package in the 28–32-pill range.

pharmacist for 2.0% and patients for 10.5%. Due to, e.g., stock-outs or pharmacies not complying with the rules, the lowest-cost alternatives' market shares on average reach just above 50 percent. Even so, the possibility to obtain half of the market provides an incentive to compete in prices.

That some patients pay extra in order to get another product than the cheapest available suggests that they do not consider the products to be identical. This is confirmed by a survey among patients at Swedish pharmacies in which 30% responded that they had experienced a weaker (medical) effect after substitution and 22% reported more side effects (Olsson et al., 2015). However, of the 282 respondents, 18% had experienced a *stronger* effect and 14% fewer side effects. Olsson et al. also report that more than half of the respondents with low trust in the bioequivalence of the products still accepted the substitution in most cases and Granlund and Rudholm (2012) report that patients agreed to substitution in 83% of the cases when they had an option.

From the early 1970s until 2009, the Swedish market for pharmacy retailing was served exclusively by a state-owned monopoly, Apoteket AB, but since July 1, 2009, also private pharmacies are allowed in Sweden (Ministry of Health and Social Affairs, 2009b). In addition, two thirds of the pharmacies were sold; the majority in blocks to private investors, while a fraction was reserved for small investors, under special conditions. The change in ownership became effective during the first months of 2010. In parallel with the privatizations the price of off-patent substances was capped at 35 percent of the price during the patent period (conditional on some criteria being met, as discussed below).

Firms wanting to sell pharmaceutical in Sweden need approval from the Swedish Medical Products Agency. Besides documentation, it costs 200 000 SEK (€ 20 000) to register a new substance and then, annually, 46 000 SEK (€ 4 600) per substance and 22 500 SEK (€ 2 250) for each additional combination of strength and form of distribution.

3. Data

The dataset used here has been compiled by IMS Sweden and covers all prescription pharmaceuticals included in the pharmaceutical benefits scheme, with active substances no longer protected by patents, and which are sold at Swedish pharmacies from January 2006 through June 2012. We dropped 4 194 observations belonging to 19 different substances, due to lack of information concerning which exchange group the product belongs to.

The products were classified by IMS as originals, generics, or belonging to the category Others. Originals (or brand-name pharmaceuticals) are products that have previously been patent protected. Generics have the same active substance but other ingredients, such as binders, flavors, and colorants, may differ. Only

0.5% of the observations belong to the category Others, which consist of, e.g., vitamins and minerals, and these are excluded from the analysis.

The data also includes a variable indicating whether the product is parallel imported or locally sourced. Parallel imported drugs are sold mainly before the first generic entry and our dataset therefore includes only 8 782 observations of parallel imports. In order to focus on the prices of locally sourced drugs, these observations are dropped leaving us with 168 188 observations of 4 730 different products in 1 303 exchange groups, 490 drugs and of 191 different substances. A drug is here defined as a unique active substance-strength-form combination, so that for each drug there can be several exchange groups differing only in packet size.

Table 1 presents the mean, standard deviation and minimum and maximum values for variables used in the estimations separately for generics and originals. The first variable, P_{it} , is the pharmacies' purchase price for product i in month t and $\ln P_{it}$ is the natural logarithm of this variable. Due to the regulatory regime and as mentioned above, P_{it} is not just the official list price but also the actual transaction price. Data (not shown in tables) reveals that 96% of generics products and 97% of original products sold one month were also sold the previous month and that generics have different prices than they had last month in 38% of the observations while the corresponding number for originals is only 3%. Still, non-stationarity of $\ln P_{it}$ is rejected on the 1% level for both categories.

Table 1. Descriptive statistics

	<i>Generics</i>				<i>Originals</i>			
	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
P	145.76	459.83	0.20	21100.24	360.22	1082.87	4.27	23660.00
$\ln P$	4.12	1.22	-1.61	9.96	4.91	1.32	1.45	10.07
$GenFirms$	3.76	2.47	1	12	2.16	2.12	0	11
$\ln GenFirms$	1.27	0.61	0.41	2.53	0.66	0.84	-0.69	2.44
$Orig$	0.48	0.50	0	1	0.93	0.25	0	1
$AddFirms$	1.92	2.15	0	12	1.27	1.81	0	12
$\ln AddFirms$	0.49	0.92	-0.69	2.53	0.15	0.90	-0.69	2.53
$ThAlt$	1.39	1.20	0	4	1.09	1.14	0	4
$\ln ThAlt$	0.39	0.76	-0.69	1.50	0.18	0.77	-0.69	1.50
$ThGenAlt$	1.16	1.19	0	4	0.96	1.12	0	4
$\ln THGenAlt$	0.23	0.77	-0.69	1.50	0.09	0.77	-0.69	1.50
$2009PriceCap$	0.52	0.50	0	1	0.43	0.49	0	1
$Months_Pat$	326.09	492.28	0	2776	469.80	719.12	0	2776
$\ln Months_Pat$	4.98	1.38	-0.69	7.93	5.03	1.66	-0.69	7.93
DDD	5.3e+05	1.3e+06	0.00	1.4e+07	3.7e+05	1.0e+06	0.00	1.4e+07
$\ln DDD$	10.88	2.55	-1.17	16.43	10.40	2.49	-1.17	16.43
Products	3,893				837			
Observations	133,667				34,521			

Note: Values for $\ln DDD$ is missing for 985 generic observations and 931 original observations.

Figure 1 shows the distribution of price changes. For originals we see that nearly 40% of the price changes are reductions smaller than 10%. That nearly 10 percent of the price changes for original are reductions by

60-70 percent is explained by the price cap introduced in July 2009, described in more detail at the end of this section. We also see that the average price changes are larger, in relative terms, for generics than for originals.

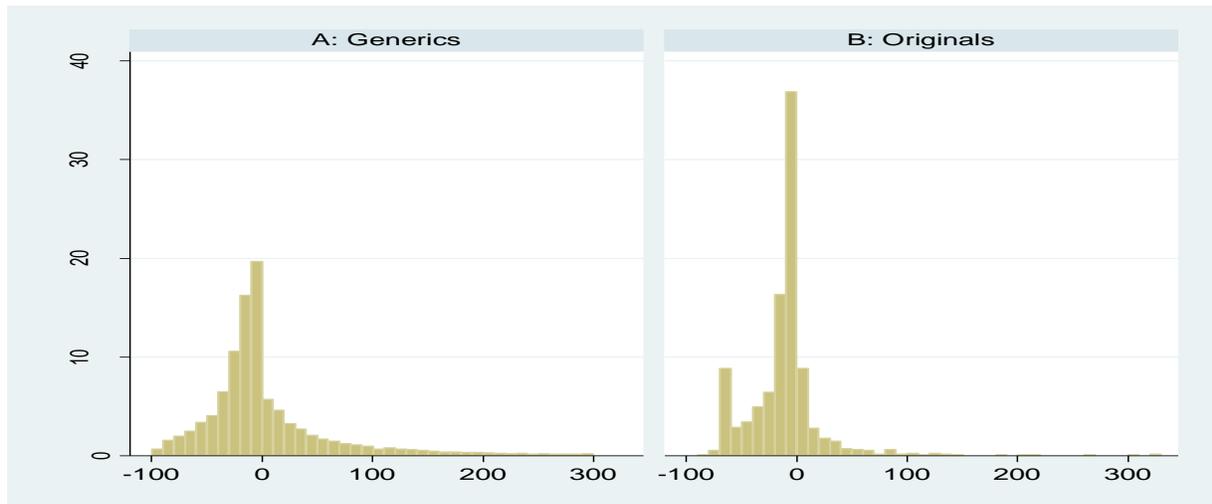


Figure 1. Distribution of price changes. The width of the bins is ten percentage points. For visual clarity, observations with no price changes are excluded; these account for 62% of the observations for generics and 97% for originals. Also, for generics the 2% of the price changes that exceed 300% are excluded.

The prices of generics are on average 53% of the price of originals. Within the exchange group and month to which they belong, 29% of generics and 88% of originals have the highest price. Also, 11% of originals are sold in exchange groups and months where two or three original products are sold. For example, both a 98 pills package and a 100 pills package may be sold or there may be sales of blisters as well as tins. These cases constitute nearly half of the cases when a given original product is not the most expensive one. Other examples occur when the original product is a 30 package that is just slightly more expensive than a 28 package, so that the 28 package is more expensive per pill, and when the original just has made a significant price cut.

Figure 2 shows how the prices, which are not the highest, compare with the highest prices within the exchange groups. We see that originals, when they are not the most expensive, most often are less than 10 percent cheaper per unit than the most expensive. For generics, in contrast, we see that it is quite common with prices lower than 15% of the highest price among exchangeable packages.

Returning to Table 1, the third variable $GenFirms_{et}$ is the number of pharmaceutical companies selling locally sourced (i.e., excluding parallel imported) generic in exchange group e in month t . For 4% of the observations this variable takes the value 0, since only original products are sold. In order to avoid missing values when creating a logarithmic version of this variable, we define a semi-logarithmic version, $lnGenFirms_{et}$, equal to $\ln(GenFirms_{et}+0.5)$. Corresponding definitions are used for $lnAddFirms_{et}$, $lnThAlt_{st}$,

$\ln ThGenAlt_{st}$, and $\ln Months_Paten_{st}$, but 0.5 is not added for $\ln P_{it}$, since P_{it} never takes the value zero, or for $\ln DDD_{et}$, discussed below.⁶

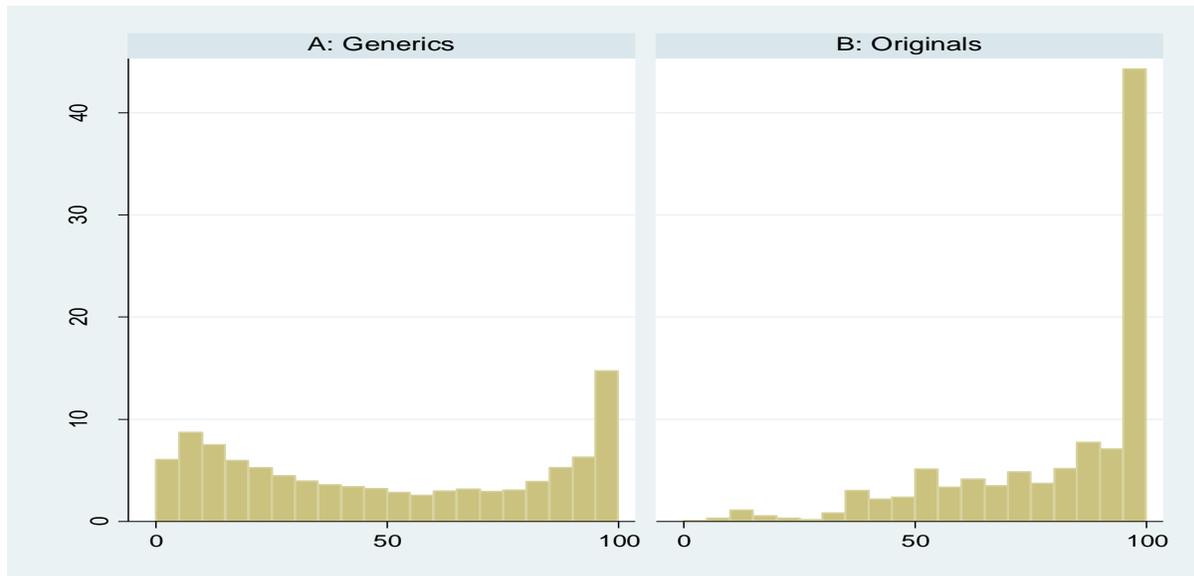


Figure 2. Distribution of price bids per unit expressed as a percentage of the highest bid within exchange group and month, excluding observations with the highest bid. The width of the bins is five percentage points. The shares of the observations having the highest bid are 29% for generics and 88% for originals.

$Orig_{et}$ is an indicator variable taking the value 1 for observations where one (or several) firms sell only original products. Hence $Orig_{et}$ takes the value 0 when no original product is sold but also in the cases when the firm selling an original product also sells a generic product, meaning that it is already included in $GenFirms_{et}$.

$AddFirms_{et}$ is defined as the number of firms selling at least one locally sourced product with the same substance as the product in question, minus the number of firms selling locally sourced products within the exchange group. In line with Brekke et al. (2009) and Pavcnik (2002), $ThAlt_{st}$ is defined as the number of other pharmaceutical substances sharing the five-digit ATC code with substance s month t . $ThGenAlt_{st}$ is defined as the number of therapeutic alternatives for which generic versions exist

The variable $2009PriceCap_{st}$ relates to the price cap at 35% of the pre-patent-expiration price which became effective in July 2009. The price cap is effective conditional on at least one generic with a market share of at least 10% having been sold at a price less than 30% of the pre-patent-expiration price, on generics having been sold in the Swedish market for more than four months, and on at least six months having passed since patent expiration. $2009PriceCap_{st}$ is an indicator variable equal to 1 for pharmaceutical substances and months after June 2009 when at least six months have passed since patent

⁶ In the appendix we show that we get the same main result if we instead add a constant of 0.5 only when the variables take the value zero and add indicator variables to control for the transformation done when the variables take the value zero.

expiration. That is, $2009PriceCap_{st}$ equals 1 for observations that could be affected by the price cap. We choose not to condition on whether the price cap is actually in effect, because this depends on generic entry and generic prices and is, therefore, endogenous.

$Months_Pat_{st}$ is defined as the number of months since the substance, according to IMS, lost its patent protection in Sweden or since the first generic product with the substance was sold, whichever came first.⁷ For 38% of the observations we lack a date for patent expiration and for these observations the variable is based only on when the first generic product with the substance was sold. This likely creates a measurement error of a few months, but since the time from the onset of generic competition on average was 122 months for these 38%, this measurement error is quite small in relative terms.

The variable DDD_{et} is the number of defined daily doses sold in exchange group e in month t and $\ln DDD_{et}$ is the natural logarithm of this variable. For 1% of the observations, DDD_{et} equals zero, but should rather be missing since daily doses for these are not defined by the World Health Organizations. For these observations, we let $\ln DDD_{et}$, which we use as an instrument in some specifications, remain missing rather than creating a semi-logarithmic variable. The last two rows of Table 1 indicate that the generics were observed on average 34 months and the originals on average 41 months.

4. Econometric specifications

As discussed in the introduction, one purpose of this paper is to distinguish between short- and long-term responses to changes in the number of competitors and to study how fast prices adjust towards new equilibria. To this end, we use a partial adjustment model.

With one lag of the dependent variable, the partial adjustment model is a special case of an error correction model. To see this, let $Y_t^* = \delta X_t + \epsilon_t$, with index t representing time, describe the long-term equilibrium relationship and let $Y_t - Y_{t-1} = (1 - \theta)(Y_t^* - Y_{t-1})$, or equivalently $Y_t = Y_{t-1} - \theta(Y_{t-1} - Y_t^*)$, describe the dynamics. Substituting for Y_t^* gives the equation to be estimated: $Y_t = \theta Y_{t-1} + \beta X_t + \epsilon_t$, where $\beta = (1 - \theta)\delta$ and $\epsilon_t = (1 - \theta)\epsilon_t$.

An error-correction model, in turn, is a reformulation of a model containing also X_{t-1} , i.e. $Y_t = \theta Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \epsilon_t$. Subtracting Y_{t-1} from both sides gives a more common representation of the error-correction model: $\Delta Y_t = \beta_0 \Delta X_t - (1 - \theta)[Y_{t-1} + \gamma X_{t-1}] + \epsilon_t$, where $\gamma = \frac{\beta_0 + \beta_1}{1 - \theta}$. Here, β_0 is the short-term effect of a change in X_t on ΔY_t and $(1 - \theta)$ is the effect on the deviation from the long-term

⁷ That generic versions were sometimes sold before patent expiration might indicate that the expiration date was disputed, but it is also possible that it was sold according to a license issued by the patent holder. As shown in the appendix, we get the same main result if we instead use $\ln Months_PatB_{st}$ – a version of the variable where only IMS's patent expiration date is used, when it is available.

equilibrium in $t - 1$ on ΔY_t . That is, the dependent variable is simultaneously affected by current changes as well as by the previous month's deviation from the long-term equilibrium.

Since we have stationary data, we can use a partial adjustment model and we choose this option mainly because it is easier to find strong instrumental variables for this model. We focus on a partial adjustment model with two lags of the dependent variable so that the dynamics is described by $Y_t = Y_t^* + \theta_1(Y_{t-1} - Y_t^*) + \theta_2(Y_{t-2} - Y_t^*)$. Substituting for Y_t^* gives the equation to be estimated: $Y_t = \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \beta X_t + \varepsilon_t$, where $\beta = (1 - \theta_1 - \theta_2)\delta$ and $\varepsilon_t = (1 - \theta_1 - \theta_2)\epsilon_t$. Hence, the long-term effect can be obtained by dividing the estimated parameters by $(1 - \theta_1 - \theta_2)$. In the appendix we show that the key OLS results are very similar if we instead use error-correction or partial adjustment models with more lags.

We first use ordinary least squares estimation for Specification 1:

$$\begin{aligned} \ln P_{i,t} = & \theta_1 \ln P_{i,t-1} + \theta_2 \ln P_{i,t-2} + \beta_1 \ln \text{GenFirms}_{e,t-1} + \beta_2 \text{Orig}_{e,t-1} + \beta_3 \ln \text{AddFirms}_{e,t-1} \quad (1) \\ & + \beta_4 \ln \text{ThAlt}_{s,t-1} + \beta_5 \ln \text{ThGenAlt}_{s,t-1} + \beta_6 2009 \text{PriceCap}_{s,t} \\ & + \beta_7 2009 \text{PriceCap}_{s,t-1} + \beta_8 2009 \text{PriceCap}_{s,t-2} + \beta_9 \ln \text{Months_Pat}_{s,t} \\ & + \sum_{m=2}^6 \gamma^m I_Months_Pat_{s,t}^m + \eta_t + \mu_i + \varepsilon_{i,t}, \end{aligned}$$

where indices i , e , s , and t represent product, exchange group, substance, and time in months, respectively. This and other specifications are estimated separately for generics and originals.

We use one-month lags for the number of generic firms ($\ln \text{GenFirms}$) and for the indicator variable for (at least) one brand-name firm selling products in the market (Orig), as well as for the variables, $\ln \text{AddFirms}$, $\ln \text{ThAlt}$ and $\ln \text{ThGenAlt}$. This is because firms, when they at the end of $t - 2$ set their prices for period t , can observe the other firms' price bids for month $t - 1$. Hence, when the prices for month t are set, firms have good information about the number of competitors they will face in month $t - 1$, but lack this information for month t .

In general, a firm's price bid for month t will depend on the expectation it has, when setting its price, about the number of competitors it will face in month t . In the Swedish generic market, this means that $\ln P_{i,t} = f(\dots, \text{GenFirms}_{e,t} | I_{t-2}, \dots)$, where I_{t-2} is the information the firm has when it in month $t - 2$ submit its price. If firms have naïve expectations in the sense of expecting the number of competitors in month t to be the number they observe for month $t - 1$, $\text{GenFirms}_{e,t} | I_{t-2}$ equals $\text{GenFirms}_{e,t-1}$, which motivates Specification 1.

Specification 1 can be estimated with OLS to study the causal effect of the number of competitors. The reason is that the error terms only depend on the current price shock, while the lags of the dependent

variable control for previous price shocks, and that firms in $t - 3$, when they choose whether or not to bid for $t - 1$, most likely cannot predict price shocks in month t . If we were using yearly averages in the estimations, rather than monthly, we could of course get endogeneity bias since firms, for example, can leave a market during a year in response to unexpectedly low prices in the beginning of the year.

If firms have rational expectations, rather than naïve, 2SLS estimation can be preferable. The reason is that an econometric prediction using information available in month $t - 2$ might be closer than the realizations in either month t or month $t - 1$ to the predictions the firms make when they, in month $t - 2$, submit their bids. Thus, we also present results from two 2SLS specifications when, instead of their lags, $\ln GenFirms_{e,t}$, $Orig_{e,t}$, $\ln AddFirms_{e,t}$, $\ln ThAlt_{s,t}$, and $\ln ThGenAlt_{s,t}$ are included. The first lags of these variables and $\ln DDD_{e,t-2}$ are used as instruments in the Specification IV 1, while their second lags and $\ln DDD_{e,t-2}$ are used in Specification IV 2.

We use a logarithm transformation for the number of generic firms because it is reasonable to think that the effect of an additional firm becomes smaller as the number of firms increases. In Section 5, we demonstrate that using indicator variables for the number of generic firms gives similar results.

$Orig_{e,t-1}$ is included in Specification 1 to study if a firm only selling original product(s) contribute at all to price competition. To study the importance of the number of additional firms selling products with the same substance as product i , but in other exchange groups, we include $\ln AddFirms_{e,t-1}$. The variables $\ln ThAlt_{s,t-1}$ and $\ln ThGenAlt_{s,t-1}$ are included to control for competition from therapeutic alternatives and therapeutic alternatives with generic products. Since we study the off-patent part of the market, we do, however, not expect large effects of these variables.

We include $2009PriceCap_{s,t}$ (taking the value 1 from July 2009 and onwards for potentially affected products) together with its first two lags. The lags are included to make it possible to offset the effect of including the lags of the dependent variable. We want to do this since it is possible that the long-term effect of the price cap equals, or is close to, the short-term effect. This should be true for many originals directly affected by the cap since they cannot choose to delay the mandatory price cut. It is, however, possible that the long-term effect of the price cap for some originals is a price below the cap, since their marginal consumer after the initial price cut might be more price sensitive. If $\beta_7 = -\theta_1\beta_6$ and $\beta_8 = -\theta_2\beta_6$ the short-term effect of the price cap equals its long-term effect.⁸

To allow for the possibility that prices fall faster in markets with recent patent expiration, we control for $\ln Months_Pat_{s,t}$ as well as for five variables indicating that two to six months have passed since the

⁸ To see this, note that the short-term effect equals β_6 while the long-term effect is given by $(\beta_6 + \beta_7 + \beta_8)/(1 - \theta_1 - \theta_2)$. Hence, if $\beta_7 = -\theta_1\beta_6$ and $\beta_8 = -\theta_2\beta_6$, then the long term effect becomes $(\beta_6 - \theta_1\beta_6 - \theta_2\beta_6)/(1 - \theta_1 - \theta_2) = \beta_6$. If $\beta_7 = -\theta_1\beta_6$, we also see that the total effect of the price cap one month after it took the value 1 are $(1 + \theta_1)\beta_6 + \beta_7 = \beta_6$, and if also $\beta_8 = -\theta_2\beta_6$ the total effect equals β_6 in all months that follows.

patent expired or generic competition began. On-patent drugs are not included in the sample and the two lags hence imply that $Months_Pat_{s,t}$ is not less than two for any observations included in the regressions. Including the indicator variables is potentially important in order to identify the effect of the price cap, since it only can affect substances whose patent expired at least six months ago.

Note that the coefficients for $lnMonths_Pat_{s,t}$ and the related indicator variables only capture non-linear effects of the time since patent expiration or onset of generic competition since we also control for time itself, using the specific effects η_t .⁹ We also control for product fixed effects, μ_i , and allow the error terms to be correlated within substances. If we instead cluster on drug, exchange group, or product level the estimated standard errors becomes, on average, 21%, 30%, and 34% smaller, respectively.

A potential problem is that serially correlated error terms could cause bias in the estimator for the lags of the dependent variable. We have therefore tested for serial correlation up to three months using a test proposed by Cumby and Huizinga (1992), as implemented by Baum and Schaffer (2013) for STATA. The test can be applied to panel dataset as ours where some regressors (such as lags of the dependent variable) are only weakly exogenous. At the 1% significance level, we cannot reject the null hypotheses of no serial correlation for any of the lags, but for generics we can reject the null of no first- and second-order serial correlation at the 5% significance level. The serial correlation is small; the estimated correlation between $\varepsilon_{i,t}$ and $\varepsilon_{i,t-1}$ is -0.155 while it is -0.003 for $\varepsilon_{i,t}$ and $\varepsilon_{i,t-2}$. Hence, the bias of the OLS estimator caused by this is likely negligible.¹⁰ Because of this and the relative low variance of the OLS estimator, we focus on OLS and IV results in the results section, but in the appendix we show that GLM regressions, allowing for first- and second-order serial correlation, gives nearly identical results as the OLS regressions.

As mentioned above, one argument for using 2SLS rather than OLS when the competition variables are not lagged, is that our econometric prediction using information available in month $t - 2$ might be closer than the realizations in period t to the predictions the firms make when they submit their bids. Another argument is that in the presence of serial correlation, $lnGenFirms_{e,t}$, $Orig_{e,t}$, $lnAddFirms_{e,t}$ can be endogenous since firms in month $t - 2$, when they decide whether or not to be active in the market in period t , have observed $\varepsilon_{i,t-1}$ (since prices are announced one month ahead). If we have serial correlation similar to that for generics in the OLS regression, also the first lags might be slightly endogenous since they can depend on $\varepsilon_{i,t-2}$. However, the second lags should be valid instruments even if we, in the IV specifications, have similar first- and second-order correlation coefficients as in the OLS regression, since

⁹ An alternative would be to include also the first and second lag for $lnMonths_Pat$ to allow the short-term effect also of this variable to equal its long-term effect. We have chosen not to do so in Specification 1, since including these variables creates so much multicollinearity, given that we also include time-specific fixed effects, while it has negligible effects on the estimates for the other variables.

¹⁰ Based on Monte Carlo studies on time-series data Keele and Kelly (2005) report biases of less than 3% for both the short- and long-term effect, when using OLS with a lagged dependent variable in a situation where the correlation coefficient is 0.2.

the correlation between $\varepsilon_{i,t}$ and $\varepsilon_{i,t-3}$ then will be below 0.01 in absolute value. We find no evidence of serial correlation in the IV 1 regressions on the 5% significance level, but do find second-order serial correlation on the 10% level and, therefore, report the result of the IV 2 specifications for comparison.¹¹

Another potential source of endogeneity bias occurs when the lagged dependent variable is included simultaneously with fixed effects. Fortunately, this bias is decreasing with the number of time periods, and (for more than a few time periods) its limit as the number of fixed-effect units approaches infinity is of the order $-(1 + \theta)/(T - 1)$ where θ is the true effect of the lagged dependent variable and T is the number of time periods (Nickell, 1981). With $\theta = 0.5$ and $T = 36$ (which is the average in our data) $-(1 + \theta)/(1 - T) = -0.04$, indicating that the bias in our case can be expected to be small. For high values of θ , like 0.9, Nickell notes that Monte Carlo simulations indicate that the bias can be considerably smaller in absolute size than suggested by this formula.

One way to avoid this Nickell bias is to take first difference to get rid of the fixed effects and instrumenting $\Delta \ln P_{i,t-1}$ that, then, becomes endogenous. When testing this, in two out of three specifications, the sum of the point estimates for the lagged coefficients became slightly lower than in the OLS specification, which is contrary to expectations since the OLS estimators should suffer from a negative bias. In the third specification the sum of the coefficients became identical on the second decimal to that obtained using OLS, but in all specification the standard errors became significantly larger than when using OLS.¹² This indicates that the bias is indeed small and that not trying to avoid this bias by taking first differences and instrumenting likely give better point estimates. In the paper, we focus on estimations where we have not accounted for this small bias, but in the appendix we show that nearly identical results are obtained when using a bias-corrected estimator, and we also report result from specifications where this bias is reduced by using firm*drug fixed effects instead of product fixed effects.

5. Results

In Table 2 we see that the adjustment speed is much faster for generics than for originals. This is expected partly because raising the price of a product that already was the most expensive one in an exchange group

¹¹ A possible explanation as to why we find significant serial correlation for generics when using OLS but not, on the 5% level, when using IV, is that the serial correlation in the first case is caused by serial correlation in the difference between the lag of one of the competition variables, e.g. $\ln \text{GenFirms}$, and firms' expectation about the value of this variable in month t .

¹² For both generics and originals we have estimated specifications where $\ln P_{i,t-3}$ and $\Delta \ln \text{GenFirms}_{e,t-2}$ are used as instruments for $\Delta \ln P_{i,t-1}$. For originals, where the second lag is less important, we also estimated an IV estimation excluding this lag which enable us to use $\ln P_{i,t-2}$ as an instrument together with $\Delta \ln \text{GenFirms}_{e,t-2}$. In all specifications, the Kleibergen-Paap rk LM statistic was above 26, allowing us to reject underidentification on the 1% significance level and we could not reject the null hypotheses of valid instruments based on the Hansen J statistic.

might result in that product being excluded from the pharmaceutical benefit scheme. In Table A1 in the appendix, we study the speed of adjustment more closely and show that, for generics, more than two thirds of the adjustment towards the new long-term equilibrium takes place within three months, irrespective of whether a partial adjustment, an error-correction or a generalized linear model accounting for second-order serial correlation is used. For originals, however, only about one fifth of the adjustment takes place within three months.

The estimates for the five competition variables, $\ln GenFirms - \ln ThGenAlt$, show that it is the number of firm selling generic product within the exchange groups that matters the most. That there is an additional firm selling the original product has no significant effect.

The estimates for $\ln AddFirms$ show that the number of firms selling locally sourced products with the same substance, but not a product in the exchange group, only has a significant effect in the OLS specification for generics. A possible explanation is that the OLS coefficient for $\ln AddFirms_{e,t-1}$ captures the firms' expectations of imminent entry into the exchange group by firms already selling the same substance. This explanation is supported by $\ln AddFirms_{e,t-1}$ having a positive effect in the first-stage regression for $\ln GenFirms_{e,t}$. To save space, the first-stage results are not presented in the paper but they are available upon request. Other results, not presented here, show that neither the number of parallel importers, nor the number of additional firms within the drug, has any significant effect on the prices. At the 5% level, we find no significant effect of therapeutic competition for these products, which all belong to substances where generic competition is possible, due to expired patent, and most often do occur.

Returning to the most important competition variable, $\ln GenFirms$, we see that for generics the point estimates are up to 4 percentage points larger in absolute size in the IV estimation.¹³ That is, firms appear to respond more strongly to an econometric prediction of the number of competitors that uses information available to firms when they submit their bids, than they appear to respond to the most recent information on $\ln GenFirms_{e,t-1}$. This is consistent with firms' expectations being rational rather than naïve. For originals we only see this pattern when comparing the OLS estimates with those for IV 1. In all IV regressions, the instruments are strong enough to reject the null hypotheses of underidentification and the null hypotheses of the instruments being uncorrelated with the error term cannot be rejected for any specifications. At the bottom of Table 2 we also report the tests for serial correlations, which show that at the 5% significance level we cannot reject the null hypotheses of no serial correlation, except for the OLS estimation for generics.

¹³ The covariance across the OLS estimate for $\ln GenFirms_{e,t-1}$ and the IV1 and IV2 estimates for $\ln GenFirms_{e,t}$ are 0.00056 and 0.00054, respectively. Together with the standard errors reported in Table 2, this implies that this difference is statistically significantly different from zero at the 1% level.

Table 2. Estimation results for $\ln P_{i,t}$

	Generics			Originals		
	OLS	IV1	IV2	OLS	IV1	IV2
$\ln P_{i,t-1}$	0.509*** (0.014)	0.508*** (0.014)	0.508*** (0.014)	0.913*** (0.017)	0.912*** (0.017)	0.912*** (0.016)
$\ln P_{i,t-2}$	0.164*** (0.014)	0.164*** (0.014)	0.164*** (0.014)	0.031** (0.015)	0.031** (0.016)	0.031** (0.015)
$\ln \text{GenFirms}_{e,t-1}$	-0.237*** (0.022)			-0.008*** (0.003)		
$\ln \text{GenFirms}_{e,t}$		-0.277*** (0.026)	-0.270*** (0.026)		-0.010*** (0.004)	-0.008** (0.003)
$\text{Orig}_{e,t-1}$	-0.022 (0.016)			-0.003 (0.005)		
$\text{Orig}_{e,t}$		-0.020 (0.019)	-0.017 (0.023)		-0.003 (0.006)	-0.005 (0.004)
$\ln \text{AddFirms}_{e,t-1}$	-0.018** (0.008)			-0.002 (0.002)		
$\ln \text{AddFirms}_{e,t}$		-0.012 (0.011)	-0.015 (0.011)		-0.001 (0.003)	0.002 (0.003)
$\ln \text{ThAlt}_{s,t-1}$	-0.014 (0.042)			0.003 (0.006)		
$\ln \text{ThAlt}_{s,t}$		-0.007 (0.045)	-0.003 (0.050)		0.004 (0.007)	0.007 (0.008)
$\ln \text{ThGenAlt}_{s,t-1}$	-0.035 (0.021)			-0.002 (0.007)		
$\ln \text{ThGenAlt}_{s,t}$		-0.042* (0.024)	-0.053* (0.027)		-0.003 (0.008)	-0.005 (0.009)
$2009\text{PriceCap}_{s,t}$	0.006 (0.054)	0.015 (0.054)	0.017 (0.054)	-0.038* (0.022)	-0.035 (0.023)	-0.036 (0.022)
$2009\text{PriceCap}_{s,t-1}$	0.025 (0.045)	0.029 (0.046)	0.028 (0.046)	-0.031 (0.051)	-0.033 (0.053)	-0.034 (0.054)
$2009\text{PriceCap}_{s,t-2}$	0.033 (0.039)	0.036 (0.039)	0.038 (0.039)	0.032 (0.040)	0.032 (0.040)	0.031 (0.039)
$\ln \text{Months_Pat}_{s,t}$	-0.081*** (0.017)	-0.085*** (0.017)	-0.085*** (0.017)	-0.012** (0.006)	-0.012** (0.006)	-0.012** (0.006)
$I_{\text{Months_Pat}}^2_{s,t}$	-0.375*** (0.086)	-0.368*** (0.084)	-0.367*** (0.084)	-0.024** (0.011)	-0.022** (0.011)	-0.021** (0.010)
$I_{\text{Months_Pat}}^3_{s,t}$	-0.198*** (0.067)	-0.190*** (0.068)	-0.189*** (0.068)	-0.030*** (0.010)	-0.030*** (0.010)	-0.029*** (0.010)
$I_{\text{Months_Pat}}^4_{s,t}$	-0.059 (0.044)	-0.049 (0.044)	-0.048 (0.043)	-0.018* (0.009)	-0.018* (0.010)	-0.018* (0.010)
$I_{\text{Months_Pat}}^5_{s,t}$	-0.047 (0.039)	-0.044 (0.038)	-0.044 (0.038)	-0.013* (0.008)	-0.013 (0.008)	-0.013 (0.008)
$I_{\text{Months_Pat}}^6_{s,t}$	0.031 (0.036)	0.032 (0.037)	0.033 (0.037)	-0.034** (0.016)	-0.034** (0.016)	-0.033** (0.016)
$d\ln P_i^*/d\ln \text{GenFirms}_e^*$	-0.725*** (0.071)	-0.845*** (0.083)	-0.824*** (0.083)	-0.146*** (0.056)	-0.172*** (0.066)	-0.135** (0.055)
Observations	121895	120924	120924	32300	31424	31424
R ²	0.450	0.449	0.449	0.916	0.918	0.918
K-P rk LM		47.403	46.198		44.494	43.342
K-P rk LM, p-value		0.000	0.000		0.000	0.000
Hansen J, p-value		0.301	0.283		0.125	0.130
Serial corr. Lag 1, p-v.	0.044	0.443	0.454	0.847	0.737	0.649
Serial corr. Lag 2, p-v.	0.012	0.099	0.109	0.248	0.542	0.581
Serial corr. Lag 3, p-v.	0.770	0.194	0.208	0.783	0.511	0.186

In all specifications, 75 indicator variables for month are included as we control for product-specific fixed effects.

$\ln \text{GenFirms}_{e,t-1}$, $\text{Orig}_{e,t-1}$, $\ln \text{AddFirms}_{e,t-1}$, $\ln \text{ThAlt}_{s,t-1}$, $\ln \text{ThGenAlt}_{s,t-1}$, and $\ln \text{DDD}_{e,t-2}$ are used as instruments in the IV 1 specification, while $\ln \text{GenFirms}_{e,t-2}$, $\text{Orig}_{e,t-2}$, $\ln \text{AddFirms}_{e,t-2}$, $\ln \text{ThAlt}_{s,t-2}$, $\ln \text{ThGenAlt}_{s,t-2}$, and $\ln \text{DDD}_{e,t-2}$ are used as instruments in the IV 2 specification. K-P rk LM is short for the Kleibergen-Paap rk LM statistic, which indicates the strength of the instruments. The null hypothesis in the Kleibergen-Paap test is the model is underidentified. The null hypotheses for the Hansen J test is that the instruments are valid, i.e., uncorrelated with the error term. The last three rows report the p-value for the serial correlation test proposed by Cumby and Huizinga (1992). The null hypotheses are no serial correlation of the first-, second-, and third-order, respectively. Standard errors, robust to correlations within substances, are given in parentheses. ***, **, * indicate that the coefficient is statistically significant different from zero on the 1%, 5% and 10% significance levels, respectively.

Figure 3 illustrates the predicted short and long-term percentage price reduction as the number of firms selling generic products within the exchange group increases, using the estimates from the IV1 specifications and the formula for calculating exact percentage changes.¹⁴ We see that an increase from one to two generic firms reduces prices by about 13% in the short term and 35% in the long term, while the effect of an increase from one to ten firms is 42% in the short term and 81% in the long term. For originals, Figure 3 illustrates the price reductions compared to having no generic competition. For example, the first generic firm reduces prices by 1% in the short-term and 17% in the long term, while the effects of an increase from zero to ten generic firms reduces prices by 3% in the short term and 41% in the long term. Using these figures, we can also infer that increasing the number of generic competitors from one to ten causes prices of originals to fall by 2% in the short term 29% in the long term.

Originals not only have a slower speed of adjustment than generics but also respond less to competition in the long term. This makes it easy in a static specification to miss that originals respond to competition. Indeed, when estimating static specifications on monthly or yearly observations, we found no significant effect of *lnGenFirms* on the prices of originals. For generics, the main advantage of using a dynamic model on monthly data, besides allowing us to distinguish between short- and long-term effects and studying the speed of adjustment, is that the instruments that can be used in the dynamic models are much stronger and that the dynamic IV results are more robust to the choice of instruments.

¹⁴ The formula $100 * [\exp(\beta) - 1]$ can be used to calculate the percentage effect of a unit increase in an explanatory variable. For small effects, however, the parameter estimates approximately equal the percentage effect. For logarithmic variables, the coefficient can be interpreted as elasticities. Since our semi-logarithmic variables are defined by adding 0.5 before taking the natural logarithm, one needs to multiply the coefficient estimate with 0.80 to approximate the elasticity of, say, *GenFirms*, when *GenFirms* is around 2. When *GenFirms* is around 3, one needs to multiply by 0.85 and as *GenFirms* increases the coefficient for *lnGenFirms* approaches the short-term elasticity of the price with respect to *GenFirms*. The exact short-term percentage effect of going from say one to two generic firms is $100 * [\exp[(\ln 2.5 - \ln 1.5) * \beta_1] - 1]$, while the long-term effect is given by $100 * [\exp[(\ln 2.5 - \ln 1.5) * \beta_1 / (1 - \theta_1 - \theta_2)] - 1]$.

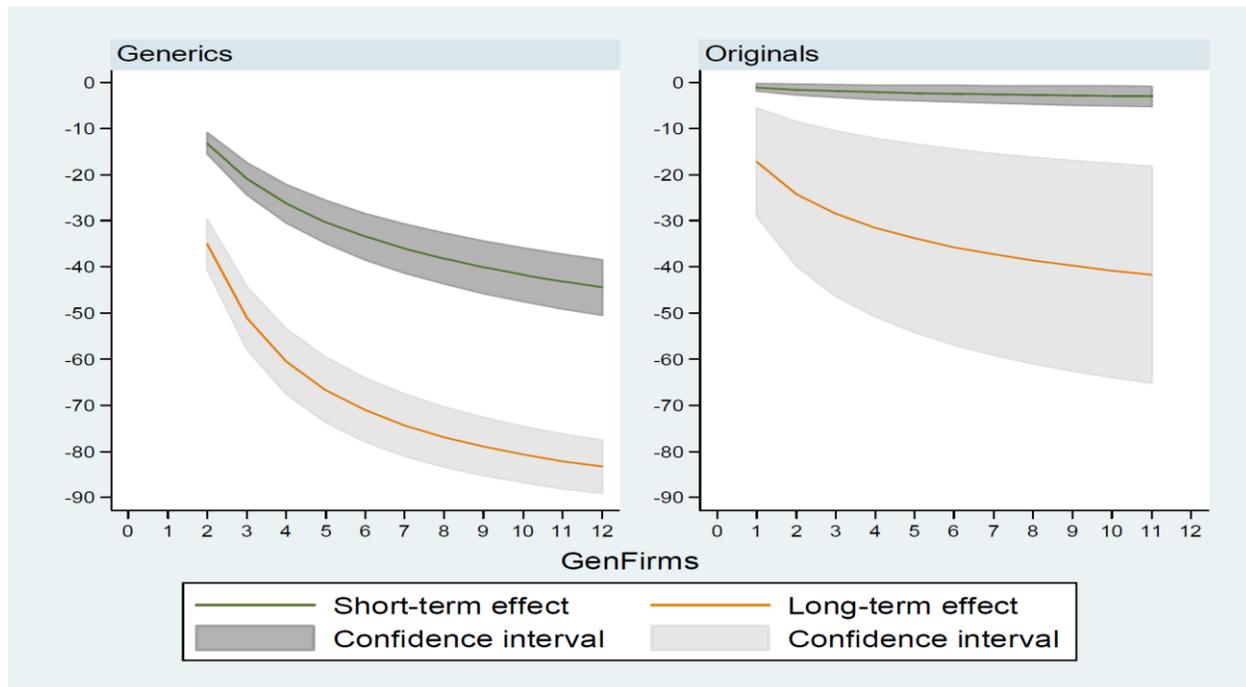


Figure 3. Estimated short and long-term price effects of the number of generic firms ($\ln GenFirms_{e,t}$) on generic and original prices, in percentages. Note that the reference point for generics is one generic firm while for originals it is no generic firm. The lines illustrate the effect of $\ln GenFirms_{e,t}$ predicted from the IV1 specifications and the gray area shows the associated 95% confidence intervals.

The estimates for $2009PriceCap_{s,t}$ and its lags indicate that the price cap has had no significant effect for generics. The point estimates are positive, which might be due to the price-cap rule giving firms an incentive to avoid selling at prices below 30% of the originals on-patent price, as this would activate the price cap. For originals, the estimate for the long-term effect of the price cap, $(\beta_7 + \beta_8 + \beta_9)/(1 - \theta_1 - \theta_2)$, is in the first two specifications -0.64^{***} (0.18), indicating average price reductions of 47%. In the IV 2 specification, the effect is two percentage points more negative. Neither for generics nor for originals can we rule out that the short-term effect of the price cap equals the long-term effect. However, this does not imply that the point estimates are very similar. Instead, this is largely a consequence of the price-cap effect being quite imprecisely estimated. This, in turn, is to be expected since we include year-month fixed effects and since the price cap variable equals 1 for 97% of the observations after July 2009. Lastly, the estimates for $\ln Months_{Pat}_{s,t}$ and the related indicator variables show that prices fall the fastest the first month after patent expiration or the onset of generic competition.

In Figure 4, we see that for generics nearly identical results are obtained when using indicator variables for the number of generic firms instead of the semi-logarithmic variable $\ln GenFirms_{e,t-1}$. Using $\ln GenFirms_{e,t-1}$ assumes e.g. that the price reduction of going from 2 to 3 firms is equally large as that of going from 7 to 10 firms, since $3.5/2.5=10.5/7.5$. When we use indicator variables, we relax this

assumption but the results show that the effects are nearly identical; the long-term effect of a third firm is a price reduction with 22% (std. err. 4.14%) while that of going from 7 to 10 firms reduces prices with 21% (std. err. 9.24%). Thus, the predicted large effects of increasing the number of firms, when that number is already large are not due to the choice of functional form.

Also for originals the estimates using indicator variables suggest that the semi-logarithmic specification is reasonable. Except for one and eleven generic firms, the point estimates using indicators falls within the fairly small confidence interval obtained using the logarithmic variable. For originals we do, however, get considerably larger confidence intervals when indicator variables are used.

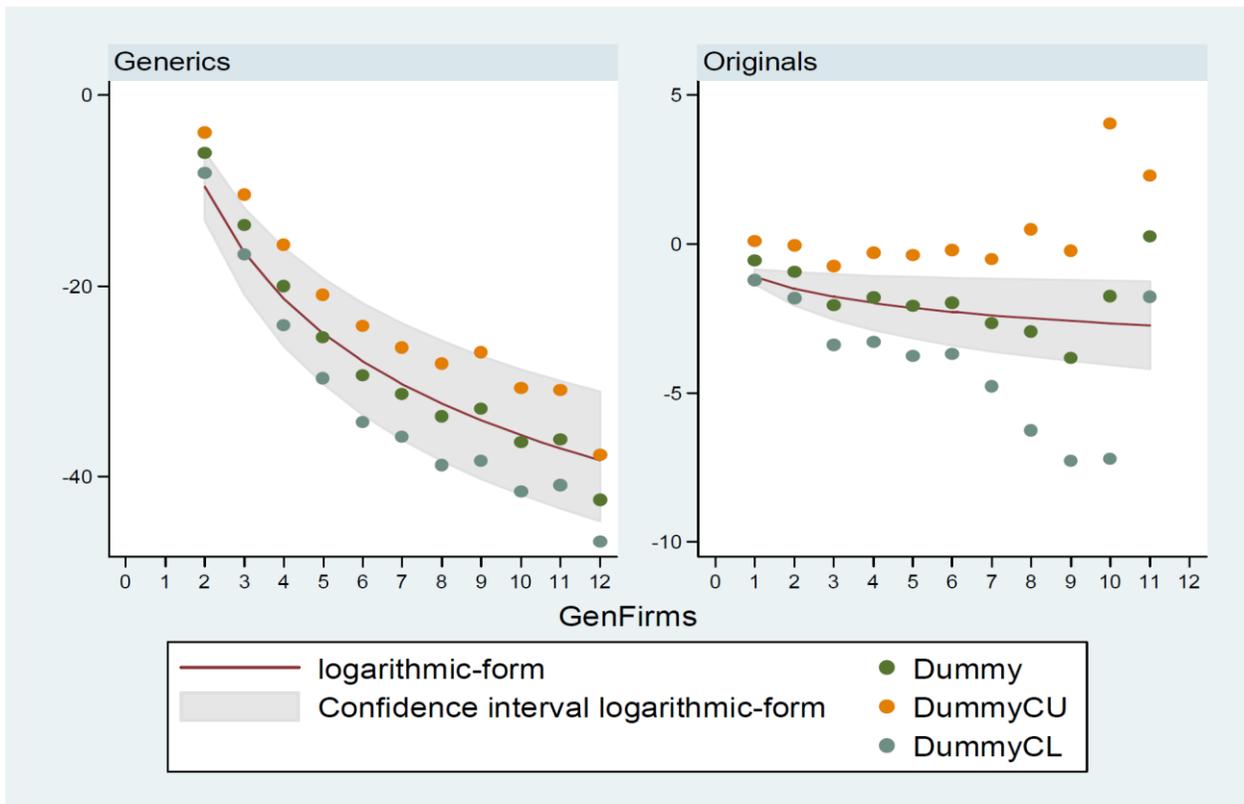


Figure 4. Estimated short-term price effects in percentages of number of generic firms previous month ($GenFirms_{e,t-1}$) on generic and original prices; comparison of logarithmic-form and flexible-form estimates. Note that different scales are used for generics and originals and that the reference point for generics is one generic firm while for originals it is no generic firm. The smooth line is the effect of $\ln GenFirms_{e,t-1}$ predicted from the OLS specification except that the difference between the point estimates for the constant between this specification and the one with indicator variables for number of generic firms previous month are added to these predictions. These differences are 0.12 for generics and -0.01 for originals. The gray area shows the associated 95% confidence interval. The dots are the point estimates of indicator variables for number of generic firms previous month and the associated 95% confidence interval. When indicator variables are used, the coefficients for $\ln P_{e,t-1}$ and $\ln P_{e,t-2}$ are for generics 0.508 and 0.163, compared with 0.509 and 0.164 when $\ln GenFirms_{e,t-1}$ is used. For originals the coefficients are 0.912 and 0.031 compared with 0.913 and 0.031 when $\ln GenFirms_{e,t-1}$ is used. This implies that the long-term effects of number of generic firms previous month are also nearly identical.

Even though the competition in this market likely best can be described as a repeated price-setting game, it might be of interest to note how the relationship between price and the number of generic firms relates to static competition models. A Cournot-competition model would predict larger relative effects of the first competitors and smaller relative effects when the number of firms is already large. A price-setting model where each generic firm has an equal number of loyal consumers and compete for consumers that are not loyal (see e.g. Varian, 1980, and Barut and Kovenock, 1998) could result in very similar price-effect for generics as those estimated with indicator variables here, if the right assumptions are imposed about the theoretical minimum price, that are part of firms' strategies, and the average price.¹⁵

In the appendix we show that the key results do not change much if we instead: use an error-correction model; use a generalized linear model accounting for first- and second-order serial correlation; include more lags of the dependent variable; use an estimator correcting for the Nickell bias; use drug*firm fixed effects instead of product fixed effects; exclude products with the package size units Dosages, Grams, and Milliliters; or use a different variable to account for the age of the market.

6. Conclusions

This paper provides an empirical investigation of the causal short- and long-term price effects of the number of sellers of pharmaceuticals. The analysis is done using panel data, allowing control for time-invariant differences in demand and costs across products and control for time-specific fixed effects. Studying this effect for pharmaceuticals has the advantage that the quality is highly controlled and not allowed to change over time. In addition, advertising directed toward consumers is banned by law for prescription pharmaceuticals in Sweden and the physical and financial conditions for delivery and payment are fixed by the market regulator, which further limits the scope for non-price competitive actions which trouble many studies trying to estimate the effect of the number of sellers on prices.

Another problem in this field of research is the simultaneity problem. Applying a dynamic model to monthly data and using the fact that the rules require firms that want to be active in the market to submit their price bids in advance, effectively solves this problem. Hence, if firms are naïve, so that they expect the number of competitors in month t to be the number they observe in month $t - 1$, we can use OLS to

¹⁵ For example, if marginal cost is set to zero, 60% of consumers buying generics are assumed to buy the cheapest one while the other generics get equally large market shares consisting of loyal (and price-insensitive) consumers and the maximum price is normalized to 1, then the minimum price that is part of the firms' strategies becomes $P_{min}=0.4/(0.6*(GenFirms-1))$ for $GenFirms>1$. The average price will be a function of this minimum price, the maximum price (P_{max}) and the number of firms. If the average price happens to be $[(GenFirms - 1)*P_{min}+P_{max}]/GenFirms*0.88+P_{max}*0.12$, we get essentially the same long-term effects of $GenFirms$ as that estimated in this paper.

estimate the effect of the number of firms with flexible-form specifications. If firms instead have rational expectations, we can use the number of firms in month $t - 1$ as a powerful predictor of the number in the subsequent month.

The results show that in the long term generic prices fall by 81% and original prices by 29% as a consequence of the number of firms selling generics in the exchange group increasing from one to ten. The effect on generic prices is larger than the effects reported by Caves et al. (1991), Frank and Salkever (1997), Wiggins and Maness (2004) and Reiffen and Ward (2005), but lower than the observed price cuts 25 months after generic entry that Bernt and Aitiken (2001) report. The effect for originals is larger than the estimates of Caves et al. (1991) for drugstore prices but in line with the estimates of Wiggins and Maness (2004).

The results also show that the effect of additional competitors is large also when the number of firms is already quite large. For example, going from seven to ten firms reduces generic prices in the long term with 21% when no functional form is imposed for the effect of the number of firms on prices. This effect is about twice as large as that found by Reiffen and Ward (2005), who, to the best of our knowledge, is the only previous study that have estimate this causal effect without imposing a functional form.

Although the hypotheses we test are different, our results can also be related to the reference-pricing literature. A typical finding is that the *introduction* of reference pricing reduces generic prices by 5-20 % in the short run, with unclear effects in the long run (López-Casanovas and Puig-Junoy, 2000; Dylst et al, 2011). We do not estimate the effect of reference pricing, but we find that *within* the Swedish reference-price system, going from one generic to two reduces prices by 35 % in the long run and going from one to four reduces prices by 60 %, with considerably smaller effects in the short run.

A possible reason for the relatively large effects we observe is that the suppliers of generics have strong incentives to reduce prices, as the low bidder in the monthly auction gets about half of the national market. Another reason may be that the rivals we observe are very close substitutes, since the reference group is defined very narrowly; more distant substitutes have a much weaker effect.¹⁶

In the theoretical literature on reference pricing, concern has been raised that the providers of generics may be able to collude. In Ghislandi (2011) multiple sellers of homogenous generics may collude under similar conditions as those that allow sellers of homogenous products in regular markets to collude: if the

¹⁶ This finding is consistent with that reported by Morrison and Winston (1987) that actual competition is much more effective than potential competition in the airline industry, but inconsistent with Bergman and Rudholm (2003) who found strong effects of potential competition in the first Swedish reference-price system.

discount factor is sufficiently high. He concludes that even though collusion is more difficult to sustain with internal reference pricing than in a regular market, generics' prices should be fixed for long periods of time, in order to make "cheating" (price competition) more tempting. In a two-period model (hence, a model without collusion concerns), Miraldo (2009) concludes that internal reference pricing is to be preferred, as it provides weaker unilateral incentives to adjust prices upward.

According to the Bertrand paradox, two sellers of a homogenous product suffice to trigger intense competition with marginal-cost pricing. However, according to the Folk theorem, when the interaction is repeated indefinitely, the sellers will be able to sustain collusion if they are sufficiently patient. Collusion is sustained by the threat of a punishment, e.g., in the form of a price war. With reasonable assumptions on discount rates and the frequency of price revisions, collusion tends to be possible for large numbers of firms, hence reversing the Bertrand paradox. The standard reasons as to why collusion is more difficult to sustain when the number of rivals increases, in combination with the presence of some perceived differentiation between different generics, may explain why we see a gradual response of prices to the number of firms, rather than a sharp decrease when a second rival enters or, conversely, no response at all even following the entry of multiple rivals, due to conditions that make collusion easy.

The paper contributes to the existing literature by estimating how fast firms adjust prices to changing number of competitors. For generic prices, 70% of the adjustment takes place within three months, but for originals the corresponding figure is only 20%. Like the other main results reported in the paper, these results are robust. To be more precise, the figures are around 70% and 20%, respectively, irrespective of whether a partial-adjustment, an error-correction or a generalized linear model accounting for first- and second-order serial correlation is used.

The large effect of the number of generic firms, also when the number is already quite large, implies that the pharmaceutical costs can be reduced substantially if this number were increased. One way to increase the number of competitors could be to increase the time between the auctions and the delivery periods. Then, firms would have more time to ensure deliveries in case their bid is the lowest one, which might induce some to place bids also for package sizes that they do not currently stock. Another option can be to increase the length of the contract period, which could attract more competitors by increasing the value of the market and make collusion more difficult. A third policy option is to reduce the administrative fees for being active in the market.

Appendix

Using another functional form for logarithmic variables

To avoid missing values when the variables take the value zero, we added 0.5 to the variables $GenFirms_{e,t}$, $lnAddFirms_{e,t}$, $lnThAlt_{st}$, $lnThGenAlt_{st}$, and $lnMonths_Paten_{st}$ before taking the natural logarithm. An alternative is to add a constant to the variables only in instances when they take the value zero and then also add an indicator variable, for each of the five variables, to control for the transformation done when the variables take the value zero. An advantage of the latter alternative is that it facilitates interpretation of the related coefficient for strictly positive values of e.g. $GenFirms_{e,t}$. A disadvantage, in our setting, is that the indicator variables potentially are endogenous and, hence, must be instrumented. For these reason we have chosen not to use this transformation in the main analysis. In Figure A1 we show that the main results are nearly unaffected by the choice of transformation and that using the alternative transformation gives marginally larger confidence intervals.

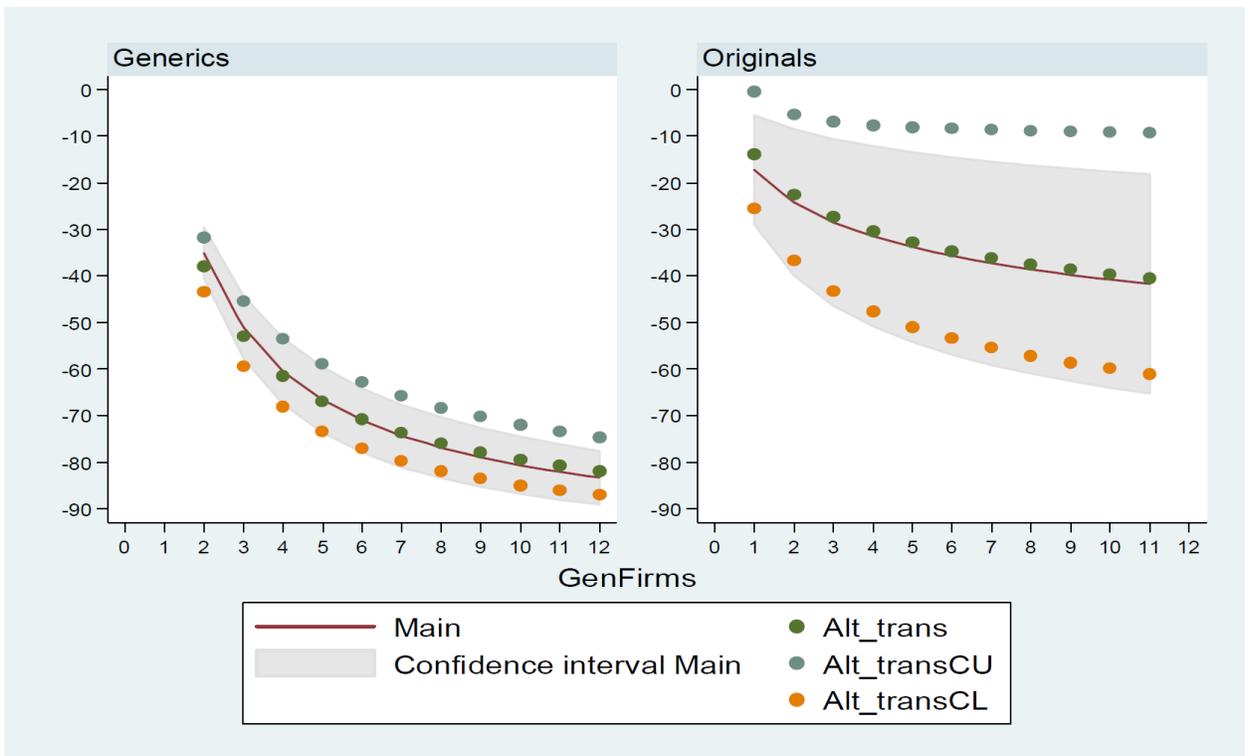


Figure A1. Estimated long-term price effects of the number of generic firms ($lnGenFirms_{e,t}$) on generic and original prices, in percentages. Note that the reference point for generics is one generic firm while for originals it is no generic firm. The smooth line illustrates the effect when predicted from the IV1 specifications and the gray area shows the associated 95% confidence intervals. The dots are the point estimates and the associated 95% confidence intervals from specifications than only differ from the IV1 specifications by instead using the alternative transformation, as discussed in this appendix, to handle zeros. The sum of the coefficients for $lnP_{e,t-1}$ and $lnP_{e,t-2}$ are identical to the third decimal for the two alternative transformations. This implies that the short-term effects of the number of generic firms are also nearly identical.

Error-correction models and models with second-order serial correlation

In Table A1 we compare the baseline OLS estimates with an error-correction model that, in order to increase the comparability with the baseline estimates, also includes a second lag of the dependent variable, as well as with a generalized linear model accounting for first- and second-order serial correlation (AR2). The standard errors for the AR2 estimations are smaller than for the other estimations, likely because that they are not robust to arbitrary correlations within substances. In the table we report the effect in month t , $t+1$, $t+2$, and $t+3$, of a persistent change in $\ln GenFirms_{e,t-1}$. We refer to the effect in month t as the direct effect. We also report the long-term effects of $\ln GenFirms$ as well as the speed of adjustment, defined as 1 minus the sum of the coefficients for the two lagged dependent variables.

Table A1. Comparisons with error-correction models and a generalized linear model

	<i>Generics</i>			<i>Originals</i>		
	OLS	Error-corr.	AR2	OLS	Error-corr.	AR2
Effect at t	-0.237*** (0.022)	-0.193*** (0.025)	-0.200*** (0.011)	-0.008*** (0.003)	-0.013 (0.010)	-0.007*** (0.002)
Effect at $t+1$	-0.357*** (0.034)	-0.341*** (0.033)	-0.331*** (0.018)	-0.016*** (0.006)	-0.020* (0.011)	-0.015*** (0.004)
Effect at $t+2$	-0.458*** (0.043)	-0.448*** (0.043)	-0.432*** (0.024)	-0.023*** (0.009)	-0.027** (0.012)	-0.022*** (0.006)
Effect at $t+3$	-0.528*** (0.050)	-0.527*** (0.050)	-0.509*** (0.028)	-0.030*** (0.011)	-0.033** (0.014)	-0.029*** (0.008)
$d\ln P_i^*$ $/d\ln GenFirms_e^*$	-0.725*** (0.071)	-0.741*** (0.073)	-0.744*** (0.041)	-0.146*** (0.056)	-0.142*** (0.050)	-0.149*** (0.040)
$1 - \theta_1 - \theta_2$	0.327*** (0.011)	0.328*** (0.011)	0.269*** (0.005)	0.057*** (0.007)	0.056*** (0.007)	0.050*** (0.003)
Observations	121895	121895	121627	32300	32300	32203
R ²	0.450	0.222		0.916	0.125	

Note: The AR2 specification includes the same variables as the OLS specifications; see Table 2. The error-correction specification in addition includes one deeper lag for each independent variable. Unlike in the other model, the standard errors in the AR2 model are not robust to arbitrarily serial correlation within substances. Standard errors are given in parentheses. ***, **, * indicate that the coefficient is statistically significant different from zero on the 1%, 5% and 10% significance levels, respectively.

For both generics and originals we see that the estimates for the long-term effect are not sensitive to the choice of model. Instead, the largest difference is found for the speed of adjustment. In the AR2 model for generics, the point estimate for this is six percentage points lower than in the baseline estimation. The slower adjustment is related to a smaller direct effect; for the AR2 model the direct effect is nearly four percentage points lower in absolute value compared to the baseline OLS estimates. The differences across models diminish fast, and are at month $t+3$ already just half as large as in month t . For originals, the most notable difference between the baseline OLS estimates and the error correction estimates is perhaps that the short-term effects are estimated far less precisely when using the error-correction model. Note that we here report AR2 results also for originals, despite the fact that the test did not indicate any serial correlation in the error terms for originals.

Effects of instead using only one, or up to six, lags

In Table A2 we see that for generics, the estimated short-term effect is five percentage points larger in absolute size when including only one lag of the dependent variable, while going from two to six lags changes the estimated short-term effect with only two percentage points. We also see that the estimated long-term effect for generics is nearly unaffected by the number of lags included. For originals, we get identical results on the third decimals for both the short- and long-term effect when having one lag of the dependent variable instead of two. When six lags are included, the estimated long-term effect becomes three percentage points smaller in absolute size, but estimations not presented in tables show that this is nearly entirely driven by the changed sample.

Table A2. Different number of lags of the dependent variable

	<i>Generics</i>			<i>Originals</i>		
	One lag	Three lags	Six lags	One lag	Three lags	Six lags
$\ln P_{i,t-1}$	0.611*** (0.012)	0.500*** (0.014)	0.487*** (0.015)	0.944*** (0.007)	0.909*** (0.017)	0.906*** (0.017)
$\ln P_{i,t-2}$		0.154*** (0.016)	0.153*** (0.015)		0.013 (0.021)	0.031** (0.015)
$\ln P_{i,t-3}$		0.030*** (0.008)	0.014* (0.008)		0.021 (0.016)	0.001 (0.021)
$\ln P_{i,t-4}$			0.022** (0.009)			0.021 (0.025)
$\ln P_{i,t-5}$			0.006 (0.006)			-0.009 (0.013)
$\ln P_{i,t-6}$			0.012** (0.006)			-0.011 (0.007)
$\ln \text{GenFirms}_{e,t-1}$	-0.280*** (0.025)	-0.227*** (0.022)	-0.217*** (0.023)	-0.008*** (0.003)	-0.008** (0.003)	-0.007** (0.003)
$\sum \ln P_{i,t-s}$	0.611*** (0.012)	0.684*** (0.011)	0.694*** (0.014)	0.944*** (0.007)	0.943*** (0.007)	0.939*** (0.007)
$d \ln P_i^*$ $/d \ln \text{GenFirms}_e^*$	-0.719*** (0.067)	-0.720*** (0.073)	-0.711*** (0.082)	-0.145*** (0.048)	-0.140** (0.057)	-0.113** (0.050)
Observations	126770	117623	106174	33223	31453	29202
R ²	0.442	0.446	0.429	0.918	0.914	0.912

Note: The specifications only differ from the OLS specification (see Table 2) by including different number of lags of $\ln P$ and 2009PriceCap (all specifications include the same number of lags of 2009PriceCap as it does of $\ln P$). Standard errors, robust to correlations within substances, are given in parentheses. ***, **, * indicate that the coefficient is statistically significant different from zero on the 1%, 5% and 10% significance levels, respectively.

Other robustness checks

As noted in Section 4, we might have a small bias in the estimators for the lagged dependent variable since we also include product fixed effects. One way to address this is to use a bias-corrected estimator, where one first obtain unbiased, but perhaps imprecise, estimates, which then are used to approximate the bias. In the last step, the approximated biases are subtracted from the OLS estimates. In columns 1 and 5 of Table A3 we report key results from such estimations done using the `xtlsdvc` routine by Bruno (2005),

where the unbiased estimates in the first stage are from Anderson-Hsiao estimations.¹⁷ This routine has been developed for the case with one lag of the dependent variable, so we exclude the second lag.

Compared with the OLS estimates with just one lag, reported in the first and fourth columns of Table A2, we see that the estimate for $\ln P_{i,t-1}$ for generics is 3 percentage points larger using the bias correction, but it is identical on the third decimal for originals. For generics, the estimated short-term effect becomes slightly closer to zero after correcting for the bias and the long-term effect becomes four percentage points larger in absolute size. For originals the short-term effect becomes marginally larger in absolute size, resulting in a one percentage point more negative estimate for the long-term effect.

An alternative to bias correction in our case is to reduce the Nickell bias by increasing the number of observations per fixed-effect unit. To study this possibility we, in Table A3, report results obtained when using drug*firm fixed effects instead of product fixed effects. We then get 1.63 as many observations per fixed-effect unit for both generics and originals. This should reduce the Nickell bias in the same order, e.g. from -0.05 to -0.03 if the formula $-(1 + \theta)/(T - 1)$ applies, but it should be noted that this formula is only a rough approximation. In these specifications, we control for four polynomials of $Rel_Package_sizes_i$, defined as $Package_sizes_i$ divided by the average package size within the drug*firm group, since differences in package sizes likely is the main difference across a firm's products based on the same drug.

Comparing the results in columns 2 and 6 of Table A3 with columns 1 and 4 in Table 2, we see that the coefficients for the lagged dependent variables become larger when using drug*firm fixed effects. For generics, these increases are larger than the expected reduction in the Nickell bias, possibly indicating that the drug*firm fixed effects together with the four polynomials for package sizes do not control for all time-invariant differences across products, thus leading to a positive bias in the estimators for the lags. For originals, however, the increase is approximately half of the increase predicted by the formula $-(1 + \theta)/(T - 1)$. Comparing these specifications with the baseline OLS estimations, we also see that for generics the changes in the coefficients for the lagged dependent variables are offset by a changed estimate for $\ln GenFirms_{e,t-1}$, leaving the estimate for the long-term effect nearly unchanged. For originals, however, the point estimate for the long-term effect is nearly one standard error larger in absolute size, compared to the OLS 1 specification.

¹⁷ The Arellano-Bond and the Blundell-Bond instrument set was found invalid according to Sargan tests, which is the reason why we here only present result using the second lag of the dependent variable as instrument for the first difference of this variable, as suggested by Anderson and Hsiao (1982).

In columns 3 and 7 of Table A3, we report key results from estimations when excluding products with the package-size units *Dosages*, *Grams*, and *Milliliters*, keeping just those with the most common measurement unit *Number of pills or capsules*. The motivation is that the definition of the package-size groups, and hence the width of the exchange groups, differ across these four categories which potentially could influence the effect of the number of firms selling generic products within the exchange group. The results show that excluding these three categories increases the absolute value of the point estimate for the long-term effect with 0.01 for generics and 0.02 for originals.

As described in the data section, $\ln\text{Months_Pat}_{st}$ is defined using the number of months since the substance (according to IMS) lost is patent protection or since the first generic product with the substance was sold, whichever came first. In columns 4 and 8 of Table A3, we report results obtained when we instead use $\ln\text{Months_PatB}_{st}$ – a version of the variable where only IMS’s patent expiration date is used, when it is available. The main results are nearly identical with the OLS results presented in Table 2, but the point estimates for $\ln\text{Months_PatB}_{st}$ are not significantly different from zero and just about a tenth of the point estimates for $\ln\text{Months_Pat}_{st}$, indicating that $\ln\text{Months_Pat}_{st}$ better explains the price variation in the data.

Table A3. Other robustness checks

	<i>Generics</i>				<i>Originals</i>			
	Bias-Cor.	Drug*firm	Pills	<i>lnM. Pat</i>	Bias-Cor.	Drug*firm	Pills	<i>lnM. PatB</i>
$\ln P_{i,t-1}$	0.643*** (0.002)	0.554*** (0.013)	0.508*** (0.014)	0.510*** (0.014)	0.944*** (0.005)	0.924*** (0.016)	0.912*** (0.017)	0.912*** (0.017)
$\ln P_{i,t-2}$		0.206*** (0.015)	0.164*** (0.014)	0.165*** (0.014)		0.034** (0.016)	0.031* (0.016)	0.032** (0.016)
$\ln\text{GenFirms}_{e,t-1}$	-0.272*** (0.008)	-0.182*** (0.015)	-0.242*** (0.023)	-0.234*** (0.023)	-0.009*** (0.002)	-0.008*** (0.002)	-0.010*** (0.003)	-0.008*** (0.003)
$\ln\text{Months_PatB}_s$				-0.009 (0.016)				-0.001 (0.004)
$d\ln P_i^*$	-0.762*** (0.022)	-0.757*** (0.057)	-0.737*** (0.072)	-0.722*** (0.073)	-0.157*** (0.029)	-0.191*** (0.058)	-0.166*** (0.057)	-0.149*** (0.047)
Observations	121896	121896	116019	121665		32300	29471	32107
R ²		0.894	0.450	0.449		0.997	0.916	0.915

Note: The specifications include the same variables as the OLS 1 specification except that: the Bias-Cor. specification does not include $\ln P_{i,t-2}$ and $2009\text{PriceCap}_{s,t-2}$; the Drug*firm specification includes $\text{Rel_Package_sizes}_i$ raised to the powers 1, 2, 3, and 4 and include drug*firms fixed effect instead on product fixed effects; and the *lnM. PatB* specification includes $\ln\text{Months_PatB}_{st}$ and five indicator variables related to this variable instead of $\ln\text{Months_Pat}_{st}$ and its five indicator variables. The Pills regressions only differ from the OLS 1 regressions by excluding observations with the package size units *Dosages*, *Grams*, and *Milliliters*. Standard errors, robust to correlations within substances, are given in parentheses. In the Bias-Cor. specification, bootstrapped standard errors using 50 repetitions are reported. ***, **, * indicate that the coefficient is statistically significant different from zero on the 1%, 5% and 10% significance levels, respectively.

References

- Anderson, T.W. and C. Hsiao (1982) Formulation and estimation of dynamic models using panel data, *Journal of Econometrics* 18(1), 47-82.
- Barut, Y. and D. Kovenock (1998) The symmetric multiple prize all-pay auction with complete information. *European Journal of Political Economy* 14(4), 627-644.
- Baum, C.F. and M.E. Schaffer (2013) actest: Stata module to perform Cumby-Huizinga general test for autocorrelation in time series. <http://ideas.repec.org/c/boc/bocode/s457668.html>
- Bergman, M.A. and N. Rudholm (2003) The relative importance of actual and potential competition: empirical evidence from the pharmaceuticals market, *The Journal of Industrial Economics* 51(4), 455-467.
- Bergman, M.A., D. Granlund, and N. Rudholm (2017) Squeezing the Last Drop Out of Your Suppliers: An Empirical Study of Market-Based Purchasing Policies for Generic Pharmaceuticals. *Oxford Bulletin of Economics and Statistics* 79(6), 969-996.
- Berndt, E.R. and M.L. Aitken (2011) Brand Loyalty, Generic Entry and Price Competition in Pharmaceuticals in the Quarter Century after the 1984 Waxman-Hatch Legislation, *International Journal of the Economics of Business* 18(2), 177–201.
- Brekke, K.R., I. Königbauer, and O.R. Straume (2007) Reference pricing of pharmaceuticals, *Journal of Health Economics* 26(3), 613-642.
- Brekke, K.R., A.L. Grasdahl, and T.H. Holmås (2009) Regulation and pricing of pharmaceuticals: Reference pricing or price cap regulation, *European Economic Review* 53(2): 170-185.
- Brekke, K.R., T.H. Holmas, and O.R. Straume (2011) Reference pricing, competition, and pharmaceutical expenditures: Theory and evidence from a natural experiment. *Journal of Public Economics* 95(7), 624-638.
- Bruno, G. (2005) *Estimation and inference in dynamic unbalanced panel data models with a small number of individuals*. Università commerciale Luigi Bocconi.
- Caves, R.E., M.D Whinston, and M.A Hurwitz (1991) Patent Expiration, Entry and Competition in the U.S. Pharmaceutical Industry: An Exploratory Analysis. *Brookings Papers on Economic Activity, Microeconomics* 1991, 1-66.
- Ching, A. (2010a) Consumer learning and heterogeneity: Dynamics of demand for prescription drugs after patent expiration. *International Journal of Industrial Organization* 28(6), 619-638.
- Ching A (2010b) A dynamic oligopoly structural model for the prescription drug market after patent expiration. *International Economic Review* 51(4), 1175–1207.
- Crockett, S., R. Oprea, and C.R. Plott (2011) Extreme walrasian dynamics: The gale example in the lab. *The American Economic Review* 101(7), 3196-3220.

Cumby, R.E. and J. Huizinga (1992) Testing the autocorrelation structure of disturbances in ordinary least squares and instrumental variables regressions. *Econometrica* 60(1), 185-195.

Danzon, P.M. and L.W. Chao (2000) Does regulation drive out competition in pharmaceutical markets? *The Journal of Law and Economics* 43(2), 311-358.

Davis, P. (2005) The Effect of Local Competition on Admission Prices in the U.S. Motion Picture Exhibition Market. *Journal of Law and Economics* 48(2), 677-707.

Dental and Pharmaceutical Benefits Agency (2009) TLVFS 2009:7. Föreskrifter om ändring i Tandvårds- och läkemedelsförmånsverkets föreskrifter och allmänna råd (2009:4) om prissättning av utbytbara läkemedel och utbyte av läkemedel m.m. [Regulation about changes in Dental and Pharmaceutical Benefit Agency's regulations and general advice (2009:4) about the price setting of exchangeable pharmaceuticals and substitution of pharmaceuticals, etc.] (in Swedish) available at <http://www.tlv.se>.

Dylst, P., A. Vulto, and S. Simoens (2011) The impact of reference-pricing systems in Europe: A literature review and case studies, *Expert Review of Pharmacoeconomics & Outcomes Research* 11(6), 729-737.

Frank, R.G. and D.S. Salkever (1997) Generic Entry and Pricing of Pharmaceuticals. *Journal of Economics and Management Strategy* 6(1), 75-90.

Galizzi, M.M., S. Ghislandi, and M. Miraldo (2011) What do we really know about reference pricing for pharmaceuticals. *Eurohealth* 17(1), 17-19.

Ghislandi, S. (2011) Competition and the Reference Pricing Scheme for pharmaceuticals. *Journal of health economics* 30(6), 1137-1149.

Granlund, D. and N. Rudholm (2012) The prescribing physician's influence on consumer choice between medically equivalent pharmaceuticals. *Review of Industrial Organization* 41(3), 207-222.

Kaiser, U., and S. Mendez (2015) How do drug prices respond to a change from external to internal reference pricing? Evidence from a Danish regulatory reform, IZA Discussion Papers, No. 8759.

Keele, L. and N.J. Kelly (2006) Dynamic models for dynamic theories: The ins and outs of lagged dependent variables. *Political analysis* 14(2), 186-205.

List, J.A. (2003) Does market experience eliminate anomalies. *Quarterly Journal of Economics* 118(1), 41-72.

López-Casasnovas, G. and J. Puig-Junoy (2000) Review of the literature on reference pricing. *Health Policy* 54(2), 87-123.

Mazzeo, M.J. (2002) Product choice and oligopoly market structure. *RAND Journal of Economics* 33(2), 1-22.

Ministry of Health and Social Affairs (2009a) Förordning (2009:659) om handel med läkemedel [Ordinance (2009:659) regarding trade of pharmaceuticals] (in Swedish) available at <http://www.riksdagen.se>.

Ministry of Health and Social Affairs (2009b) Lag (2009:366) om handel med läkemedel [Law (2009:366) regarding trade of pharmaceuticals] (in Swedish) available at <http://www.riksdagen.se>.

Miraldo, M. (2009) Reference pricing and firms' pricing strategies, *Journal of Health Economics* 28(1), 176-197.

Morrison, S.A. and C. Winston (1987) Empirical Implications and Tests of the Contestability Hypothesis, *Journal of Law and Economics* 30(1), 53-66.

National Board on Health and Welfare (2013) Läkemedelsförsäljningen i Sverige – Analys och prognos [Pharmaceutical sales in Sweden: Analysis and forecast], Stockholm (in Swedish).

Nickell, S. (1981) Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society* 49(6), 1417-1426.

Olsson, E., H. Wallach Kildemoes, E. Carlsson, C. Hällkvist, S. Kaae, and S. Källemark Sporrang (2015) The case of generic substitution: What factors influence the patients' trust in bioequivalence? In Olsson, E. Generic substitution in Swedish Community Pharmacies Understanding the influence of a pharmaceutical policy on pharmacy practice, PhD thesis Graduate School of The Faculty of Health and Medical Sciences, University of Copenhagen.

Pavcnik, N. (2002) Do pharmaceutical prices respond to potential patient out of pocket expenses, *RAND Journal of Economics* 33(3), 469-487.

Plott, C. and G. George (1992) Marshallian vs. Walrasian stability in an experimental market. *The Economic Journal* 102(412), 437-460.

Reiffen, D. and M.R. Ward (2005) Generic drug industry dynamics. *Review of Economics and Statistics*, 87(1), 37-49.

Regan, T.L. (2008) Generic entry, price competition, and market segmentation in the prescription drug market. *International Journal of Industrial Organization* 26(4), 930-948.

Saha, A., H. Grabowski, H. Birnbaum, P. Greenberg, and O. Bizan (2006) Generic competition in the US pharmaceutical industry. *International Journal of the Economics of Business* 13(1), 15-38.

Singh, V. and T. Zhu (2008) Pricing and Market Concentration in Oligopoly Markets. *Marketing Science* 27(6), 1020-1035.

Smith, V.L. (1962) An experimental study of competitive market behavior. *The Journal of Political Economy* 70(2), 111-137.

Stargardt, T. (2011) Modelling pharmaceutical price changes in Germany: a function of competition and regulation. *Applied Economics* 43(29), 4515-4526.

Varian, H.R. (1980) A Model of Sales. *American Economic Review* 70(4), 651–659.

Weiss, L.W. (Ed., 1989) *Concentration and Price*. MIT Press, Cambridge, Mass.

Wiggins, S.N. and R. Maness (2004) Price Competition in Pharmaceuticals: The Case of Anti-infectives. *Economic Inquiry* 42(2), 247–263.