# When the baby cries at night.

## Uninformed and hurried buyers in non-competitive markets*

Giacomo Calzolari[§] Andrea Ichino[§] Francesco Manaresi[¶]
Viki Nellas[§]

October 3, 2011

**Abstract**

We study the entry effect of less informed and hurried consumers in retail markets. Theory predicts that firms react by increasing prices to expand surplus extraction, but this effect weakens as market competition increases. High frequency data from Italian pharmacies confirm these predictions. Monthly variation in the number of newborns at the city level generates exogenous changes in the number of less informed and hurried buyers (the parents) who consume a basket of hygiene products demanded by more informed consumers as well. We estimate that the number of newborns has a positive effect on equilibrium price even if marginal costs are decreasing. We exploit exogenous variation in market competition generated by the Italian legislation concerning how many pharmacies should operate in a city as a function of the existing population. Using a RD design we find that an increase in competition, with respect to monopoly, has a significant and negative effect on the capacity of sellers to extract surplus from uninformed buyers. However, three pharmacies in a city of approximately 7500 inhabitants are not sufficient to prevent surplus extraction.

JEL-Code: D43, D83 L13.
Keywords:Informed and uninformed consumers, search, price competition, pharmacies, Regression discontinuity

*We gratefully acknowledge New Line for providing the data that made this research possible. We also thank Elena Folpini, Gabriele Pierani and Gianfranco Pieretti, of New Line, for endless discussions that helped us understanding the complex market studied in this paper. We also thank ANPI (notably, Massimo Brunetti) for the data on "Parafarmacie", ISTAT for the demographic data, and Qualitï¿½ in Farmacia for the data on pricing schedule of wholesalers. We received further insightful comments from Paul Heidhues, Peter Kuhn and Giulio Zanella to whom our gratitude goes.

# 1 Introduction

The role of consumers' information on market prices has been the object of a vast theoretical literature in industrial organization since the seminal works of George Stigler (1961) and Hal Varian (1980). When consumers cannot easily observe and compare prices, the underlying structure of the market can affect the level and the dispersion of prices unexpectedly. On the empirical side, economists have focused so far mainly on estimating the effect of competition and market structure on price dispersion, showing that the number of active firms may indeed affect prices in non standard ways.[1] Much less empirical analysis has instead been devoted to the investigation of the role of consumers' heterogeneity in terms of information at their disposal. Some consumers are more informed than others for both exogenous factors, such as the availability of better search technologies like the Internet, or because of endogenous determinants, such as previous search and consumption experiences. An almost universal conclusion of the theoretical analysis is that more informed consumers force firms to reduce their prices.

In this paper we are precisely interested in measuring the effects of changes in the shares of informed and uninformed consumers, a question that so far has been relatively neglected in the empirical literature despite its relevance. We provide a novel and clean identification strategy to study what happens to the equilibrium prices (and quantities) when exogenous waves of hurried and less informed buyers enter a market. We then estimate how the *composition effect* generated by an increase in the share of these buyers is affected by an exogenous modification in the degree of competition faced by firms.

Our analysis rests on the assumption that immediately after child birth, parents suddenly enter as buyers in the market of the goods that are necessary

[§]University of Bologna, Department of Economics. Giacomo Calzolari is affiliate of CEPR, Andrea Ichino is fellow of CEPR, IZA and CESifo.

[5]Bank of Italy.

[1]See among others Lach (2002), DeCicca, Kenkel, and Li (2010), Martin-Oliver, Salas-Fumas, and Saurina (2008), Gerardi and Shapiro (2009), Lewis (2008), Lach and Moraga-Gonzalez (2009). Fewer papers have recently addressed also the estimation of consumers' search costs, Hortaçsu and Syverson (2004) and Moraga-Gonzalez, Sandor, and Wildenbeest (2009).

to raise their babies, but are relatively less informed than other consumers of those products, including cohorts of parents of older babies. They are also more likely to be under pressure (... when the kid cries!), being less able than more experienced parents in the assessment of the urgency of children's needs and claims.[2]

For each of the 8,092 Italian municipalities (henceforth, cities), we have data on the number of newborn babies at the monthly frequency between 2006 and June 2010. We have also identified a set of hygiene products demanded by parents of small babies as well as by other consumers and we are able to access monthly data on the quantities of these goods sold by a large number of pharmacies in these cities, together with charged prices. Thanks to these unique sets of data, under relatively mild identification assumptions[3], we are able to estimate the elasticity of the equilibrium price and of the equilibrium quantity with respect to a shock in the monthly number of newborns. Consistently with the theoretical predictions, we find that an increase in the number of newborns significantly raises the average price at the city level. There are other possible interpretations for this result (most notably increasing marginal costs) but we are able to show that none of these alternative interpretations is consistent with the other available pieces of evidence. We also find that quantities sold at pharmacies increase as a result of the demand shock, even if prices are higher in the presence of a larger number of newborns.

The insights of our theoretical model invites us to search for exogenous sources of variation in the number of sellers. We find these sources by concentrating the analysis on cities whose maximum population during the last 45 years is in a neighborhood of the 7500 units threshold. Indeed, the Italian law prescribes that cities with a population lower than this threshold should have only one pharmacy, while an additional pharmacy should be opened in cities above the threshold. With respect to current population, there is sub-

[2]In the rest of the paper, we will typically use "less informed consumers" as short for "less informed and hurried".

[3]Controlling for city and time fixed effects, the variation in newborns at the monthly frequency is arguably random.

stantial non-compliance with this rule, partly because of geographic reasons[4], but more importantly because during the post-war period, when population grew above the threshold, pharmacies were opened but later they were not closed if population declined under the threshold. Precisely for this reason, the maximum population size reached historically by cities generates a partially fuzzy assignment mechanism for the current number of pharmacies. We exploit this assignment mechanism within a Regression Discontinuity Design to study how the number of sellers influences the effect of an increase in the share of less informed consumers.

Using this identification strategy we show that cities immediately above the thresholds (in terms of maximum historical population) have, on average, a larger number of pharmacies than cities immediately below. As a consequence, where the number of competing pharmacies is larger for this exogenous reason, the elasticity of equilibrium prices to newborns is significantly smaller, although we show that three pharmacies in a city of approximately 7500 are not sufficient to completely eliminate the price increase. We interpret this finding as evidence that in less competitive environments sellers can exploit to their advantage increases of demand originating from less informed consumers, as the model predicts. Competition, however, limits severely this sellers' advantage.

Although there has been a recent surge of empirical investigations about consumers' information heterogeneity, these studies investigate only *indirectly* the composition effects which emerge when, for some reasons, the fraction of informed and uninformed consumers changes. The analysis of these *composition effects* is instead, precisely, the main contribution of this paper.

The advent of Internet has been seen as one leading factor reducing search costs, increasing the fraction of more informed consumers, and ultimately implying a reduction of prices. This as been documented, for example, by Brown and Goolsbee (2002) illustrating the effect of Internet comparison shopping sites on the prices of life insurance in the 1990s.

Using scanner data, Aguiar and Hurst (2007) have shown that older in-

[4]The presence of remote areas or valleys within the city boundaries is the most common motivation for being allowed to have more pharmacies than what the Law would prescribe.

dividuals, facing a lower opportunity cost, shop more frequently, looking for temporary discounts, and thus pay lower prices than younger ones for exactly the same products. They are able to calculate the implicit opportunity cost of time, showing that it is hump shaped with respect to age, with a peak at mid thirty, much earlier than the peak of the typical wage profile. A plausible interpretation of this observation is that child care becomes a crucial ingredient in the household production function at that age, thus rising the opportunity cost of shopping for given wage. This empirical observation is consistent with our findings but, differently from their paper, we do not take shops' pricing strategies as given and we are rather interested in precisely verifying if and how shops endogenously modify prices when they observe a change in the composition of their customers.

In the mutual funds industry, Hortaçsu and Syverson (2004) recover estimates of the search cost distributions for heterogeneous (uninformed) investors and show that the observed large dispersion of fund participation fees (i.e. the price to join a fund) can be explained by the non-negligeable search costs they are able to estimate. They also document an upward shift of the estimated search costs distribution occurred between 1996 and 2000 and suggest, with indirect evidence, that this observation may be the result of entry of novice investors.[5]

Similarly to these papers, we are interested in measuring the *composition effects* in markets with consumers characterized by different amount of information on prices. However, and differently from these papers, we address this analysis with *a direct measure* of an *exogenous* change in the composition of consumers, by explicitly accounting for the number of inexperienced newborns' parents entering the market for childcare products. We further qualify this composition effect by interacting it with an exogenous source of variation in the market structure, i.e. the number of pharmacies available to parents as implied by the law.

The rest of the paper is organized as follows. Section 2 provides the theoretical background that guides our empirical exercise. Section 3 describes

[5]They state (p. 441) "*We emphasize that our model's implication of such a composition shift is only suggestive–we would need investor-level data to test it definitively*".

the data and justifies the identification strategy. Section 4 presents and discusses the econometric results. The effect of newborns on equilibrium price and quantity is estimated in Section 4.1. Section 4.2 shows instead how competition affects the elasticities estimated in the previous section. Finally, Section 5 concludes.

# 2 Less informed buyers and competition: theoretical insights

In a market for a good there are $S$ shops and consumers who may perceive the same good purchased at different shops as differentiated. At any period $t$ consumers are divided into two groups. A first group is composed of $N_t^I$ "regular" consumers each of whom has an individual demand $q_i^I(p, S)$ for the good purchased at shop $i$ with $p = (p_i, p_{-i})$, where $p_i$ is the prices of shop $i$ and $p_{-i}$ the vector of prices of all other shops. These regular consumers know all the details of the market and will then be identified as the *informed* consumers (hence the apex $I$). The second group is composed of $N_t^U$ individuals who just entered the market and may thus express a different individual demand. To account that they may have limited information, we indicate their demand for the good at shop $i$ with $q_i^U(\mathcal{P}, \mathcal{S})$ where $\mathcal{P}$ and $\mathcal{S}$ respectively indicate with compact notation the information consumers have about prices and shops. Although they may be simply less informed than the informed consumers, for brevity we will indicate them as the *uninformed* consumers (hence the apex $U$).

In our empirical analysis we will consider products that are used for various purposes among which, in particular, childcare. Parents of newborns are typically very pressed and may well have limited information, especially if they are at their first baby and just after delivery. They will be our type $U$ uninformed consumers. All the others consumers (comprising possibly also parents after some months of experience) will be instead our type $I$ informed consumers.[6] The main difference between these two types of consumers that

[6]Ointments for child skin protection are largely used also by sportsmen; shampoos, bath foams, and barrier creams for children are extensively used by adults as well.

we will exploit is that the uninformed have an individual demand for the good purchased at any shop $i$ which is less elastic than that of informed consumers.

At any period $t$, we assume that there is an inflow of uninformed which is IID over time. In general, the number of informed $N_t^I$ at date $t$ may also depend on uninformed inflows of previous periods because some uninformed may learn enough information about the market and or have been able to better organize their purchases (after some time parents may become child-care experts) in which case they exit group $U$ and enter in group $I$. Although for simplicity here we do not explicitly model this transition, we will consider this possibility in our empirical analysis.

We assume that shops cannot price discriminate but in Section 2.2 we consider whether this assumption is plausible in our case and what would happen if it were relaxed. For the time being we also assume that any shop $i$ has constant (and time invariant) marginal cost $c_i$ and will consider other possibilities in the sequel. The profit of shop $i$ at time $t$ finally is:[7]

$$\pi_{it} = (p_i - c_i) \left[ q_i^I(p, S) N_t^I + q_i^U(\mathcal{P}, \mathcal{S}) N_t^U \right]. \tag{1}$$

## 2.1 The composition effect

If all consumers were identical, the optimal pricing condition

$$\left[ q_i^I + (p_i - c_i) \frac{\partial q_i^I}{\partial p_i} \right] N_t^I + \left[ q_i^U + (p_i - c_i) \frac{\partial q_i^U}{\partial p_i} \right] N_t^U = 0 \tag{2}$$

would become

$$\left[ q_i + (p_i - c_i) \frac{\partial q_i}{\partial p_i} \right] (N_t^I + N_t^U) = 0 \tag{3}$$

which clearly implies the following.

**Remark 1 (Identical consumers)** *If the individual demand of all consumers is the same, then the equilibrium price of any shop is independent of the number of consumers* $N_t^I + N_t^U$.

[7] To simplify notation we suppress the time-index except for profit and the numbers of consumers

This is not the case, instead, when demands are different. In particular, and as argued above, we may well expect that for given prices, the uninformed are less responsive to price changes than the informed, i.e. $\eta_i^I \geq \eta_i^U$ where the two are respectively the individual-demand price elasticity of informed and uninformed consumers. To illustrate this point, assume for the moment that there is a single monopolist shop. From the pricing condition we can simply obtain

$$\frac{\partial p_i}{\partial (N_t^U/N_t^I)} = \left(\eta_i^I - \eta_i^U\right) \frac{c_i q_i^I q_i^U}{\left[q_i^I N_t^I(\eta_i^I - 1) + q_i^U N_t^U(\eta_i^U - 1)\right]^2} \geq 0. \tag{4}$$

Intuitively, when the fraction of uninformed consumers increases, the shop finds it optimal to increase its price since the associated (possibly large) demand reduction of the more-elastic informed consumers is more than compensated by the total demand increase of the less-elastic uninformed consumers.

To deepen our understanding of a change in the composition of the population of consumers, we now further specify our environment assuming that uninformed consumers do not react to price changes (i.e. $\eta_i^U = 0$) and the informed perceive the good purchased at different shops as perfect substitutes (so that $q_i^I(p, S) = 0$ if $p_i > p_{-i}$). It is well known since Varian (1980) that in this case there are no equilibria in pure strategies. The reason is that each shop faces a trade-off between extracting the maximal rent from uninformed consumers, thus increasing the price (rent extraction effect), and selling to all the informed which requires to be the shop with the lowest price (business stealing effect). To simplify exposition we assume unitary demand for any consumer, with a value $v$ for the good and zero marginal cost.

Let $G(p)$ be the probability that any shop $j$ sets a price lower than $p$. The profit of shop $i$ associated with a price $p_i$ is then

$$\pi_i = p_i \left[\frac{N_t^U}{S} + (1 - G(p_i))^{S-1} N_t^I\right]. \tag{5}$$

The first term in the squared parenthesis is the demand of uninformed $q_i^U(\mathcal{P}, \mathcal{S})N_t^U$ and shows the fraction of uninformed who randomly enter shop at $i$. The second term is the demand of informed consumers $q_i^I(p, S)N_t^I$ and

$(1 - G(p_i))^{S-1}$ is the probability that price $p_i$ is actually the lowest price in the market and the shops then sells to all informed consumers.

The symmetric mixed strategy equilibrium is (for details see Janssen and Moraga-Gonzalez (2004))

$$G(p) = 1 - \left( \frac{v - p}{Sp} \frac{N_t^U}{N_t^I} \right)^{\frac{1}{S-1}} \quad \text{on} \quad [p_0, v], \text{ with } p_0 \equiv v \frac{N_t^U}{N_t^U + SN_t^I}. \tag{6}$$

and the expected price

$$E[p] = \left( \frac{N_t^U}{N_t^I} \right)^{\frac{1}{S-1}} \int_{p_0}^{v} \frac{v}{S(S-1)p} \left[ \frac{v-p}{Sp} \right]^{\frac{2-S}{S-1}} dp. \tag{7}$$

From this expression one can then derive the following.[8]

**Remark 2 (Composition effect and competition)** *With unitary demand and perfectly substitutable shops, for the informed consumers:*

*(i) the expected price $E[p]$ is increasing in $N_t^U/N_t^I$, the composition effect;*

*(ii) the composition effect is decreasing in $S$.*

When more consumers are uninformed, the surplus extraction generates larger profits which translate into a right shift of the (equilibrium) probability distribution of prices $G$ and a consequent increase of the expected price (point (i) in the remark). This is the composition effect that we want to measure empirically. We also want to test whether this effect is weakened by the presence of more shops around, i.e. whether the expected price increase due to more uninformed consumers is mitigated when $S$ is large (point (ii) in the remark).

It is important to notice that, as intuitively illustrated with equation (4), the composition effect is independent of shops being perceived as perfect substitutes and thus mixing over prices. Similarly, also the result that the composition effect is mitigated by competition can be shown in an environment in which imperfectly substitutable shops set deterministic pricing strategies.[9]

[8] Point (i) in the remark is immediate from (7). Point (ii) follows from the fact that $\frac{\partial E[p]}{\partial (N_t^U/N_t^I)}$ can be written as $\frac{N_t^U}{N_t^I(S-1)}E[p]$ and that the change induced by more shops in the first term of this product is always larger than that on the second term.

[9] A Salop model of this type and with these properties is available upon request.

An increase in the number of newborns is also likely to have an effect on equilibrium quantities sold by pharmacies, an effect which is less interesting for the purpose of this paper, but that still needs to be mentioned. In the simplified theoretical framework described above, where we assume a unitary demand of equal value $v$ for all buyers, if $X$ new customers enter the market and the equilibrium price does not increase above $v$, the equilibrium quantity should trivially increase exactly by $X$ units. The market would instead collapse if the equilibrium price grew above $v$, for example because of a very strong composition effect. The real world is however more complex in at least two dimensions that are relevant for the effect of newborns on quantities. First, parents may buy more than one unit of product, second, they may value differently the same quantity of product. In this more general and realistic context, the increase in the equilibrium price may dampen the increase in the equilibrium quantity because some customers may exit the market or buy fewer units than those they would have demanded at the price prevailing before the shock. As a result of these considerations, theory suggests that an increase in the number of newborns should induce an increase in equilibrium quantities, but, differently than what would be predicted by the unitary demand model, the elasticity may be significantly lower than one. This also because, in our empirical application, we do not observe all the sellers in the market: some of the products that we consider could in fact be bought also in supermarkets. In Section 4 we will report and discuss also estimates of the effect of newborns on equilibrium quantities, but we will be less interested in them since they do not contain information on the main questions addressed in this paper, i.e. whether the presence of uninformed buyers allows sellers to extract more surplus from trade and whether this capacity is limited by competition between sellers.

So far we have studied how competition (measured by $S$) changes the elasticity of prices to the number of uninformed consumers, but we have not considered how it affects price levels directly. Janssen and Moraga-Gonzalez (2004) have shown that, with perfect substitutability, more shops induce a larger expected price (which instead remains unaffected if the uninformed optimally randomize between buying or not buying, an optimal behavior if

$S$ is large). With imperfect substitutability instead more shops may reduce prices, as normally expected. Our data do not allow us to measure the degree of substitutability of the products that we consider, and are therefore not particularly useful to study the potentially ambiguous effect of competition on price and quantity levels.[10] We think instead that our data may provide a more significant value added if used to measure the *composition effect* highlighted in this section and how it changes at different levels of competition between sellers. This will therefore be the main focus of our empirical analysis, but before moving to it we need to consider the possibility of more conventional explanations for the effect of a demand shock on prices.

## 2.2 Alternative interpretations

An increase of uninformed consumers may have a positive effect on the equilibrium price because of at least two alternative explanations that deserve to be mentioned even if in the empirical analysis we will provide evidence suggesting that they are implausible in our application.

### Price discrimination

The results derived above have been obtained under the assumption that pharmacies cannot discriminate between informed and uninformed consumers. However, in the context of the specific empirical application that we consider, it is possible that pharmacist might identify the parents of newborns when they enter their shops. For example, mothers of newborns have probably shopped at the local pharmacy while pregnant before the birthdate and newborns, parents are likely to talk with their pharmacist about the recent change in their family life.

Suppose that pharmacist are indeed able to identify the uninformed customers and charge them with a higher price for products that the informed

[10] Our empirical evidence on this issue, that we do not report here given the different focus of this paper but is available upon request, suggests a close to zero overall effect of competition on the level of prices and quantitities. This result probabily hides the two opposite effects mentioned in the text that may be induced by the coexistence of consumers for whom products in different pharmacies are substitutable with consumers for whom instead they are not.

customers can buy for less. In this case, we would essentially have two separate markets in each of which buyers would be identical, even if they differ across markets. Separately in each market, Remark 1 would hold: pharmacies would charge different prices for the two types of consumers, but these equilibrium prices would not change with the respective number of buyers.

If we could as well identify which transactions involve informed or uninformed consumers, we could test directly for price discrimination based on the information available to buyers, and this would generate a completely different kind of paper. But this is not the case. We observe the prices of each single transaction but we do not know if the buyer is an inexperienced newborn parent or an informed customer. Nevertheless, also in the case of price discrimination, an increase in the number of newborns would raise the *observed average price* of each pharmacy. This because if the number of uninformed consumers increases, the price charged to them would contribute to the average price with a larger weight.

In other words, a model with perfect price discrimination between informed and uninformed consumers would deliver the same empirical prediction of the model described in this paper, but the driving mechanism behind the prediction would be different. If price discrimination is impossible, pharmacies charge all customers with the same price and they raise it if they expect a demand shock originating from uninformed consumers. If price discrimination is possible, pharmacies charge the uninformed with a higher price, but the average price observed by the econometrician would still increase when the share of uninformed becomes larger, even if the two prices faced by the two types of customers do not change.

### Non-constant marginal costs

An increase of uninformed consumers may also have effects on prices due to more conventional explanations related to non-constant marginal costs. It is clear that if marginal costs are decreasing, then an increase of $N_t^U$ trivially implies a reduction of prices, even if the demand of informed is identical to that of uninformed consumers. On the other hand, an increasing marginal cost may imply an effect on prices that is observationally equivalent to our composition effect.

Wholesale contracts clearly affect shops' costs: quantity discounts induce decreasing marginal cost, quantity premia induce instead increasing marginal costs. In the following we will investigate this possibilities and show that for the type of products we are considering in the empirical analysis, shops are actually offered quantity discounts. Hence, the composition effects cannot be explained by the type of wholesale contracts which actually should instead mitigate it.

Alternatively, increasing marginal costs could be the consequence of congestion in the shop both directly (it is proportionally more costly to serve more consumers) and indirectly (having customers queueing in the shop may discourage future visits thus inducing lower future demand). Again we will explore empirically this possibility and show that the change in demand that we observe generated by the increase of uninformed consumers, is too small to imply any significant congestion.

Related to the possibility of congestion, we finally mention the potential effects of capacity constraints. On one hand a shop may be constrained by space so that more (uninformed) consumers increase congestion, a possibility that we will be able to discard as discussed above. On the other hand, the shop may run out of stock. With this respect, we will explain that wholesalers promptly supply pharmacies so that this possibility is unlikely. Moreover, the effect of limited product availability on prices is not obvious. Indeed, in a model with perfect substitutable shops, as the one we have illustrated above, Lester (2010) has shown that an increase of uninformed consumer may actually decrease prices when shops are capacity constrained. To see this, consider the converse effect of an increase of informed consumers. The limited availability of products makes the shops posting the lowest prices less interesting for these more consumers since they risk not being served. This competition (for being served) among informed consumer, in some cases, relaxes competition among shops and induces an increase of prices. Conversely, an increase of uninformed consumers reduces the competition among themselves and shops may have to adapt reducing their prices, contrary to the composition effect illustrated above.

## 3 The data and the empirical strategy

We use information on a large sample of Italian pharmacies, collected by New Line, a company producing software for pharmacies. With the consent of its clients, we were given access to the details of every item sold by each pharmacy in the New Line database for the period from January 2007 to July 2010. The dataset originates from each single sale receipt. During the period under study, New Line collected data from 3,331 Italian pharmacies, corresponding to 18.6% of the universe of pharmacies in Italy. For 60% of them, we have complete information for the entire period; for 28.7% we have information starting from January 2009; and for the remaining 11.26% data is available only for the period January 2007-December 2008. The pharmacies in the New Line database are located in almost all the Italian regions (with the exception of Basilicata), but their concentration is higher in the North since the company is located near Milan.[11]

Our goal is to use this dataset to test the theoretical predictions of Section 2 concerning how, in a market, prices and quantities are affected by a demand shock deriving from a change in the number of less informed consumers. We argue that a measure of this kind of shock for a subset of products sold by these pharmacies, is represented by changes at the monthly frequency in the number of newborns in the neighborhood where a pharmacy is located. Monthly data on newborns are obtained at the city level from the National Statistical Office (ISTAT). The left panel of Figure 1 plots the temporal evolution of the number of newborns in the cities where the pharmacies of the New Line sample operate. There is a significant seasonality in newborns: the most relevant peaks are typically in the summer, while the lowest levels are more frequent in the winter. The right panel of the figure plots the residuals of a regression of (log) newborns on city fixed effects. These residuals show a substantial within-city and over time variability in the number of newborns.

Ideally we would like to measure the monthly number of newborns in some neighborhood of each pharmacy, but we can only measure it at the level of

[11]Specifically 19% of these pharmacies are in the north east of Italy, 45% in the north west, 9% in the center, 16% in the south and 11% in the islands.

a city. Therefore in the empirical analysis we aggregate all the pharmacies of the New Line data set in each municipality and consider as a unit of observation the average price and the average quantity of these pharmacies in each city. Note that unfortunately we do not observe the quantity and the price of the pharmacies that, within each city, are not in the New Line sample. This drawback of our dataset is in principle problematic but we will report results restricted to cities in which we observe all the existing pharmacies (i.e. cities in which New Line has a full market coverage), to show that our tests of the theoretical predictions remain unaffected.

We select child hygiene products as the ones for which changes in newborns may be considered as a proxy of exogenous demand shocks originating from variations in the composition of informed and uninformed consumers in the market. With respect to these products the parents of newborns, like type $U$ consumers of the theoretical models described in Section 2, are less informed and more pressed than other buyers in the same market. For this reason we focus on a set of 2925 hygiene products that are used for children immediately *after* birth and then extensively during the first years of their life. This set includes items like: bath foams and shampoos for babies; cleansers for babies; cold and barrier creams and oils for babies; baby wipes; talcum and other after-bath products for babies. Table 1 describes a sample of items in this basket: the upper panel shows the five products sold in largest quantity during the period 2007-2010, while the lower panel shows the ones that featured the highest unit price over the same period. For each item, we have the quantity sold by each pharmacy in each month and the price charged.[12]

Moreover, these products are consumed also by other customers, like for example bikers and runners, to prevent skin rash. While these customers can be seen as experienced buyers of these products, newborns' parents (unless they are themselves bikers and runners) are new in this market and informa-

[12]For items that have not been sold for an entire month, the price imputed is the price reported by the pharmacy in the New Line software database. Instead when the sold quantity is positive, the monthly price is the weighted average of the (possibly) different prices actually charged over the month, with weights equal to the number of items sold at each price level.

tion about prices is available to them through costly trips to pharmacies.[13] Parents of newborns are also likely to be under pressure when buying for their babies, because they lack experience in the evaluation of the urgency of their offsprings' needs. Of course parents may learn relevant information after some months of purchases and may be able to smooth out their organization of parental activities, thus becoming less pressed. So, after a while, they are likely to join the stock of pre-existing consumers that are reasonably more informed and less hurried than parents of just newborns.

We will therefore consider newborns' parents as uninformed or less informed than experienced parents and other customers like sports-persons although they all buy the same set of products. Note that our theoretical results require the existence of both informed and uninformed consumers, and focuses precisely on the effect of changes in their proportions. Other items consumed by newborns, like age-specific diapers, do not have alternative demand sources and are therefore less suitable for the purposes of this study. For these products, the fraction of informed buyers would be severely restricted (possibly to zero) if parents are slow in learning about the market and by the fact that, when children grow, parents leave the market perhaps precisely when they have acquired all the relevant information.

To show that the set of hygiene products on which we focus is indeed demanded by both types of informed and uninformed consumers we regress the total number of units sold in each city by the pharmacies under study, on the number of newborns and on time and city fixed effects.[14] This regression allows us to decompose the variability of the number of sold boxes in the part attributable to newborns (plus time and city effects to be conservative) and the residual part that can be attributed to other customers. This decomposition indicates that 88% of the variability in the number of boxes sold from

[13] In principle, consumers may use the internet to learn the expected price. However, even the acquisition of information from the internet is far from being costless. Moreover, in the Italian context, very few pharmacies have a website.

[14] Later we will construct Laspeyeres quantity and price indexes for the basket of goods in which we are interested, but for the purpose of assessing whether these goods are demanded by other customers beyond newborns, we can focus now on the actual total number of "boxes" sold in a month for each item

month to month is generated by consumers other than newborns.[15]

We model the number of experienced buyers of child hygiene products as a specific characteristic of each city and month of the year, that can be captured by city and time fixed effects. Our identification strategy therefore hinges on the assumption that *at the monthly frequency and controlling for city and time fixed effects* a change in the number of newborns is a random event not correlated with other city characteristics.

For the basket of products that is relevant for our study, we constructed Laspeyres indexes of prices and quantities. Denoting with $h \in 1, \ldots, H$ each product in the basket, the price and quantity indexes (hereafter, price and quantity) for pharmacy $i$ in month $t$ are defined by:

$$p_{it} = \frac{\sum_h p_{iht}\bar{q}_h}{\sum_h \bar{p}_h\bar{q}_h} \tag{8}$$

$$q_{it} = \frac{\sum_h \bar{p}_h q_{iht}}{\sum_h \bar{p}_h\bar{q}_h} \tag{9}$$

where $\bar{q}_h$ and $\bar{p}_h$ are the quantity and the price for product $h$, respectively sold and charged on average by all pharmacies in all months. In other words, $p_{it}$ is the weighted average price charged by pharmacy $i$ in month $t$ for the entire basket, where the weights are based on the quantities of each item sold on average in the entire market over all months. So this price index is independent of the quantities sold by pharmacy $i$ and changes over time (and with respect to any pharmacy $j$) if and only if the price of at least one item changes in pharmacy $i$ (or $j$). The quantity index $q_{it}$ is instead the weighted average quantity sold by pharmacy $i$ in month $t$ for the entire basket, where the weights are based on the prices of each item charged on average in the entire market over all months. So this quantity index is independent of the prices charged by pharmacy $i$ and changes over time (and with respect to any pharmacy $j$) only if the quantity of at least one item changes in pharmacy $i$ (or $j$).

The temporal evolutions of these two indexes for the pharmacies in the New Line dataset are plotted in the left panels of Figure 2. Quantities are

[15] Using the quantity index described below, the decomposition shows that around 85% of the variability can be attribued to customers other than newborns.

characterized by seasonality (with the most relevant peaks during the summer) and by a weak downward trend. Conversely, prices are characterized by a more robust upward trend. Our empirical strategy exploits within pharmacy variability of both these variables. The right panels plot the residuals of a regression of (log) quantity and (log) price on city fixed effects. These residuals show that both the quantity and the price change substantially over time at the intra-pharmacy level. Although the intra-pharmacy variability of quantity index is larger, there is substantial variability also in the price index.

The theoretical model has also predictions concerning the effects of changes in the degree of competition between sellers in the market under study. To study these effects we need an exogenous source of variation in the number of pharmacies. Here, we exploit the rules that regulate the Italian pharmacy market. In Italy, entry in and exit from this market are regulated by the Law 475/1968. This Law establishes (as in many other countries) the so-called "demographic criterion" to define the number of pharmacies authorized to operate in each city. Specifically, the law generates a set of population thresholds at which the number of existing pharmacies that should operate in a city changes discontinuously. Leaving the details to Section 4.2, for our purposes this law generates a Regression Discontinuity design that allows for the possibility to estimate the causal effect of a change in the number of competing pharmacies at each threshold.

Descriptive statistics for the variables used in the econometric analyses are displayed in Table 2.

# 4 Effects of changes of uninformed consumers in a market

We exploit the data described in the previous section to estimate, with different empirical models, the parameters of the following linear regression, which allows us to test the predictions of the theory:

$$Y_{ct} = \alpha + \beta S_c + \delta N^U_{ct} + \phi_c + \mu_t + \varepsilon_{ct} \qquad (10)$$

where $c \in \{1, \ldots, C\}$ denotes cities and $t \in \{1, \ldots, T\}$ denotes months. $Y_{ct} = \{q_{ct}, p_{ct}\}$ are the two dependent variables in which we are interested: the (log) quantity index $q_{ct}$ and the (log) price index $p_{ct}$ respectively sold and charged by the $S_c$ pharmacies in the city. $N_{ct}^U$ is the (log) number of newborns in city $c$ at time $t$. $\phi_c$ and $\mu_t$ are city and month fixed-effects which capture relevant characteristics of city markets, like the distance between pharmacies, or of calendar months, like seasonal effects.

As explained in Section 3, our identifying assumption requires also that these fixed effects capture the (log) number of informed consumers $N_{ct}^I$. Therefore, by measuring the effect of $N_{ct}^U$ controlling for $\phi_c$ and $\mu_t$ we are effectively measuring the composition effect of the ratio $\frac{N_{ct}^U}{N_{ct}^I}$ described in Remark 2. $\varepsilon_{ct}$ is an error term, which is allowed to display heteroskedasticity and serial correlation at the city level. Note that serial correlation may originate from the fact that parents of newborns, who are uninformed at time $t$, may later become informed. Inasmuch as this process is not fully captured by city and time fixed effects, it generates serial correlation in the error term. This, however, is not a threat for our identification strategy, since it does not affect the randomness of the number of newborns $N_{ct}^U$.

The parameters $\{\alpha, \beta, \delta, \lambda, \phi_c, \mu_t\}$ should have a $Y$ subscript to indicate that they differ according to whether the dependent variable is the price or the quantity but, to simplify notation, we omit the subscript unless it is necessary for the exposition.

## 4.1 Uninformed consumers, price and quantity

We first exploit within-city variation of newborns for the estimation of the following standard fixed-effect model:

$$Y_{ct} = \alpha + \delta N_{ct}^U + h_c + \mu_t + \varepsilon_{ct} \tag{11}$$

where, given equation 10, $h_c = \beta S_c + \phi_c$. Note that if our identifying assumptions are valid, $N_{ct}^U$ is randomly assigned conditioning on city and time fixed effects. Thus, the parameter $\delta$, that captures the *composition effect* derived in Remark 2 of Section 2.1, can be interpreted as a causal parameter and its OLS estimate is consistent.

We are mainly interested in the effect of newborns on the equilibrium price and we therefore start by focusing on Table 3, which reports estimates of equation 11 when the dependent variable is the price ($Y_{ct} = p_{ct}$). Let's first assume that marginal costs are constant or decreasing. Under this assumption, if parents of newborns are less informed than other consumers we expect a positive estimate for $\delta$. If instead, newborns' parents are as informed as other consumers, the effect on price should be nil (or negative in case of decreasing marginal costs). Thus, a positive parameter $\delta$ should signal the presence of uninformed consumers and measure the associated *composition effect* discussed in our theoretical analysis (see Remark 2).

All estimates of $\delta$ in Table 3 are positive and highly significant. To help the interpretation of their size, we have again standardized coefficients and standard errors by the standard deviation of the dependent variable. The first column in the top panel, for example, indicates that a 100% increase in newborns (which corresponds to two third of a standard deviation of this variable) causes an increase of 1.6% of a standard deviation of the (log) price. It is particularly important to consider that the estimates of the top panel may be confounded by the fact that we do not observe the pharmacies not covered by the New Line dataset. However, the estimates in the bottom panel show that the effects are essentially unchanged when we consider only the cities in which New Line has full coverage.

The second and third columns of Table 3 measure the size of the demand shock by summing the number of newborns over current and previous months. The size of the estimated coefficients increases significantly over the three columns. Although we do not know exactly if and when parents of newborn enter the pool of informed consumers, this finding indicates that they are still relatively uninformed after two months from delivery. Indeed estimates that we do not report to save on space suggest that approximately one year is necessary for the positive effect of newborns on prices to fade away.

However, before concluding that the estimates of Table 3 are positive because the parents of newborns are uninformed consumers, it is necessary to dismiss the possibility that this finding is driven instead by alternative explanations discussed in Section 2.2.

The first alternative explanation argues that if price discrimination is possible, pharmacies may charge newborns' parents with a higher price, but the average price observed by the econometrician would still increase when the share of uninformed customers becomes larger, even if the two prices faced by each type of buyers do not change. *TO BE COMPLETED with evidence showing that there is no variability in unit prices from till receipts issued by pharmacies within the same week for the same product.*

The second alternative explanation argues that a positive estimate for $\delta$ could be the consequence of increasing marginal costs even if all customers were equally and fully informed. We present three pieces of evidence that exclude this possibility.

First, it may be marginally more costly for a pharmacy to sell larger quantities if wholesalers charge higher prices for larger orders, in which case an increase in newborns may obviously translate into higher prices for consumers. The evidence, however, points in the opposite direction. From Qualitï¿½ in Farmacia, a software house specialized in managing information systems for pharmacies, we obtained the pricing schedule adopted by an Italian wholesaler in the period 2010-2011. This wholesaler sells 727 child hygiene products (24.8% of our basket). For none of them the wholesale price schedule as a function of quantity is increasing. Figure 3 shows the average price charged by the wholesaler for different quantities of the four main categories of child hygiene products (bath foam and shampoo, after bath products, barrier creams, wipes). What emerges is a clear decreasing pattern supporting the theoretical assumption of non-increasing marginal costs. If anything, the presence of these quantity discounts should have reduced the composition effect that we document in Table 3.

Second, we can also exclude that problems of congestion, caused by the increase of newborns, indirectly raise pharmacies' marginal costs. If the increase of newborns has the potential to generate a queue of hurried parents in the pharmacy and if expected revenues from them are lower than the ones that can be expected from other consumers (of any product), the pharmacist may react by increasing prices on child hygiene products in order to reduce the undesired queue of newborns' parent. This possibility is extremely unlikely in our environment because child hygiene products represent on average a tiny percentage of the monthly transactions of a pharmacy: evidence from till receipts emitted by the pharmacies in our sample show that those containing at least one child hygiene product are on average around 130, while the average monthly number of till receipts emitted is around 5,800. A 100% increase in monthly sales of child hygiene products would thus yield an increase of around 2.2% of total demand, which would not be enough to generate substantial queueing in the pharmacy.

Third, pharmacies in our dataset can receive supplies from wholesalers more than once a day so that there is basically no effective shortage of inventories that might be binding for more than few hours.[16]

We can therefore conclude with sufficient confidence, that the positive effect of the number of newborns on prices estimated in Table 3 is evidence that the parents of newborns are relatively less informed than other consumers of child hygiene products and that pharmacies are capable to extract surplus exploiting their lack of information. We now can explore whether this capacity to extract surplus from the uninformed consumers changes according to different levels of competition faced by pharmacies in this market.

Before doing so a word should be said concerning equilibrium quantities. As already mentioned in Section 2, we expect that an increase in the number of newborns raises the equilibrium quantity sold by pharmacies, although this results does not contain specific information concerning the capacity of firms to extract surplus from uninformed consumers, which is the main question addressed by this paper. It is, nevertheless reassuring to see that this expectation is confirmed by our evidence. This is shown in Table 4 which reports estimates of equation 11 when the dependent variable is the (log) equilibrium quantity ($Y_{ct} = q_{ct}$). The estimated coefficients and the standard errors have been again standardized by the standard deviation of the dependent variable. In the first column of the top panel, all cities are used

[16] Lester (2011) has shown that if firms have binding capacity constraints, then an increase of the share of uninformed consumers may actually lead to a decrease of the average price. We do not observe this possibility also because our pharmacies almost never face these constraints, as explained in the text.

and the demand shock is measured by the (log) number of newborns in month $t$. The standardized coefficient indicates that a 100% increase in newborns (which corresponds to two third of a standard deviation of this variable) causes an increase of 3% of a standard deviation of the (log) quantity. The bottom panel of the table restricts the analysis to cities in which New Line has full coverage of the existing pharmacies. The first column of the bottom panel confirms the estimates of the top panel. The remaining columns of the table measure the demand shock using the total number of newborns in $t$ and $t-1$ (column 2) or in $t$, $t-1$ and $t-2$ (column 3). As expected, we see that when we measure a larger shock, the coefficient associated with newborns increases with respect to the first column.

## 4.2 Can competition limit the capacity of pharmacies to extract surplus from consumers?

As previously discussed, entry and exit in the pharmacy market is regulated by law 475/1968 which establishes how many pharmacies should operate in a city as a function of the existing population. Below 7500 inhabitants there should be only one pharmacy. From 7500 to 12500 there should be two pharmacies. Above this threshold a new pharmacy should be added every 4000 inhabitants. Compliance with this theoretical rule is however imperfect for at least two reasons. First, cities that are composed by differentiated geographical areas between which communication is difficult (e.g. because of mountain ridges or rivers), are allowed to have more pharmacies than what would be implied by a strict application of the law. Second, the evidence suggests that it is easier to open a pharmacy than to close one, probably because of the difficulty of "deciding" who should exit the market when pharmacies are too many (the law being silent on this issue). In some rare occasions market forces induce the bankruptcy of the weakest pharmacy in a city in which demand is no longer sufficient to sustain positive profits for all the existing ones. But otherwise, the evidence suggests that, given the rents that a pharmacy probably grants to its owners in a highly regulated market like the Italian one, new sellers enter immediately whenever possible, but very few later exit if and when the city population declines.

This historical asymmetry in the likelihood that pharmacies are opened or closed, generates, nevertheless, an exogenous source of variation in the current number of pharmacies based not on the current population but on the population peak reached since 1971.[17] Consider the threshold of 7,500 inhabitants at which the number of existing pharmacies should theoretically increase from 1 to 2. The left panel of Figure 4 shows local polynomial smoothing (LPS) regression estimates of the number of pharmacies as a function of the current city population, together with the 95% confidence intervals. No discontinuity in the number of competitors can be appreciated.[18] The right panel of the same Figure shows instead analogous LPS regression estimates of the number of pharmacies against the maximum level reached by the city population since 1971. Here the discontinuity is large and statistically significant.

As it can be seen, there are cities in which the population never went above 7500 units since 1971 and nevertheless have more than one pharmacy for the already mentioned historical or geographic reasons. Similarly, on the right of the threshold, the average number of pharmacies is larger than two, in contrast with what the law would require. But even in the presence of this generalized "upward non-compliance", a significant discontinuity of approximately half a pharmacy emerges at the threshold.

Table 5 reports estimates of the equation

$$S_c = \omega + \varphi K_c + g(|Pop_c - \kappa|) + \zeta_c \tag{12}$$

where $c$ denotes a city, $S_c$ is the number of pharmacies in a city; $Pop_c$ is the maximum historical population in a city; $K_c = 1(Pop_c \geq \kappa)$ is a dummy taking value 1 for cities on the right hand side of the $k-$threshold. The parameter $\varphi$ in this equation measures the discontinuity at the threshold. Independently of the specification of the polynomial $g(.)$, the first panel of

[17] The 1971 Census is the first reliable population measure at the city level after the date of enactment of Law 475/1968.

[18] The ratio between the number of pharmacies and population is a city is re-evaluated every two years. For this reason, we define 'current population' as the one measured two years before the city enters in the New Line panel.

Table 5 confirms the visual impression of Figure 4, suggesting an even larger discontinuity at the threshold (we report local linear and polynomial regressions for different population windows around the threshold, from $\pm 1,500$ to $\pm 4,000$ inhabitants). The same happens in Panel B of the same table in which the analysis is restricted to the cities with a 100% New Line coverage and in Panel C where observable controls are included in the specification to improve efficiency.[19]

At higher thresholds, involving larger cities with more pharmacies, even the compliance with the rule based on the population peak is more blurred, so that we are forced to use only the first threshold of 7500 units for our analysis. This however is enough to test in a clean way the predictions of our theoretical model concerning the effects of competition in this market.[20]

Having shown that the number of pharmacies effectively changes discontinuously at this threshold, we now have to provide evidence supporting the identifying assumption for a RD design, requiring that nothing else relevant changes discontinuously at the same threshold. Figure 5 shows the local polynomial smoothing (LPS) regressions of four observable pre-treatment factors on the maximum historical population since 1971: the average monthly number of newborns, a dummy taking value 1 if the city is in a urban area, a dummy taking value 1 if the city is in Northern Italy and per capita disposable income (measured in 2008) at the city level. For none of these variables a quantitatively or statistically significant discontinuity should be observed at the threshold and this is precisely the evidence emerging from the figure.[21] Nonetheless, in some empirical specifications we include these variables as

[19] The included controls are: the average monthly number of newborns, a dummy taking value 1 if the city is in a urban area, a dummy taking value 1 if the city is in Northern Italy, and per capita disposable income at the city level.

[20] It would instead not be enough for a complete policy design since we only have insights concerning changes from approximately 1 to approximately 2 in relatively small cities.

[21] We have also tested the existence of discontinuity at the threshold for these variables using local linear and polynomial regressions for different windows around the threshold (as suggested by Imbens and Lemieux (2008)). Results uniformly fail to identify any significant discontinuity. Additional covariates for which the unconfoundedness hypothesis has been tested include the population growth rate since 1971, per capita consumption, and per capita expenditure on pharmaceuticals, all at the city level. Results are available upon request.

regressors to increase efficiency.

Since the conditions for a RD design are satisfied we can now describe what we learn from estimates based on it, concerning the effect of competition on the capacity of firms to extract surplus from uninformed consumers. To this end, we test whether the *composition effect* on prices and quantities of an increase in the number of uninformed consumers is different at different levels of competition between pharmacies. In other words, we are now interested in the interaction between $N_{ct}^U$ and $S_c$. The theoretical model in fact suggests that when competition is tougher, an inflow of uninformed consumers should have a less positive effect on prices than when competition is weaker. To gather evidence on this prediction we proceed in three steps.

1. We regress prices, quantities, and the number of newborns on city and time fixed effects to obtain their values demeaned with respect to space and time. That is, we estimate the following model for the set of dependent variables $H_{ct} = \{N_{ct}^U, Y_{ct}\} = \{N_{ct}^U, q_{ct}, p_{ct}\}$:

$$H_{ct} = \phi_c + \mu_t + \eta_{ct}, \tag{13}$$

   and we retrieve the residuals (i.e. the demeaned values) for each dependent variable that we denote as: $\tilde{H}_{ct}$.

2. *Separately for each city*, we regress demeaned prices and quantities on demeaned newborns obtained from step 1. That is, we run a total of $C$ regressions like:

$$\tilde{Y}_{ct} = \alpha_c + \delta_c^Y \tilde{N}_{ct}^U + \tilde{\varepsilon}_{ct} \tag{14}$$

   Each regression yields a city specific estimate $\hat{\delta}_c^Y$ of the elasticity of the dependent variable $Y$ (price $p_{ct}$ or quantity $q_{ct}$) to a change in newborns.

3. We then use the RD design, to test whether these city-specific elasticities differ on the two sides of the population threshold $\kappa = 7500$, keeping in mind that on the left of this threshold the number of pharmacies is significantly lower than on the right. Therefore we estimate the

following equation restricting the analysis to cities in a neighborhood of the threshold:

$$\hat{\delta}_c^Y = \omega + \gamma^Y K_c + g(|Pop_c - \kappa|) + \eta_c \tag{15}$$

where $K_c = 1(Pop \geq \kappa)$ is a dummy equal to 1 on the right of the threshold.

A negative estimate for $\gamma^p$ in the RD regression 15 (i.e., when $Y_c = p_c$) would confirm the prediction of the model (point (ii) of Remark 2) according to which an increase in the number of competitors should reduce the *composition effect* i.e. the capacity of pharmacies to extract surplus from less informed consumers and therefore that it should reduce the positive effect of newborns on prices.

As far as quantities are concerned, note that the growth rate of newborns is identical on the two sides of the threshold (as it should be for the validity of the RD design). In this context the elasticity of the equilibrium quantity to the number of newborns at the pharmacy level could be higher on the right of the threshold, only if the price level were lower. But we find a similar price level on the two sides of the threshold (see footnote 10), because, as we explained in Section 2, the number of competitors has an ambiguous overall effect on the price level, if the degree of product substitutability is perceived in different ways by different consumers. We therefore do not expect $\gamma^q > 0$ to differ significantly from zero, because the elasticity of quantity to newborns should be the same on the two sides of the threshold. As already mentioned, however, these results are irrelevant for the main questions addressed in this paper.

Figures 6 plots the LPS regression estimates of the elasticity of price and quantity with respect to the number of newborns on the two sides of the threshold. As far as prices are concerned (left panel), in line with the theoretical model we see that the elasticity declines down to zero when competition increases, suggesting that pharmacies facing larger competition are less able to extract surplus from uninformed consumers. Similarly expected is also the absence of differences concerning quantities at the threshold (right panel).

Table 6 reports estimates of the coefficient $\gamma^Y$ in equation 15, confirming the visual impression of Figure 6. In Column 1 we find no significant effect of the number of competitors on the elasticity of quantity with respect to newborns, independently of the specification of the polynomial in the distance from the threshold. Column 2 reports similar estimates for effect of competition on the elasticity of price to newborns. This elasticity is significantly lower on the right of the threshold than on the left. This finding is consistent with the theoretical model that predicts that an increase in the number of competitors reduces the *composition effect*, i.e. the capacity of firms to extract surplus from uninformed buyers.

Panel A in Table 7 shows that these results are robust with respect to the failure of observing all the pharmacies in a city. Estimates are in fact qualitatively and quantitatively similar when we restrict the analysis to cities with a 100% New Line coverage.[22] Panel B of Table 7 reports, instead, results obtained controlling for observables to increase efficiency.[23] As expected, point estimates remain practically unchanged.

Finally, our setting allow us to use also Instrumental Variable methods for the purpose of measuring the effect of competition on the pharmacies' ability to extract surplus from uninformed consumers. We estimate the equation

$$\hat{\delta}_c^Y = \chi + \psi^Y s_c + g(|Pop_c - \kappa|) + \nu_c \tag{16}$$

using the threshold dummy $K_c = 1(Pop \geq \kappa)$ as an instrument for the number of competitors perceived by a pharmacy, which is given by $s_c = S_c - 1$. Therefore, the measure of competition used in this regression is the number $S_c$ of pharmacies in the market minus 1. In this way, *near the threshold* where $Pop_c \approx \kappa$, the constant captures what happens when in a city there is only one monopolist pharmacy. Table 8 reports estimates of the effect $\psi^Y$ of the real number of competitors (*Treatment Effect*), as opposed to estimates

[22]Note, however, that focusing on this subsample reduces the number of observations, so that we have to enlarge the window up to 4,000 inhabitants and rely on lower order polynomials to obtain precise estimates.

[23]These controls are, again, the average monthly number of newborns, a dummy taking value 1 if the city is in a urban area, a dummy taking value 1 if the city is in Northern Italy and per capita disposable income (measured in 2008) at the city level.

of the effect of their theoretical number (*Intention to Treat Effect*), which were reported in Tables 6 and 7. Here, controls have been added to increase efficiency. Panel B, C and D enlarge the window around the threshold up to ± 4,000 inhabitants and this increases markedly the precision of the estimates: both the F-statistics of the excluded instrument and the significance of the estimated coefficients in the second stage increase. This set of estimates shows that, near the threshold, when there is just one pharmacy in the market (i.e. a pharmacy with $s_c = 0$ competitors), the elasticity of equilibrium price to the number of newborn is positive and significant. A second competitor roughly halves this elasticity (from 0.10 to 0.05, according to the estimate in Panel D) but the elasticity remains positive and significant. According to these estimates, 2 competitors are not sufficient to prevent the extraction of surplus from uninformed consumers by pharmacies when the city population is around 7500 units.

# 5 Conclusions

In this paper we have provided new evidence on the role of consumers' information in the retail sector, and on its interplay with competition among sellers. Theory predicts that an inflow of less informed and more hurried consumers should have a positive effect on the average price charged by sellers. This composition effect (generated by sellers being able to extract larger rents from uninformed consumers through higher prices) should also decline as the number of competitors increases.

We gather data for a large sample of Italian pharmacies and estimate the effect of a positive shock in the number of newborns on the average price at the city level, for a basket of child hygiene products. We argue that parents of newborns are less informed than other consumers of the same set of products and that an increase of newborns is thus a source of exogenous variation in the number of uninformed consumers. Consistently with the theoretical prediction, an increase in newborns has a positive effect on price. We provide evidence allowing us to exclude that this positive effect might be driven by increasing marginal costs and/or by congestion at the pharmacy level.

To study the effect of competition on the elasticity of price to the presence of uninformed consumers, we exploit a legislative feature of the Italian pharmacy market: the law imposing that municipalities under 7,500 inhabitants should have a single pharmacy, while those right above this threshold should have two. Despite the presence of non-compliance with this law, we are able to exploit it within a Regression Discontinuity Design and show that the elasticity of prices to the number of newborns declines to zero in cities where the number of pharmacies is higher. These results confirm the theoretical prediction that competition reduces the capacity of firms to extract surplus from uninformed buyers. However, three pharmacies in a city of approximately 7500 are not sufficient to prevent surplus extraction.

# References

Aguiar, M. and E. Hurst. 2007. "Life-cycle prices and production." *The American Economic Review* 97 (5):1533–1559.

Brown, J. and A. Goolsbee. 2002. "Does the internet make markets more competitive? Evidence from life insurance industry." *Journal of Political Economy* CX:481–507.

DeCicca, P., D. Kenkel, and F. Li. 2010. "Who pays for cigarette taxes? The impact of consumer price search."

Gerardi, K.S. and A.H. Shapiro. 2009. "Does competition reduce price dispersion? New evidence from the airline industry." *Journal of Political Economy* 117 (1):1–37.

Hortaçsu, A. and C. Syverson. 2004. "Product differentiation, search costs, and competition in the mutual fund industry: a case study of S&P 500 Index funds." *Quartely Journal of Economics* 119 (2):403–456.

Imbens, G. and T. Lemieux. 2008. "Regression discontinuity design: a guide to practice." *Journal of Econometrics* 142 (2):615–635.

Janssen, M. and J.L. Moraga-Gonzalez. 2004. "Strategic pricing, consumer search and the number of firms." *Review of Economic Studies* 71:1089–1118.

Lach, S. 2002. "Existence and persistence of price dispersion: an empirical analysis." *The Review of Economics and Statistics* 84 (3):433–444.

Lach, S. and J.L. Moraga-Gonzalez. 2009. "Asymmetric price effects of competition." *CEPR Discussion Paper* 7319.

Lewis, M. 2008. "Price dispersion and competition with differentiated sellers." *The Journal of Industrial Economics* 56 (3):654–678.

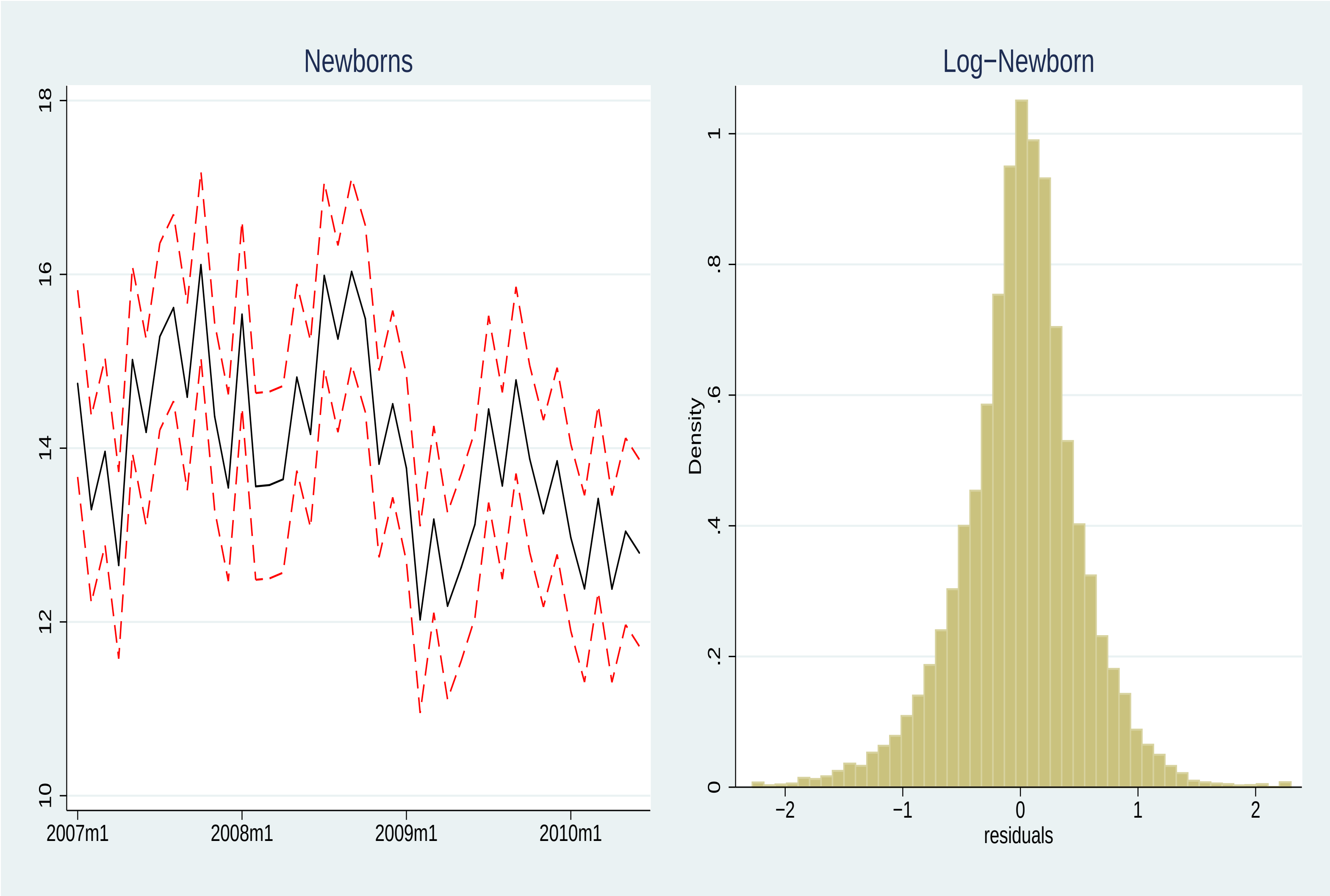
Martin-Oliver, A., V. Salas-Fumas, and J. Saurina. 2008. "Search cost and price dispersion in vertically related markets: the case of bank loans and deposits." *Review of Industrial Organization* 33:297–323.

Moraga-Gonzalez, J., Z. Sandor, and M. Wildenbeest. 2009. "Consumer search and prices in the automobile market."

Stigler, G. 1961. "The economics of information." *Journal of Political Economy* 69:213–225.

Varian, H.R. 1980. "A model of sales." *The American Economic Review* 70 (4):651–659.

Figure 1: Temporal evolution and within city variability of the number of newborns




*Notes:* Temporal evolution of the average number newborns per city (left panel), and histograms of the residuals of a regression of log-newborns on city fixed effects (right panel). Dashed lines delimit the 95% confidence interval.

Table 1: Top items in the basket of hygiene products

| Name | Description | Price | Quantity |
|---|---|---|---|
| *Top-5 by Sold Quantity* | | | |
| Salviette Assorbello | Hygienic Towels | 2.04 | 39.94 |
| GP Baby Pasta all'Ossido di Zinco | Zinc-Oxyde Paste | 4.91 | 23.3 |
| Bluedermin Pasta BB 100ml | Diaper Change Ointment | 5.83 | 17.21 |
| Trudi Baby Care Salviettine | Hygienic Towels | 2.07 | 16.53 |
| GP Baby Detergente | Cleansing Cream | 5.02 | 15.3 |
| *Top-5 by Price* | | | |
| Soin de Fee 24-Hour Baby Cream 50ml | Barrier Cream | 45 | 0.21 |
| Vidermina 3 Soluzione 1000ml | Cleansing Cream | 40.32 | 0.01 |
| Buba Shampoo e Bagno | Shampoo and Bath Foam | 37.61 | 0.04 |
| Unilen Gel 15ml | Barrier Cream | 36.06 | 0.11 |
| Protezione Solare Bambini Vichy | Suntan Cream | 30.9 | 0.02 |

*Notes:* Our calculations based on the New Line database

Table 2: Descriptive statistics of the variables used in the econometric anlysis

| | Mean | Standard deviation | Minimum | Maximum | Number of Observations |
|---|---|---|---|---|---|
| Price Index | 1 | 0.03 | 0.8 | 1.25 | 56798 |
| Quantity Index | 1 | 0.49 | 0.004 | 7.3 | 56798 |
| Sold boxes of hygiene products | 68 | 56 | 0 | 808 | 56798 |
| Number of newborns in t | 14 | 22 | 0 | 208 | 56798 |
| Number of newborns in t and t-1 | 28 | 44 | 1 | 598 | 56798 |
| Number of newborns in t, t-1 and t-2 | 42 | 66 | 1 | 831 | 56798 |
| Number of pharmacies per city | 5 | 8 | 1 | 125 | 56798 |

Notes: Price and quantity information concerns 2925 hygiene products sold by the 3331 pharmacies in the New Line dataset. Information on newborns refers to the 1665 cities in which the pharmacies of the New Line dataset operate. One observation is a city in a month.

Figure 2: Temporal evolution and within city variability of the quantity and price indexes


*Notes:* Temporal evolution of the average quantity and price indexes of hygiene products (left panels), and histograms of the residuals of a regression of the (log) quantity and (log) price indexes on city fixed effects (right panels). Dashed lines delimit the 95% confidence interval.

Table 3: Effect of the monthly number of newborns on the equilibrium price

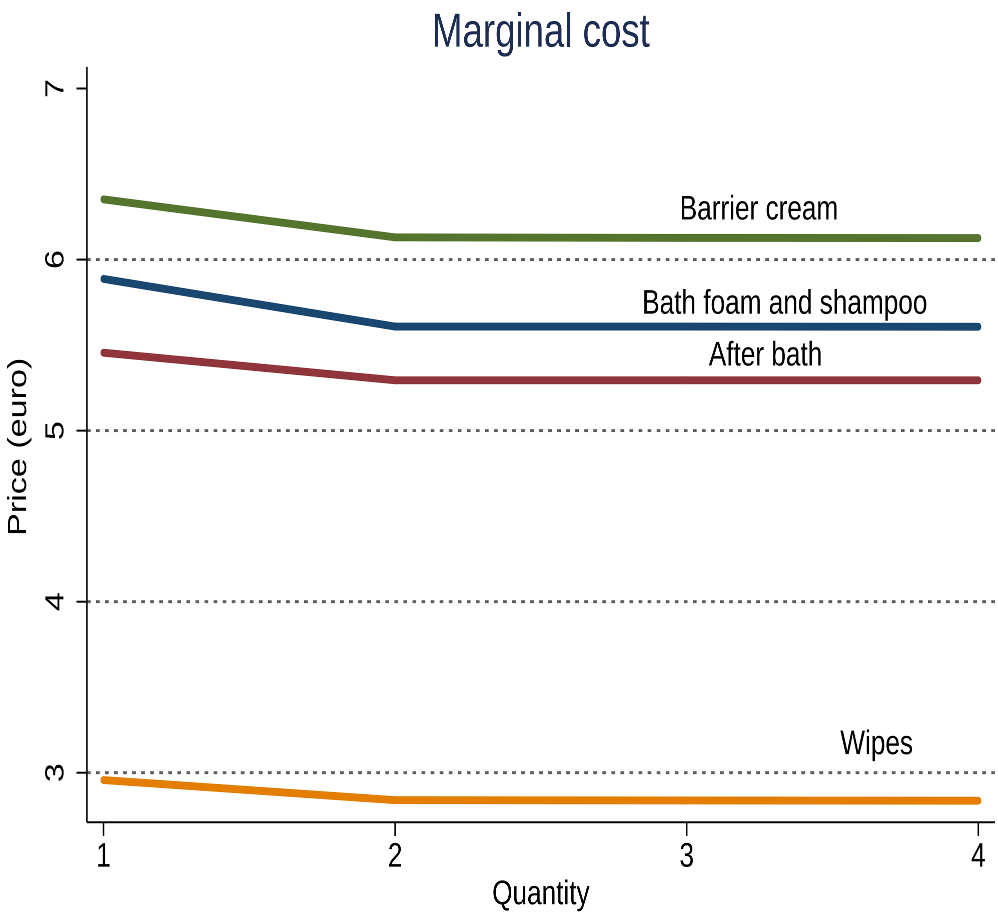
| | Newborns in $t$ | Average newborns in $t$ and $t-1$ | Average newborns in $t$, $t-1$ and $t-2$ |
|---|---|---|---|
| *All cities* | | | |
| Log Newborns | 0.0166 | 0.0266 | 0.0366 |
| | (0.0033)*** | (0.01)*** | (0.0133)*** |
| Number of observations | 56798 | 56798 | 56798 |
| Number of cities | 1665 | 1665 | 1665 |
| *Cities with 100% New Line coverage* | | | |
| Log Newborns | 0.02 | 0.0366 | 0.05 |
| | (0.0066)*** | (0.01)*** | (0.0166)*** |
| Number of observations | 24743 | 24723 | 24723 |
| Number of cities | 792 | 792 | 792 |
| Time effects | Yes | Yes | Yes |
| Cities effects | Yes | Yes | Yes |

*Notes:* OLS estimates of the regression

$$p_{ct} = \alpha + \delta N_{ct}^{U} + h_c + \mu_t + \varepsilon_{ct}$$

where all variables are in logs, $c$ denots a city, $t$ a month, $p_{ct}$ is the equilibrium (log) price and $N_{ct}^{U}$ is the (log) number of newborns. Robust standard error, clustered at the city level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Reported coefficients and standard error have been standardized by the standard deviation of the dependent variable. To evaluate the size of the estimates note that a standard deviation is equal to 1.39 for the (log) number of newborns at time t, 1.31 for the (log) number of newborns at time t and ti-1, and 1.28 for the (log) number of newborns at time t, ti-1 and t-2, and to 0.03 for the (log) price.

Figure 3: Marginal cost faced by pharmacies for child hygiene products.




**Source:** Authors' calculations on data from Qualità in Farmacia

Table 4: Effect of the monthly number of newborns on the equilibrium quantity

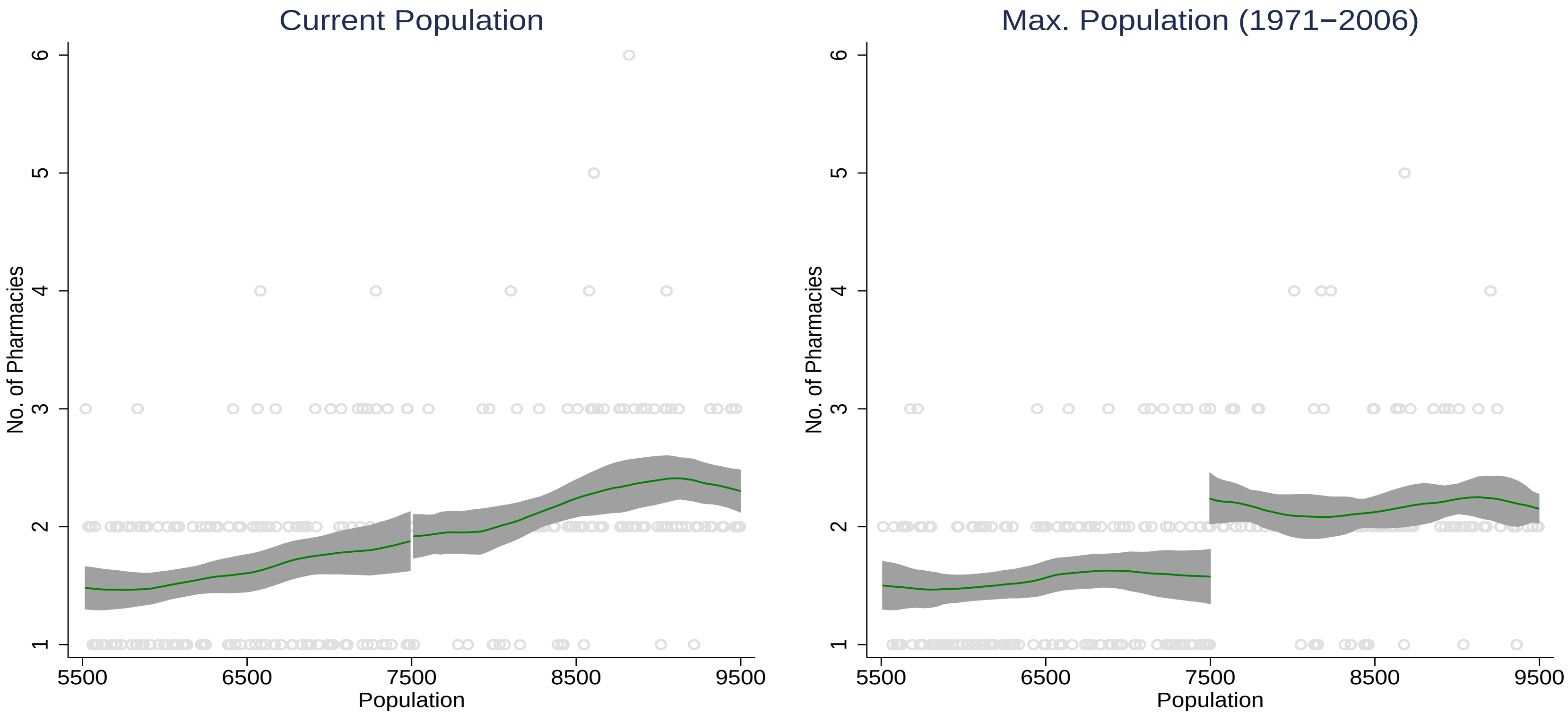
| | Newborns in $t$ | Average newborns in $t$ and $t-1$ | Average newborns in $t$, $t-1$ and $t-2$ |
|---|---|---|---|
| *All cities* | | | |
| Log Newborns | 0.0301 | 0.0525 | 0.0622 |
| | (0.0049)*** | (0.0081)*** | (0.0110)*** |
| Number of observations | 56798 | 56798 | 56798 |
| Number of cities | 1665 | 1665 | 1665 |
| *Cities with 100% New Line coverage* | | | |
| Log Newborns | 0.0325 | 0.0539 | 0.0686 |
| | (0.0061)*** | (0.0098)*** | (0.0132)*** |
| Number of observations | 24723 | 24723 | 24723 |
| Number of cities | 792 | 792 | 792 |
| Time effects | Yes | Yes | Yes |
| Cities effects | Yes | Yes | Yes |

*Notes:* OLS estimates of the regression

$$q_{ct} = \alpha + \delta N_{ct} + h_c + \mu_t + \varepsilon_{ct}$$

where all variables are in logs, $c$ denotes a city, $t$ a month, $q_{ct}$ is the equilibrium (log) quantity and $N_{ct}$ is the (log) number of newborns. Robust standard error, clustered at the city level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Reported coefficients and standard error have been standardized by the standard deviation of the dependent variable. To evaluate the size of the estimates note that a standard deviation is equal to 1.39 for the (log) number of newborns at time t, 1.31 for the (log) number of newborns at time t and t-1, and 1.28 for the (log) number of newborns at time t, t-1 and t-2, and to 0.59 for the (log) quantity.

Figure 4: Current population, maximum population and competition at the threshold




*Notes:* Scatter plot and local polynomial smoothing regressions of the number of pharmacies with respect to current and maximum historical population. Current population is measured at 12-31-2005 for municipalities observed since January 2007, at 12-31-2007 for municipalities observed since January 2009.

Table 5: Competing pharmacies on the two sides of the maximum hystorical population threshold.

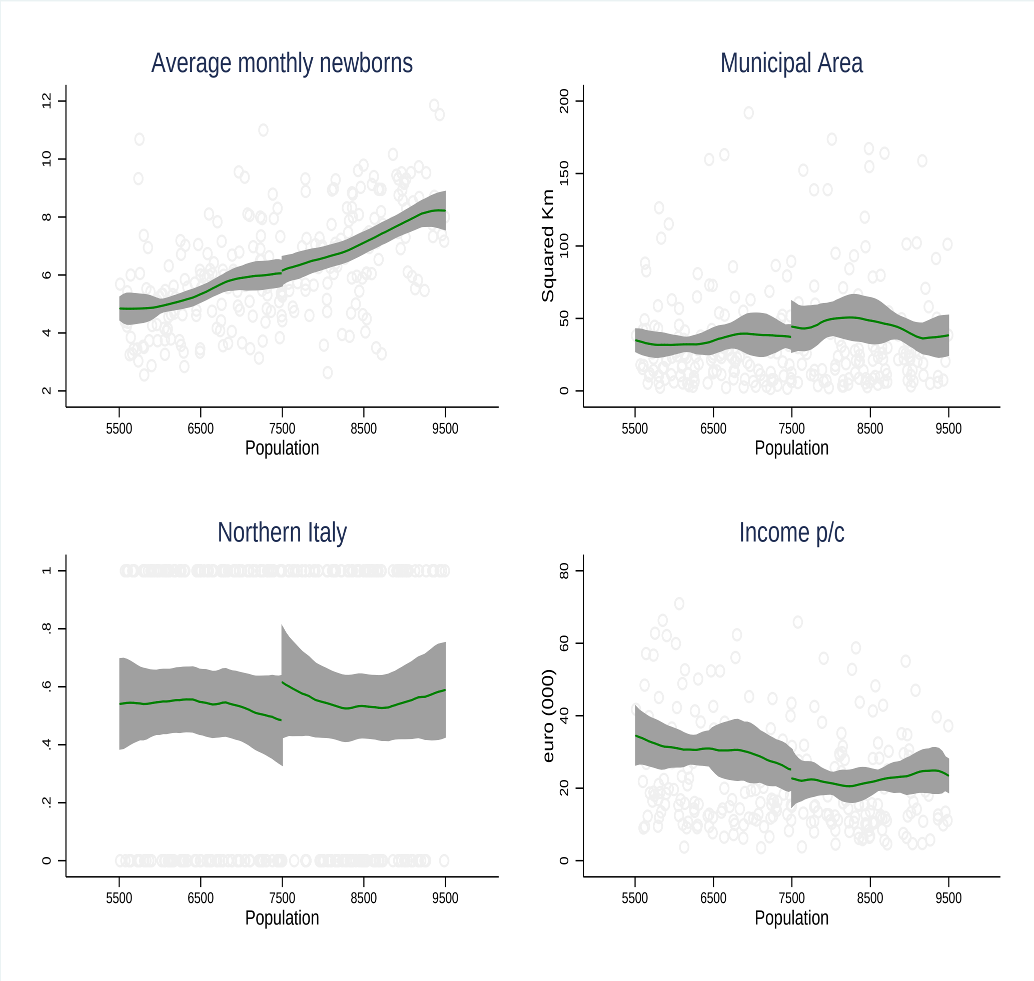
| | Local linear<br>± 1,500 inhabs. | Polynom. $2^{n}d$<br>± 2,000 inhabs. | Polynom. $3^{r}d$<br>± 3,000 inhabs. | Polynom. $4^{t}h$<br>± 4,000 inhabs. |
|---|---|---|---|---|
| *Panel A: All cities* | | | | |
| Right side of the threshold | 0.470 | 0.628 | 0.774 | 0.891 |
| | (0.183)** | (0.226)*** | (0.240)*** | (0.259)*** |
| Constant | 0.640 | 0.615 | 0.544 | 0.519 |
| | (0.126)*** | (0.159)*** | (0.166)*** | (0.175)*** |
| No. of Obs. | 215 | 272 | 433 | 591 |
| *Panel B: 100% New Line coverage* | | | | |
| Right side of the threshold | 0.888 | 1.248 | 1.335 | 1.247 |
| | (0.241)*** | (0.244)*** | (0.249)*** | (0.236)*** |
| Constant | 0.113 | 0.088 | 0.069 | 0.086 |
| | (0.080) | (0.101) | (0.105) | (0.109) |
| No. of Obs. | 89 | 114 | 197 | 297 |
| *Panel C: Controls included* | | | | |
| Right side of the threshold | 0.384 | 0.607 | 0.766 | 0.902 |
| | (0.178)** | (0.214)*** | (0.220)*** | (0.236)*** |
| Constant | 1.464 | 1.544 | 1.491 | 1.349 |
| | (0.343)*** | (0.343)*** | (0.287)*** | (0.270)*** |
| No. of Obs. | 215 | 272 | 433 | 591 |

*Notes:* OLS estimates of the regression:

$$S_c = \omega + \varphi K_c + g(|Pop_c - \kappa|) + \zeta_c$$

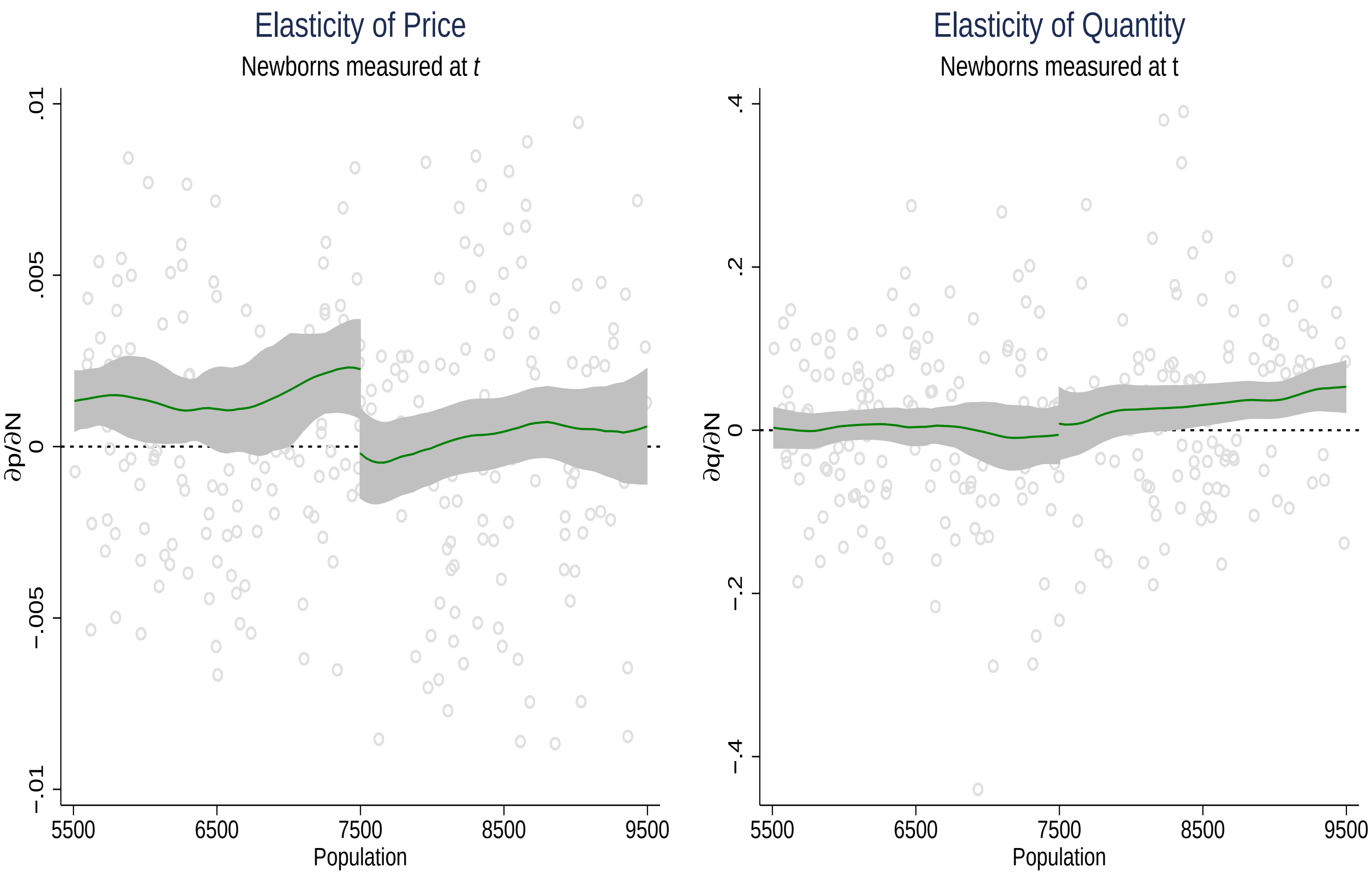
where $c$ denotes a city, $S_c$ is the number of pharmacies in a city; $pop_c$ is the maximum historical population in a city; $K_c = 1(Pop_c >= \kappa)$ is a dummy taking value 1 for cities on the right side of the threshold. Robust standard error, in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Panel B includes cities in which New Line has a 100% coverage on pharmacies. Regressions in Panel C includes the average monthly number of newborns, a dummy taking value 1 if the city is in a urban area, a dummy taking value 1 if the city is in Northern Italy and per capita disposable income at the city level as controls.

Figure 5: Continuity tests for covariates




*Notes:* Scatter plot and local polynomial smoothing regressions of four observable "pre-treatment" city characteris with respect to maximum historical population. The four characteristics are: the average monthly number of newborns, a dummy taking value 1 if the city is in a urban area, a dummy taking value 1 if the city is in Northern Italy and per capita disposable income in the city.

Figure 6: The effects of newborns on quantity and price at the threshold




*Notes:* Scatter plot and local polynomial smoothing regressions of the city specific elasticities $\hat{\delta}_c^Y$ of quantity ($Y_c = q_c$) and price ($Y_c = q_c$) with respect to the monthly number of newborns.

Table 6: Elasticities of the quantity and price indexes with respect to the monthly number of newborns on the two sides of the threshold.

| | Local Linear Regression ± 1,500 inhabs. | | Polynomial $2^{nd}$ ± 2,000 inhabs. | | Poynomial $3^{rd}$ ± 3,000 inhabs. | | Polynomial $4^{th}$ ± 4,000 inhabs. | |
|---|---|---|---|---|---|---|---|---|
| | Quantity | Price | Quantity | Price | Quantity | Price | Quantity | Price |
| Threshold Dummy | -0.001 | -0.003 | -0.014 | -0.004 | -0.005 | -0.005 | -0.003 | -0.005 |
| | (0.035) | (0.001)** | (0.046) | (0.001)** | (0.051) | (0.002)*** | (0.054) | (0.002)*** |
| Constant | -0.004 | 0.002 | -0.006 | 0.003 | -0.009 | 0.003 | -0.016 | 0.003 |
| | (0.023) | (0.001)*** | (0.025) | (0.001)*** | (0.025) | (0.001)*** | (0.024) | (0.001)*** |
| Number of Observations | 217 | 217 | 277 | 277 | 439 | 439 | 597 | 597 |

*Notes:* OLS estimates of the regression

$$\hat{\delta}_c^Y = \omega + \gamma^Y K_c + g(|Pop_c - \kappa|) + \eta_c$$

where $c$ denots a city; $Y_c$ is equal to $q_c$ in column 1 and to $p_c$ in column 2; $pop_c$ is the maximum historical population in a city; $K_c = 1(Pop_c >= \kappa)$ is a dummy taking value 1 for cities on the right side of the threshold. The dependent variable $\hat{\delta}_c^Y$ is obtained in two steps: first we regress $Y_c$ on city and time fixed effects, and then we regress the residuals from the fist step on the (log)number of newborns separately for each city, exploiting the within city time variability. Robust standard error are in parentheses with *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Elasticities of the quantity and the price indexes with respect to the number of newborns on the two sides of the threshold - Special sub-samples

| | Quantity | Price |
|---|---|---|
| *Panel A: 100% New Line coverage - local linear regression - ± 4,000 inhab.* | | |
| Threshold Dummy | 0.023 | -0.003 |
| | (0.028) | (0.001)*** |
| Constant | -0.013 | 0.002 |
| | (0.017) | (0.001)*** |
| Number of Observations | 301 | 301 |
| *Panel B: Controls included - polynom. $4^{th}$ - ± 4,000 inhab.* | | |
| Threshold Dummy | -0.006 | -0.005 |
| | (0.054) | (0.002)*** |
| Constant | -0.039 | 0.003 |
| | (0.042) | (0.002)* |
| Number of Observations | 597 | 597 |

*Notes:* OLS estimates of the regression

$$\hat{\delta}_c^Y = \omega + \gamma^Y K_c + g(|Pop_c - \kappa|) + \eta_c$$

where $c$ denotes a city; $Y_c$ is equal to $q_c$ in column 1 and to $p_c$ in column 2; $pop_c$ is the maximum historical population in a city; $K_c = 1(Pop_c >= \kappa)$ is a dummy taking value 1 for cities on the right and side of the threshold. The dependent variable $\delta_c^Y$ is obtained in two steps: first we regress $Y_c$ on city and time fixed effects, and then we regress the residuals from the fist step on the (log)number of newborns separately for each city, exploiting the within city time variability. Panel A includes cities in which New Line has a 100% coverage on pharmacies. Panel B includes the average monthly number of newborns, a dummy taking value 1 if the city is in a municipal area, a dummy taking value 1 if the city is in Northern Italy dummy, and per capita disposable income at the city level as controls. Robust standard error are in parentheses with *** p<0.01, ** p<0.05, * p<0.1.

Table 8: IV estimates of the effect of a change in the number of competitors on the elasticities of quantity and price with respect to the number of newborns.

| | Local Linear Reg. $\pm$ 1,500 inhabs. | | Polynomial $2^{nd}$ $\pm$ 2,000 inhabs. | | Poynomial $3^{rd}$ $\pm$ 3,000 inhabs. | | Polynomial $4^{th}$ $\pm$ 4,000 inhabs. | |
|---|---|---|---|---|---|---|---|---|
| | Quantity | Price | Quantity | Price | Quantity | Price | Quantity | Price |
| Number of competitors | 0.005 | -0.005 | -0.029 | -0.005 | -0.012 | -0.006 | -0.012 | -0.005 |
| | (0.070) | (0.003)* | (0.075) | (0.003)** | (0.066) | (0.003)** | (0.058) | (0.002)** |
| Const. (1 pharmacy, 0 competitors) | -0.013 | 0.006 | 0.014 | 0.012 | 0.004 | 0.013 | -0.019 | 0.010 |
| | (0.072) | (0.003)* | (0.136) | (0.005)** | (0.115) | (0.005)** | (0.097) | (0.004)** |
| Observations | 215 | 215 | 272 | 272 | 433 | 433 | 591 | 591 |
| F-stat of excluded instrument | 7.19 | 7.19 | 8.16 | 8.16 | 12.09 | 12.09 | 14.66 | 14.66 |

*Notes:* IV estimates of the regression

$$\hat{\delta}_c^Y = \omega + \gamma^Y s_c + g(|Pop_c - \kappa|) + \eta_c$$

where $c$ denotes a city; $Y_c$ is equal to $q_c$ in column 1 and to $p_c$ in column 2; $s_c$ is the number of competitors (equal to the number $S_c$ of pharmacies in a cities minus 1) and is instrumented with $K_c = 1(Pop_c > \kappa)$; $Pop_c$ is the maximum historical population in a city; $K_c = 1(Pop_c >= \kappa)$ is a dummy taking value 1 for cities on the right and side of the threshold. The dependent variable $\delta_c^Y$ is obtained in two steps: first we regress $Y_c$ on city and time fixed effects, and then we regress the residuals from the fist step on the (log)number of newborns separately for each city, exploiting the within city time variability. All regressions include the following controls: the average monthly number of newborns, a dummy taking value 1 if the city is in a urban area, a dummy taking value 1 if the city is in Northern Italy and per capita disposable income in the city. Robust standard error are in parentheses with *** p<0.01, ** p<0.05, * p<0.1.