Departamento de Economía
Facultad de Ciencias Sociales
Universidad de la República

Documentos de Trabajo

Distance and Political Boundaries: Estimating Border Effects under Inequality Constraints

Fernando Borraz, Alberto Cavallo, Roberto Rigobon y Leandro Zipitría

Documento No. 10/12
Agosto 2012

ISSN 0797-7484

Distance and Political Boundaries: Estimating Border Effects under Inequality Constraints.*

Fernando Borraz
Banco Central del Uruguay
dECON-FCS-UdelaR

Alberto Cavallo
MIT

Roberto Rigobon
MIT
NBER

Leandro Zipitría
Universidad de San Andrés
Universidad de Montevideo

August 15, 2012

Abstract

La literatura sobre el efecto frontera encuentra que las fronteras políticas tienen un impacto muy grande en los precios relativos, agregando de manera implícita varios miles de kilómetros al comercio. En este trabajo se muestra que la especificación empírica tradicional adolece de sesgo de selección, y se propone una nueva metodología basada en regresiones por cuantiles. Usando una base de datos para Uruguay, aplicamos nuestro procedimiento para medir la segmentación introducida por las fronteras de la ciudad. Las fronteras de la ciudad deberían importar muy poco para el comercio. Encontramos que cuando se utiliza la metodología estándar las fronteras políticas de las ciudades triplican la distancia. Cuando se utiliza nuestra metodología, el efecto frontera de la ciudad se vuelve insignificante. Asimismo, utilizamos la metodología para probar la "frontera en línea", utilizando los precios en línea para una gran cadena de supermercados del país, y demostrar que es equivalente a la distancia media desde el almacén en línea para cada una de las tiendas fuera de línea.

Keywords: Border effect, political borders, price dispersion, quantile regression.
JEL Codes: C14, D40, E31, F40.

*We are grateful to Fernando Antía and Bruno Delgado at the General Directorate of Commerce of Uruguay for providing us with the data. We thank seminar participants at the University of Wisconsin - Madison, MIT, the Central Bank of Peru, the Central Bank of Uruguay, Universidad de San Andres, dECON, and Universita di Venezia for all their comments and suggestions.

Distance and Political Boundaries: Estimating Border Effects under Inequality Constraints.*

Fernando Borraz
Banco Central del Uruguay
dECON-FCS-UdelaR

Alberto Cavallo
MIT

Roberto Rigobon
MIT
NBER

Leandro Zipitría
Universidad de San Andrés
Universidad de Montevideo

August 14, 2012

Abstract

The "border effect" literature finds that political boundaries have a large impact on relative prices, implicitly adding several thousands of miles to trade. In this paper we show that the standard empirical specification suffers from selection bias, and propose a new methodology based on quantile regressions. Using a novel data set from Uruguay, we apply our procedure to measure the segmentation introduced by city borders. We find that when the standard methodology is used, two supermarkets separated by 10 kilometers across two different cities have the same price dispersion as two supermarkets separated by 30 kilometers within the same city; so the city border triples the distance. When our methodology is used, the city border effect becomes insignificant. We further test our method using online prices for the largest supermarket chain in the country, and show that the "online border" is equivalent to the average distance from the online warehouse to the offline stores.

Keywords: Border effect, political borders, price dispersion, quantile regression.
JEL Codes: C14, D40, E31, F40.

*We are grateful to Fernando Antía and Bruno Delgado at the General Directorate of Commerce of Uruguay for providing us with the data. We thank seminar participants at the University of Wisconsin - Madison, MIT, the Central Bank of Peru, the Central Bank of Uruguay, Universidad de San Andres, dECON, and Universita di Venezia for all their comments and suggestions.

1 Introduction

Political boundaries can have a significant impact on relative prices and welfare. The degree of price segmentation caused by such boundaries was first documented in a seminal paper by Engel and Rogers (1996), who showed that the dispersion of prices within a country is orders of magnitude smaller than across countries, and estimated that the US - Canadian border was equivalent to a distance of 75,000 miles. Their work spurred a large literature documenting the sizable and distortionary implications of the "border effect" on prices.[1] For example, Parsley and Wei (2001) found that the border effect between US and Japan is equivalent to several hundred thousands miles, while Ceglowski (2003) reported that provincial borders in Canada are equivalent to 5 thousand miles. Although this type of results have been heavily scrutinized, the degree of segmentation induced by political borders and the economic reasons behind it are still open questions in the literature.[2]

In this paper we argue that the standard regression in the literature is subject to a selection bias that affects both the distance and border coefficients, and propose an alternative approach using quantile regressions that controls for this bias. We apply our method to estimate the impact of distance and political borders on price dispersion across different cities in Uruguay. Our dataset has daily prices collected by the Uruguayan government for a set of 202 UPC-level products sold in 333 supermarkets across 47 cities. We use data at the city level –within a single country– to ensure that we are comparing prices for identical goods and

[1]Engel and Rogers (2004), Frankel, Stein, and Wei (1995), Nitsch (2000), Anderson and van Wincoop (2001), Helliwell (1997), Helliwell and Verdier (2001), Helliwell and Schembri (2005), Engel, Rogers, and Wang (2003), Parsley and Wei (2001), Crucini, Shintani, and Tsuruga (2010), and Gopinath, Gourinchas, Hsieh, and Li (2011) to name a few papers that have documented the border effect. Goldberg and Knetter (1997) presents a very nice survey of the earlier literature. Finally, Wolf (2000) and Ceglowski (2003) present evidence that the border effect exists across provinces and cities.

[2]Some papers have argued that (i) the distances have been mismeasured (see Head and Mayer (2002)), (ii) that the regressions suffer from aggregation and sample selection bias of the traded products (see Evans (2001) and Broda and Weinstein (2008)) (iii) that the gravity equation implied in the standard specification has been misspecified (see Anderson and van Wincoop (2003) and Hillberry and Hummels (2003)) , (iv) and that the regressions do not have a proper benchmark due to the fact that country distributions of prices are very different across countries (see Gorodnichenko and Tesar (2009)). Dani Rodrik pointed out in his discussion of Anderson and van Wincoop (2001) that "there is convergence in the literature that border effects are very large, while explicit trade barriers in the form of trade policies, tariffs and quotas, are generally small."

are able to control for retailer effects, therefore isolating the impact of the political boundary. Our method also can be applied to measure price dispersion across all kinds of borders, as we illustrate by measuring the degree of segmentation between online and offline markets in Montevideo.

Following Engel and Rogers (1996), we start with a simple framework where price dispersion is bounded by the existence of a no-arbitrage condition. Factors such as heterogeneous demands, different productivity shocks, and price stickiness may increase the degree of price dispersion across locations, but firms are subject to an arbitrage constraint. For simplicity we assume that the consumer is doing the arbitrage, so that the price of a good in one location cannot be higher than the price in another location plus the trade cost. If the trade cost between two establishments (i and j) is τ , and p denotes the log price in each location, then the constraint can be expressed as a simple inequality:

$$\left|p_i - p_j\right| \leq \tau \tag{1}$$

The distance between locations and the existence of a border have a direct impact on τ. The distance adds a transportation cost, while the border introduces other costs related to tariffs, market regulations, differences in product packages, and languages. All other things equal, if the distance increases or a border exists, then τ rises and we should find a greater price dispersion across locations. Following this logic, most papers in the literature have estimated τ, and its determinants, regressing the absolute value (or the standard deviation) of the observed price differences in locations i and j, on the distance $D_{i,j}$, a border dummy B and a series of additional controls $X_{i,j,t}$.

$$|p_i - p_j| = \alpha + \beta D_{i,j} + \gamma B + \delta X_{i,j} + \varepsilon_{i,j} \tag{2}$$

The border effect is the equivalent number of miles that would produce the same dispersion as the estimated border dummy coefficient γ. In its simplest form, it is the ratio γ/β, so a bias in either (or both) of these coefficients will have an impact on the estimate for the border effect.

We argue that the estimation of τ cannot be done using a simple OLS regression because prices in the two locations are an optimal choice subject to a constraint that may not be binding. If the optimal prices of the two stores lie within the constraint, then their difference is smaller than τ and this observation is not relevant to estimate the trade costs. To illustrate this, consider two markets that are highly segmented but have identical supply and demand characteristics. Goods will have the same price across the two locations, but this price gap tells us nothing about the trade costs or the degree of segmentation between the markets. In fact, all observations within the no-arbitrage range suffer from selection bias, and estimates that use the mean or the standard deviation of $|p_1 - p_2|$ are going to be biased downward.

Given the inequality constraints, the only observations that are not subject to a bias are the ones lying on the boundary.[3] This means that τ is better estimated using the maximum of the observed absolute deviations. This maximum, however, is sensitive to the possibility of errors-in-variables, so we estimate instead a series of quantile regressions, starting with the mean (to replicate the methodology commonly used in the literature), then using the 80th, 90th, 95th, 99th percentiles, and finally the maximum observed price difference. As we move to higher percentiles, our estimates are less affected by the sample selection bias. They are also more sensitive to the errors in variables, but if these errors are small, then the regression coefficients should be monotonically increasing with higher percentiles.

We first estimate the border effect in our data using standard methods. We find that the city border between two stores separated by 10 kilometers is larger than 20 kilometers wide, and statistically different from zero. Hence, the border triples the implied distance of stores

[3]The estimation problem is equivalent to estimating using inequality moment as opposed to equality moments. Recently, there has been significant research in the area of estimation under moment inequalities. See Andrews, Berry, and Jia (2004), Andrews and Guggenberger (2009), Andrews and Soares (2010), Andrews and Shi (2010), Ponomareva and Tamer (2011), and Rosen (2008) for some of the best theoretical papers in this area.

across the city borders. Using our quantile method to perform the same exercise, however, the border declines until it is not significantly different from zero.

In our results, the city border effect is smaller because distance matters much more. In fact, both the distance and border dummy coefficients are downward biased in the standard regression, but the bias is strongest on the distance parameter. The reason is that price gaps within the arbitrage constraint are less common for observations across cities, and therefore the border coefficient is less affected by the selection bias. Within cities, by contrast, small price gaps are very frequent and can greatly bias the distance coefficient. This could happen, for example, if a supermarket has a single-price policy within a city. This would not mean that the distance within the city does not create some market segmentation; it would simply mean that those differences are responding to other pricing optimality decisions not affected by distance, and therefore are not meaningful for the computation of trade costs. In fact, when we are comparing across political borders there are many unobserved factors that can impact the relative importance of these non-binding price gaps, creating all sorts of distortions on the traditional border estimates.[4] In our data, as we use higher percentiles of the price-gap distribution, the selection bias falls, the distance parameter β rises more than the border dummy coefficient γ, and therefore the border effect falls (almost) monotonically towards zero.

We run several robustness exercises, correcting for outliers, product mix, and changing the specification to include non-linearity and interaction terms. In all of them, the city-border effect tends to disappear when the higher percentiles are used. Furthermore, the results are similar for the 99th, 99.5th, 99.9th percentile, and the maximum, suggesting that the errors in variables problem is small.

We further validate our methodology by estimating the degree of segmentation between online and offline prices in Montevideo. We use daily prices from the largest supermarket

[4] The existence of this heterogeneity in price dispersions was discussed at the country level by Gorodnichenko and Tesar (2009)

chain in the city, comparing its online prices to those in each of the offline stores in the city. We measure the online border effect, which is the implied "distance" between the offline stores and the online stores. If the usual procedure is used, online and offline markets appear to be very closely integrated, with an equivalent border of 1.6 kilometers. However, when use the 95th percentile of the price gap distribution, the online border effect becomes 8.8 kilometers. This is very close to the actual physical average distance between the online warehouse (where the online goods are delivered from) and each of the offline stores in the city.[5] In this case, the standard methodology *underestimates* the border effect because the online price is usually somewhere in between the prices of the other offline stores, creating small price gaps that are no longer relevant when higher quantiles of the distribution are used.

Our paper is related to a large literature on the border effects measured with price dispersions. We attempt to deal with the most important critiques that have been raised on the original Engel-Rogers regression. First, we use product-level data with identical goods across all locations. As suggested by Goldberg and Knetter (1997), product-level data is crucial to understand the deviations of the LOP. Indeed, Evans (2001) and Broda and Weinstein (2008) argue that a significant problem in the border effect literature is the aggregation bias induced by price indexes. Second, we use retail prices. Hillberry and Hummels (2003) have argued that business-to-business data tends to overestimate trade flows and underestimate price differences within countries. Third, we have the exact location of each store. As pointed out by Head and Mayer (2002), using approximate distances (such as from one country capital to the other) can greatly overestimate the border effect. Finally, all the stores in our sample sell the same set of products. As Evans (2001) points out, the mix of products sold across countries is much smaller than the mix of products traded within countries, which might lead to a bias

[5]The supermarket website says that the online prices match the prices of the store where the orders are sent from. To identify it, we compared online prices one-to-one with each store and found a location where they were identical in 97.3 percent of the daily observations. That store has an average distance to all the other stores in the city of Montevideo of 7.2 kilometers – close to our estimates of the online border when using the upper quantiles in the regression

in the standard regressions. Our results are consistent with Gorodnichenko and Tesar (2009), who argue that with "cross-country heterogeneity in the distribution of within-country price differentials, there is no clear benchmark from which to gauge the effect of the border." We agree with this statement, but show that even in the absence of a structural model it is still possible to obtain a simple and reliable estimate for the magnitude of the border effect using quantile regressions. Our paper is also complementary to the work of Gopinath, Gourinchas, Hsieh, and Li (2011). They point out that "the logic of using price gaps to infer trade costs implicitly assumes that markets remain integrated despite these transaction costs." This is precisely why we use observations that are at the extremes of the price-gap distribution to estimate the lower bound of the trade costs.

2 Methodology

Most papers estimating border effects run one of the following two regressions:[6]

$$|p_{i,t} - p_{j,t}| = \alpha + \beta D_{i,j} + \gamma B + \delta X_{i,j,t} + \varepsilon_{i,j,t} \tag{3}$$

$$\sigma\left(p_{i,t} - p_{j,t}\right) = \alpha + \beta D_{i,j} + \gamma B + \delta X_{i,j,t} + \varepsilon_{i,j,t} \tag{4}$$

where $p_{i,t} - p_{j,t}$ is the log price difference between locations i and j at time t. The locations can be countries, provinces, cities or establishments. $D_{i,j}$ is the distance between the two locations, and B is a dummy if a border between the two locations exists. $X_{i,j,t}$ are some additional controls. Regression 3 estimates how distance and border impact the average absolute deviation of prices, while regression 4 estimates their impact on the dispersion of prices (measured by their standard deviation). The objective is to estimate the degree of segmentation introduced by trade costs – where it is assumed to depend on distance, border, and other controls.

[6] See section IV in Broda and Weinstein (2008) for a very good summary of the papers using these two regressions.

These regressions have been widely used in the literature. Papers that have supported the existence of border effects, and those that have criticize it, use the same specification. In this section, we show that if these regressions are used, the estimated coefficients are biased downward.

The intuition for why the bias arises can be easily derived from the no-arbitrage pricing region that Samuelson's Iceberg costs generate.[7] Assume that there is a trade cost between two locations that can be described as follows:

$$\tau_{i,j,t} = \alpha + \beta D_{i,j} + \gamma B + \delta X_{i,j,t} \tag{5}$$

where the variables are defined as before. This trade cost represents the proportion of the item that is lost when a customer transports one unit from i to j. For simplicity in the exposition it is assumed that the agent performing the arbitrage is the customer itself.[8] Under this form of trade costs, prices need to lie within the range $\left|p_i - p_j\right| \leq \tau_{i,j,t}$ to avoid the possibility that a customer arbitrates among the locations. Assume that p_i is set. The second store, when deciding its price, maximizes profits subject to the no-arbitrage constraint. If the optimal price is such that the difference between p_i and p_j is smaller than τ then the constraint is not binding and the price difference is a biased estimate of τ. But if the difference is bigger, then the store sets the price at the corner solution and the constraint is binding.

This simple behavior implies that the absolute difference of log prices satisfies inequality 1, which can be rewritten here as

$$\left|p_i - p_j\right| \leq \tau_{i,j,t} = \alpha + \beta D_{i,j} + \gamma B + \delta X_{i,j,t}$$

Note that this inequality implies that in equation 3 all the residuals $\left(\varepsilon_{i,j,t}\right)$ are either zero

[7] See Samuelson (1954).

[8] So, the trade cost can be interpreted not only the loss of physical items, but also the loss in terms of utility that the customer experience, or that it would have to incur, if it were forced to travel from one location to the other.

or negative, so $E\left[\varepsilon_{i,j,t}\right] \leq 0$. In general, the estimation by OLS produces biased estimates, because of the failure of orthogonality conditions. Nevertheless, there is one case in which the estimates are unbiased. If the price deviations are exactly equal to the trade cost – i.e. the constraint is always binding – then the residuals are identical to zero and estimation by OLS produces an unbiased estimates. The intuition can be easily explained if we start by assuming that prices do not have errors-in-variables.[9] This is a strong assumption that we relax immediately below. Under this assumption the extreme in the distribution of price differences is the closest estimator to the trade costs. It is indeed the best estimator of the lower bound of the trade cost.

The estimation procedure is as follows:

1. Compute the absolute price differences for all possible location pairs

2. Define distance-border-bins according to some discrete spacing that depends on the unit of observation (city vs countries) and the availability of enough observations within each bin. For example, for the city effect stores are assigned to bins of a few miles apart. If the unit of analysis is countries, bins should be larger to contain stores that are separated by bigger distances. The distance in each bin does not have to be set in linear increments. Assume there are N bins and denote each bin as b_n. Each bin is defined by a distance D_n, whether there is a border between the two stores ($B_n = 1$), and with additional controls X_n such as chain dummies, and interaction terms.

3. For each bin, compute the extremum statistic of the absolute price differences. Denote the statistic as $Q(n, \theta)$, where $Q(\bullet, 1)$ is the maximum and $Q(\bullet, q)$ is for the q^{th} percentile.

[9] This requires two assumptions: First, assume that prices are observed and/or reported without errors. Second, assume that stores do not make errors in their pricing decision. In other words, stores never post prices outside the no-arbitrage range.

4. Using the $\{Q(n,\theta)\}_1^N$, estimate

$$Q(n,\theta) = \alpha + \beta D_n + \gamma B_n + \delta X_n + \varepsilon_n$$

In Figure 1 we depict the source of the bias and the intuition behind our methodology. On the horizontal axis the bins for each distance is shown. The vertical axis is the absolute price difference. For each “bin”, all the absolute differences from the data are shown (the dots). The thick black line reflects the price difference implied by the no-arbitrage constraint. Because all the observed price differences are less or equal to the thick line, the estimation in the standard regression – which implicitly uses the mean within each bin – is downward biased (denoted as the thin black line). In small samples, the true maximum for each bin might not be observed, and therefore estimating using the sample maxima will also be biased downward. However, the bias is smaller than using the mean. In other words, it is possible that there is no realization on the black line, but using the maximum within each bin gets closer to the true one. This is why we interpret our results as a lower bound estimate of the degree of segmentation.

[Figure 1 here]

2.1 Dealing with errors-in-variables

One important aspect is how to deal with the possibility of errors in variables (EIV). These errors can arise either because prices are misreported, or because stores make mistakes and post prices outside the no-arbitrage range. The biggest challenge is that the maximum price difference within each bin is significantly affected if prices are mismeasured. We describe the data we use in Section 3 and it will become clear that the errors from misreporting are very small, given the way the data is collected. However, there still exists the possibility that the prices are incorrectly reported and concentrating the estimates on the maximum within each bin would exacerbate the impact of any errors-in-variables.

This situation is depicted in Figure 2. The black thick line is still the "true" upper bound of the no-arbitrage band. This is the true degree of segmentation. Notice that now, because of EIV, some price differences might even be above the no-arbitrage range. In this case, using the maximum within each bin also produces a bias in the estimation.

[Figure 2 here]

We deal with errors-in-variables in two ways. One is to eliminate outliers from the distribution. As we discuss below, the type of errors that are likely to be present in our data are misplacement of the decimal point, or flipping digits. Both are likely to produce large price changes at the item level that we can observe. We evaluate the robustness of the estimates to the elimination of price change outliers. This approach, however, does not provide a definite answer. For example, if the estimates change little then we could conclude that either the EIV had a small impact, or that not enough observations were eliminated. We decided, therefore, to estimate the regression using quantiles. Within each bin we compute several quantiles – the median, 80th, 90th, 95th, and 99th percentiles. The 50th and 80th percentiles are clearly less affected by the EIV than the maximum, but those estimates will be affected by the sample selection of prices within the no-arbitrage range. As we move to higher and higher percentiles, the estimates are less affected by the sample selection, and more affected by the EIV. If the EIV is small, it should be the case that the estimates are monotonically increasing. We evaluate the robustness and sensitivity of our estimates to several quantiles below.

3 Data

We use a micro dataset of daily prices compiled by The General Directorate of Commerce (DGC) which includes grocery stores all over the country. The DGC is the authority responsible for the enforcement of the Consumer Protection Law at the Ministry of Economy and Finance. In 2006 a new tax law was passed by the legislature which changed the tax base and rates of the value added tax (VAT). The Ministry of Economy and Finance was concerned about incomplete pass-through from tax reductions to consumer prices, so it decided to collect and publish a dataset of prices in different grocery stores and supermarkets across the country. The DGC issued Resolution Number 061/006 which mandates that grocery stores and supermarkets must report the daily prices for a list of products if they fulfill the following two conditions: i) they sell more than 70% of the products listed, and ii) they have more than four grocery stores under the same name, or have more than three cashiers in a store. The information sent by each supermarket is a sworn statement, and they are subject to penalties in case of misreporting

The data include the daily prices for 202 UPC corresponding to 61 product categories.[10] The products in the sample represent 16.34% of the goods and services in the CPI basket. The DGC requires large retailers to report their daily prices once a month using an electronic survey. The three highest-selling brands are reported for each product category. Most items had to be homogenized in order to be comparable, and each supermarket must always report the same item. For example, sparkling water of the SALUS brand is reported in its 2.25 liter variety by all stores. If this specific variety is not available at a store, then no price is reported.

The DGC makes the information public through a web page that publishes the average monthly prices of each product for each store in the defined basket.[11] This information is available within the first ten days of the next month. There is no further use for the information; e.g. no price control, nor are any further policies implemented to control supermarkets or producers.

Each item is defined by its universal product code (UPC) with the exception of meat,

[10] The same data set is used in Borraz and Zipitría (2012).

[11] See http://www.precios.gub.uy/publico.

eggs, ham, some types of cheese, and bread. In some instances, as in the case of meat and various types of cheese, general definitions were set, but because of the nature of the products they could not be homogenized. In the case of bread, most grocery stores buy frozen bread and bake it, rather than produce it at the store. Grocery stores differ in the kind of bread they sell, so in some cases the reported bread does not coincide with the exact definition, so grocery stores estimate the equivalent price submitted to the DGC; i.e. if the store sells bread that is 450 grams per unit, and the requested bread is 225 grams, it submits half the price of its own bread.

Within four working days of the end of the month, each supermarket uploads its price information to the DGC. After that, a process of "price checking" begins. This process starts by calculating the average price for each item in the basket. Each price 50% greater or less than the average price is identified. Then the supermarket is contacted in order to check whether the submitted price is correct. We eliminated items that were not correctly categorized (marked as 'XXX' and '0') and some products that mistakenly share the same UPC.[12]. We also eliminated all observations for March 2007, which were labeled as preliminary data.

We end up with 202 products at the UPC level in 333 grocery stores from 47 cities in the 19 Uruguayan departments. See Figure 3 for a map with the cities covered in the dataset. These cities represent more than 80% of the total population of Uruguay. Montevideo, with 45% of the population contains 58% of the supermarkets in the sample. As our approach is based on dealing with the largest price differences between one good, we need to carefully account for outliers. So we work with two different databases; one with the complete sample, and a second one in which we delete those prices higher than 3 times the median price, or those that are less than a third of the median daily price. The deleted prices account for a tinny 0.034% of the whole database.

For computing the distances, we use information on the exact geographical location of each supermarket provided by Ciudata, an industry organization. We use it to calculate

[12] The list of products can be found at http://www.precios.gub.uy/publico

the linear distance between each of supermarket in our sample. The maximum distance between two supermarkets is 526 kilometers. Using this distances we construct bins using a geometric sequence starting from 0.1 kilometers and having increments of $((550/0.1)^{1/X})$%. Our preferred estimation uses 500 bins, but we estimated using 50, 100, and 1,000 as well. We calculate the distance between all supermarkets in the sample (333) and assign each pair of supermarkets (55,278) to its proper bin according to their distance range.

We also calculate two dummies that take the value of one if the two supermarkets are in the same city (B_n), or if they belong to the same firm ($Chain_n$). Finally, we include interaction term: distance and the city dummy, as well as non-linearities.

We have two specifications:

$$
\begin{aligned}
Q\left(|p_{i,t}-p_{j,t}|_n,\theta\right) &= \alpha+\beta D_n+\gamma B_n+\delta B_n\times D_n \\
&\quad +\gamma Firm_n+\varepsilon_n \qquad (6)\\
Q\left(|p_{i,t}-p_{j,t}|_n,\theta\right) &= \alpha+\beta D_n+\gamma B_n+\delta B_n\times D_n \\
&\quad +\beta_1 D_n^2+\beta_2 D_n^3+\delta_1 B_n\times D_n^2+\delta_2 B_n\times D_n^3 \\
&\quad +\gamma Firm_n+\varepsilon_n \qquad (7)
\end{aligned}
$$

where Q estimates the quantile θ of the absolute price differences for all store pairs i and j that have distances that belong to bin n; D_n measure the distance between stores that belong to bin n; B_n is a dummy that takes the value 1 if supermarkets are in different cities; $Firm$ is a dummy variable that takes the value 1 if price difference in that bin came from the same supermarket chain. We also add to the equation a fixed effect for each good. In the end, for each bin and for all the stores that belong to the same city, there is only one observation used in the regression: the quantile θ of its distribution. For the stores with the same distances but across cities, there is another bin and another quantile from that distribution.

Figure 4 shows the distribution of observations for each of the 500 bins for the same city and different cities. The horizontal axis is the log distance starting at 100 meters to a maximum of 550 km. The black line are the number of observations in each bin for stares within the same city boundaries, while the blue line are the observations for stores in different cities. Notice that there is a non-trivial range in which stores are separated exactly by the same distance within cities and across cities - although almost all of them within 10 to 15 kilometers.

[Figure 4 here]

4 Results

In this section we present the main results. As was said, we pooled all the data inside each bin and estimate the distribution of price differences. We picked the mean, median, 50, 80, 85, 90, 95, 97.5, 99, 99.5 and 99.9th percentile. For each one, we estimate equation (6) and (7) by weighted least squares to account for the number of observations. The price differences are in percentage terms, while distance is measured in hundreds of kilometers (this is just for normalization purposes).

[Table 1 about here]

The results are presented in Table 1. For each coefficient we present the point estimate and its standard errors. The first panel in the Table show the results from estimating the linear specification, while the second panel shows the coefficients from the non-linear regression. The first coefficient is the segmentation generated by distance. The second and third estimate the effect of the city boundaries (the constant) as well as the interaction term (how the effect of distance changes once the stores are in different cities.) The fourth coefficient is the impact of belonging to the same chain, and the last one is the constant term. In the non-linear specification the interaction term $City * Dist^3$ was always dropped from all models we estimated.

Each column reflects a different regression. The first one computes the mean within each bin which replicates the regressions in the literature. After that we present the results for the quantile moving from the 50th until 99.9th and the maximum.

Using the mean, the city border triples the implied distance of two stores separated by 10 kilometers. To compute the border effect, we need to find the equvalent number of miles that would create the same price dispersion as the border coefficient. Since most of our estimates imply non-linear relationships, the computation of the border effect has to be done for a specific distance. We show the results for 10 kilometers but results are qualitatively the same for stores 15 and 20 kilometers apart. Given our data, it makes no sense to go beyond this point because in the city of Montevideo there are very few observations with stores farther than 20 kilometers.

For a given distance of 10 kilometers we calculate the degree of price dispersion when the two stores are located in different cities. Then we solve for the distance that would be needed within the same city for two stores to have the same degree of price dispersion. The following example clarifies the analysis. Using the results from Table 1 for the estimation using the average, we can compute the price dispersion of two cities across the border that are 10 km apart. The price dispersion is $5.081 + 4.188 * 0.1 + 1.260 - 4.049 * 0.1 = 6.355$. Two stores in the same city exhibit a segmentation equal to $5.081 + 4.188 * X$. Solving for X to make the within city segmentation equal to 6.355 gives 30.5 km. So, the border adds 20 kilometers to two stores 10 kilometers apart – it triples its distance.

We then re-estimate everything using higher quantiles instead of the mean in each bin. All the individual coefficients increase – in line with the intuition we discussed before. This pattern can be easily appreciated in Figures 5 and 6 where we plot the coefficient on distance, and the city dummy. We plot these two coefficients because they are central to the discussion of the border effect, but all the point estimates exhibit this pattern.

[Figure 5 and Figure 6 here]

The exact same pattern occurs in the non-linear specification – shown in the bottom panel in Table 1. In absolute value, all the coefficients become bigger as the estimation is performed over the higher quantiles. In Figure 7 we compute the additional distance implied by the border effect for each of the quantiles and specifications – linear and non-linear. Panel (a) shows the additional kilometers for a pair of stores that are 10 kilometers apart.

As can be easily seen, the computation of the border effect – measured in kilometers – collapses towards zero around the 97.5th percentile when the non-linear specification is used and when the 99.5th quantile in the linear regression is estimated. Also notice the (almost) monotonicity in which the effects are being reduced. This is encouraging from the errors-in-variables point of view. If the maximum of the distribution were the result of large errors-in-variables, there is no reason to expect that the estimates and the impact of the border effect could be similar to the upper percentiles.

[Figure 7 here]

The next step is to evaluate the significance of the border effect. Panel (a) in Figure 7 shows that the effect in kilometers comes down to be close to zero – even negative after some quantiles. To evaluate the significance of the estimates we compute the standard deviation of the relative increase in the price dispersion – rather than concentrating on the individual significance of each coefficient. The exercise we ask is the following: how large and significant is the implied degree of segmentation for a pit of stores separated by 10 kilometers across cities, relative to the degree of segmentation of a pair of stores separated by 10 kilometers within the same city. In other words, we compute the estimated segmentation for a distance of 0.1 with the city dummy equal to one, and then estimate the segmentation for the same distance and the city dummy at zero. All for stores that are not in the same chain. For example, for the estimates of the average, the price dispersion for $D_n = 100$ and $B_n = 1$ is as before $5.081 + 4.188 * 0.1 + 1.260 - 4.049 * 0.1 = 6.355$. The price dispersion when $B_n = 0$ is $5.081 + 4.188 * 0.1 = 5.499$. The border implies a 15.57 percent higher degree of segmentation. In Panel (b) in Figure 7 we present this relative increase in the degree of segmentation, together with its standard deviation, only for the linear specification. The figure shows the point estimate and the 95th percent confidence band.

These results show that the degree of segmentation is overestimated when the average price deviations are used, and that it becomes small and insignificant when the upper quantiles of the distribution within each bin are used. The change in the estimates is exclusively the outcome of running the quantile regressions.

The result that the degree of segmentation falls is not a spurious result of the methodology. The estimation using the upper quantiles should increase the absolute value of all coefficients – because all coefficients are affected by the sample selection problem. Our results, however, are the outcome of the bias being larger in one coefficient (distance) than in the other (border). Therefore, *ex-ante*, it is impossible to anticipate whether the border effect was going to increase or decrease.

4.1 Robustness

We run several robustness tests and present some of the results. In all our estimates we found the exact same message: the border effect becomes smaller and insignificant when the upper percentiles are used.

The first exercise is to eliminate products in which the matching across stores is not perfect. We eliminated meat, bread, among others. The results are presented in Table 2 and Figure 8. The exact same pattern as when using the full data is found.

[Table 2 and Figure 8 about here]

The second exercise uses all products but eliminates the outliers. We exclude all prices that are above three times or a third below the median price. This approach is more conservative that the one typically used in the literature. For example, Gopinath and Rigobon (2008) and Klenow and Kryvtsov (2008) eliminate prices that are more than 10 times higher or less that a tenth of the median price. In fact, we have just a few prices that are above three times or a third below the median daily price: 11.2 thousand in 32.8 million, or just 0.034%. The results are presented in Table 3 and Figure 9. Again, the patterns are almost identical to the ones from using the whole data set. The only difference is that the border effects at all percentiles gets closer to zero in absolute terms. In other words, in Panel (b) of Figure 9, the point estimates are smaller than those in Panel (b) in Figure 7. Other than this small effect, the estimates and patterns are identical.

[Table 3 and Figure 9 about here]

Third, just for completeness, we estimated eliminating the products in which the matchings are difficult and also the outliers. The results are presented in Table 4 and Figure 10 and there are no differences in our findings.

[Table 4 and Figure 10 about here]

We performed other robustness tests. We estimated everything using 50, 100 and 1000 bins. The advantage of larger number of bins is that each pair of stores is allocated to a very specific distance bin and the distance representing the bin is closer to the real distance across the stores. The disadvantage is that the number of observations within each bin decreases. In the limit, if the bins are so narrow that each store pair belongs to a single bin, then the problem is that the estimation of the 99.9 percentile becomes very noisy.[13]

[13]Future research should define the optimal bandwidth of these estimation procedure. For the moment we

5 Online Market Segmentation

In this section we use online and offline prices for a supermarket chain in Uruguay to estimate an "online border" effect that can help us validate our distance and border effect methodology.

The degree of segmentation between online and offline markets is an interesting topic by itself. Unfortunately, it has received little attention in the literature because the online and offline prices are hard to collect simultaneously.[14] Our goal in this paper is to use our new border methodology to measure the degree of online-offline segmentation and understand how much is driven by the real physical distance between the offline stores and the warehouse where the online orders are delivered from.

The online data was collected by the Billion Prices Project at MIT, using a method that scans the HyperText Markup Language (HTML) code of public retailer's website, identifying relevant price and product information to store in a database. HTML is a structured coding language that uses small pieces of code, called "tags", that can be used to automatically locate relevant pieces of information in the page.[15] We used this method to collect prices for all products sold online by this particular retailer in Uruguay, every day, between October 2007 and December 2010.

We matched each product id in the online and offline samples, and compared the daily prices across stores. Figure 11 provides an example of the prices posted in for a single product in all stores, including the online prices. On most dates, the online price is within the range of prices observed in offline stores. This pattern is typical for most goods in the sample.

---

compare the results across different specification and because the results are virtually identical we did not explore further. It is possible that if the estimation is done month by month, or in a much smaller data set, then the issue of the bandwidth becomes more important. In our application this was not the case.

[14]For example, using a small survey of 400 goods in four countries, Cavallo (2010) found that in some retailers prices are identical while in others online prices appear to have a stable markup over those in offline stores.

[15]For more details on the data scraping methodology, see Cavallo (2010).

[Figure 11 here]

The retailer lists a series of offline stores where the items sold online could be sent from, stating that the online prices are the same as those available at the offline store that fills the order at the time it is shipped. This means that, after controlling for this distance, there should be no additional price differences caused by the fact that a product is being bought online.

To get an idea of the real distance between online and offline stores, we first identified the most likely location where the online goods are shipped from. This was done by computing a "matching probability" between the online store prices and each of the offline stores. This is simply the average probability that the online and offline price are identical on a given day. We constructed this probability in two steps at the store level. First, for each product, we compute the share of days that the online price is identical to the offline price. Second, we take the mean (or median) across all products in that stores. Table 5 shows that online prices most closely resemble those of offline store number 22. The last column in the table shows the physical distance between store 22 and each of the other offline stores. Then, if we used the observed price dispersion to estimate an online border effect, the result should be close to the average distance to store 22, which is approximately 8 kilometers.

[Table 5 here]

To test the implied distance from the online store, we computed the online border effect using both the traditional and the quantile regression methods. Since we do not want to make any ex-ante assumptions on the distance, we do this in two steps. First, we estimated a simple regression for each quantile θ using only the offline prices for all stores in Montevideo.

$$Q\left(|p_{i,t} - p_{j,t}|_n, \theta\right) = \lambda + \beta D_n + \varepsilon_n \tag{8}$$

This is the equal to equation 6 with $B_n = 0$ (same city), $Firm_n = 1$ (same retailer), and $\lambda = \alpha + \gamma$. The coefficient β provides an estimate for the effect of distance on the dispersion of prices across stores, when only offline prices are used. Second, we then compute the average online-offline price dispersion (using all pairs of online-offline stores), subtract the constant λ and divide it by β to compute the "online border" effect. Results are shown in Table 6.

[Table 6 here]

Our main finding is that the traditional method appears to underestimate the online border, with an implied distance of just 1.6 kilometers. By contrast, if we use the 90th percentile we obtain an implied distance of 8.78 kilometers, very close to the average distance of 7.22 kilometers (and median of 8.04km) shown in Table 5.

Why does the traditional method underestimate the border effect in this case? It is, once again, coming from the fact that there is a bias in the distance coefficient in equation 8, which is different when we consider different subsets of stores. The online store, in particular, tends to have prices that are somewhere in between the prices of the offline stores, so that the observed price differences are smaller on average. This smaller within-sample dispersion increases the bias on the effect of distance, making the online border appear smaller when the standard method is used. The quantile method, by contrast, is not affected by how similar the online store is on average to the other stores. By focusing on the maximum difference within a quantile range, the quantile regression is providing a better estimate for the effect of distance on price dispersion.

6 Conclusions

The literature estimating the degree of segmentation introduced by political borders is a vast and important literature in international economics. The literature has continuously reported extremely large transaction costs introduced by country, province, and even city borders. In this paper we argue that some of those estimates have been overstated because the empirical approach has not taken into account the selection problem in posted prices: when a firm faces the possibility of arbitrage due to the existence of a transaction cost, the firm decides prices subject to a no-arbitrage constraint. If the optimal price falls into the no-arbitrage range, the difference in prices does not reflect the tightness of the constraint. This implies that the estimation using average absolute price differences or standard deviations of price differences do not capture the size of the trade or arbitrage cost.

This paper has two contributions. One is methodological, and the other is a contribution to the border effect literature. First, it offers an alternative methodology of estimation of transactions costs – which not only can be used in the estimation of transaction costs in international trade, but also can be used in other areas. For instance, in empirical finance and the measurement of liquidity, or the cost of regulatory restrictions. Second, we show that the political border matters little for price dispersion across cities. Although the border effect of a city should be small from an intuitive point of view, when the standard methodologies are used a very wide border is found (20 kilometers in a country where the largest city is less than 40 kilometers wide). By contrast, when our methodology is used, the border becomes insignificant. The logic applies to any border that is measured using the standard regressions in the literature. Of course, country borders could still remain wide even after our methodology is used, but they should be smaller than previously estimated.

Further research should advance in two dimensions. First, from the methodological point of view, it is important to determine procedures that define the optimal bandwidth. In our paper we used different bin sizes and because the results were consistent across all specifications we were not concerned with this issue. But other applications might need a different strategy. Second, a similar data needs to be collected across countries, and the same estimation should be performed to determine the actual width of international borders, and their determinants.

References

ANDERSON, J., AND E. VAN WINCOOP (2001): "Borders, Trade, and Welfare," *Brookings Trade Forum*.

——— (2003): "Gravity with Gravitas: A Solution to the Border Puzzle," *The American Economic Review*.

ANDREWS, D. W. K., S. BERRY, AND P. JIA (2004): "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location," *NBER*.

ANDREWS, D. W. K., AND P. GUGGENBERGER (2009): "Validity of Sampling and "Plug-in Asymptotic" Inference for parameters defined by moment inequalitites," *Economic Theory*, pp. 1–41.

ANDREWS, D. W. K., AND X. SHI (2010): "Inference Based on Conditional Moment Inequalities," *NBER*.

ANDREWS, D. W. K., AND G. SOARES (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78(1), 119–157.

BORRAZ, F., AND L. ZIPITRÍA (2012): "Retail Price Setting in Uruguay," *Economia*, 12(2), 77–109.

BRODA, C., AND D. E. WEINSTEIN (2008): "Understanding International Price Differences Using Barcode Data," *NBER*.

CAVALLO, A. (2010): "Scraped Data and Sticky Prices," *MIT Sloan Working Paper*.

CEGLOWSKI, J. (2003): "The Law of One Price: Intranational Evidence for Canada," *The Canadian Journal of Economics*.

CRUCINI, M. J., M. SHINTANI, AND T. TSURUGA (2010): “The Law of One Price without the Border: The Role of Distance versus Sticky Prices*,” *The Economic Journal*, 120(544), 462–480.

ENGEL, C., AND J. H. ROGERS (1996): “How wide is the border?,” *American Economic Review*, 86, 1112–1125.

——— (2004): “European Product Market Integration after the Euro,” *Economic Policy*.

ENGEL, C., J. H. ROGERS, AND S.-Y. B. WANG (2003): “Revisiting the Border: an assessment of the law of one price using very disaggregated consumer price data,” International Finance Discussion Papers 777, Board of Governors of the Federal Reserve System (U.S.).

EVANS, C. L. (2001): “The Economic Significance of National Border Effects,” *The American Economic Review*.

FRANKEL, J., E. STEIN, AND S.-J. WEI (1995): “Trading Blocs and the Americas: The Natural, the Unnatural and the Super-natural.,” *Journal of Development Economics*.

GOLDBERG, P. K., AND M. M. KNETTER (1997): “Goods Prices and Exchange Rates: What Have We Learned?,” *Journal of Economic Literature*.

GOPINATH, G., P.-O. GOURINCHAS, C.-T. HSIEH, AND N. LI (2011): “International Prices, Costs, and Markup Differences,” *American Economic Review*, 101(6), 2450–86.

GOPINATH, G., AND R. RIGOBON (2008): “Sticky Borders,” *The Quarterly Journal of Economics*, 123(2), 531–575.

GORODNICHENKO, Y., AND L. TESAR (2009): “Border Effect or Country Effect? Seattle may not be so far from Vancouver after all.,” *American Economic Journal – Macroeconomics*, 1, 219–241.

HEAD, K., AND T. MAYER (2002): “Illusory Border Effects: Distance mismeasurement inflates estimates of home bias in trade,” *CEPII research center, Working Paper 2002-01*.

HELLIWELL, J. F. (1997): “National Borders, trade and immigration,” *Pacific Economic Review*, 3, 165–185.

HELLIWELL, J. F., AND L. L. SCHEMBRI (2005): “Borders, Common Currencies, Trade, and Welfare: What Can We Learn from the Evidence?,” *Bank of Canada Review*.

HELLIWELL, J. F., AND G. VERDIER (2001): “Measuring Internal Trade Distances: A New Method Applied to Estimate Provincial Border Effects in Canada,” *The Canadian Journal of Economics*.

HILLBERRY, R., AND D. HUMMELS (2003): “Intranational Home Bias: Some Explanations,” *The Review of Economics and Statistics*.

KLENOW, P. J., AND O. KRYVTSOV (2008): “State-Dependent or Time-Dependent Pricing: Does It Matter for Recent U.S. Inflation?,” *The Quarterly Journal of Economics*, 123(3), 863–904.

NITSCH, V. (2000): “National Borders and International Trade: Evidence from the European Union,” *The Canadian Journal of Economics*, 33(4), 1091–1105.

PARSLEY, D. C., AND S.-J. WEI (2001): “Explaining The Border Effect: The Role Of Exchange Rae Variability, Shipping Costs, And Geography,” *Journal of International Economics*, 55, 87–105.

PONOMAREVA, M., AND E. TAMER (2011): “Misspecification in moment inequality models: back to moment equalities?,” *The Econometrics Journal*, 14(2), 186–203.

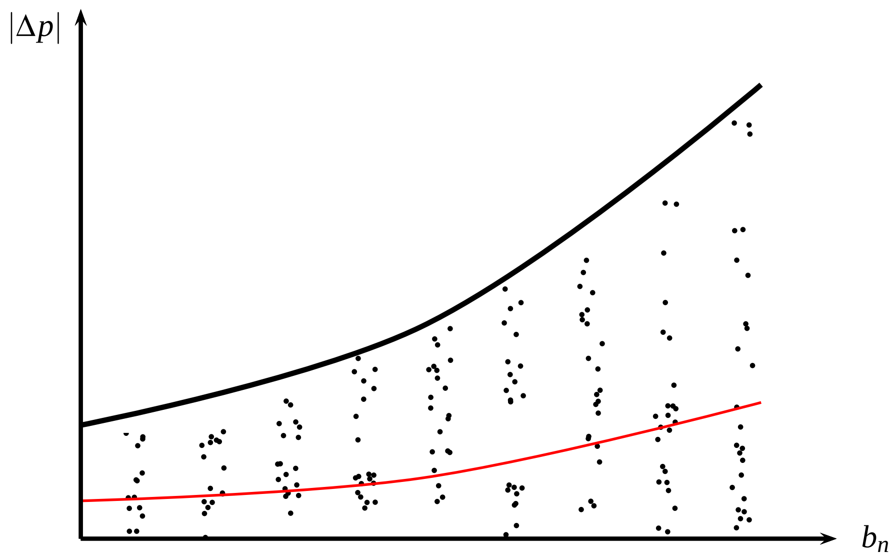
ROSEN, A. M. (2008): “Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities,” *Journal of Econometrics*, 146(1), 107–117.

SAMUELSON, P. A. (1954): “The Transfer Problem and Transport Costs, II: Analysis of Effects of Trade Impediments,” *The Economic Journal*, 64(254), pp. 264–289.

WOLF, H. (2000): “Intranational Home Bias in Trade,” *Review of Economics and Statistics*, 82(4), 555–563.

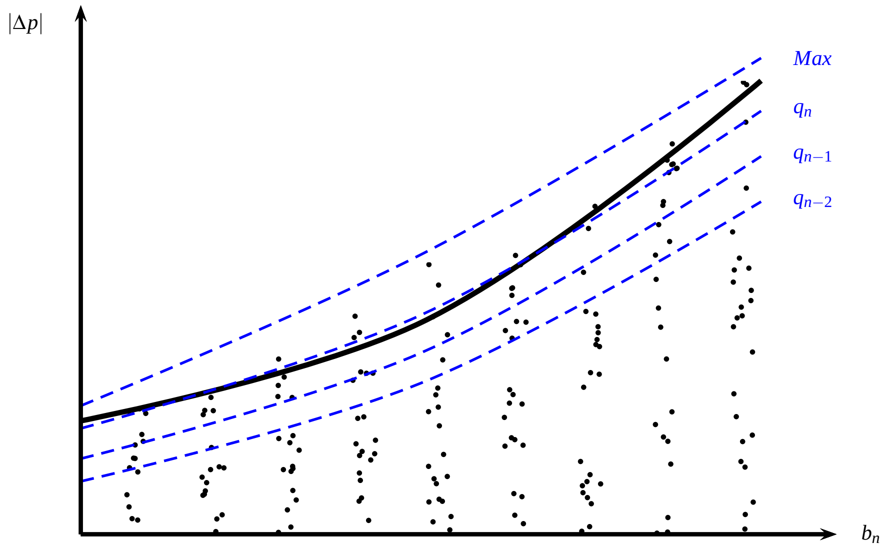
7 Figures and Tables

Figure 1: Bias in Standard Regressions



Note: This figure illustrates the source of the selection bias. The horizontal axis shows the bins for a range of distances. The vertical axis is the absolute price difference across locations. For each bin, all the absolute differences from the data are shown as the black dots. The thick black line reflects the price difference implied by the no-arbitrage constraint. Because all the observed price differences are less or equal to the thick line, the estimation in the standard regression which implicitly uses the mean within each bin (thin red line) is downward biased .

Figure 2: Bias in Standard Regression in the presence of EIV

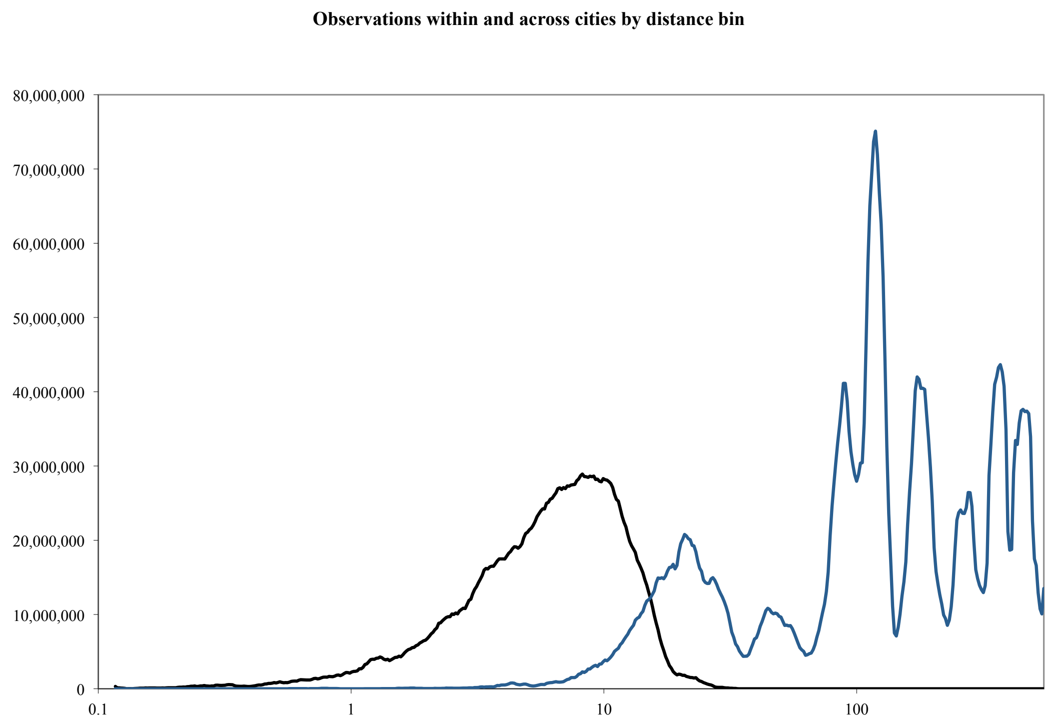


Note: The black thick line is still the true upper bound of the no-arbitrage band. This is the true degree of segmentation. Notice that now, because of EIV, some price differences might even be above the no-arbitrage range. In this case, using the maximum within each bin also produces a bias in the estimation. For this reason we use a series of quantile regressions instead.

Figure 3: Cities covered in the sample

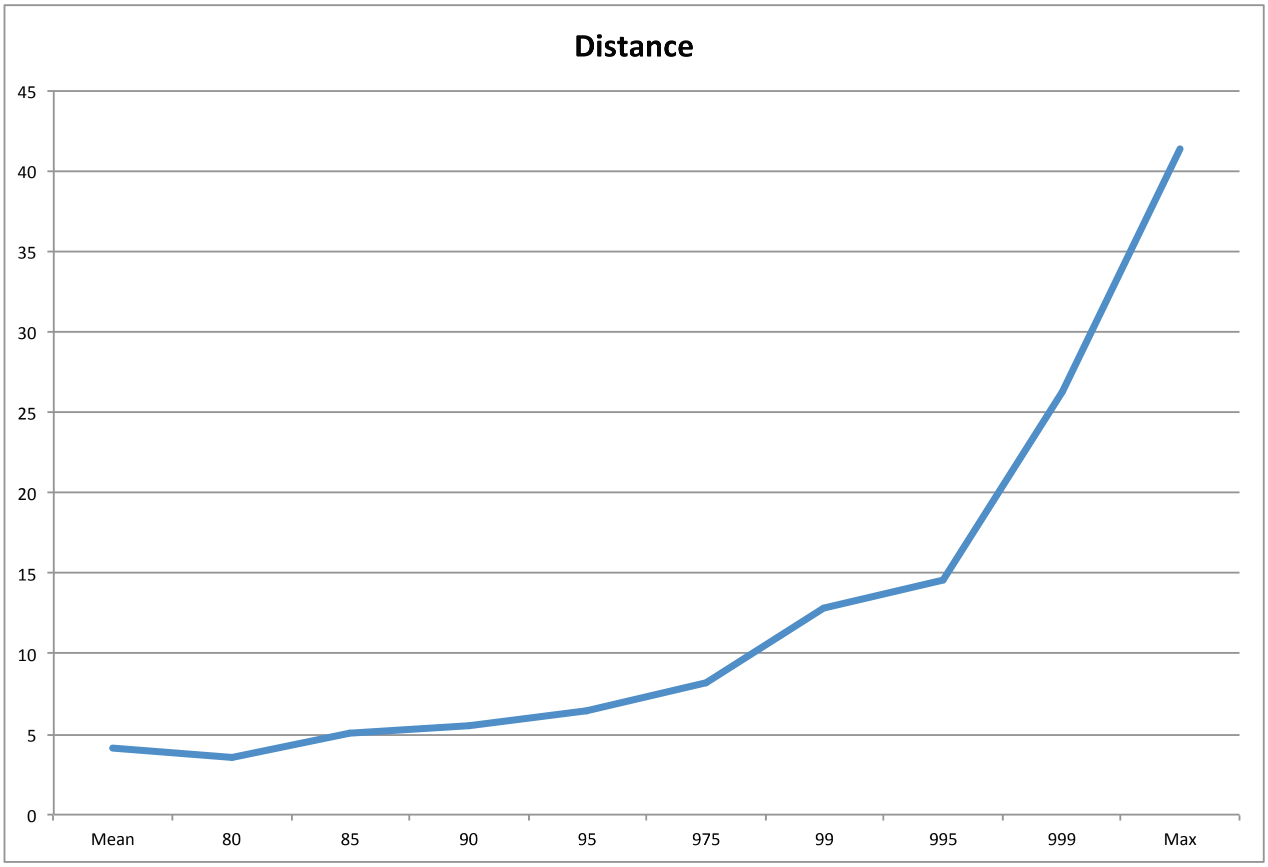
Note: Each dot represents a store location.

Figure 4: Distribution of observations for 500 bins in the same city and between cities



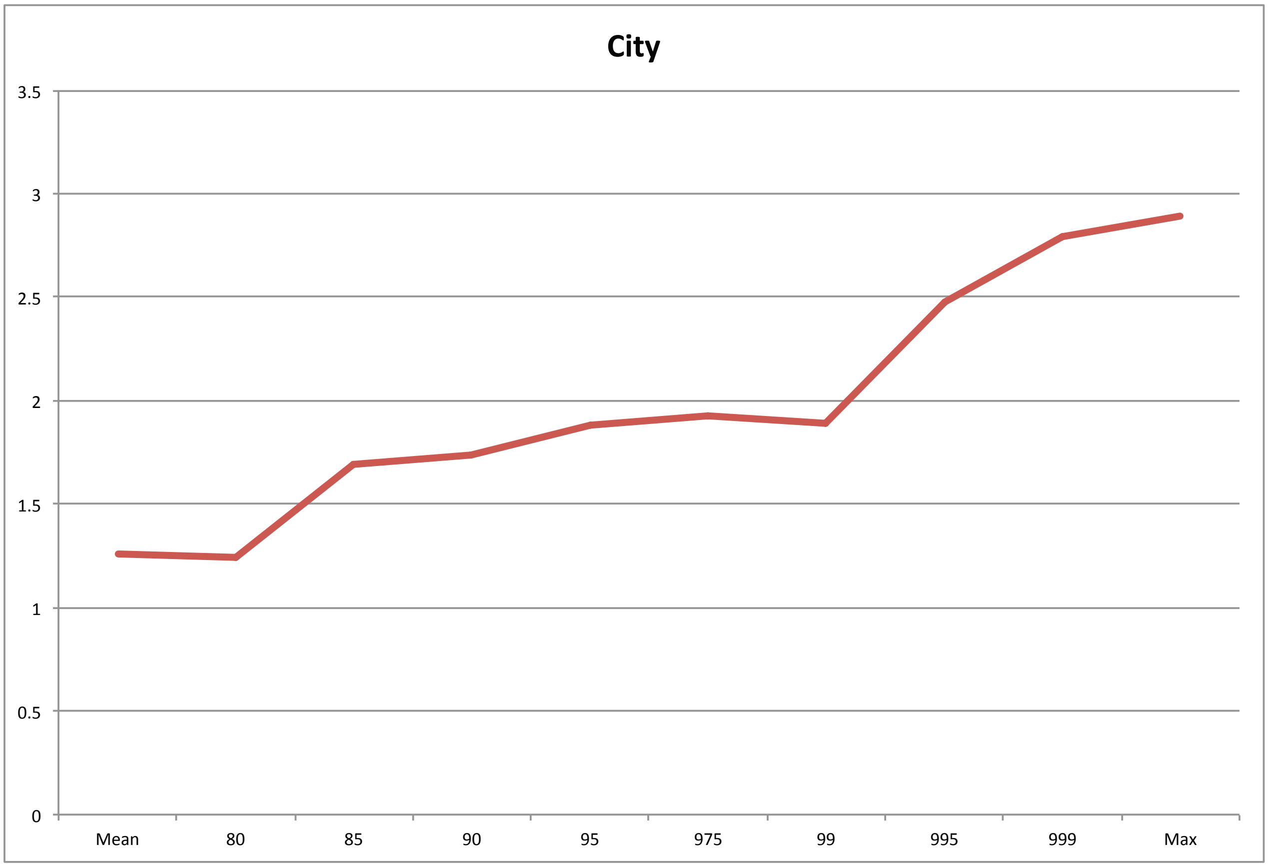
Note: The black line shows the distribution of bilateral observations within cities, while the blue line (extending to the right, with multiple peaks) shows the distribution across cities.

Figure 5: Estimation of coefficient of distance by quantile.



Note: This is the estimated distance coefficient when different quantiles are used for the baseline regression.

Figure 6: Estimation of coefficient of city by quantile.

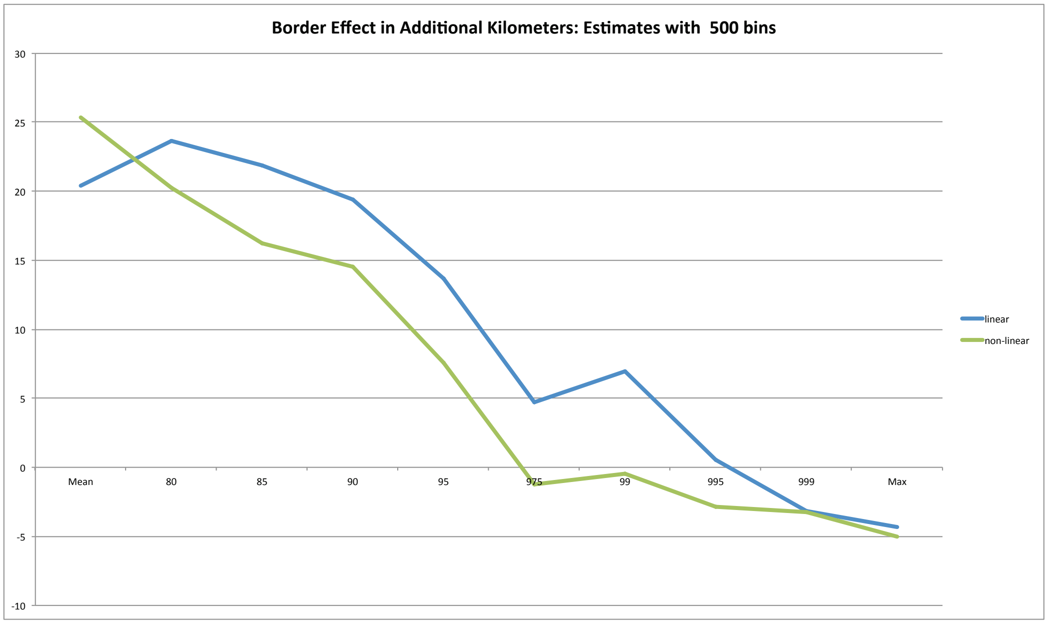


Note: This is the estimated city dummy coefficient when different quantiles are used for the baseline regression.

Figure 7: Estimation of city border effect. All data. 500 bins

(a) Implied Kilometers

Additional Km implied by City Border Effect for Stores 10 Km Apart



(b) Relative Increase in Price Dispersion

of City Borders for Stores 10 Km Apart.

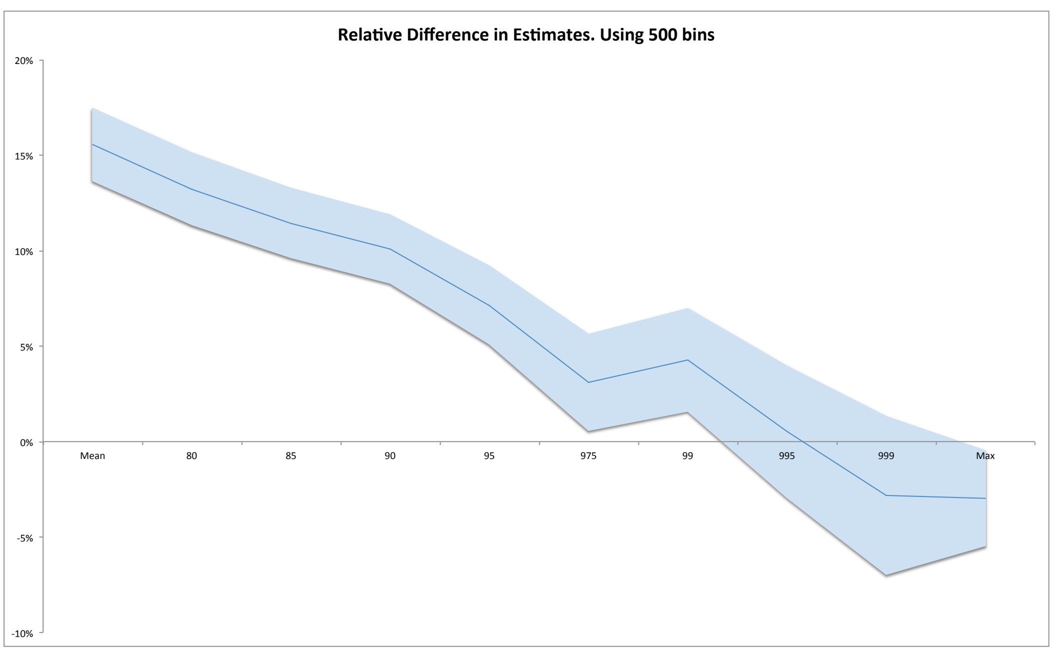
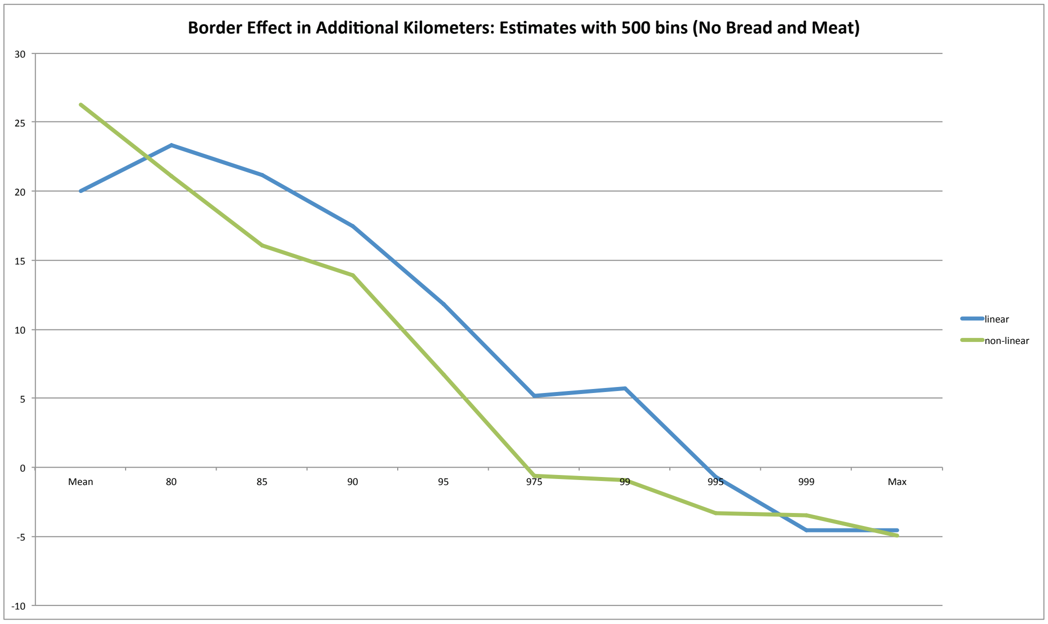


Figure 8: Estimation of city border effect. Excluding Meat and Bread. 500 bins

(a) Implied Kilometers

Additional Km implied by City Border Effect for Stores 10 Km Apart



(b) Relative Increase in Price Dispersion

of City Borders for Stores 10 Km Apart.

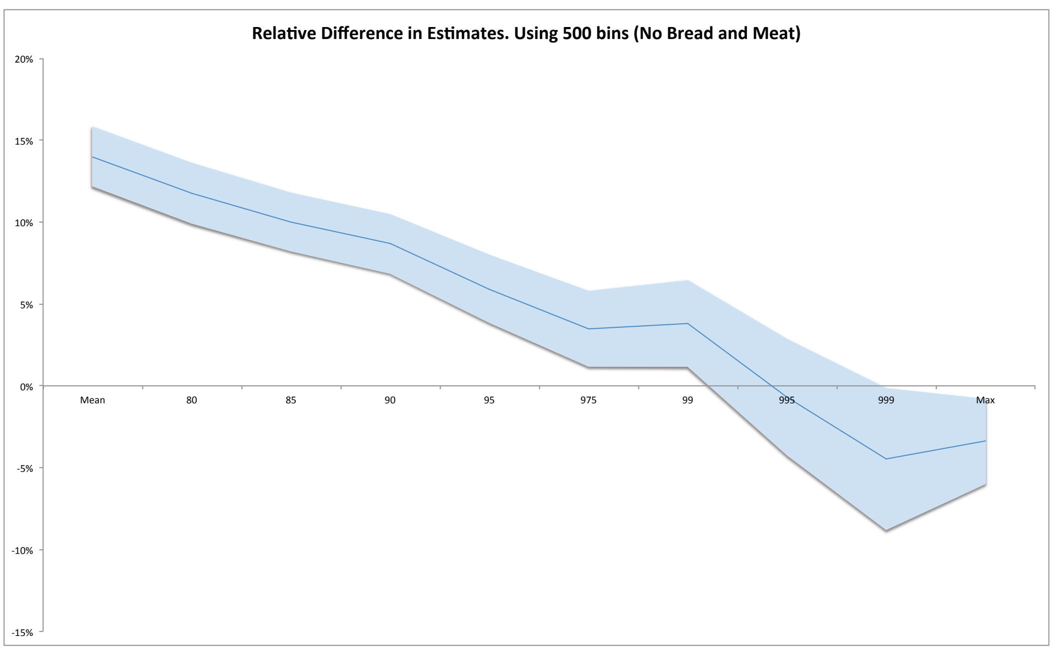
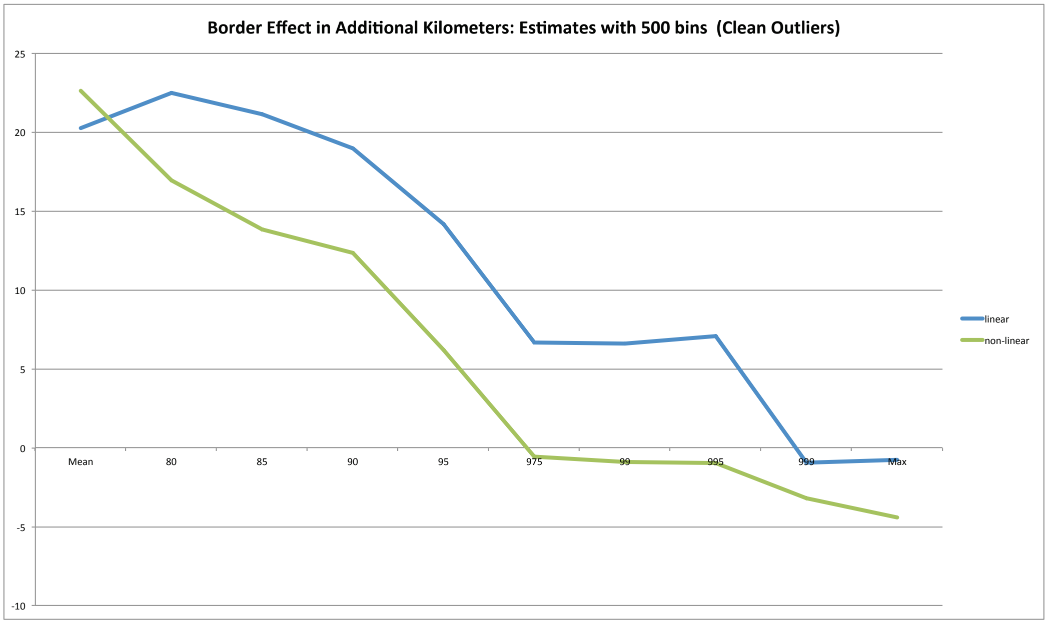


Figure 9: Estimation of city border effect. All data. Excluding Outliers. 500 bins

(a) Implied Kilometers

Additional Km implied by City Border Effect for Stores 10 Km Apart



(b) Relative Increase in Price Dispersion

of City Borders for Stores 10 Km Apart.

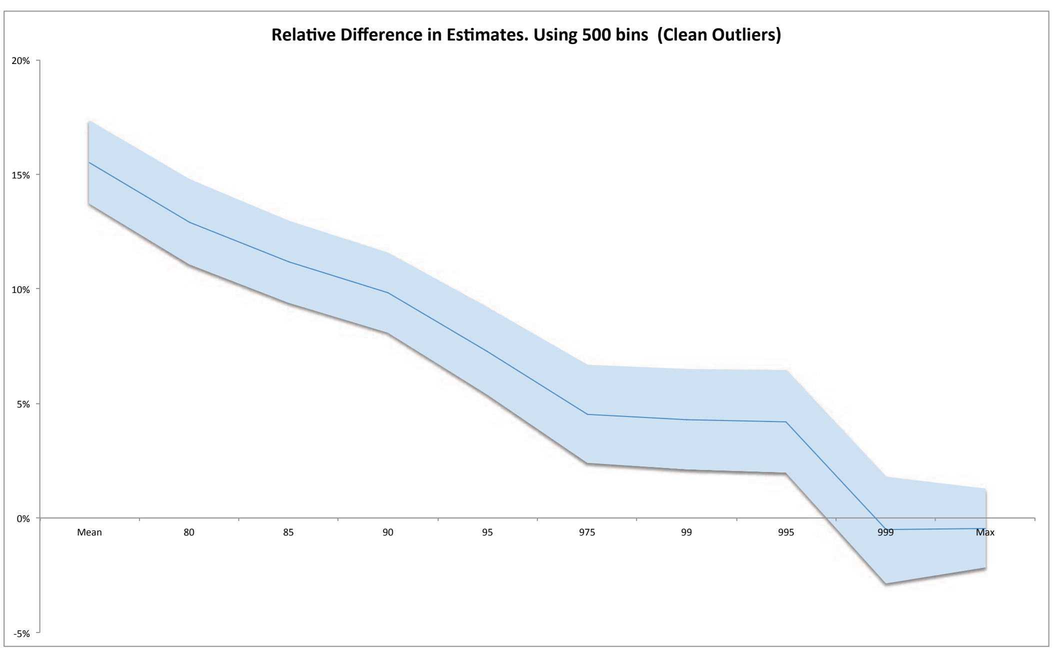
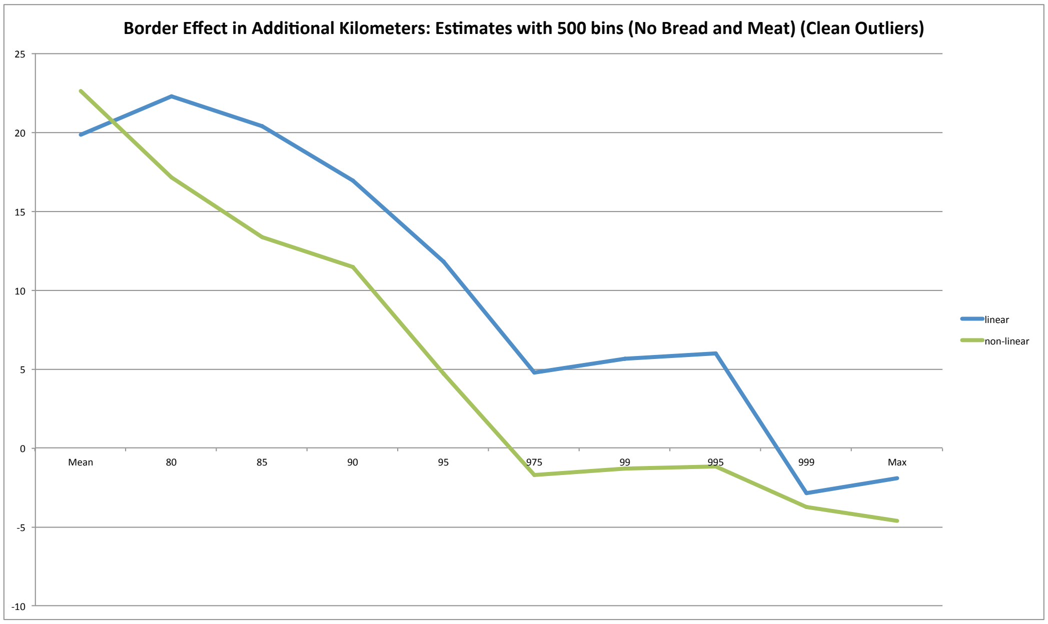


Figure 10: Estimation of city border effect. Excluding Mean and Bread. Excluding Outliers. 500 bins

(a) Implied Kilometers

Additional Km implied by City Border Effect for Stores 10 Km Apart



(b) Relative Increase in Price Dispersion

of City Borders for Stores 10 Km Apart.

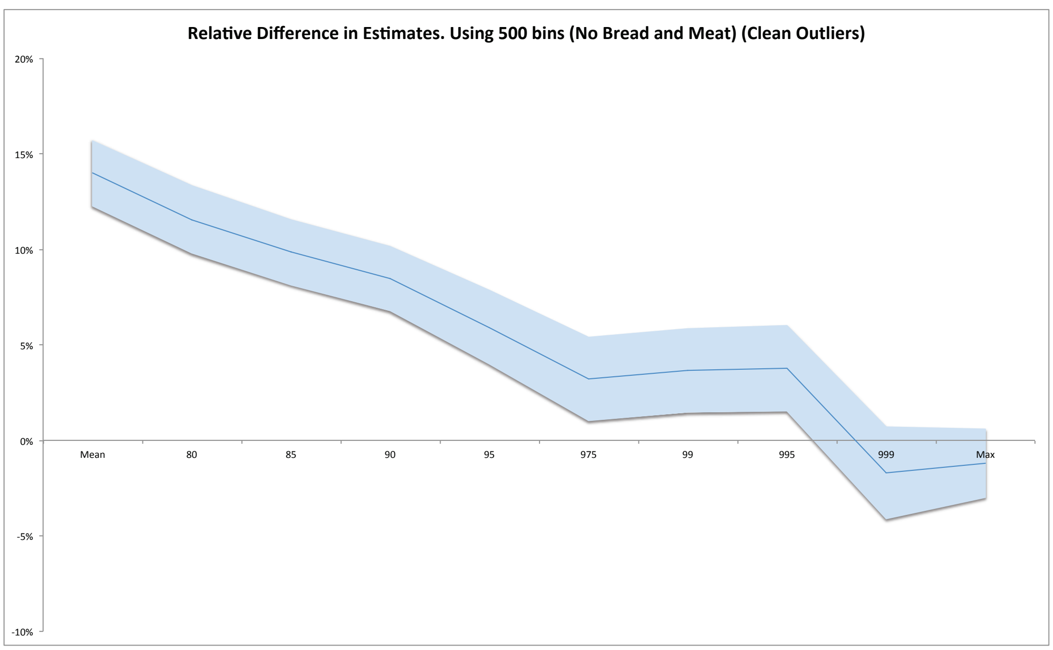
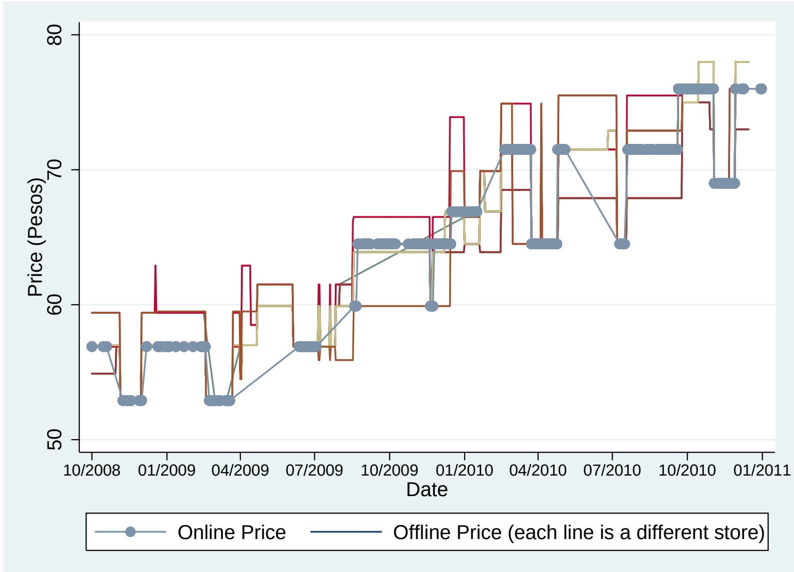


Figure 11: Example of Online and Offline Prices: Cocoa - 0.5Kg



Note: This is an example of the typical time series pattern of online prices compared to offline prices in the same city of Montevideo. Each line is a different store. The online price is marked with a dotted line, and tends to lie in-between the prices of the offline stores.

Table 1: Weighted Least Square Estimation of Price Differential using the whole database and for 500 bins.

Linear Specification

	Average	50	80	85	90	95	97.5	99	99.5	99.9	Max
Distance	4.188***	3.528***	5.073***	5.504***	6.428***	8.153***	12.822***	14.618***	26.287***	41.329***	95.596***
	(0.186)	(0.202)	(0.297)	(0.323)	(0.368)	(0.510)	(0.733)	(0.967)	(1.414)	(2.601)	(4.677)
City	1.260***	1.243***	1.691***	1.738***	1.880***	1.926***	1.890***	2.478***	2.794***	2.889***	5.105***
	(0.017)	(0.018)	(0.026)	(0.029)	(0.033)	(0.045)	(0.065)	(0.086)	(0.126)	(0.232)	(0.417)
City*Dist	-4.049***	-3.350***	-4.930***	-5.364***	-6.323***	-8.083***	-12.880***	-14.670***	-26.460***	-41.833***	-92.579***
	(0.186)	(0.202)	(0.297)	(0.323)	(0.368)	(0.510)	(0.733)	(0.967)	(1.414)	(2.602)	(4.678)
Chain	-6.012***	-5.196***	-9.652***	-10.738***	-12.101***	-14.642***	-18.086***	-22.188***	-25.305***	-38.565***	-68.955***
	(0.020)	(0.022)	(0.032)	(0.035)	(0.040)	(0.056)	(0.080)	(0.106)	(0.155)	(0.285)	(0.508)
Const	5.081***	3.782***	8.541***	9.956***	11.745***	14.832***	18.150***	22.106***	25.807***	41.789***	130.155***
	(0.038)	(0.041)	(0.061)	(0.066)	(0.075)	(0.104)	(0.150)	(0.198)	(0.289)	(0.532)	(0.975)
N	179215	179215	179215	179215	179215	179215	179215	179215	179215	179215	184328
R2	0.752	0.645	0.749	0.761	0.762	0.725	0.744	0.766	0.68	0.58	0.492

Non Linear Specification

	Average	50	80	85	90	95	97.5	99	99.5	99.9	Max
Distance	-1.396**	-4.359***	1.854**	3.213***	3.132***	4.005***	24.295***	43.301***	94.219***	221.741***	946.493***
	(0.542)	(0.589)	(0.870)	(0.947)	(1.081)	(1.498)	(2.153)	(2.840)	(4.152)	(7.636)	(13.542)
City	0.610***	0.486***	1.042***	1.091***	1.132***	0.982***	1.355***	2.276***	3.575***	7.527***	18.414***
	(0.025)	(0.028)	(0.041)	(0.044)	(0.051)	(0.070)	(0.101)	(0.133)	(0.195)	(0.358)	(0.640)
City*Dist	2.467***	5.559***	-0.669	-1.974**	-1.780*	-2.262	-22.579***	-41.457***	-92.334***	-221.689***	-919.021***
	(0.543)	(0.590)	(0.872)	(0.948)	(1.082)	(1.500)	(2.156)	(2.844)	(4.158)	(7.647)	(13.564)
Chain	-6.001***	-5.186***	-9.637***	-10.721***	-12.083***	-14.618***	-18.044***	-22.122***	-25.194***	-38.355***	-67.614***
	(0.020)	(0.022)	(0.032)	(0.035)	(0.040)	(0.056)	(0.080)	(0.106)	(0.155)	(0.284)	(0.500)
$Dist^2$	34.149***	48.223***	19.697***	14.030***	20.175***	25.390***	-70.068***	-175.224***	-415.058***	-1102.431***	-5203.545***
	(3.117)	(3.387)	(5.002)	(5.440)	(6.210)	(8.607)	(12.373)	(16.321)	(23.862)	(43.883)	(77.900)
$Dist^3$	0.026***	0.031***	0.025***	0.028***	0.034***	0.061***	0.062***	0.032***	0.008	-0.112***	0.870***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)	(0.005)	(0.008)	(0.010)	(0.015)	(0.027)	(0.060)
City*$Dist^2$	-34.466***	-48.581***	-20.037***	-14.394***	-20.597***	-26.020***	69.408***	174.667***	414.569***	1102.788***	5194.370***
	(3.117)	(3.387)	(5.002)	(5.440)	(6.210)	(8.607)	(12.373)	(16.321)	(23.862)	(43.883)	(77.901)
Const	5.241***	4.010***	8.630***	10.017***	11.835***	14.946***	17.800***	21.239***	23.768***	36.406***	104.760***
	(0.041)	(0.044)	(0.065)	(0.071)	(0.081)	(0.112)	(0.161)	(0.213)	(0.311)	(0.572)	(1.031)
N	179215	179215	179215	179215	179215	179215	179215	179215	179215	179215	184328
R2	0.755	0.649	0.751	0.763	0.763	0.726	0.745	0.767	0.681	0.581	0.509

Note: * significant at 10%; ** significant at 5%; *** significant at 1%

Table 2: Weighted Least Square Estimation of Price Differential excluding Meat and Bread and for 500 bins.

Linear Specification

	Average	50	80	85	90	95	97.5	99	99.5	99.9	Max
Distance	3.892***	3.118***	4.602***	5.029***	6.207***	7.902***	13.154***	15.652***	28.285***	45.346***	104.739***
	(0.182)	(0.199)	(0.291)	(0.317)	(0.365)	(0.513)	(0.684)	(0.968)	(1.466)	(2.724)	(4.906)
City	1.156***	1.136***	1.523***	1.554***	1.695***	1.718***	1.999***	2.472***	2.650***	2.513***	5.423***
	(0.016)	(0.018)	(0.026)	(0.028)	(0.033)	(0.046)	(0.061)	(0.086)	(0.131)	(0.242)	(0.437)
City*Dist	-3.773***	-2.966***	-4.484***	-4.913***	-6.119***	-7.846***	-13.222***	-15.725***	-28.474***	-45.858***	-101.768***
	(0.182)	(0.199)	(0.291)	(0.317)	(0.366)	(0.513)	(0.684)	(0.968)	(1.466)	(2.724)	(4.907)
Chain	-5.831***	-5.020***	-9.383***	-10.459***	-11.807***	-14.329***	-17.488***	-21.293***	-24.357***	-37.604***	-68.527***
	(0.020)	(0.022)	(0.032)	(0.035)	(0.040)	(0.056)	(0.075)	(0.106)	(0.161)	(0.298)	(0.534)
Const	5.170***	3.884***	8.680***	10.100***	11.878***	14.975***	18.057***	22.072***	25.835***	41.914***	129.772***
	(0.036)	(0.039)	(0.058)	(0.063)	(0.073)	(0.102)	(0.136)	(0.192)	(0.291)	(0.541)	(0.995)
N	159566	159566	159566	159566	159566	159566	159566	159566	159566	159566	165297
R2	0.705	0.593	0.712	0.726	0.724	0.684	0.71	0.679	0.576	0.523	0.479

Non Linear Specification

	Average	50	80	85	90	95	97.5	99	99.5	99.9	Max
Distance	-1.721***	-4.583***	0.963	2.260**	1.818*	1.948	19.575***	38.915***	91.935***	220.458***	960.037***
	(0.534)	(0.582)	(0.853)	(0.931)	(1.073)	(1.508)	(2.011)	(2.845)	(4.309)	(8.002)	(14.213)
City	0.521***	0.389***	0.878***	0.912***	0.949***	0.775***	1.299***	2.045***	3.363***	7.466***	19.508***
	(0.025)	(0.027)	(0.040)	(0.044)	(0.050)	(0.071)	(0.094)	(0.133)	(0.202)	(0.375)	(0.672)
City*Dist	2.776***	5.783***	0.214	-1.029	-0.501	-0.27	-17.751***	-36.835***	-90.069***	-221.470***	-934.595***
	(0.534)	(0.583)	(0.855)	(0.932)	(1.075)	(1.510)	(2.013)	(2.849)	(4.315)	(8.013)	(14.236)
Chain	-5.821***	-5.010***	-9.369***	-10.443***	-11.791***	-14.309***	-17.450***	-21.232***	-24.252***	-37.419***	-67.215***
	(0.020)	(0.022)	(0.032)	(0.035)	(0.040)	(0.056)	(0.075)	(0.106)	(0.160)	(0.298)	(0.525)
$Dist^2$	34.381***	47.169***	22.303***	16.981***	26.898***	36.483***	-39.254***	-142.333***	-389.547***	-1071.870***	-5238.861***
	(3.072)	(3.351)	(4.913)	(5.358)	(6.178)	(8.679)	(11.575)	(16.379)	(24.805)	(46.066)	(81.902)
$Dist^3$	0.031***	0.037***	0.034***	0.037***	0.041***	0.068***	0.080***	0.061***	0.027*	-0.159***	0.725***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)	(0.005)	(0.007)	(0.010)	(0.015)	(0.029)	(0.063)
City*$Dist^2$	-34.723***	-47.560***	-22.682***	-17.386***	-27.347***	-37.132***	38.493***	141.596***	388.977***	1072.661***	5230.745***
	(3.072)	(3.351)	(4.913)	(5.358)	(6.178)	(8.679)	(11.575)	(16.379)	(24.805)	(46.066)	(81.903)
Const	5.332***	4.107***	8.782***	10.177***	12.001***	15.144***	17.858***	21.370***	23.931***	36.707***	104.313***
	(0.039)	(0.042)	(0.062)	(0.068)	(0.078)	(0.110)	(0.147)	(0.208)	(0.314)	(0.584)	(1.055)
N	159566	159566	159566	159566	159566	159566	159566	159566	159566	159566	165297
R2	0.708	0.599	0.714	0.728	0.726	0.685	0.711	0.68	0.577	0.525	0.497

Note: * significant at 10%; ** significant at 5%; *** significant at 1%

Table 3: Weighted Least Square Estimation of Price Differential excluding Outliers and for 500 bins.

Linear Specification

	Average	50	80	85	90	95	97.5	99	99.5	99.9	Max
Distance	4.188***	3.684***	5.205***	5.564***	6.442***	8.025***	13.066***	15.412***	16.448***	21.987***	43.223***
	-0.173	-0.189	-0.285	-0.31	-0.35	-0.469	-0.617	-0.775	-0.925	-1.313	(1.719)
City	1.252***	1.242***	1.674***	1.715***	1.851***	1.933***	2.178***	2.554***	2.814***	2.036***	3.885***
	-0.015	-0.017	-0.025	-0.028	-0.031	-0.042	-0.055	-0.069	-0.083	-0.117	(0.153)
City*Dist	-4.031***	-3.489***	-5.040***	-5.404***	-6.316***	-7.945***	-13.075***	-15.427***	-16.567***	-22.461***	-42.125***
	-0.173	-0.189	-0.285	-0.31	-0.35	-0.469	-0.617	-0.775	-0.925	-1.313	(1.720)
Chain	-6.114***	-5.327***	-9.843***	-10.971***	-12.380***	-14.943***	-18.049***	-21.893***	-24.775***	-34.202***	-48.398***
	-0.02	-0.021	-0.032	-0.035	-0.039	-0.053	-0.069	-0.087	-0.104	-0.148	(0.187)
Const	5.046***	3.773***	8.537***	9.959***	11.755***	14.834***	17.891***	21.944***	25.812***	38.182***	67.798***
	-0.036	-0.039	-0.059	-0.064	-0.073	-0.097	-0.128	-0.161	-0.192	-0.273	(0.358)
N	183341	183341	183341	183341	183341	183341	183341	183341	183341	183341	184277
R2	0.752	0.657	0.749	0.761	0.764	0.733	0.736	0.749	0.735	0.687	0.645

Non Linear Specification

	Average	50	80	85	90	95	97.5	99	99.5	99.9	Max
Distance	-1.220**	-3.840***	2.562***	3.890***	3.888***	5.123***	23.404***	44.674***	67.140***	190.276***	496.586***
	-0.506	-0.553	-0.835	-0.909	-1.027	-1.376	-1.812	-2.276	-2.716	-3.838	(4.885)
City	0.544***	0.439***	0.973***	1.014***	1.045***	0.886***	1.347***	2.148***	2.999***	5.676***	10.710***
	-0.024	-0.026	-0.039	-0.043	-0.049	-0.065	-0.086	-0.108	-0.128	-0.181	(0.231)
City*Dist	2.632***	5.349***	-0.989	-2.257**	-2.080**	-2.727**	-20.608***	-41.664***	-64.293***	-187.771***	-481.214***
	-0.507	-0.554	-0.836	-0.91	-1.029	-1.378	-1.815	-2.28	-2.72	-3.845	(4.893)
Chain	-6.099***	-5.312***	-9.823***	-10.949***	-12.355***	-14.911***	-18.000***	-21.820***	-24.680***	-33.997***	-47.680***
	-0.019	-0.021	-0.032	-0.035	-0.039	-0.053	-0.069	-0.087	-0.104	-0.147	(0.180)
$Dist^2$	33.048***	45.971***	16.168***	10.248**	15.630***	17.760**	-63.107***	-178.667***	-309.537***	-1027.722***	-2772.443***
	-2.908	-3.175	-4.797	-5.221	-5.9	-7.905	-10.406	-13.076	-15.599	-22.048	(28.102)
$Dist^3$	0.067***	0.068***	0.072***	0.075***	0.090***	0.139***	0.176***	0.174***	0.147***	0.134***	0.619***
	-0.002	-0.002	-0.004	-0.004	-0.005	-0.006	-0.008	-0.01	-0.012	-0.017	(0.022)
City*$Dist^2$	-33.610***	-46.552***	-16.786***	-10.896**	-16.387***	-18.860**	61.740***	177.259***	308.252***	1026.488***	2766.629***
	-2.908	-3.175	-4.797	-5.221	-5.9	-7.905	-10.406	-13.076	-15.599	-22.048	(28.102)
Const	5.202***	3.992***	8.610***	10.003***	11.825***	14.912***	17.574***	21.062***	24.291***	33.159***	54.267***
	-0.038	-0.042	-0.063	-0.069	-0.078	-0.104	-0.138	-0.173	-0.206	-0.291	(0.372)
N	183341	183341	183341	183341	183341	183341	183341	183341	183341	183341	184277
R2	0.755	0.661	0.751	0.762	0.766	0.734	0.737	0.75	0.736	0.691	0.669

Note: * significant at 10%; ** significant at 5%; *** significant at 1%

Table 4: Weighted Least Square Estimation of Price Differential excluding Meat and Bread, excluding Outliers, and for 500 bins.

Linear Specification

	Average	50	80	85	90	95	97.5	99	99.5	99.9	Max
Distance	3.892***	3.291***	4.753***	5.128***	6.249***	7.930***	13.006***	15.211***	17.272***	24.002***	45.887***
	-0.169	-0.185	-0.279	-0.305	-0.348	-0.478	-0.637	-0.789	-0.939	-1.374	(1.820)
City	1.149***	1.143***	1.520***	1.545***	1.674***	1.721***	1.923***	2.382***	2.772***	1.754***	3.603***
	-0.015	-0.016	-0.025	-0.027	-0.031	-0.043	-0.057	-0.07	-0.084	-0.123	(0.162)
City*Dist	-3.760***	-3.128***	-4.621***	-5.000***	-6.149***	-7.872***	-13.035***	-15.237***	-17.375***	-24.488***	-44.766***
	-0.169	-0.185	-0.279	-0.305	-0.348	-0.478	-0.638	-0.79	-0.939	-1.374	(1.821)
Chain	-5.954***	-5.151***	-9.579***	-10.700***	-12.104***	-14.697***	-17.884***	-21.690***	-24.328***	-33.595***	-48.239***
	-0.019	-0.021	-0.031	-0.034	-0.039	-0.054	-0.072	-0.089	-0.106	-0.155	(0.198)
Const	5.140***	3.873***	8.674***	10.103***	11.893***	14.988***	18.073***	22.063***	25.784***	38.307***	67.890***
	-0.034	-0.037	-0.056	-0.061	-0.07	-0.097	-0.129	-0.159	-0.19	-0.277	(0.369)
N	163729	163729	163729	163729	163729	163729	163729	163729	163729	163729	165257
R2	0.718	0.611	0.715	0.728	0.731	0.698	0.718	0.74	0.714	0.65	0.62

Non Linear Specification

	Average	50	80	85	90	95	97.5	99	99.5	99.9	Max
Distance	-1.484***	-4.044***	1.702**	2.972***	2.602**	3.160**	21.621***	41.226***	61.861***	189.136***	503.975***
	-0.496	-0.541	-0.818	-0.895	-1.021	-1.406	-1.874	-2.32	-2.759	-4.02	(5.173)
City	0.458***	0.352***	0.824***	0.852***	0.872***	0.677***	1.103***	1.907***	2.762***	5.301***	10.530***
	-0.023	-0.026	-0.039	-0.042	-0.048	-0.066	-0.089	-0.11	-0.13	-0.19	(0.244)
City*Dist	2.850***	5.523***	-0.18	-1.398	-0.876	-0.885	-18.941***	-38.225***	-58.906***	-186.650***	-488.696***
	-0.496	-0.542	-0.819	-0.896	-1.023	-1.408	-1.877	-2.324	-2.763	-4.026	(5.181)
Chain	-5.940***	-5.137***	-9.560***	-10.679***	-12.083***	-14.670***	-17.840***	-21.623***	-24.239***	-33.397***	-47.521***
	-0.019	-0.021	-0.031	-0.034	-0.039	-0.054	-0.072	-0.089	-0.106	-0.154	(0.191)
$Dist^2$	32.907***	44.890***	18.684***	13.212**	22.336***	29.204***	-52.666***	-159.090***	-272.702***	-1010.075***	-2805.846***
	-2.851	-3.113	-4.705	-5.148	-5.876	-8.088	-10.782	-13.35	-15.875	-23.13	(29.808)
$Dist^3$	0.068***	0.071***	0.074***	0.078***	0.092***	0.139***	0.178***	0.184***	0.163***	0.137***	0.600***
	-0.002	-0.002	-0.004	-0.004	-0.005	-0.006	-0.008	-0.01	-0.012	-0.018	(0.023)
City*$Dist^2$	-33.471***	-45.482***	-19.309***	-13.863***	-23.085***	-30.284***	51.313***	157.639***	271.333***	1008.830***	2800.138***
	-2.852	-3.113	-4.705	-5.148	-5.876	-8.088	-10.782	-13.35	-15.875	-23.13	(29.808)
Const	5.296***	4.086***	8.760***	10.162***	11.996***	15.122***	17.809***	21.281***	24.448***	33.390***	54.252***
	-0.037	-0.04	-0.061	-0.066	-0.076	-0.104	-0.139	-0.172	-0.204	-0.298	(0.384)
N	163729	163729	163729	163729	163729	163729	163729	163729	163729	163729	165257
R2	0.721	0.616	0.716	0.73	0.733	0.699	0.719	0.741	0.715	0.655	0.646

Note: * significant at 10%; ** significant at 5%; *** significant at 1%

Table 5: Online vs Offline stores

Store	City	Online Match Probability	Distance to Store 22
22	Montevideo	97.34	0.00
31	Montevideo	96.59	1.28
39	Montevideo	96.59	1.88
41	Montevideo	96.83	2.32
21	Montevideo	96.83	2.72
38	Montevideo	96.58	3.32
33	Montevideo	81.85	5.66
34	Montevideo	96.96	6.50
35	Montevideo	96.70	8.04
32	Montevideo	81.702	8.84
43	Montevideo	81.18	8.96
28	Montevideo	81.68	9.23
30	Montevideo	96.54	10.58
27	Montevideo	81.73	11.81
23	Montevideo	81.57	12.87
36	Montevideo	81.56	13.29
42	Montevideo	81.37	15.42
	Mean	89.62	7.22
	Median	96.54	8.04

Note: The ”Online Match Probability” shows the percentage of days in which the online price is identical to the price observed offline in a particular store. Distance from store 22 to the other offline stores is measured in kilometers.

Table 6: The Online-Offline Border

Percentile	Mean	95th
Differences Online-Offline (%)	0.60	4.55
Implied Distance (In Kilometers)	1.60	8.78

Note: We measure the online border effect, which is the implied distance between the offline stores and the online stores. If the usual procedure is used, online and offline markets appear to be very closely integrated, with an equivalent border of 1.6 kilometers. When use the 95th percentile of the price gap distribution, the online border effect becomes 8.8 kilometers. This is very close to the actual physical average distance between the online warehouse (store 22, where the online goods appear to be delivered from) and each of the offline stores in the city.