INTERNATIONAL JOURNAL OF FINANCE AND ECONOMICS
Int. J. Fin. Econ. **21**: 3–35 (2016)
Published online 19 April 2015 in Wiley Online Library
(wileyonlinelibrary.com). DOI: 10.1002/ijfe.1517

DISTANCE AND POLITICAL BOUNDARIES: ESTIMATING BORDER EFFECTS UNDER INEQUALITY CONSTRAINTS

FERNANDO BORRAZ[a,*], ALBERTO CAVALLO[b], ROBERTO RIGOBON[c] and LEANDRO ZIPITRIA[d]
[a]*Banco Central del Uruguay, Montevideo, Uruguay and Department of Economics, FCS, UdelaR, Montevideo, Uruguay*
[b]*MIT, Boston, MA, USA*
[c]*MIT NBER, Boston MA, USA*
[d]*Department of Economics, FCS, UdelaR, Montevideo, Uruguay and Universidad de San Andrés, Buenos Aires, Argentina*

ABSTRACT

The 'border effects' literature finds that political boundaries have a large impact on relative prices across locations. In this paper, we show that the standard empirical specification suffers from selection bias and propose a new methodology based on binned-quantile regressions. We use a novel micro-price dataset from Uruguay and focus on city borders. We find that, when the standard methodology is used, two supermarkets separated by 10 km across two different cities have the same price dispersion as two supermarkets separated by 30 km within the same city, implying that crossing a city border is equivalent to tripling the distance. By contrast, when upper quantiles are used the city border effect disappears. These findings imply that transport cost have been systematically underestimated by the previous literature. Our methodology can be applied to measure any kind of border effect. We illustrate this in the context of online–offline price dispersion to measure an 'online-border' effect in the city of Montevideo. Copyright © 2015 John Wiley & Sons, Ltd

Received 6 October 2014; Accepted 15 March 2015

JEL CODE: C14; D40; E31; F40

KEY WORDS: Border effect; political borders; price dispersion; quantile regression

1. INTRODUCTION

Political borders can have a significant impact on relative prices. The degree of price segmentation caused by such boundaries was empirically documented in a seminal paper by Engel and Rogers (1996), who showed with consumer price index (CPI) data that the US–Canadian border had an effect on price dispersion equivalent to adding a distance of at least 1780 miles between locations (approximately the distance between Miami and Quebec). Their work spurred a large literature that found similarly large 'border effects' across countries, states and even cities.[1] These results have been heavily debated over the years. Some papers have argued that (i) the distances have been mis-measured (Head and Mayer, 2002), (ii) the regressions suffer from aggregation bias (Evans, 2003; Broda and Weinstein, 2008), (iii) the gravity equation implied in the standard specification has been misspecified (Anderson and van Wincoop, 2003; Broda and Weinstein, 2008), and that (iv) the regressions do not have a proper benchmark because of the fact that country distributions of prices are very different across countries (Gorodnichenko and Tesar, 2009). Despite all this work, the magnitude and reasons behind the segmentation introduced by political borders is still an open question in the literature.

In this paper, we propose a simple method to estimate the size of the 'border effect' based on Samuelson's iceberg cost model. This methodology imply that largest price differences observed between locations are relevant for transport cost estimation. We first argue that the standard regression is based on an arbitrage

*Correspondence to: Fernando Borraz, Banco Central del Uruguay, Montevideo, Uruguay and Department of Economics, FCS, UdelaR, Montevideo, Uruguay.
†E-mail: fborraz@bcu.gub.uy

Copyright © 2015 John Wiley & Sons, Ltd

inequality constraint, and that using all price observations creates a selection bias that affects both the distance and border coefficients (and, therefore, the estimates of the 'border effect'). We propose an alternative approach based on quantile regressions that corrects for the selection bias while simultaneously controlling for potential measurement errors.

Our method can be explained using a very simple framework along the lines of Engel and Rogers (1996). Consider the problem of a firm that sets a price bounded by the existence of an arbitrage constraint. If the arbitrage cost between two establishments (i and j) is τ and p denotes the log price in each location, then the arbitrage constraint can be expressed as a simple inequality:

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

where τ is a function of distance, political boundaries, and other regional and product characteristics. The literature typically estimates τ and the border effects by running the following regression on price dispersion:[2]

$$|p_{i,t} - p_{j,t}| = \alpha + \beta D_{i,j} + \gamma B_{i,j} + \delta X_{i,j,t} + \epsilon_{i,j,t} \tag{2}$$

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, $B_{i,j}$ is a dummy that takes value 1 if a border exists between locations i and j, and $X_{i,j,t}$ is a series of additional controls. In this context, 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, which means that a bias in either (or both) of these coefficients will have an impact on the estimate of the border effect.

We argue that τ and its determinants cannot be estimated through a simple ordinary least squares (OLS) regression because prices in the two locations are an optimal choice subject to an inequality constraint that is not necessarily binding. If the optimal prices of the two stores lie within the constraint, then their difference is smaller than τ; these observations are not relevant to estimate the arbitrage 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 arbitrage costs or the degree of segmentation between the markets. In fact, all observations within the no-arbitrage range suffer from selection bias; estimates that use the mean or the standard deviation of $|p_1 - p_2|$ will be biased downward as well.

The arbitrage cost τ is better estimated when we use only the largest observed price differences between locations. Those are the observations that provide information about the limit that arbitrage imposes of the magnitude of price dispersion.[3] Ideally, we would like to use the maximum observed price gap between locations, but it is potentially sensitive to measurement errors.[4] Instead, we estimate a series of binned-quantile regressions that allow us to measure the sensitivity of our estimates to the errors-in-variables (EIV). We start with the mean price gap between locations (equivalent to the method typically used in the literature) and then use only the observations in the 80th, 90th, 95th, 99th percentiles, and the maximum observed price difference.

We apply this method to study the impact of city borders on price dispersion in Uruguay. We use a novel good-level dataset composed by daily prices from 202 Universal Product Code (UPC)-level products sold in 333 supermarkets across 47 cities collected between 2007 and 2010. When we first estimate the border effect using standard methods, we find that the city border between two stores separated by 10 km is larger than 20 km wide, and statistically different from zero. This implies that the border triples the distance of stores across the city borders. However, when we re-estimate using distance-binned quantile regressions, the border declines until it is not significantly different from zero. As expected from our discussion, both the distance and border dummy coefficients are downward biased in the standard regression, but the bias is largest on the distance parameter.[5] As a result, the net impact is that the implied border effect (in kilometers) falls.

We perform robustness tests to correct for outliers and product mix, and we change the specification to include non-linearity and interaction terms. In all of them, the city-border effect measured in kilometers tends to disappear when higher percentiles are used. Furthermore, the results are similar at the 99th, 99.5th, 99.9th percentile and the maximum, suggesting that the estimates are not significantly affected by potential errors in the data.

Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe

We further illustrate our methodology to study the dispersion between online and offline prices. We use daily prices collected from the website of the largest grocery retailer in Montevideo and compare them with the prices in all offline locations of the same retailer. The offline store's data provide an estimate of the impact of distance on price dispersion across locations. The 'online border' is simply the distance that would generate the same effect on online–offline price dispersion observed in the data. When the standard procedure is used, online and offline markets appear to be very closely integrated, with a border effect of just 1.6 km. However, when we use the 95th percentile of the price gap distribution, the online border effect becomes 8.8 km, a number close to the actual average physical distance between the online warehouse (where the online goods are delivered from) and each of the offline stores in the city.[6]

Our approach and the nature of the data address four additional sources of concerns that have been raised since the original Engel–Rogers regressions. 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 deviations from the Law of One Price. Indeed, Evans (2003) 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 to underestimate price differences within countries. Third, we know the exact location of each store. As pointed out by Head and Mayer (2002), using approximate distances (such as from one country capital to another) can greatly overestimate the border effect. Finally, all the stores in our sample sell the same set of products. As Evans (2003) points out, the mix of products sold across borders can lead to a bias in the standard regressions.

Compared with recent papers in the literature, 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 we 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 Gourinchas, Hsieh, and Li (2011) who estimate the border effect by studying the response of average prices in one market to cost shocks in another market. An advantage of our approach is that it does not require any cost data.

2. METHODOLOGY

In this section, we present a simple model of price-setting across locations that provides the inequality we use to estimate arbitrage costs and the border effect. In particular, we propose a model where the firms' pricing decision is constrained by the ability of the consumer to arbitrage the price gap between two locations. Standard OLS regressions used in the literature consider all pairs of prices, including those that lie within the arbitrage constraint, which introduces a bias in the estimates of factors that affect the cost to arbitrage, such as distance and political borders. We propose an alternative methodology that focuses on the largest observed price differences between locations using binned-quantile regressions.

2.1. A simple model of price-setting with arbitrage

2.1.1 Consumers

Consider an economy with a mass of consumers uniformly distributed along a line. This line encompasses two cities (A, B) of equal distance. There are J stores in the economy, J_{A} stores in city A and J_{B} stores in city B. There is also a 'border' between A and B, in the sense that consumers pay a cost whenever they cross to another city. This border cost may arise because of differences in taxes, convenience in shopping hours and other characteristics associated with the city but not driven by distance. A consumer located on point ℓ and buying in store i has an indirect utility function represented by

$$u_{\ell}(i) = v - \theta p_i - \tilde{\beta} \mid \ell - \ell_i \mid - \tilde{\gamma} b_i - \tilde{\delta} I_i \tag{3}$$

Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe

where v is the reservation price of the consumer, and θ captures how sensitive the consumer is to prices. The rest of the parameters measure transaction costs are the following: $\tilde{\beta}$ measures unit transportation costs, multiplied by the distance between the consumer location (ℓ) and the store position (ℓ_i) (including information costs about the store, such as knowing the distance to the store or its prices); b_i is a dummy that takes value 1 if the consumer and the store are in different cities; and I_i measures additional store-specific costs not related with distance, such as learning the layout and sale events of a given store.

Because the consumer buys the one item that maximizes his utility, we can compare the price each consumer pays across all possible pairs of stores.[7] The consumer ℓ weakly prefers store i to store j if $u_\ell(i) \geq u_\ell(j)$, for each $i, j \in J = (J_A + J_B)$, $i \neq j$. For simplicity, assume the price elasticity; the transportation cost are symmetric in all locations. This implies that

$$v - \theta p_i - \tilde{\beta} \mid \ell - \ell_i \mid - \tilde{\gamma} b_i - \tilde{\delta} I_i \geq v - \theta p_j - \tilde{\beta} \mid \ell - \ell_j \mid - \tilde{\gamma} b_j - \tilde{\delta} I_j \tag{4}$$

By rearranging terms, we obtain

$$p_i - p_j \leq \frac{\tilde{\beta}}{\theta} \left(\mid \ell - \ell_j \mid - \mid \ell - \ell_i \mid \right) + \frac{\tilde{\gamma}}{\theta} \triangle b_{i,j} + \frac{\tilde{\delta}}{\theta} \triangle I_{i,j} \tag{5}$$

where $\triangle b_{ij}$ is equal to 1 if both stores are located in different cities and 0 otherwise, and $\triangle I_{i,j}$ measures the incremental information cost incurred by changing the store. Thus for each pair of stores, the consumption decision can be expressed as the result of inequality (6). The value of the distance terms depends on which store is further away from the consumer. If the difference between $\mid \ell - \ell_j \mid - \mid \ell - \ell_i \mid$ is negative, the price difference could simply be defined as $\left(p_j - p_i\right)$. Therefore, the expression is simplified to the absolute difference of the location between stores:

$$\mid p_i - p_j \mid \leq \frac{\tilde{\beta}}{\theta} \left|\ell_i - \ell_j\right| + \frac{\tilde{\gamma}}{\theta} \triangle b_{i,j} + \frac{\tilde{\delta}}{\theta} \triangle I_{i,j} \tag{6}$$

Comparing across different pairs, if the distance increases, a border exists between the stores, or if there is a positive cost of switching stores, the level of price dispersion rises. The opposite occurs if consumers are more sensitive to prices. It can be shown that the price space is not empty and that the inequality constraint is binding only for the indifferent consumer; ie. the one indifferent between buying in two different stores.[8] This implies that the rest of the consumers are not indifferent between two stores and always prefer to buy on a particular one. In the end, the indifferent consumer is the one for which the inequality is binding and defines the demand for each store.

2.1.2 Producers

Assume there are J_A and J_B identical producers (or stores) in each city that sell an identical good at price p_j, where $j \in J = (J_A + J_B)$. Each producer maximizes profits, given the prices of the other stores and subject to the participation constraint of the consumers. Suppose all producers, except for j, are in equilibrium. Then firm j sets its price subject to the participation constraint of consumer ℓ:

$$\max_{p_j} \prod_j \left(p_j / p_{-j}\right)$$

$$st \quad p_j \in \Re_+$$

and to the other $J - 1$ consumer constraints

$$\mid p_i - p_j \mid \leq \frac{\tilde{\beta}}{\theta} \left|\ell_i - \ell_j\right| + \frac{\tilde{\gamma}}{\theta} \triangle b_{ij} + \frac{\tilde{\delta}}{\theta} \triangle I_{i,j}, \ \forall i \in J, \ i \neq j$$

where this condition applies to all the J firms in the sample.

Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe

Firms maximize profits when setting the maximum price for the indifferent consumer, as shown in Appendix 6. This in turn implies that the inequality will not be binding for the rest of the consumers. Therefore, comparing pairs of prices for all stores to estimate equation 6 will not result in the correct measure of the consumers' parameters, as only those where the restriction is binding are valid.

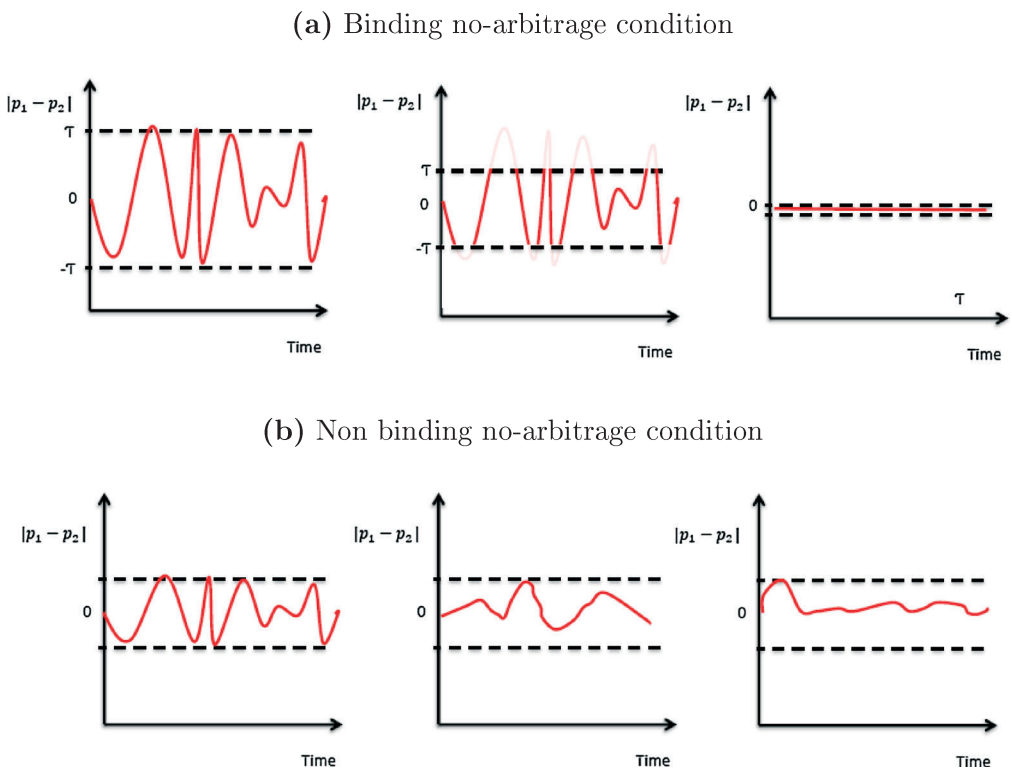
2.2. *Binned-Quantile Regressions*

This model of inequality constraints provides an equation that can be estimated as any other regression in the literature of border effects.[9] In this case, the specification can be defined as follows:

$$\mid p_{i,t} - p_{j,t} \mid \leq \beta D_{i,j} + \gamma B_{i,j} + \delta X_{i,j,t} + \varepsilon_{i,j,t} \tag{7}$$

where $\beta D_{i,j} \equiv \frac{\tilde{\beta}}{\theta} \left| \ell_i - \ell_j \right|$, $\gamma B_{i,j} \equiv \frac{\tilde{\gamma}}{\theta} \triangle b_{i,j}$, and $\delta X_{i,j,t} \equiv \frac{\tilde{\delta}}{\theta} \triangle I_{i,j}$.

Notice that this inequality implies that all the residuals $\left(\epsilon_{i,j,t}\right)$ in equation 7 are either zero or negative, in which case $E\left[\epsilon_{i,j,t}\right] \leq 0$. When this happens, the estimation by OLS is expected to produce biased estimates because of the failure of the orthogonality conditions, where the bias is downward. There is only one case in which the estimates remain unbiased and is if the price deviations are exactly equal to the arbitrage cost—that is, the constraint is always binding. The residuals are identical to zero and OLS produces unbiased estimates. Intuitively,



Note: Panel (a) shows cases where the no-arbitrage condition constrains the price dispersion. Panel (b) cases where the price dispersion is not correlated with the arbitrage cost.

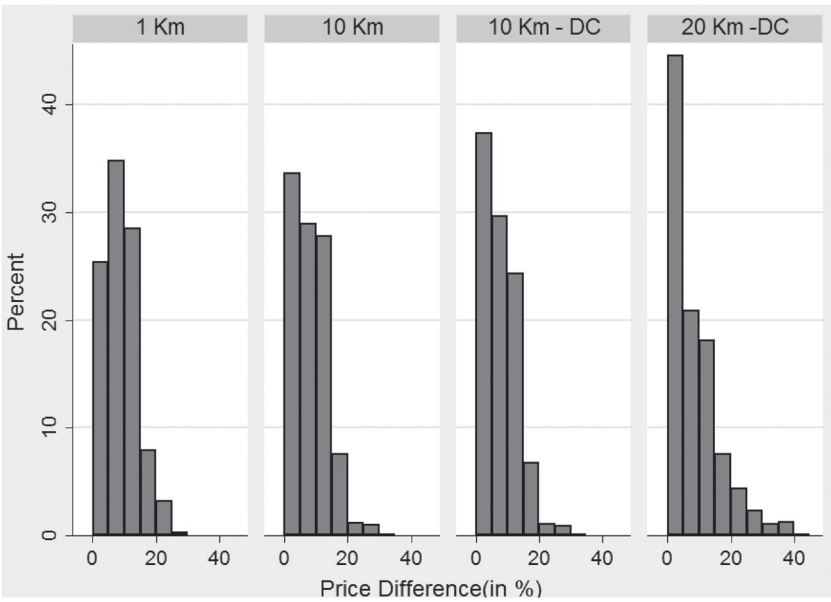
Figure 1. Price dispersion and the arbitrage cost. Note: Panel (a) shows cases where the no-arbitrage condition constrains the price dispersion. Panel (b) cases where the price dispersion is not correlated with the arbitrage cost.

Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe

prices are assumed free of EIV, so that the extreme in the distribution of price differences is the closest estimator to the arbitrage costs. It is important to mention that if all prices are optimally chosen to lie within the no-arbitrage region, then even the estimation using the extreme of the price distribution will produce downward biased estimates. However, by construction, the biases will be smaller. Formally, the expected value of the errors at the medium are more negative than the errors at the 95th quantile: $E\left[\epsilon_{i,j,t}|50th\right] \leq E\left[\epsilon_{i,j,t}|95th\right] \leq 0$. Nevertheless, as pointed out by Simonovska and Waugh (2014), the maximum price difference could introduce additional biases; so we use, instead, the upper quintiles for our estimation.[10]

Figure 1 illustrates how the observed price dispersion may not be informative of the arbitrage cost. Panel (a) in Figure 1 shows hypothetical prices over time of one product in two locations. If the no-arbitrage condition is binding, then as the arbitrage cost τ decreases, so does the price dispersion. However, if the condition is not binding,



Distance (Km)	Same City	Share at 0% (in %)	Price Gaps (in %) Mean	90th	99th	99.9th
1	yes	96.5	0.3	0	11.1	20.5
10	yes	76.6	1.9	8.64	16.9	25.1
10	no	73.6	2.0	9.46	16.9	25.1
20	no	69.3	2.6	10.6	26.9	39.3

Note: We calculate the price gaps (in absolute value) for all goods sold in a single retailer across two locations. We picked a random store from the largest retailer in the country and compared its prices to those of other stores from the same retailer located at 1 km, 10 km (same city), 10 km (different city), and 20 km (different city). Where "DC" denotes pairs in different cities. The graph excludes the mass at 0% to facilitate the comparison of positive gaps. The table shows the distributional statistics for all price gaps, including those at 0% (identical prices).

Figure 2. Example of price gaps in different store pairs. Note: We calculate the price gaps (in absolute value) for all goods sold in a single retailer across two locations. We picked a random store from the largest retailer in the country and compared its prices with those of other stores from the same retailer located at 1 km, 10 km (same city), 10 km (different city) and 20 km (different city). Where 'DC' denotes pairs in different cities. The graph excludes the mass at 0% to facilitate the comparison of positive gaps. The table shows the distributional statistics for all price gaps, including those at 0% (identical prices).

Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe

as shown in Panel (b), distributional statistics such as the mean or the standard deviation will not be associated with the estimate of τ. In both cases, however, we would be able to obtain good estimates of τ by using only the maximum observed price difference.

Figure 2 makes a similar point with real data. We plot the distribution of price differences for all goods between two locations of a given retailer and compare the results for stores located at 1 km, 10 km (same city), 10 km (different cities) and 20 km (different cities) of each other. As expected, as the distance increases the share of price gaps at 0% (see table at bottom in Figure 2), and the mass between 1% and 20% increases. Interestingly, when we compare the two pairs of stores located at 10 km from each other (one of which is for stores in different cities), we find that crossing the city border has an effect on the mean and 90th percentile. The two pairs, however, have exactly the same gaps at the 99th and 99.9th percentiles of the distribution. This last result is consistent with the idea that city borders should not affect the cost to arbitrage across locations. Using the mean and lower percentiles of the price gaps that lie within the arbitrage constraint can therefore lead to biased results.

In order to address this selection bias, we propose a new method to estimate border effects using distance binned-quantile regressions. The methodology can be described as follows:

First, compute the absolute price difference for all possible location (stores) pairs. Repeat this exercise across time and all goods, and pool all observations.

Second, define distance-border-bins according to a discrete spacing criteria that depends on the unit of observation (city versus countries) and the availability of enough observations within each bin. In the case of 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 larger distances. The distance in each bin does not have to be set in linear increments. For simplicity, denote each bin as n, where $n = \{1, \ldots, N\}$, and N is the number of bins. Each bin is defined by a distance D_n, a dummy $B_n = 1$ if there is a border between the two stores and additional controls X_n. In our case, X_n includes a chain dummy and an interaction term between distance and city dummy.

Third, compute the relevant quantile statistic of the absolute price differences for each bin. Denote the statistic as $Q_n\left(|p_{i,t} - p_{j,t}|, q\right)$ for the qth percentile of bin n.

Finally, estimate the following equation:

$$Q_n\left(|p_{i,t} - p_{j,t}|, q\right) = \alpha + \beta D_n + \gamma B_n + \delta X_n + \epsilon_n \tag{8}$$

In Figure 3, we depict the impact of the bias and the intuition behind our methodology. The horizontal axis shows the bins for a range of distances, and the vertical axis is the absolute price difference. The dots mark the absolute differences in the data for each of the selected 'bins'. The thick black line reflects the price difference implied by the no-arbitrage constraint. Because all the observed price differences are less than or equal to the thick line, the estimation in the standard regression—which implicitly uses the mean within each bin—is downward biased as denoted by the red line. In small samples, the true maximum per bin might not be observed; therefore estimating via the sample maximum will also be downward biased. Still, in this case, 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' line. This explains why we interpret our results as a lower bound estimate of the degree of segmentation.

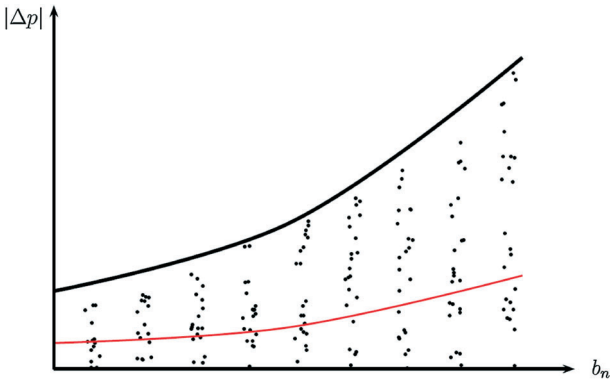
2.3. Dealing with errors-in-variables

One of the reasons we use quantile regressions to estimate arbitrage costs, instead of only the maximum, is to relax the assumption of no (EIV). In particular, the maximum of the price difference distribution within each bin can be significantly affected if prices are mis-measured. These errors can arise either because prices are observed and/or reported with errors or because stores make mistakes and post prices outside the no-arbitrage range. When we describe our data in Section 3, it will become clear that the errors from misreporting are very small because of the way the data is collected. However, there is still the possibility that the prices are incorrectly reported; thus, concentrating the estimates on the maximum within each bin would exacerbate the impact of any EIV.

This case is depicted in Figure 4. The black thick line is still the 'true' upper bound of the no-arbitrage band; that is the true degree of segmentation. However, 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.

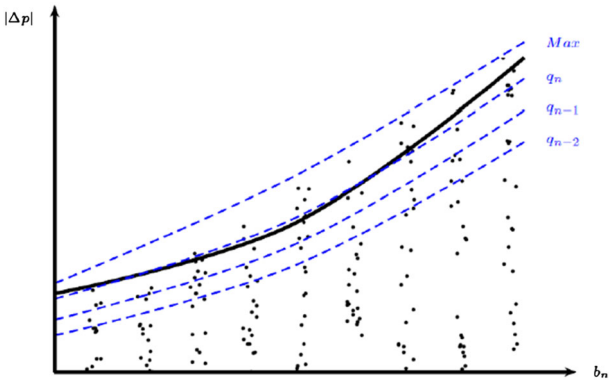
Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe



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 (red line) is downward biased.

Figure 3. 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 (red line) is downward biased.



Note: The black thick line is still the "true" upper bound of the no-arbitrage band, i.e. the true degree of segmentation. However due to 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 4. Bias in standard regression in the presence of errors-in-variables. Note: The black thick line is still the 'true' upper bound of the no-arbitrage band, that is, the true degree of segmentation. However, because of errors-in-variables, 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.

We address EIV in two ways. One is to eliminate outliers from the distribution. As we discuss in the succeeding paragraphs, the type of errors that are likely to be present in our data are misplacement of the decimal point or flipping digits, both of which are likely to produce large price changes at the item level that we can identify. This approach, however, does not provide a definite answer. For example, if the estimates change little, then it is not clear whether the EIV had a small impact, or not enough observations were eliminated to remove the bias.

Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe

The alternative is we propose it to estimate the regression using different quantiles. Within each bin, we compute several quantiles—the median, 80th, 90th, 95th and 99th percentiles.[11] 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 towards 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 in Section 4.1.

3. DATA

We use a good-level dataset of daily prices compiled by the General Directorate of Commerce (DGC), which comprises grocery stores all over the country.[12] 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 Uruguayan legislature that changed the tax base and rates of the value added tax. The Ministry of Economy and Finance was concerned about incomplete pass-through from tax reductions to consumer prices and hence 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 grocery stores and supermarkets to report its daily prices for a list of products if they meet the following two conditions: (i) they sell more than 70% of the products listed and (ii) they either have more than four grocery stores under the same name or have more than three cashiers in a store. The information sent by each retailer is a sworn statement, and they are subject to penalties in case of misreporting. The objective of the DGC is to ensure that prices posted reflect real posted prices by stores. In this regard, stores are free to set the prices they optimally choose, but they face a penalty only if they try to misreport them

The data includes daily prices from 1 April 2007 to 31 December 2010 for 202 items corresponding to 61 product categories, where each item is defined by its UPC.[13] See table A.4 in Appendix A.4 for a detailed description of the products in the data.

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 L variety by all stores. If this specific variety is not available at a store, then no price is reported. Whenever prices are 50% greater (or less) than the average price, the retailer

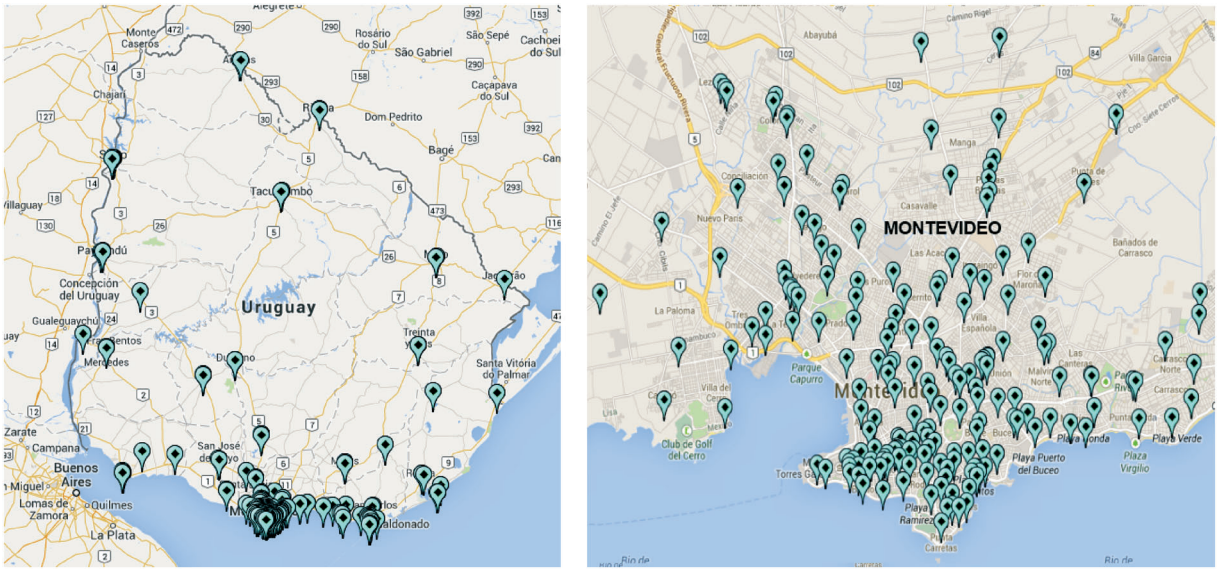
Table 1. Product, time and regional coverage in the data

		All stores
(i)	Retailers	136
(ii)	Stores	333
(iii)	Products	202
(iv)	Categories	61
(v)	Country	Uruguay
(vi)	Cities	47
(vii)	Departments	19
(viii)	Time Period	1 April 2007 to 31 December 2010
(ix)	Days	1154
(x)	Observations (bins)	179,215
(xi)	Observations (pairs)	32,159,865

Note: Summary statistics of the data compiled by the General Directorate of Commerce (DGC).

Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe



Note: Each dot represents a store location across the 19 Uruguayan departments.

Figure 5. Cities covered in the sample. Note: Each dot represents a store location across the 19 Uruguayan departments.

is contacted to confirm whether the submitted price is correct. The data are then used in a public web site that allows consumers to check prices in different stores or cities and to compute the cost of different baskets of goods across locations.[14]

After the exclusion of observations labeled as 'preliminary' as well as wrongly categorized or unidentified data (e.g. products that share the same UPC), our final dataset is composed by 202 products at the UPC level in 333 grocery stores from 47 cities. Table 1 describes the summary statistics of the coverage in the data and Appendix 6 provides a detailed list of the products. In addition, see Figure 5 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, accounts for 58% of the supermarkets in the sample. The maximum distance between two supermarkets is 526 km.[15]

We consider two datasets separately to account for outliers that may have a greater impact on the largest price differences between one good. We construct a baseline case with the complete sample, and a second case in which those price higher than three times (or less than a third) of the median daily price were excluded. However, deleted prices only account for a small 0.034% of the whole database.

In order to compute the linear distance between each pair of stores in our sample, we use information on the exact geographical location of each supermarket as provided by Ciudata, an industry organization. We then construct distance bins using a geometric sequence starting from 0.1 km and incrementing by $((526/0.1)^{1/N})\%$. Our baseline estimation uses N=500 bins, but we re-estimated our results using 50, 100 and 1,000 bins as well. We then 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.

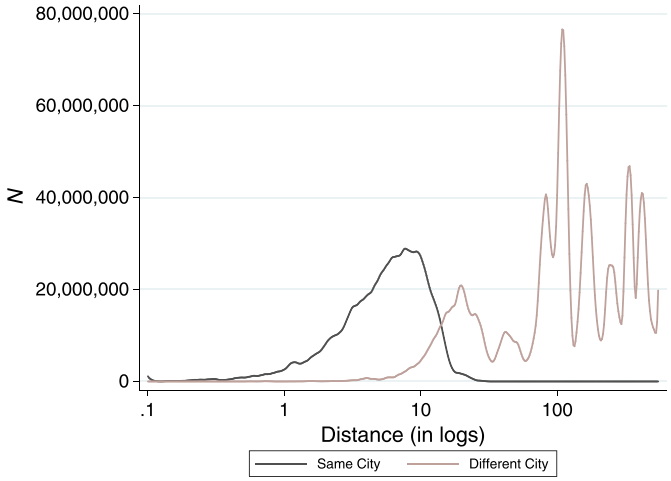
Finally, we define the following specification:

$$Q_n\left(|p_{i,t} - p_{j,t}|, q\right) = \alpha + \beta D_n + \gamma B_n + \delta_1 B_n \times D_n + \delta_2 Firm_n + \epsilon_n \tag{9}$$

where $Q_n\left(|p_{i,t} - p_{j,t}|, q\right)$ is the qth quantile of the absolute price differences for all store pairs i and j whose distance belongs to bin n; D_n measures the distance between stores that belong to bin n; B_n is a dummy that takes the value 1 if the supermarkets are in different cities; $Firm_n$ is a dummy variable that takes the value 1 if the price difference in bin n comes from the same supermarket chain. We also add an interaction dummy between distance and the city border dummy, and fixed effects for each good.

Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe



Note: The black line shows the distribution of bilateral observations for each of the 500 bins within cities, while the grey line (extending to the right, with multiple peaks) shows the distribution across cities. Lines are smoothed for better visualization.

Figure 6. Distribution of observations for 500 bins in the same city and between cities. Note: The black line shows the distribution of bilateral observations for each of the 500 bins within cities, while the grey line (extending to the right, with multiple peaks) shows the distribution across cities. Lines are smoothed for better visualization.

Notice that this regression requires that we have both observations within and across cities that overlap in distance bins. Figure 6 shows the distribution of observations for each of the 500 bins for the same city pairs and the different city pairs. The horizontal axis is the log distance starting at 100 m to a maximum of 526 km. The black line is the number of observations per bin for the stores within the same city boundaries, while the grey line are the observations for the stores in different cities. 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 between 5 and 15 km. This is the source of the variation where the city-border effect is actually estimated.

4. RESULTS

As described in Section 2.2., we pool all the prices into each corresponding bin and estimate the distribution of price differences. We select the mean, median, 80, 85, 90, 95, 97.5, 99, 99.5 and 99.9th percentiles. For each of these, we estimate equation 9 by weighted least squares to account for differences in the number of observations inside each bin. Price differences are expressed in percentage terms, while distance is measured in hundreds of kilometers.

The results are presented in Table 2. The first coefficient is the segmentation generated by distance. The second and third estimate the effect of the city boundaries (border dummy) and the interaction term (how the effect of distance changes once the stores are in different cities), respectively. The fourth coefficient is the impact of both stores belonging to the same retailer, and the last one is the constant term. Each column reflects a different regression. The first one uses the mean within each bin, which replicates the standard regressions in the literature. After that, we present the results for the quantiles moving from the 50th until 99.9th percentiles and finally the maximum.

Notice that, as we increase the percentile, all individual coefficients increase—in line with the intuition we discussed before. This pattern can be easily appreciated in Figures 7 and 8, which show the coefficient on distance and city dummy, respectively, as a function of the percentile.

Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe

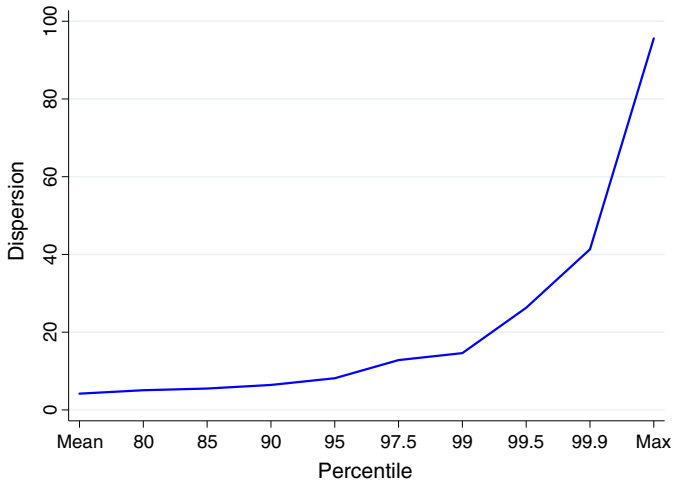
Table 2. Weighted least square estimation of price differential using the complete database and for 500 bins

	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	179,215	179,215	179,215	179,215	179,215	179,215	179,215	179,215	179,215	179,215	184,328
*R*2	0.752	0.645	0.749	0.761	0.762	0.725	0.744	0.766	0.68	0.58	0.492

Note: * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors in parenthesis.

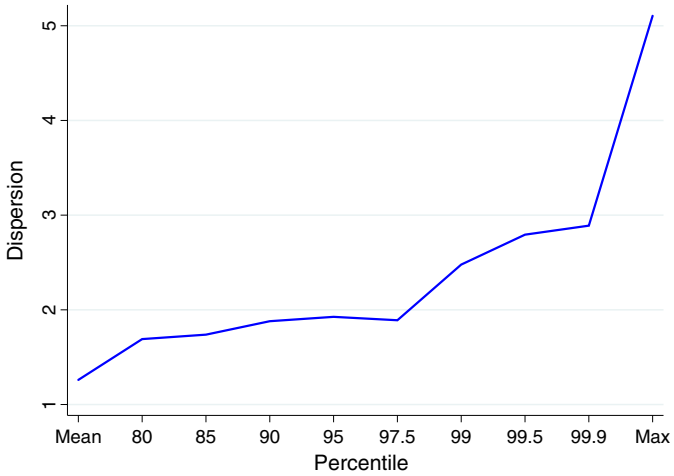
Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe



Note: Estimated distance coefficient when different quantiles are used for the baseline regression.

Figure 7. Estimation of the distance coefficient by quantile. Note: Estimated distance coefficient when different quantiles are used for the baseline regression.



Note: Estimated city dummy coefficient when different quantiles are used for the baseline regression.

Figure 8. Estimation of the city coefficient by quantile. Note: Estimated city dummy coefficient when different quantiles are used for the baseline regression.

There are two alternatives to compute the border effect. One way is to base the effect upon a specific distance. First, 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 for two stores within the same city to have the same degree of price dispersion. The following example clarifies the analysis.[16] Using the results in the first column (average) in Table 2, we 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 yields 30.5 km. Therefore, the border adds 20 km to two stores 10 km apart—that is, the city border triples the distance. Although the literature simply uses the ratio of the two coefficients to compute the border effect, our specification also allows for non-linearities. Therefore, the implied border effect needs to be estimated conditional to a given distance.

Copyright © 2015 John Wiley & Sons, Ltd

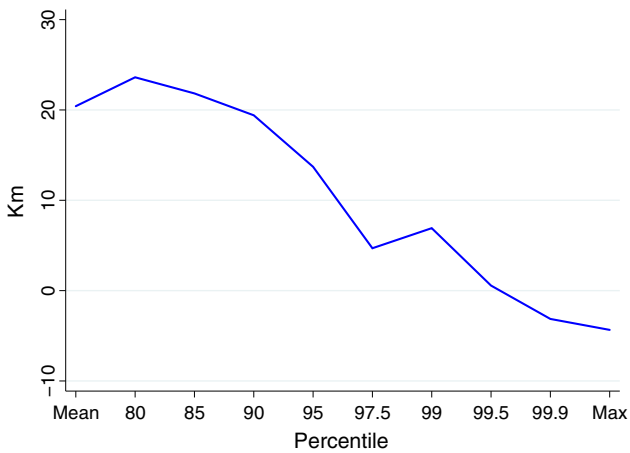
DOI: 10.1002/ijfe

In Panel (a) of Figure 9, we compute this implied additional distance for two stores 10 km apart for each of the quantiles. The border effect, as measured in kilometers, collapses towards zero around the 99.5th percentile. Interestingly, the effect is even found negative at the highest percentiles. In addition, notice the (almost) monotonic decrease in the estimates. This is encouraging from an EIV point of view. If the maximum of the distribution were the result of large EIV, there is no reason to expect the estimates and the impact of the border effect to remain similar to the upper percentiles.

The second way to compute the border effect is to focus on the relative price dispersion for a given distance. In other words, we compute how large is the implied degree of segmentation for a pair of stores 10 km apart across two cities, relative to another pair of stores 10 km apart within the same city. In both cases, we consider

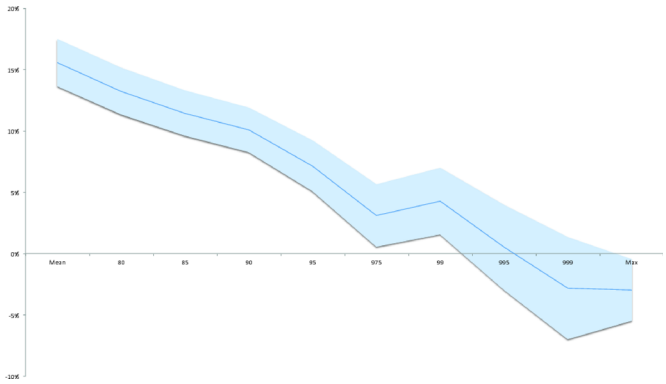
(a) Implied Kilometers Additional Km

implied by City Border Effect for Stores 10 Km Apart



(b) Relative Increase in Price Dispersion

of City Borders for Stores 10 Km Apart.



Note: Panel (a) shows the implied additional km for the linear specification using all data and 500 bins. Panel (b) shows the relative increase in the degree of segmentation for the baseline linear specification, with its 95th percent confidence band.

Figure 9. Estimation of the city border effect using all data. Note: Panel (a) shows the implied additional kilometre for the linear specification using all data and 500 bins. Panel (b) shows the relative increase in the degree of segmentation for the baseline linear specification, with its 95th percent confidence band.

Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe

all stores that do not belong to the same retailer. For instance, in the average case (column 1 in Table 2), the price dispersion for $D_n = 0.1$ 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 $5.081 + 4.188 * 0.1 = 5.499$. Thus, the border implies a 15.57 percent higher degree of segmentation. However, this relative effect becomes small and insignificant using higher quantiles. In Panel (b) in Figure 9, we present the decreasing pattern in the relative degree of segmentation, together with its 95th confidence band.

Panels (a) and (b) show that the degree of segmentation is overestimated—and the impact of distance is underestimated—when the average price deviations are used. By contrast, the effect of the border becomes not significantly different from zero when the upper quantiles of the distribution within each bin are used. Notice that the reduction in the border effect is not a mechanical consequence of the methodology. The estimation using upper quantiles should in fact increase the absolute value of all coefficients—simply because there is less sample selection. The decrease in the final border effect, however, comes from the fact that the bias in the distance coefficient is larger than in the border coefficient.

4.1. Robustness

In this section, we test the sensibility of the baseline estimates to changing the specification of the regression; to different subsamples of product mix, to elimination of outliers and to different number of bins. In all cases, we find that the results are qualitatively similar. That is, the traditional regression (average price dispersion) estimates a large and significant city border effect, whereas quantile regressions show that the city border becomes insignificant using upper quantiles of the distribution. Furthermore, the results are similar at the highest percentiles and the maximum, suggesting that the estimates are not significantly affected by measurement errors.

First, we modify the equation to the following non-linear specification:

$$\begin{aligned} Q_n\left(|p_{i,t} - p_{j,t}|, q\right) = \alpha + \beta D_n + \gamma B_n + \delta_1 B_n \times D_n + \delta_2 D_n^2 + \delta_3 D_n^3 \\ + \delta_4 B_n \times D_n^2 + \delta_5 B_n \times D_n^3 + \delta_6 Firm_n + \epsilon_n \end{aligned} \tag{10}$$

where the variables are defined as in equation 9.

The results, presented in Table 3, yield the same patterns we described earlier for the baseline estimation. In absolute value, all point estimates increase as the estimation is performed over the higher quantiles. Furthermore, if one computes the implied additional distance, the results remain qualitatively the same as those in Panel (a) in Figure 9. The border effect, as measured in kilometers, is close to 25 km using the traditional regression (average). However, it decreases with higher percentiles until it becomes small and insignificant at the 97.5th percentile.

In addition, we perform three robustness tests using different subsamples. Results are presented in the Appendix A.3 (Tables A.1, A.2, A.3 and Figures A.1, A.2, A.3) for both the linear and non-linear specifications. First, we eliminate products in which the matching across stores is not perfect. In particular, we exclude meat, bread, and among other categories. Quantile regressions yield identical patterns as when using the complete dataset. Second, we use all products but eliminate the outliers; defined here as those whose price is 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 10th of the median price. Still, our rule only excludes 11.2 thousand in 32.8 million or just 0.034% of the observations. Once again, the patterns are almost identical to the ones obtained using the complete number of observations. The only minor difference is that, for a given percentile, the border effects are smaller in absolute terms. In other words, the estimated implied distances are smaller than those in Panel (a) in Figure 9.

Third, we further combine the two cases earlier and estimate quantile regressions excluding goods with imperfect matching as well as those defined as outliers. The results do not yield significant differences with respect to our baseline estimation.

Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe

Table 3. Weighted least square estimation of price differential using the complete database and for 500 bins

		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	179,215	179,215	179,215	179,215	179,215	179,215	179,215	179,215	179,215	179,215	184,328
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%. Standard errors in parenthesis.

Copyright © 2015 John Wiley & Sons, Ltd

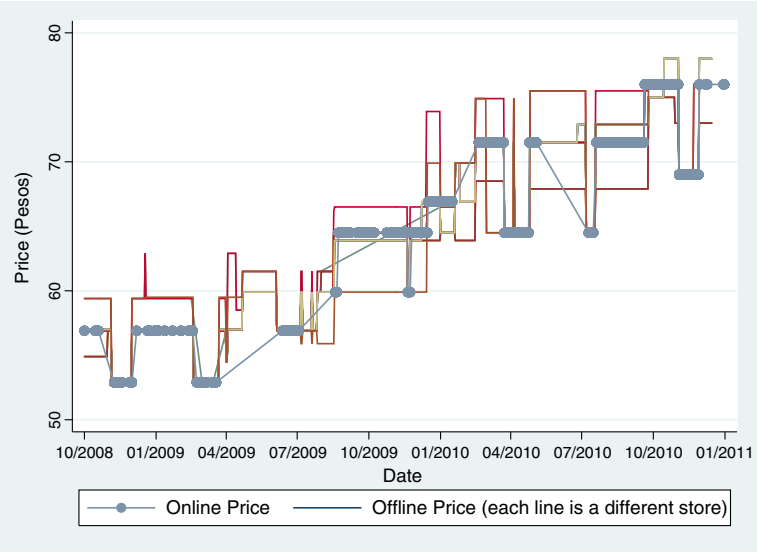
Lastly, we tested the linear and non-linear specifications to the sensibility of the number of distance bins. Instead of 500 bins, we re-estimated using 50, 100 and 1000 bins. Notice the trade-off in the selection of bins. The advantage of a larger number of bins is that each pair of stores is allocated 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 at the 99.9 percentile becomes very noisy.[17] The results are qualitatively the same to the baseline estimation. The city-border effect measured in kilometers falls and becomes insignificant when the upper quantiles are used in the estimation.

5. THE ONLINE BORDER EFFECT

We now use online and offline prices from a retailer in Uruguay to estimate an 'online border' effect. The degree of segmentation between online and offline markets is an interesting topic by itself but has received little attention because of the data limitations. Selling online allows a retailer to price discriminate among consumers who have time to travel to a store and those who prefer the convenience of online shopping. By deciding to buy online, the consumer may therefore be paying a cost in terms of price dispersion (in addition to any delivery charges).

We use data collected by the Billion Prices Project at MIT using a method that scans the HyperText Markup Language code of public retailers' website, identifying and storing all relevant price and product information on a daily basis.[18] The largest grocery retailer in Uruguay sells products in dozens of offline locations, as well as online, in the city of Montevideo. We compared the daily prices of all goods in the DGC offline data with their corresponding online price on the same date. Both datasets contain daily prices for the period between 1 October 2007 and 31 December 2010.

Figure 10 provides an example of the prices posted in the different stores (including the online store) for a given product over time. On most dates, the online price is within the range of prices observed in offline stores. This feature is observed in most goods in the sample.



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.

Figure 10. Example of online and offline prices: Cocoa – 0.5 kg. Note: This is an example of the typical time series pattern of online prices compared with 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.

Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe

Table 4. The online–offline border

	Price difference	
	(1)	(2)
Variables	Mean	95th Percentile
Log distance	0.100***	0.156***
	(0.005)	(0.010)
Constant	0.439***	3.177***
	(0.174)	(0.362)
Observations	2300	2300
Difference online-offline (%)	0.60	4.55
Implied distance (Km)	1.60	8.78

Note: *** significant at 1%. Robust standard errors in parenthesis. We measure the online border effect, defined as 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 the 95th percentile of the price gap distribution is used, 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.

Table 5. Online versus 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.

Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe

Suppose a consumer decides to buy a good from the retailer's website instead of walking to an offline location. What is the effect of crossing this 'online border' on prices? We can calculate an online border effect by simply estimating the implied distance that would produce the same degree of online–offline price dispersion observed in the data. We calculate this effect in two steps. First, we estimate the following regression for each quantile q using only data from the offline stores:

$$Q_n\left(|p_{i,t} - p_{j,t}|, q\right) = \lambda + \beta D_n + \epsilon_n \tag{11}$$

Notice that this is equal to equation 9 with $B_n = 0$ (same city), $Firm_n = 1$ (same retailer) and $\lambda = \alpha + \gamma$. The coefficient β therefore provides an estimate of the effect of distance on the dispersion of prices across locations. Finally, to calculate the implied online border effect, we simply compute the average online–offline price dispersion (between all pairs of online–offline stores), subtract the constant λ and divide by β.

We estimated regression 11 using both the traditional and the binned-quantile methods and computed the online border in both cases. Table 4 shows that using the traditional OLS regression provides an implied distance of 1.6 km. By contrast, if we use the 95th percentile, we obtain an implied distance of 8.78 km.

How do we know which estimate is better? We can compare the results in both methods with the actual 'distance' of the online store. Although the warehouse of the online store is not known, the website of this retailer states that the online prices are identical to those available in an offline store that fills the orders. It fails to name it explicitly, but we can compute a simple 'matching probability' between the online prices and each of the offline stores to identify it. This matching probability is just the average probability that the online price is identical to the price in an offline store on any given day. We calculate it at the store level in two steps. First, for each product, we compute the share of days that the online price is identical to the offline price. And second, we get the mean (or median) across all products in that store.

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. The average distance is 7.22 km (and a median of 8.04 km). This number is very close to the estimate we obtained by using the 95th percentile in the regression. In this case, the traditional regression greatly *underestimates* the degree of online–offline market segmentation.

This example illustrates why our method will not mechanically cause a reduction in border effects every time. By using the largest observed price differences, we know that all coefficients and constant in 9 will increase because the standard OLS method creates a downward bias in all of them. But whether the border *effect* rises or falls depends on the magnitude of those changes. In the case of the online border in equation 11, the distance coefficient β rises less than the constant, as shown in Table 4. This is because we are only looking at stores within the city of Montevideo, where the share of identical prices is very large. When the traditional OLS regression is used, the average price dispersion captured by λ is extremely small. When those identical prices are ignored, as we do with our methodology, λ rises significantly more than the β coefficient on distance; the "border effect" increases.

6. CONCLUSIONS

The extensive literature on the degree of segmentation resulting from political borders has 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 strategy has not taken into account the selection problem in posted prices: when a firm faces the possibility of arbitrage because of the existence of a transaction cost, the firm sets prices subject to a no-arbitrage constraint. However, if the optimal price falls within the no-arbitrage range, the dispersion in prices is not informative of the tightness of the constraint. A firm may set the same price in two locations, but it does not mean that the arbitrage cost for the consumer is zero. This implies that the estimation using average absolute price differences or standard deviations of price differences will not capture the size of the trade or arbitrage cost.

Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe

This paper builds on the existing literature with two main contributions. In the first place, it offers an alternative methodology to estimate transactions costs—which can be applied not only in international trade but also in other areas as in empirical finance, measurement of liquidity or the cost of regulatory restrictions. In the second place, we show that city borders matter little for price dispersion within a country. Although the border effect of a city should be small from an intuitive point of view, the traditional methods still estimated a very wide border effect (20 additional kilometers to two stores separated 10 km apart, that is, the border triples the distance). This is particularly large in a country where the largest city is less than 40 km wide, and there are no significant differences between cities in terms of taxes, language and the like. By contrast, the border becomes insignificant once we estimate using our method of distance binned-quantile regressions. We illustrate our method to measure border effects for the online–offline border in Montevideo and showed that we can correctly approximate the true average distance between the offline stores and the location where online purchases are sourced by simply using price gap data.

Finally, we believe further research should advance in at least two dimensions. From a methodological point of view, it is important to further examine the definition of optimal bandwidths. Although, in our paper, we used different bin sizes and results remained consistent across all specifications, this may not be the case in other applications in economics. And second, similar micro-level data needs to be collected across several countries to shed light on the actual width of international borders.

APPENDIX A

A.1 Proofs to the model of price-setting with arbitrage

A.1.1 Consumers

Lemma 1

The price space is non empty.

Proof

Given prices p_i and p_j the right-hand side of the inequality is non negative for the consumer located on store ℓ_i. In this case, equation 6 is $\mid p_i - p_j \mid \leq \frac{\tilde{\beta}}{\theta}\left|\ell - \ell_j\right| + \frac{\tilde{\gamma}}{\theta} b_j + \frac{\tilde{\delta}}{\theta} I_j$, as $\ell = \ell_i$. This implies that there must exist at least one price in order for the consumer to compare its bundles. Thus, both the right-hand and left-hand side of the inequality are positive. □

Lemma 2

The inequality constraint is binding only for the indifferent consumer.

Proof

The indifferent consumer can be defined as the one obtaining the same utility from buying in both stores, that is, $u_\ell(i) = u_\ell(j)$. This in turn implies that $v - \theta p_i - \tilde{\beta}\,|\ell - \ell_i| - \tilde{\gamma} b_i - \tilde{\delta} I_i = v - \theta p_j - \tilde{\beta} \mid \ell - \ell_j \mid - \tilde{\gamma} b_j - \tilde{\delta} I_j$. By rearranging terms, we obtain that $\mid p_i - p_j \mid = \frac{\tilde{\beta}}{\theta}\left|\ell_j - \ell_i\right| + \frac{\tilde{\gamma}}{\theta} \triangle b_{i,j} + \frac{\tilde{\delta}}{\theta} \triangle I_{i,j}$. □

A.1.2 Producers

Lemma 3

Firms maximize profits by setting the price that binds the participation constraint for consumer i.

Proof

Given the prices of all stores except for j, and given the right-hand side of the equation in terms of p_j, Kuhn Tucker conditions determine that the price difference should be maximal, that is, when the consumer restriction is binding. At the same time, notice that the indifferent consumer for firm j determines the demand for its products. □

Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe

Proposition 4
The consumer that maximizes profits is the indifferent consumer.

Proof
From Lemma 2 and Lemma 3, the firm sets the maximum price for the indifferent consumer. □

Proposition 5
For any given two stores (locations), the parameters can only be estimated for the indifferent consumer, that is, where the inequality is binding.

Lemma 6
If transportation costs increase (beta) or a border exists between two stores, or if the sensitivity of the consumer to price changes decrease, then the price dispersion increase.

Proof
From Lemma 2 and Proposition 4, the firm sets its price for the indifferent consumer such that equation 6 is binding. Therefore, the price differences will be maximum given the store location and other exogenous variables for the consumer. □

Proof
Take partial derivatives of each coefficient on the last equation of Lemma 2. □

A.1.3 Consumer heterogeneity: Discussion

So far, we have assumed that consumers only differ in their location on the line. However, consumers can also differ in their valuation of the good. This feature can be introduced to the original model in either two ways. First, consumers can differ in their maximum valuation of the good, in which case $v \in [\underline{v}, \overline{v}]$. In this case, previous results are easily maintained as well, although now satisfying two conditions for the indifferent consumer: indifference in distance and in valuation. Recall that previous results are for the medium consumer. Second, consumers can differ in their disposition to pay for the good, that is, $\theta \in \left[\underline{\theta}, \overline{\theta}\right]$. Similarly as before, there are two constraints to estimate the demand for each store: the distance constraint and the valuation constraint.

Therefore, adding heterogeneity to consumers' taste does not change the estimation procedure. In order to estimate the demand for each store, we must still solve the model for the indifferent consumer. As previously shown, the price inequality should only be binding for this consumer and slack for non-indifferent consumers.

A.2 Samuelson's Iceberg Costs

The results from our model of product arbitrage is also related to the no-arbitrage pricing region generated in Samuelson's Iceberg costs.[19] Assume that there is an arbitrage cost between two locations that can be described as follows:

$$\tau_{i,j,t} = \alpha + \beta D_{i,j} + \gamma B_{i,j} + \delta X_{i,j,t} \tag{12}$$

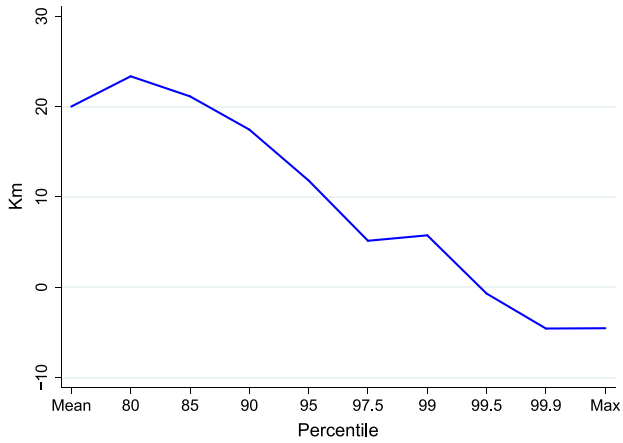
where the variables are defined as before. This arbitrage cost τ represents the proportion of the item that is lost when a customer transports one unit from i to j.[20] Under this form of arbitrage costs, prices need to lie within the range $\left|p_i - p_j\right| \leq \tau_{i,j,t}$ to avoid the possibility that a customer arbitrates across locations. In particular, 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; the price difference is a biased estimate of τ. But if the difference is larger, 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 as $\left|p_i - p_j\right| \leq \tau_{i,j,t} = \alpha + \beta D_{i,j} + \gamma B + \delta X_{i,j,t}$.

Copyright © 2015 John Wiley & Sons, Ltd

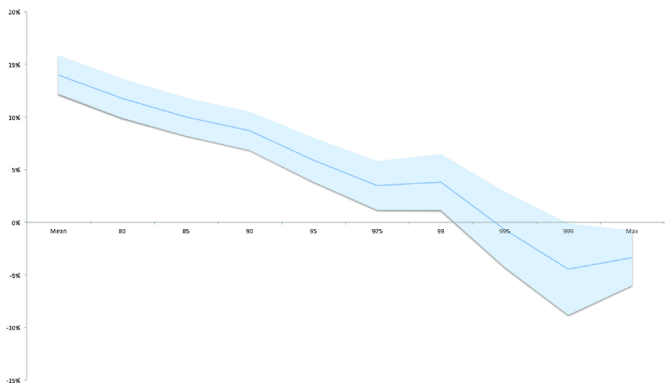
DOI: 10.1002/ijfe

A.3 Alternative Specifications

(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



Note: Panel (a) shows the implied additional km for the linear specification using all data and 500 bins. Panel (b) shows the relative increase in the degree of segmentation for the baseline linear specification, with its 95th percent confidence band.

Figure A.1. Estimation of the city border effect excluding meat and bread. Note: Panel (a) shows the additional kilometre implied by the city border effect for the linear specification, excluding meat and bread, using 500 bins. Panel (b) shows the relative increase in the degree of segmentation, with its 95th percent confidence band, for the same specification.

Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe

Table A.1. Price differential (excluding meat and bread, 500 bins and weighted least square)

	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
*R*2	0.705	0.593	0.712	0.726	0.724	0.684	0.71	0.679	0.576	0.523	0.479

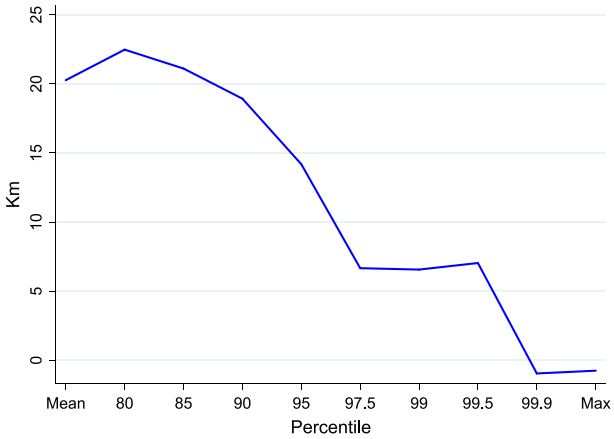
	Non-linear specification										
	Average	50	80	85	90	95	97.5	99	99.5	99.9	Max
Distance	−1.721***	−4.583***	0.963	2.260**	1.818*	1.948	19.575***	38.915***	91.935***	220.458***	960.037***
	(0.534)	(0.582)	(0.853)	(0.931)	(1.073)	(1.508)	(2.011)	(2.845)	(4.309)	(8.002)	(14.213)
City	0.521***	0.389***	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
*R*2	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%. Standard errors in parenthesis.

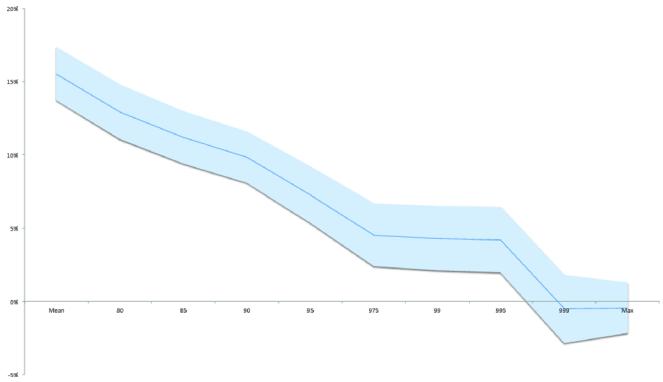
Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe

(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



Note: Panel (a) shows the additional km implied by the city border effect for the linear specification, excluding outliers, and using 500 bins. Panel (b) shows the relative increase in the degree of segmentation, with its 95th percent confidence band, for the same specification.

Figure A.2. Estimation of the city border effect using all data and excluding outliers. Note: Panel (a) shows the additional kilometre implied by the city border effect for the linear specification, excluding outliers, and using 500 bins. Panel (b) shows the relative increase in the degree of segmentation, with its 95th percent confidence band, for the same specification.

Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe

Table A.2. Price differential (excluding outliers, 500 bins and weighted least square)

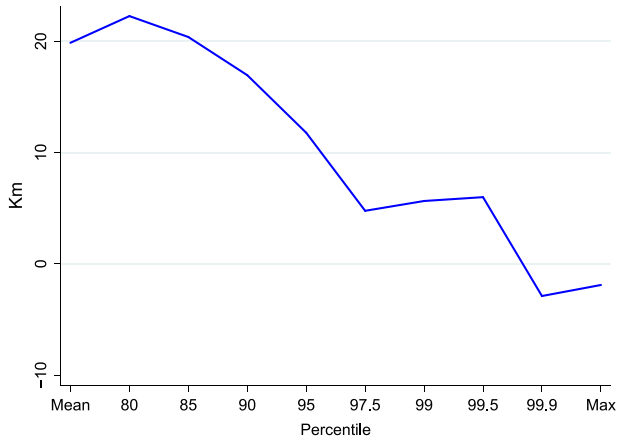
	Average	50	80	85	90	95	97.5	99	99.5	99.9	Max
Linear Specification											
Distance	4.188***	3.684***	5.205***	5.564***	6.442***	8.025***	13.066***	15.412***	16.448***	21.987***	43.223***
	(0.173)	(0.189)	(0.285)	(0.31)	(0.35)	(0.469)	(0.617)	(0.775)	(0.925)	(1.313)	(1.719)
City	1.252***	1.242***	1.674***	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%. Standard errors in parenthesis.

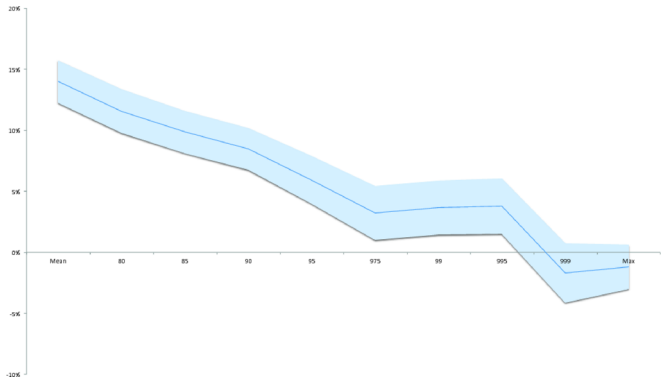
Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe

(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



Note: Panel (a) shows the additional km implied by the city border effect for the linear specification, excluding meat, bread, as well as outliers, using 500 bins. Panel (b) shows the relative increase in the degree of segmentation, with its 95th percent confidence band, for the same specification.

Figure A.3. Estimation of the city border effect excluding meat, bread and outliers. Note: Panel (a) shows the additional kilometre implied by the city border effect for the linear specification, excluding meat, bread and outliers, using 500 bins. Panel (b) shows the relative increase in the degree of segmentation, with its 95th percent confidence band, for the same specification.

Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe

Table A.3. Price differential (excluding meat, bread and outliers, 500 bins, weighted least square)

	Average	50	80	85	90	95	97.5	99	99.5	99.9	Max
Linear specification											
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%. Standard errors in parenthesis.

Copyright © 2015 John Wiley & Sons, Ltd

A.4 Data Details

Table A.4. Description of products in the database and share in consumer price index

Product	Brand	Specification	Share in CPI (%)
Beer	Pilsen	0.96 L	0.38
Beer	Patricia	0.96 L	0.38
Beer	Zillertal	1 L	0.38
Bleach	Agua Jane	1 L	n/i
Bleach	Solucion Cristal	1 L	n/i
Bleach	Sello Rojo	1 L	n/i
Bovine Beef 'Aguja'	No Brand – Cow	1 kg	0.23
Bovine Beef 'Aguja'	No Brand	1 kg	0.23
Bovine Beef 'Aguja'	With Bone – No Brand	1 kg	0.23
Bovine Beef 'Aguja'	Boneless – No Brand	1 kg	0.23
Bovine Beef 'Aguja'	With Bone – No Brand – Cow	1 kg	0.23
Bovine Beef 'Aguja'	Boneless – No Brand	1 kg	0.23
Bovine Beef 'Nalga'	No Brand – Novillo	1 kg	0.32
Bovine Beef 'Nalga'	With Bone – No Brand	1 kg	0.32
Bovine Beef 'Nalga'	Boneless – No Brand	1 kg	0.32
Bovine Beef 'Nalga'	Boneless – No Brand – Cow	1 kg	0.32
Bovine Beef 'Nalga'	With Bone – No Brand – Cow	1 kg	0.32
Bovine Beef 'Paleta'	With Bone – No Brand – Cow	1 kg	0.20
Bovine Beef 'Paleta'	Boneless – No Brand	1 kg	0.20
Bovine Beef 'Paleta'	With Bone – No Brand	1 kg	0.20
Bovine Beef 'Peceto'	No Brand	1 kg	0.16
Bovine Beef 'Peceto'	No Brand – Cow	1 kg	0.16
Bovine Beef 'Rueda'	With Bone – No Brand	1 kg	0.17
Bovine Beef 'Rueda'	With Bone – No Brand – Cow	1 kg	0.17
Bread	Bimbo	0.33 kg	0.06
Bread	Los Sorchantes	0.33 kg	0.06
Bread	Pan Catalan	0.33 kg	0.06
Bread	No Brand	1 Unit approx. 0.215 kg	1.14
Brown Eggs	El Ecologito	1/2 dozen	0.46
Brown Eggs	El Jefe	1/2 dozen	0.46
Brown Eggs	Prodhin	1/2 dozen	0.46
Brown Eggs	Super huevo	1/2 dozen	0.46
Brown Eggs	El Ecologito	1 dozen	0.46
Brown Eggs	El Jefe	1 dozen	0.46
Brown Eggs	Prodhin	1 dozen	0.46
Burgers	Burgy	3 units	0.17
Burgers	Schneck	2 units	0.17
Burgers	Paty	2 units	0.17
Butter	Lacteria	0.2 kg	0.23
Butter	Conaprole sin sal	0.2 kg	0.23
Butter	Calcar	0.2 kg	0.23
Butter	Kasdorf	0.2 kg	0.23
Cacao	Copacabana	0.5 kg	0.08
Cacao	Aguila	0.5 kg	0.08
Cacao	Vascolet	0.5 kg	0.08
Cacao	Saint	0.5 kg	0.08
Cheese	Cerros del Este	1 kg	0.21

Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe

Table A.4. Continued.

Product	Brand	Specification	Share in CPI (%)
Cheese	Dispnat	1 kg	0.21
Chicken	Avesur	1 kg	0.83
Chicken	Tenent	1 kg	0.83
Chicken	Avicola del Oeste	1 kg	0.83
Chicken	Melilla	1 kg	0.83
Chicken	Tres Arroyos	1 kg	0.83
Coffee	Aguila	0.25 kg	0.09
Coffee	Chana	0.25 kg	0.09
Coffee	Saint	0.25 kg	0.09
Coffee	Tropical	0.2 kg	0.09
Cola	Coca Cola	1.5 L	1.23
Cola	Pepsi	1.5 L	1.23
Cola	Nix	1.5 L	1.23
Cola	Coca Cola	2.25 L	1.23
Cola	Pepsi	2 L	1.23
Corn oil	Delicia	0.9 L	n/i
Corn oil	Rio de la Plata	0.9 L	n/i
Corn oil	Salad	1 kg	n/i
Cornmeal	Gourmet	0.45 kg	n/i
Cornmeal	Presto Pronta Arcor	0.5 kg	n/i
Cornmeal	Puritas	0.45 kg	n/i
Crakers	Famosa	0.14 kg	0.28
Crakers	Maestro Cubano	0.12 kg	0.28
Crakers	El Trigal	0.15 kg	0.28
Deodorant	Axe Musk	0.113 kg	0.34
Deodorant	Dove Original	0.1 kg	0.34
Deodorant	Rexona Active Emotion	0.105 kg	0.34
Diswashing detergent	Hurra Nevex Limon	1.25 L	0.13
Diswashing detergent	Deterjane	1.25 L	0.13
Diswashing detergent	Protergente limon	1 L	0.13
Dulce de leche	Conaprole	1 kg	0.14
Dulce de leche	Manjar	1 kg	0.14
Dulce de leche	Los Nietitos	1 kg	0.14
Fish	No Brand	1 kg	0.43
Flour	Cololo	1 kg	0.21
Flour	Canuelas	1 kg	0.21
Flour	Cololo	1 kg	0.21
Flour	Puritas	1 kg	0.21
Frankfurters	Cattivelli	8 Units – approx. 0.340 kg	0.23
Frankfurters	Ottonello	8 Units – approx. 0.330 kg	0.23
Frankfurters	Schneck	8 Units – approx. 0.330 kg	0.23
Frankfurters	Centenario	8 Units – approx. 0.33 kg	0.23
Frankfurters	Sarubbi	8 Units	0.23
Frozen fish	No Brand	1 kg	n/i
Grated Cheese	Conaprole	0.08 kg	0.16
Grated Cheese	El Trebol	0.08 kg	0.16
Grated Cheese	Milky	0.08 kg	0.16
Grated Cheese	Artesano	0.08 kg	0.16

Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe

Table A.4. Continued.

Product	Brand	Specification	Share in CPI (%)
Grit Noodles	Adria	0.5 kg	0.43
Grit Noodles	Las Acacias	0.5 kg	0.43
Grit Noodles	Puritas	0.5 kg	0.43
Ground Beef	No Brand - Cow	1 kg	0.98
Ground Beef	No Brand	1 kg	0.98
Ground Beef	Up to 20% Fat	1 kg	0.98
Ground Beef	Up to 5% Fat	1 kg	0.14
Ham	La Constancia	1 kg	0.16
Ham	Schneck	1 kg	0.16
Ham	Centenario	1 kg	0.16
Ice cream	Crufi	1 L	0.22
Ice cream	Conaprole	1 L	0.22
Ice cream	Gebetto	1 L	0.22
Laundry Soap	Nevex	0.8 kg	0.45
Laundry Soap	Drive	0.8 kg	0.45
Laundry Soap	Skip – Paquete azul	0.8 kg	0.45
Laundry Soap in Bar	Bull Dog	0.3 kg – 1 Unit	n/i
Laundry Soap in Bar	Nevex	0.2 kg – 1 Unit	n/i
Margarine	Doriana nueva	0.25 kg	n/i
Margarine	Primor	0.25 kg	n/i
Margarine	Danica dorada	0.2 kg	n/i
Margarine	Flor	0.25 kg	n/i
Mayonnaise	Hellmans	0.5 kg	0.21
Mayonnaise	Natura	0.5 kg	0.21
Mayonnaise	Fanacoa	0.5 kg	0.21
Mayonnaise	Uruguay	0.5 kg	0.21
Noodles	Cololo	0.5 kg	0.43
Peach Jam	Los Nietitos	0.5 kg	n/i
Peach Jam	Dulciora	0.5 kg	n/i
Peach Jam	Limay	0.5 kg	n/i
Peach jam	El Hogar	0.5 kg	n/i
Peas	Arcor	0.35 kg	0.09
Peas	El Hogar	0.35 kg	0.09
Peas	Trofeo	0.35 kg	0.09
Peas	Campero	0.3 kg	0.09
Peas	Cololo	0.38 kg	0.09
Peas	Nidemar	0.3 kg	0.09
Quince jam	Los Nietitos	0.4 kg	0.13
Quince jam	Limay	0.4 kg	0.13
Rice	Green Chef	1 kg	0.38
Rice	Blue Patna	1 kg	0.38
Rice	Aruba tipo Patna	1 kg	0.38
Rice	Pony	1 kg	0.38
Rice	Vidarroz	1 kg	0.38
Rice	Saman Blanco	1 kg	0.38
Salt	Sek	0.5 kg	0.09
Salt	Urusal	0.5 kg	0.09
Salt	Torrevieja	0.5 kg	0.09

Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe

Table A.4. Continued.

Product	Brand	Specification	Share in CPI (%)
Sausage	Cattivelli – Extra	1 kg	0.37
Sausage	La Familia – Hilo amarillo	1 kg	0.37
Sausage	Centenario – Extra	1 kg	0.37
Semolina Noodles	Adria	0.5 kg	0.43
Semolina Noodles	Las Acacias – franja celeste	0.5 kg	0.43
Shampoo	Sedal	0.35 L	0.36
Shampoo	Suave	0.93 L	0.36
Shampoo	Fructis	0.35 L	0.36
Soap	Astral	0.125 kg	0.16
Soap	Palmolive	0.125 kg	0.16
Soap	Suave	0.125 kg	0.16
Soap	Astral plata	0.125 kg	0.16
Soap	Rexona	0.125 kg	0.16
Soap	Primor	0.3 kg	n/i
Soybean Oil	Condesa	0.9 L	0.11
Soybean oil	Rio de la Plata	0.9 L	0.11
Soybean oil	Salad	0.9 L	0.11
Sparkling Water	Salus	2.25 L	0.82
Sparkling Water	Matutina	2 L	0.82
Sparkling Water	Nativa	2 L	0.82
Sugar	Azucarlito	1 kg	0.35
Sugar	Bella Union	1 kg	0.35
Sunflower oil	Optimo	0.9 L	0.37
Sunflower oil	Uruguay	0.9 L	0.37
Sunflower oil	Rio de la Plata	0.9 L	0.37
Tea	Hornimans	Box 10 Units	0.07
Tea	La Virginia	Box 10 Units	0.07
Tea	Lipton	Box 10 Units	0.07
Tea	President	10 Units	0.07
Toilet paper	Higienol Export	4 Unit – 25 m each	0.24
Toilet paper	Sin Fin	4 Unit – 25 m each	0.24
Toilet paper	Personal	4 Unit – 25 m each	0.24
Toilet paper	Elite	4 Units – 30 m each	0.24
Tomate pulp	Gourmet	1 kg	0.16
Tomato Paste	Qualitas	1 L	0.16
Tomato Paste	Conaprole	1 L	0.16
Tomato Paste	De Ley	1 L	0.16
Toothpaste	Colgate Total	0.09 kg	0.19
Toothpaste	Kolynos	0.09 kg	0.19
Toothpaste	Colgate Herbal Blanqueador	0.09 kg	0.19
Toothpaste	Closeup Triple	0.09 kg	0.19
Toothpaste	Kolynos Triple accion	0.09 kg	0.19
Toothpaste	Pico Jenner Plus	0.09 kg	0.19
Wheat Flour	Canuelas	1 kg	0.21
Wheat Flour	Primor	1 kg	0.21
Wine	Santa Teresa Clasico	1 L	0.79
Wine	Tango	1 L	0.79
Wine	Roses	1 L	0.79

Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe

Table A.4. Continued.

Product	Brand	Specification	Share in CPI (%)
Wine	Faisan	1 L	0.79
Yerba	Canarias	1 kg	0.64
Yerba	Sara	1 kg	0.64
Yerba	Envase Del Cebador	1 kg	0.64
Yerba	Del Cebador	1 kg	0.64
Yerba	Baldo	1 kg	0.64
Yogurt	Conaprole	0.5 kg	0.13
Yogurt	Parmalat (Skim)	0.5 kg	0.14
Yogurt	Calcar	1 L	0.14
Yogurt	Conaprole BIO TOP	1.2 L	0.14
Yogurt	Parmalat BIO YOGUR	1 L	0.14

Note: n/i means not included in the CPI, consumer price index; kg, kilograms; L, liters; and m, meters.
Source: Own elaboration from data of the Uruguayan Ministry of Economy and Finance.

ACKNOWLEDGEMENTS

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 Diego Aparicio for excellent research assistance and 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, and the 2012 Latin American Meeting of the Econometric Society in Peru for all their comments and suggestions. Borraz and Zipitría gratefully acknowledge the financial support of the 'Agencia Nacional de Investigación e Innovación (ANII)' under Grant # 'FMV_2_2011_1_6318'.

NOTES

1. For example, see Parsley and Wei (2001) for results between the US and Japan and Ceglowski (2003) for the effects of provincial borders in Canada.
2. A common alternative specification used by papers such as Engel and Rogers (1996) has the standard deviation $\sigma\left(p_{i,t} - p_{j,t}\right)$ instead. In both cases, the objective is to measure the effect of the right-hand side variables on price dispersion, which can be done either through the mean of the absolute value or the standard deviation of the price differences. Our results do not change if we use the standard deviation. See Broda and Weinstein (2008) for an overview of the papers that use these two regressions in the literature.
3. The estimation problem is analogous to estimating using inequality moments as opposed to equality moments. This area has received significant attention recently. See, for example, Andrews, Berry, and Jia (2004), Andrews and Guggenberger (2009), Andrews and Soares (2010), Andrews and Shi (2014), Ponomareva and Tamer (2011), and Rosen (2008).
4. A related idea in the context of trade can be found in Eaton and Kortum (2002) who propose estimating trade friction using the maximum price difference. Simonovska and Waugh (2014) criticizes the use of the maximum price difference in the estimating strategies, based on the possibility of bias of the estimator on finite samples.
5. The reason is that price gaps within the arbitrage constraint are less common for observations across cities; 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.
6. The retailer's website indicates that the online prices match those of the store where the orders are sent from, but it does not provide details on what specific store it is. In order to identify the most likely candidate, we compared daily online prices for all products with each offline store in the city and found a location where prices were identical 97.3 percent of the time. That location has an average distance of 7.2 km to all the other stores.
7. We require that v is large enough so that $u_\ell(i)$ is positive in at least one store.
8. See the proof in Appendix 6
9. The results from our model are also related to Samuelson (1954). Appendix 6.
10. Note that Simonovska and Waugh (2014) argues that the estimation of the transport costs can be downward biased if the maximum price difference is used, but we find a monotonic increase in this parameter as we use move from lower quintiles to the maximum price difference.
11. We also evaluate the robustness of our estimates to the elimination of price change outliers.
12. The same dataset is used in Borraz and Zipitria (2012).
13. The only exceptions are meat, eggs, ham, some types of cheese, and bread. However, as we later show, the exclusion of these goods, which could potentially be affected by an imperfect matching, does not modify the results.
14. http://www.precios.uy/servicios/ciudadanos.html.

Copyright © 2015 John Wiley & Sons, Ltd

15. See Borraz and Zipitria (2012) for a detailed description of the database and an analysis on its price stickiness.
16. We show the results for 10 km, but results remain qualitatively the same for stores 15 and 20 km apart. Given the characteristics of our data, it makes no sense to go beyond that distance because, in the city of Montevideo, there are very few observations with stores more than 20 km apart.
17. Future research should formally address the optimal bandwidth selection. For the moment, we compare the results across different specifications and do not explore the issue further because the results remain essentially identical. It is possible that if the estimation is done using less frequent data such as month by month or using a much smaller dataset; then the issue of bandwidth selection becomes more important. This was not the case in our application.
18. See Cavallo (2010) for additional details on the online data scraping methodology.
19. Samuelson (1954).
20. For simplicity in the exposition, it is assumed that the arbitrageur is the customer itself. Thus, the arbitrage cost can be interpreted not only as the loss of physical items but also as the loss in terms of utility that the customer would experience if were forced to travel from one location to another.

REFERENCES

Anderson J, van Wincoop E. 2003. Gravity with gravitas: a solution to the border puzzle. *American Economic Review* **93**(1): 170–192.

Andrews DWK, Berry S, Jia P. 2004. *Working Paper, Yale University*, Confidence regions for parameters in discrete games with multiple equilibria, with an application to discount chain store location.

Andrews DWK, Guggenberger P. 2009. Validity of subsampling and "plug-in asymptotic" inference for parameters defined by moment inequalities. *Economic Theory* **25**(03): 1–41.

Andrews DWK, Shi X. 2014. Nonparametric inference based on conditional moment inequalities. *Journal of Econometrics* **179**(1): 31–45.

Andrews DWK, Soares G. 2010. Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica* **78**(1): 119–157.

Borraz F, Zipitria L. 2012. Retail price setting in uruguay. *Economia* **12**(2): 77–109.

Broda C, Weinstein DE. 2008. *NBER Working Paper 14017*, Understanding international price differences using barcode data.

Cavallo A. 2010. *MIT Sloan Working Paper 4976-12*, Scraped data and sticky prices.

Ceglowski J. 2003. The law of one price: intranational evidence for Canada. *The Canadian Journal of Economics* **36**(2): 373.

Eaton J, Kortum S. 2002. Technology, geography, and trade. *Econometrica* **70**(5): 1741–1779.

Engel C, Rogers JH. 1996. How wide is the border? *American Economic Review* **86**(5): 1112–1125.

Evans CL. 2003. The economic significance of national border effects. *American Economic Review* **93**(4): 1291–1312.

Goldberg PK, Knetter MM. 1997. Goods prices and exchange rates: what have we learned? *Journal of Economic Literature* **35**(3): 1243–1272.

Gopinath G, Gourinchas PO, Hsieh CT, Li N. 2011. International prices, costs, and markup differences. *American Economic Review* **101**(6): 2450–86.

Gopinath G, Rigobon R. 2008. Sticky borders. *The quarterly journal of economics* **123**(2): 531–575.

Gorodnichenko Y, Tesar L. 2009. Border effect or country effect? Seattle may not be so far from Vancouver after all. *American Economic Journal – Macroeconomics* **1**(1): 219–241.

Head K, Mayer T. 2002. *CEPII Research Center, Working Paper 2002-01*, Illusory border effects: distance mismeasurement inflates estimates of home bias in trade.

Hillberry R, Hummels D. 2003. Intranational home bias: some explanations. *The Review of Economics and Statistics* **85**(4): 1089–1092.

Klenow PJ, Kryvtsov O. 2008. State-dependent or time-dependent pricing: does it matter for recent U.S. inflation? *The Quarterly Journal of Economics* **123**(3): 863–904.

Parsley DC, Wei SJ. 2001. Explaining the border effect: the role of exchange ratee variability, shipping costs, and geography. *Journal of International Economics* **55**(1): 87–105.

Ponomareva M, Tamer E. 2011. Misspecification in moment inequality models: back to moment equalities?. *The Econometrics Journal* **14**(2): 186–203.

Rosen AM. 2008. Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities. *Journal of Econometrics* **146**(1): 107–117.

Samuelson PA. 1954. The transfer problem and transport costs, ii: analysis of effects of trade impediments. *The Economic Journal* **64**(254): 264–289.

Simonovska I, Waugh ME. 2014. The elasticity of trade: estimates and evidence. *Journal of International Economics* **92**(1): 34–50.

Copyright © 2015 John Wiley & Sons, Ltd

DOI: 10.1002/ijfe