Estimating the Border Effect: Some New Evidence*

Gita Gopinath
Harvard University

Pierre-Olivier Gourinchas
University of California at Berkeley

Chang-Tai Hsieh
University of Chicago

Nicholas Li
University of California at Berkeley

This Version: April 21, 2009

Abstract

To what extent do national borders and national currencies impose costs that segment markets across countries? To answer this question we use a dataset with product level retail prices and wholesale costs for a large grocery chain with stores in the U.S. and Canada. We develop a model of pricing by location and employ a regression discontinuity approach to estimate and interpret the border effect. We report three main facts: 1) The median absolute retail price and whole-sale cost discontinuity between adjacent stores on either side of the U.S.-Canada border is as high as 21%. In contrast, within-country border discontinuity is close to 0%; 2) The variation in the retail price gap at the border is almost entirely driven by variation in wholesale costs, not by variation in markups; 3) The border gap in prices and costs co-move almost one to one with changes in the U.S.-Canada nominal exchange rate. We show these facts suggest that the price gaps we estimate provide only a lower bound on border costs.

*We are grateful to Paul Bergin, Stefano della Vigna, Jack Duane, Charles Engel, Edward Glaeser, Yuriy Gorodnichenko, Guido Imbens, Ariel Pakes, John Rogers and David Sraer for valuable discussions. We thank Kevin Devereux, Michal Fabinger, Robert Johnson, Lorenz Küng, Gloria Sheu, Kelly Shue and Synuhe Uribe for excellent research assistance. We gratefully acknowledge financial support from the National Science Foundation through grants SES0820468 and SES0820241. Contact email: gopinath@harvard.edu.

1 Introduction

To what extent do national borders and national currencies impose costs that segment markets across countries? Some of the central questions in international economics, ranging from the transmission of shocks across borders to the gains from regional integration hinge on the answer to this question.

There is little doubt that borders generate additional transaction costs, from the use of different monies to the regulatory costs of obtaining national permits, that can translate into price differences. In addition, national borders delineate different economic environments: variation in national tastes, market conditions, wages, transportation infrastructures among others can generate additional sources of price differences. Further, the effect of border costs varies by market. From a consumer's perspective fairly small transaction costs can effectively segment markets. By contrast, at the whole-sale level, given the large volumes involved, the gains to arbitraging even small price gaps are large. The relevant question then is about identifying the factors that generate the 'border effect' and the magnitude of this effect.

We address these questions, by making the following contributions. First, we use unique weekly data on retail prices and wholesale costs for detailed products at the barcode level from a large grocery chain with stores in the US and Canada to measure the effect of the border on market segmentation. We present evidence of the impact of border costs at both the consumer and whole-sale level. Second, we develop a model of pricing by stores located on a circle to document possible patterns of cross-border prices. We then employ a regression discontinuity approach to estimate the border effect, exploiting the information we have about the geographical location of stores.

More specifically, our dataset has information on retail prices and wholesale costs for 250 U.S. stores (in 19 states) and 75 Canadian stores (in 5 provinces). Prices and costs are observed at the Universal Product Code level (UPC or barcode) for 178 weeks between January 2004 and June 2007. This alleviates concerns about compositional effects that arise with more aggregated price index data. From this dataset, we extract a sample of 4,221 identical products sold by this retailer in the two countries. We find that the median gap (across the UPCs) between the average price and cost in Canada and the US increased from

-5% in June 2004 to 15% in June 2007, and that the variation in this gap closely track the U.S.-Canada nominal exchange rate. By contrast, the median mark-up deviation remains largely unchanged over this period.

While these raw facts are indicative of a significant economic effect of the border and departure from the law of one price, a comparison of averages masks potential differences in market conditions and arbitrage costs for US and Canadian stores close to and far away from the border. We address this issue with a stylized model of pricing by stores located on a circle, along the lines of Salop (1979)'s circular city model. The model endogenously determines the distribution of prices within and across countries in the presence of a border cost and heterogeneity in marginal costs across countries. It delivers two important insights. First, the impact of border costs is observed only through changes in prices "close" to the border, and has little effect on pricing decisions "far" away from the border, a distinction often overlooked in the empirical literature. Second, when border costs become sufficiently large, markets are fully segmented across countries, and the magnitude of border costs does not affect pricing decisions any longer. In that case, price differences at the border provide a *lower bound* for border costs and move one-to-one with cost differences, while markup differences remain almost unchanged. Thus, the model has the potential to account for the stylized facts exhibited in the data.

We then exploit the central prediction of the model – that the impact of the border on prices depends on the distance to the border – to estimate the effect of the border using a regression discontinuity (RD) design. The RD approach has been popularized in recent years to estimate the causal effect of treatment in a variety of settings, and we apply it to the question of border effects. We use the precise geographic location of the stores in our data to answer the following questions: Do we see deviations from the law of one price between stores located right across the border from each other? To illustrate, Figure 1 plots the (log) average price across stores (in 50 km. bins) for Perrier Sparkling Natural Mineral Water, 25fl. oz against the distance of the store from the border in the first week of 2004. As is evident, there is a clear discontinuity at the border that is indicative of the treatment of the border. The RD design controls for the fact that stores located far apart can face very different market conditions or arbitrage costs compared to stores located close to one another. A significant price discontinuity as we cross the border is then interpreted as the local effect of the border. We then decompose the border discontinuity in prices into a discontinuity in costs and in markups.

We report three main findings from the RD estimates. First, we observe large and heterogenous discontinuities at the border across products for retail prices, wholesale costs and markups. The median price discontinuity (across UPC's) is as high as 15% for prices and 17% for wholesale costs while the median absolute price discontinuity is 21% for prices and 21% for costs. The standard deviation across UPC's is large, indicating that the discontinuity at the border varies from large and positive to large and negative across goods. Second, the median retail and wholesale price discontinuities at the border move one to one with the U.S.-Canada nominal exchange rate. The Canadian dollar appreciated (in cumulative terms) by 16% over our sample period. Over the same period both the median retail price and wholesale cost discontinuities increased by almost 12%. Third, the mark-up discontinuity remained mostly unchanged over the sample period. These last two findings are consistent with a full segmentation of retail markets between the U.S. and Canada over the period of our study and the set of goods in our sample.

We probe the robustness of our results in four ways. First, we restrict the sample to stores located in Oregon, Washington and British Columbia, and find that the estimates are unchanged. Second, we expand the number of products we consider by creating store-level price indices calculated over finely disaggregated sub-categories of goods. Here, we compare the discontinuous change in the price indices at the border and find similar results. Third, we compare the behavior of costs for store-branded products to our benchmark estimates to examine whether our cost data is allocative. As expected, we find much less co-movement between relative costs and the exchange rate for the store branded products. Fourth, we contrast our results for the U.S.-Canada border with similar estimates within country. We focus on the Washington-Oregon border and find almost no evidence of a discontinuity at the border.

This paper builds on the large literature on the segmentation of retail markets across countries. This literature has generally found deviations from the law of one price that are large, volatile, and remarkably correlated with the nominal exchange rate.[1] In particular, a seminal paper by Engel and Rogers (1996) shows that the volatility of changes in price indices for disaggregated product categories between U.S. and Canadian cities are much larger than that observed across cities within the same country. A large literature has followed up on Engel and Rogers (1996)'s influential paper by studying goods at a more disaggregated level.[2] In this respect, our paper is related to the work of Broda and Weinstein (2007) who use a large amount of barcode-level price data collected at the consumer level and find a similar degree of price segmentation across and within borders. While the level of disaggregation in Broda and Weinstein (2007) is similar to that in the data we use, a key difference is that our data captures prices charged by the *same* retailer in all locations, while the Broda and Weinstein (2007) data contains prices at which consumers purchase a particular good in a given location without any control for retailer heterogeneity.[3]

Another main difference from these papers and many others in the literature is the use of the regression discontinuity approach, which directly addresses an important critique raised by Gorodnichenko and Tesar (2009). That paper points out that estimates of the border effect in regressions like those used by Engel and Rogers (1996) are not identified. Heterogeneity in price determining factors, such as variation in demand can generate price dispersion that have little to do with border costs. Standard regressions will incorrectly attribute the difference to border costs. Our paper directly addresses the critique laid out in Gorodnichenko and Tesar (2009) as it both develops a stylized model of price determination across the border and employs a regression discontinuity approach that exploits critical information about the geographical location of stores.[4]

Sections 2 and 3 describe the data and present preliminary evidence on the pattern of

[1]See Rogoff (1996) and Goldberg and Knetter (1997) for comprehensive reviews of this literature.

[2]Crucini and Shintani (2006) and Crucini, Telmer and Zachariadis (2005) for instance examine the retail price of narrowly defined product categories such as "Washing Powder", across countries within the European Union. Others focused on specific goods, such as *The Economist* magazine (Ghosh and Wolf (1994)) *Ikea*'s furniture products (Haskel and Wolf (2001); Hassink and Schettkat (2001)), or Scandinavian *duty-free* outlets (Asplund and Friberg (2001)). Parsley and Wei (2007) decompose the price of a *Big Mac* across countries into variation in marginal costs and variation in markups. Goldberg and Verboven (2005) study the automobile car market in Europe. See Goldberg and Verboven (2001) for a survey.

[3]See Nakamura (2008) for novel evidence on pricing across and within retailer at the upc level.

[4]In contemporaneous work, Burstein and Jaimovich (2008), also examine the pattern of prices in the U.S. and Canada using the same dataset. Unlike us, Burstein and Jaimovich (2008) take as given that markets are segmented and do not address the question of measuring border costs.

prices, costs and mark-ups in the U.S. and Canada. Section 4 describes the circular world model while section 5 discusses the regression discontinuity design and the estimates of border costs across countries. Section 6 presents additional results and Section 7 concludes. All proofs are relegated to the Appendix.

2 Data Source

We have access to weekly store-level data for a sample of 325 grocery stores from a large retail chain that operates in the United States and Canada. This chain is one of the leading food and drug retailers in North America. It operates directly or through its subsidiaries nearly 1800 stores in a broad range of geographic locations and socio-economic neighborhoods (1400 stores in the US and 400 in Canada).[5]

The data set contains weekly total sales, quantities sold, wholesale unit cost as well as a measure of per-unit gross profit for 125,048 unique Universal Product Codes (UPC) in 61 distinct product groups, for 250 stores in 19 US states, and 75 stores in 5 Canadian provinces, during 178 weeks between January 2004 and June 2007.[6]

Figure 2 plots the location of the stores in our data. Most US stores are located in the Western and Eastern corridors, in the Chicago area, Colorado and Texas, while most Canadian stores are located along a relatively narrow horizontal band running close to the border with the US.

The total number of observations across stores and time is close to 40 million. Most of these observations are concentrated in the processed and unprocessed food and beverages categories, housekeeping supplies, books and magazines, and personal care products. Column 1 of table 4 in the Appendix reports a breakdown of available UPCs by product categories. This level of disaggregation allows for a very precise identification of products. For instance, in our data, a 25 fl.oz Perrier Mineral Water with a Lemon Twist and a 25 fl.oz Perrier

[5] The data sharing agreement between this retailer and the research community is managed through the SIEPR-Giannini data center (http://are.berkeley.edu/SGDC/).

[6] All UPCs fall within the following structural hierarchy: (a) Business Group (e.g. DRF, Dairy, Refrigerated, Frozen Foods); (b) Business Unit (Dairy and Refrigerated Foods); (c) Product Group or 2-digit SMIC (36 Refrigerated Dairy); (d) Category or 4 digit SMIC (3601 - Milk/Milk Substitutes); (e) Class or 6 digit SMIC (3601 01 - Mainstream Milk); (f) Subclass or 8 digit SMIC (3601 01 01 - Whole Milk); (g) Subsubclass or 10 digit SMIC (3601 01 01 05 - 1/2 Gallon Whole Milk).

Mineral Water with a Lime Twist are two separate members of the Soft Beverages product group.

Of the 125,048 unique UPCs in our dataset, our first task is to find the set of 'matched UPCs', i.e. the set of products that appear in identical form in at least one Canadian and one US store, in at least one week.[7] It represents the set of goods for which we can evaluate deviations from the law of one price. This dataset of matched UPCs contains 4,221 unique products, or about 3.3% of the original dataset.[8] This decline in matched products across the border is an important effect emphasized in Broda and Weinstein (2007) that carries across to our dataset. It underlies the importance of working with unique products, as identified by their UPCs. When comparing price indices across countries at higher levels of aggregation, one is likely to suffer from a serious composition bias. One concern is that otherwise identical goods have different UPC's because of different labeling requirements in the U.S. and Canada (e.g. language, nutrition). To assess this we conducted a physical survey of the matched UPC's in one store in the U.S. (Oakland) and in Canada (Vancouver). We found that for most of them the labeling was indeed different but the UPC was the same. Consequently, different labeling does not necessarily imply different UPC's, but it could still be a factor behind the low match rate.[9]

The set of matched UPCs are concentrated in books and magazine (2,505), alcoholic beverages (403), Ethnic & gourmet food (306) and household cleaning products (159).[10] The distribution of products across product groups is very skewed, with a median around 11 products and a mean of 97.[11]

Table 1 reports information on the number of distinct products (among matched goods)

---

[7] There are 98,430 unique UPCs in the U.S., and 32,961 unique UPCs in Canada. The total number of price observations across stores and time is close to 40 million.

[8] We arrive at this number in the following way. We start with the set of unique UPCs that appear in at least one US and one Canadian store (6,343). We check the product descriptions to ensure that the products are identical (6,283). We further drop UPCs with less than 10 digit since these are generated internally by the retail chain and may not be consistent across countries (5,900). We further eliminate products in the fresh bread/baked goods, deli, food service, produce, seafood, meat and floral arrangements categories since these goods contain a higher local labor content and are not available in identical form in different stores (4,221 goods).

[9] Matching goods that do not share the same UPC is an impossible task given the limited product information we have.

[10] 'Books and Magazines' have a printed sale price that is sticky in the local currency. We find that all our results hold similarly for the sample that excludes this category of goods.

[11] See column (4) of table 4 for a breakdown of matched upcs by product groups.

per store-week and per store-pair-week in our data. The average store in the data carries 492 distinct matched products for which we have data in a typical week. We find about 272 (243) matched products for a typical within country store pair in the U.S. (Canada) in a given week, and 164 for a cross-border store-pair.

The data set contains information on "Gross" and "Net" sales. We construct corresponding gross and net prices for each UPC by dividing sales by quantities. Both retail prices exclude (US) sales and (Canadian) federal value-added taxes and provincial sales taxes.[12] The net price can differ from the gross price when there are sales (coupons, promotions). It is always smaller than or equal to the gross price and exhibits significantly more variability.

We also have information on the "Whole-sale Cost" which refers to the list price at which the retailer purchases a given product from the wholesaler. These costs need not represent the true cost to the retailer given that there are typically freight and transport costs as well as retail allowances, i.e. rebates provided by the wholesaler to the retailer or vice versa. To correct for this, we use data on "Adjusted Gross Profits" per unit to back out the "Net Cost", or imputed cost of goods. The precise link between the whole-sale cost and the net cost is as follows:[13]

$$\begin{aligned} \text{Net cost} &= \text{Whole-sale cost} - \text{Allowances} \qquad (1) \\ &= \text{Net Price} - \text{Adjusted Gross Profits} \end{aligned}$$

At short horizons, with rent, capital and labor taken as given, it is natural to interpret the net cost as a measure of the marginal cost of goods. Equivalently, "Adjusted Gross Profits" measure the mark-up at the product and store level. At longer horizons, adjusted gross profits represent an upper bound for the product mark-up.

[12]From a consumer's perspective the relevant price is the price inclusive of sales and V.A.T. tax. We do not have this tax information which varies by UPC and location both within and across countries. For instance, many food products are exempt from sales tax both in the US and Canada. On the other hand, we found that sales and value added taxes are higher in British Columbia (13%) as compared to Washington State (around 8%). To the extent that before tax prices are higher in Canada than in the US, as we find for most goods, this implies that the after tax price gap is even larger than what we measure.

[13]According to information provided by our retailer, allowances consist of the sum of shipping allowances, scan allowances, direct-store-delivery case billback allowances, header flat allowances, late flat allowances, new item allowances, *minus* the sum of buying allowances, freight allowances, overseas freight and distress and other allowances.

It is natural to question whether our measure of net costs is allocative. Among other things, this requires that freight, transport and retail allowances are measured correctly at the product and store level. There are a number of reasons why this might be difficult, potentially affecting our measure of marginal costs.[14] Since we don't have a breakdown of allowances between their different components, we cannot directly address this question. Nevertheless we propose in section 6.2 an indirect way to assess whether our cost measure is allocative by comparing whole-sale costs and net costs for store brand products and other products. Since our retailer controls most of the supply chain for store-brand products, we would expect –as we indeed find– a very different behavior of cross-border relative costs for these two categories of products, especially for whole-sale costs. This is consistent with the view that our cost measures are indeed allocative.

3 Preliminary analysis of LOP deviations at the border

3.1 Median deviations over time

As a first pass at the data, the top left part of Figure 3 reports the median average cross-border price gap over time. That is, for each week and each UPC, we compute the log-deviation between the average Canadian and US net prices across stores. The figure reports the median of that distribution across UPCs, over time. When positive, this number indicates that more Canadian goods have a higher average price than the corresponding US good. The figure indicates that the median price gap has increased over time from roughly -5% in June 2004 to 15% in June 2007. The figure also reports (the dashed line on the right-axis) the (log) US/CAN nominal exchange rate. The overall correlation between the two series is striking: the evolution over time in the median price gap mirrors almost perfectly the evolution of the nominal exchange rate.

Using the identity:

$$\ln pk = \ln\left(\text{cost}^{k}\right) + \ln\left(\text{markup}k\right),$$

[14]E.g., one may imagine that a soft drink manufacturer negotiates global allowances on a broad range of drinks sold to the retailer; similarly, it may be difficult to assess the transport cost & freight component for a single bottle of milk.

where p^k is the price of good k,$cost^k$ is the marginal cost and markup^k denotes the markup, the top-right and bottom left panels perform the same exercise for the imputed (net) cost and the resulting markup. Looking at the two figures side by side reveals a striking fact: most of the movements in the median price gap result from corresponding movements in relative costs, while relative mark-ups show barely any response to the fluctuations in the exchange rate. This result is robust to the definition of the price (gross versus net) or of the costs (wholesale versus imputed).

Prices in our sample change very frequently. The median frequency across UPC's is 0.41 for net prices (0.22 for gross prices), implying a median duration of 2.4 (4.5) weeks.[15] Despite the high median frequency, a significant fraction of goods do not change price during the entire sample. To ensure that these goods do not drive the results we divided the sample into high- and low-frequency adjusters depending on whether their frequency of price adjustment is above or below the median. In both cases, we found that the median price gap increases over time.[16]

Overall, the evidence indicates that the median price gap moves closely with the nominal exchange rate and that cost differences play an important role.

3.2 Dispersion Across UPC's

Figure 4 sheds light on the extent of dispersion of price gaps across UPCs at a point in time. Figure 4(a) reports the distribution of the cross-border net price gap across UPCs for the first week of 2004 (2242 UPC's) and the twenty first week of 2007 (2267 UPC's).[17] The figure shows that there is a large dispersion of price gaps across UPCs at any given point in time. Hence, while the median moves closely with the exchange rate, the price gap for any individual UPC is likely to be dominated by idiosyncratic factors, a feature also documented in Crucini and Shintani (2007).

[15] The frequency number was arrived at as follows: we estimated the frequency of price adjustment for each UPC-store combination; Then we estimated the average frequency across these store combinations for each UPC. We then estimated the median within each category and the median across these categories.

[16] The contribution of imputed costs is smaller for the frequent adjusters. Finally, the median markup gap movements are small relative to prices and costs. These additional results are available from the authors upon request.

[17] This corresponds to the beginning and (end-1) weeks of our sample. In the 22nd week of 2007 there is a significant drop in the number of upc's given to us, which is why we use the 21st week.

Figures 4(b) and 4(c) report the same distribution for the cross-border average imputed cost gap and markup gap. The figures indicate significant dispersion in relative costs across the border, but a much tighter distribution of markup differences across the border. The distributions for price and cost shift to the right between 2004 and 2007 alongside the appreciation of the Canadian dollar.

3.3 Dispersion Across Stores

Finally, table 2 reports some raw statistics for the extent of price dispersion within and across US and Canadian stores. Panel A reports statistics for the net price charged across stores in the US and Canada in the first week of 2005. USA-USA (resp. CAN-CAN) reports prices for store-pairs located within the US (resp. Canada), while CAN-US examines prices for cross-border store pairs. With 250 US stores and 75 Canadian stores, there are 31,125 US-US store-pairs, 2,775 CAN-CAN ones and 18,750 cross border pairs. Define p_i^k as the gross US dollar price of product k in store i. We construct the (log) price gap between two stores i and h for good k as $\ln\left(p_i^k/p_h^k\right)$.

The median number of common UPCs for store pairs is 373 (405) within the US (Canada) and 248 for cross-border pairs.[18] Columns (1)-(3) report the mean, median and standard deviation of price differences for store-pairs for the first week of 2005. Statistics of this distribution are reported in the rows. The median across store-pairs of the median price gap is 0 for this week both within US and within Canada pairs. This result corroborates the evidence in Crucini et al. (2005) and Broda and Weinstein (2007) that price differentials are centered around zero within countries in some periods. Cross-border store pairs however have a large median gap of 12 percentage points. Since the U.S. store is always treated as store of reference, this implies that Canadian retail prices were 12 percent higher than US prices in the first week of 2005.[19]

Similarly, the median absolute price gap (column 5) is larger for cross-border stores (15 percent) as compared to either the within-US (3.7 percent) or the within-Canada (0 percent)

[18]The median number of UPCs differs from the numbers in table 1 because we are only looking at a single week of data.

[19]Since these are pre-tax prices, the 7% Canadian value-added tax (or GST) cannot account for the result.

pairs. The fact that there is less price dispersion within-Canada as compared to within-US is also consistent with the evidence in Engel and Rogers (1996) and Gorodnichenko and Tesar (2009) and unlike Broda and Weinstein (2007). Panel B indicates similar results for the median absolute imputed cost gap: it is much larger for cross-border store pairs (18 percent) as compared to within-U.S. store pairs (1 percent) and within-Canada pairs (0 percent). This difference is small for mark-ups.

While these raw statistics of the failure of the law of one price are indicative of border costs, there are other reasons for these differences that one needs to control for. One popular approach to estimating the border effect consists in estimating regressions of some measure of deviation from the law of price across store-pairs against distance between stores and a border dummy, along the lines of Engel and Rogers (1996).[20] As argued by Gorodnichenko and Tesar (2009), estimates of the border effect from these regressions are generically not identified. In particular, cross-country heterogeneity in price determinants can generate price dispersion that have little to do with border costs. Standard regressions will incorrectly attribute these differences to border costs. Another important issue is that market conditions and arbitrage costs may be very different for US and Canadian stores located close or far away from the border, a feature that is not captured by the median price gaps described above, or the usual border regressions.

In the next sections we address both issues by presenting a model of price determination as a function of the distance to the border, along with other usual price determinants (such as costs, demand, market structure). The analysis both motivates a departure from the standard regressions used in the literature to one that uses a regression discontinuity design and helps interpret the estimated 'border effect'.

[20] For comparison with the previous literature we estimated border regressions similar to Engel and Rogers (1996) and Broda and Weinstein (2007). We find that the 'border coefficient' is both sizeable and robust to the exclusion of within-country store pairs in Canada or in the U.S. However, we depart from this regression framework in the rest of the paper for reasons discussed below.

4 Circular World

We present a stylized model that endogenizes the distribution of prices across locations in the presence of border costs. The model is a two-country version of Salop (1979) circular city model of horizontal differentiation. We define a location as a position indexed by $\omega \in [0, 1]$ on a circle of unit length. A border splits the circle into two countries (country A and country B). Figure 5 provides a graphical representation of this circular world.

4.1 Stores

There are $N_{AB} = N_A + N_B$ retail stores located at exogenous equidistant intervals along the circle, with N_A stores in country A and N_B stores in country B. The borders are located at $\omega = 0$ and $\omega = N_A/N_{AB}$. We refer to stores by their location, parameterized by the variable ω_i where $i \in \{1, .., N_{AB}\}$, with $\omega_i = (2i - 1)/2N_{AB}$. The stores closest to the border are $i = 1,\ N_A$ for country A and $i = N_A + 1,\ N_{AB}$ for country B. We further assume that each store sells a homogenous good (same upc) and sets the price of this good independently.[21]

4.2 Consumers

We assume that a unit mass of consumers is uniformly distributed on the unit circle. Each consumer buys one unit of the good and, all else equal, strictly prefer to shop in stores that are closer to their location. They incur a cost $t \geq 0$ per unit of distance traveled that reflects transportation costs or the consumer's value of time, as well as a cost $b \geq 0$ when crossing the border. The utility of a consumer located at ω and shopping in store i is given by

$$u(\omega) = \nu - \theta p - t|\omega_i - \omega| + bI(\omega_i, \omega)$$

Here, $I(\omega_i, \omega)$ is an indicator function for whether the consumer and store are in different countries, θ captures the own price elasticity of demand and t is inversely related to the

[21] This assumption may seem at odds with our data, which consists of stores operated by a single retail chain. Yet this is a reasonable assumption that captures the notion that pricing decisions in any given location are more influenced by the pricing decision of competitors in the immediate vicinity than by pricing decisions of stores belonging to the same chain located further apart. In our model, if we assume that the particular retail chain we have data from operates every other store along the circle, then each store in the chain behaves exactly as an independent store.

degree of substitutability across store locations. We assume that ν is large enough so that all consumers purchase one unit of the good in equilibrium.

4.3 Costs

The marginal cost of goods in location i is

$$c_i = \begin{cases} \min\{\chi_A, \chi_B + b_c\}, & \text{if} \quad i \in A \\ \min\{\chi_B, \chi_A + b_c\}, & \text{if} \quad i \in B \end{cases}$$

Here, χ_j denotes the wholesale price of the good in country j and $b_c \geq 0$ is the border cost to the retailer. Note that it will always be the case that c_i is the same for all stores in the same region.

4.4 Demand Functions

We solve for the equilibrium distribution of prices in the following manner. We first solve for the profit maximizing price for interior stores, defined as stores not adjacent to the border. We then consider the profit maximizing prices of the border stores. If we assume that the parameters of the model are such that all stores earn positive profits in equilibrium, this implies that consumers will not shop at stores that are further than $1/N_{AB}$ from their own location. In particular, between any pair of stores i and $i-1$, there will be a marginal consumer indifferent between shopping at either store.

4.4.1 Interior Stores

Consider an interior store i in country j. Given the previous discussion, the total demand for products at that store is[22]

$$D_i(p_{i-1}, p_i, p_{i+1}) = \frac{1}{N_{AB}} + \frac{p_{i+1} - 2p_i + p_{i-1}}{2t}.$$

[22] This is derived by finding the location of the marginal consumers between store i and $i-1$ and between stores i and $i+1$.

That store chooses its price p_i to maximize static profits:

$$\pi_i = (p_i - c_j) D_i(p_{i-1}, p_i, p_{i+1}), \tag{2}$$

taking p_{i-1} and p_{i+1} as given. The following proposition characterizes the distribution of interior prices.

Proposition 1 *The distribution of interior prices takes the following form*

1. *For stores in the interior of country* A :

$$p_i = \left(\hat{p}_A - c_A - \frac{t}{N_{AB}}\right) \cdot \frac{\cosh\left(\kappa(i - \frac{N_A+1}{2})\right)}{\cosh\left(\kappa(\frac{N_A-1}{2})\right)} + c_A + \frac{t}{N_{AB}}, \tag{3}$$

2. *For stores in the interior of country* B :

$$p_i = \left(\hat{p}_B - c_B - \frac{t}{N_{AB}}\right) \cdot \frac{\cosh\left(\kappa(i - N_A - \frac{N_B+1}{2})\right)}{\cosh(\kappa(\frac{N_B-1}{2}))} + c_B + \frac{t}{N_{AB}}. \tag{4}$$

In the expressions above, cosh denotes the hyperbolic cosine function, $\kappa \equiv \cosh^{-1} 2 \approx 1.317$ is a constant, $\hat{p}_A = p_1 = p_{N_A}$ represents the price in the border store in country A and $\hat{p}_B = p_{N_{AB}} = p_{N_A+1}$ represents the price in the border store in country B.[23]

As equations (3) and (4) indicate, prices are increasing in marginal costs c_i, decreasing in the elasticity of substitution across locations $(1/t)$ and the total number of stores N_{AB}, and increasing in the price of the store located at the border $\hat{p}_A$ and $\hat{p}_B$. Importantly, the border cost only affects prices of interior stores through its effect on prices at the border stores, and this effect decreases with the distance from the border.

4.4.2 Border Stores

The final step is to characterize the prices of the border stores, $\hat{p}_A$ and $\hat{p}_B$. We consider two cases: (a) full market segmentation, for the case where border costs are large enough relative to the equilibrium price gap across the border such that consumers do not cross the border; (b) partial market segmentation, for the case when some consumers cross the border.

[23] The hyperbolic cosine function is given by $cosh(x) = (e^x + e^{-x})/2$.

The following set of propositions characterizes border prices in these two cases, if such equilibria exists.[24]

Proposition 2 *[Full Segmentation] If the marginal consumer is at the border, that is*

$$|\hat{p}_A - \hat{p}_B| < b$$

then national markets are fully segmented and:

(i) the prices of stores at the border are given by

$$\hat{p}_A = c_A + \frac{t}{N_{AB}} \frac{3 - \nu_A}{2 - \nu_A}; \qquad \hat{p}_B = c_B + \frac{t}{N_{AB}} \frac{3 - \nu_B}{2 - \nu_B} \tag{5}$$

where

$$\nu_A = \frac{\cosh \kappa(\frac{N_A - 3}{2})}{\cosh \kappa(\frac{N_A - 1}{2})}; \qquad \nu_B = \frac{\cosh \kappa(\frac{N_B - 3}{2})}{\cosh \kappa(\frac{N_B - 1}{2})}$$

(ii) The difference in prices of border stores moves one to one with the difference in costs, i.e. $\partial(\hat{p}_A - \hat{p}_B)/\partial(c_A - c_B) = 1$.

Proposition 2 corresponds to the case where the difference in prices between border stores, $|\hat{p}_A - \hat{p}_B|$, is smaller than the border cost b. In this case the demand functions are independent of prices on the other side of the border, and markets are completed segmented. The observed difference in prices at the border is also independent from the border cost b, and only provides a lower bound on its true value.

Proposition 3 *[Partial Segmentation]*

(i)If the marginal consumer for the border stores is located in country A, that is

$$\hat{p}_A - \hat{p}_B > b, \tag{6}$$

[24] Whether partial or full market segmentation exists in equilibrium depends in a nontrivial way on the parameters of the model. Checking for all the conditions for a particular equilibrium to exist in our multi store set-up is a complicated theoretical problem, largely orthogonal to our main purpose. In a simpler environment with two symmetric stores located on a line and linear transportation costs d'Aspremont, Gabszewicz and Thisse (1979) show that the profit function has two discontinuities and under some conditions an equilibrium may not exist. Existence requires that both firms not be located too close to each other. In our setting the number of possible demand scenarios faced by a firm increases relative to the case analyzed in d'Aspremont et al. (1979) because there are more than two firms, the shape of the profit function varies with the border cost parameter and with differences in costs.

then markets are partially segmented and the prices of stores at the border are given by

$$\hat{p}_A = \frac{(4-\nu_B)(j_A+b)+(j_B-b)}{(4-\nu_A)(4-\nu_B)-1}; \qquad \hat{p}_B = \frac{(4-\nu_A)(j_B-b)+(j_A+b)}{(4-\nu_A)(4-\nu_B)-1} \tag{7}$$

where ν_A and ν_B are as before and

$$j_A = (3-\nu_A)\left(c_A + \frac{t}{N_{AB}}\right); \qquad j_B = (3-\nu_B)\left(c_B + \frac{t}{N_{AB}}\right).$$

(ii) If the marginal consumer for the border stores is located in country B, that is

$$\hat{p}_B - \hat{p}_A > b,$$

then markets are partially segmented and the prices of stores at the border are given by

$$\hat{p}_A = \frac{(4-\nu_B)(j_A-b)+(j_B+b)}{(4-\nu_A)(4-\nu_B)-1}; \qquad \hat{p}_B = \frac{(4-\nu_A)(j_B+b)+(j_A-b)}{(4-\nu_A)(4-\nu_B)-1} \tag{8}$$

The last proposition illustrates the case when $|\hat{p}_A - \hat{p}_B| > b$. In this case, the demand functions depend on prices on the other side of the border, the border parameter b enters the pricing equations and changes in relative costs affects both relative prices of stores at the border as well as relative markups of the border stores.

4.4.3 Graphical Illustration

Figures 6 and 7 illustrate the qualitative features of the model. These figures depict prices as a function of the distance to the border, where the border is represented by the solid vertical line at 0. Prices for region A (region B) are to the right (left) of the border. In Figure 6 we consider the case where the border parameter b is high enough that markets are fully segmented (Proposition 2). We assume that the number of stores is the same in the two countries and set $N_A = N_B = 20$.

For the left figure we assume that $c_A = c_B$.[25] Since markets are entirely segmented, stores close to the border are shielded from competition from stores across the border and charge

[25] This will be the case if either $\chi_A = \chi_B$ or $\chi_A \neq \chi_B$ and $b_c = 0$. In the latter case, there is no border cost at the whole-sale level and retailers' wholesale cost is $c_i = \min\langle\chi_A, \chi_B\rangle$.

a higher price than interior stores. Given the symmetry, however, there is no difference in border prices: $\hat{p}_A = \hat{p}_B$. As stated earlier, this does not imply that there is no border cost ($b = 0$), simply that it cannot be estimated from price differences across borders. For the panel on the right, we consider the case where costs differ on each side of the border, with costs in region A being greater than costs in region B: $c_A > c_B$. This difference in wholesale costs generates a price discontinuity at the border but the discontinuity is unrelated to the border cost. As we will see in the empirical section, this seems to be the relevant case.

In Figure 7 the border parameter b is set to 0 so we know that markets must be integrated across borders. All else is the same as in Figure 6. In the panel to the right, there is still a price discontinuity, that arises purely from the differences in costs. The magnitude of this discontinuity is always smaller than the difference in costs.[26] Intuitively, since markets are integrated, stores compete for customers on the other side of the border. This explains why, in the case when $c_A - c_B > 0$, the border store in country A charges a lower price compared to the interior stores in country A, while the border store in country B charges a higher price than its interior stores.

Finally, in figure 8, we report the gap in price ($\ln(\hat{p}_A/\hat{p}_B)$) and markups ($\ln((\hat{p}_A/c_A)/(\hat{p}_B/c_B))$) at the border as a function of the gap in marginal costs ($\ln(c_A/c_B)$). The parameters are the same as in figure 6, except that $c_B = 0.02$ and c_A varies from 0.01 to 0.03. In panel (a), the border cost b is high enough to ensure full segmentation. We observe that retail markups are inversely related to relative costs, offsetting some of the cross-border cost differential. However, most of the variation in relative costs translates into relative border prices. In panel (b), $b = 0$ and thus full market integration. Here, we observe that relative costs have a smaller effect on relative border prices and a larger effect on relative markups.

4.4.4 Discussion

The model presented in the previous section delivers the following insights. First, if countries are completely symmetric, the endogenous distribution of prices is identical across countries and there are no border price discontinuities, regardless of the size of border costs.

[26] When $N_A = N_B = 20$, $\hat{p}_A - \hat{p}_B = [(3-\nu)/(5-\nu)](c_A - c_B)$. The derivative of $(\hat{p}_A - \hat{p}_B)$ relative to $(c_A - c_B)$ is $(3-\nu)/(5-\nu)$, strictly less than 1.

Consequently, the border cost cannot be estimated by comparing price differences across borders alone: regressions along the lines of Engel and Rogers (1996) and Broda and Weinstein (2007) reveal no information about the extent of the border effect.[27] A related point is that if border costs are sufficiently high, markets are perfectly segmented and the magnitude of border costs does not affect pricing decisions. In that case, price differences at the border provide only a lower bound on the true size of border costs.

Second, prices of stores that are far from the border are minimally affected by the size of the border cost b. As seen in the right panel of Figures 6 and 7 prices of stores far from the border barely change even when we move from full segmentation to $b = 0$. The effect of the border is observed mainly for stores close to the border.[28] In most of the existing literature, owing to lack of data, no distinction is made between stores that are close to the border and stores that are far from it. Our dataset allows us to use the precise geographical location of stores to make this important distinction.

Third, the behavior of relative prices and relative markups is very different in situations of full and partial segmentation. As figure 8 demonstrates, when markets are fully segmented, fluctuations in relative costs are reflected mostly in relative prices, with minimal impact on relative markups. By contrast, when markets are partially segmented, fluctuations in relative costs impact both relative prices and relative markups. We will exploit the time series dimension of our dataset and the movements in the US-Canada nominal exchange rate, interpreted as exogenous shocks to the relative costs, to explore this implication of the model.

Lastly, equilibrium prices depend on many factors such as the degree of substitutability across locations, the number of competitors, the own price elasticity of demand all of which can vary with location. If this heterogeneity is not taken into account price differences can be attributed to border costs even when these costs are zero, a point made by Gorodnichenko and Tesar (2009). The next section details how the regression discontinuity approach will address this concern.

[27] This point is distinct from the one made in Gorodnichenko and Tesar (2009) who emphasize the problems that arise with heterogeneity across countries.

[28] It follows straightforwardly that this will also be true for wholesale costs if we extend the model to allow for transportation costs that increase with distance.

5 Regression Discontinuity Design

This section implements a regression discontinuity design to measure the effect of the US-Canadian border.[29] The central difficulty with estimating the border effect is that border costs affect mostly stores close to the border while market conditions and arbitrage costs may be very different for stores located far away from the border. We will address this difficulty by exploiting the precise geographic location of each store in our dataset. We will compare the price of identical products sold in adjacent stores located on different sides of the border and measure the discontinuous change in prices as one crosses the border. [30]

Consider the following empirical model of the relationship between the U.S. dollar price p_i^k of product k in store i and various covariates:

$$\ln p_i^k = \alpha^k + \gamma^k C_i + \beta^k X_i + \epsilon_i^k \tag{9}$$

where C_i is a dummy variable that is equal to 1 if store i is located in Canada, X_i measures other observable characteristics of market i, and ϵ_i^k captures unobserved characteristics that are store and good specific. The parameter of interest is γ^k. The problem for inference is that the unobserved characteristics may not be independent from the location of store i, that is $E\left[\epsilon_i^k | C_i\right] \neq 0$, which can bias simple border regression estimates.

However, if the unobserved characteristics are a continuous function of the distance between the stores, we can control for these characteristics by introducing the distance between stores as an additional regressor. Define D_i as the distance (in kilometers) from store i to the border. By convention, stores located in the U.S. are at a positive distance from the border ($D_i > 0$) while stores located in Canada are at a negative distance ($D_i < 0$). With this convention, a store exactly on the border would have $D_i = 0$. The key identifying assumption then is that the unobserved characteristics do not change discontinuously at the border:

$$\lim_{\varepsilon \uparrow 0} E\left[\epsilon_i^k | D_i = \varepsilon\right] = \lim_{\varepsilon \downarrow 0} E\left[\epsilon_i^k | D_i = \varepsilon\right]$$

[29] See Imbens and Lemieux (2007) for a practical guide to the RD framework. See also the February 2008 special issue of the Journal of Econometrics.

[30] Holmes (1998) uses a similar approach to estimate the effect of right-to-work laws on employment across US states.

The effect of the border is then estimated as:

$$\gamma^k = \lim_{\varepsilon \uparrow 0} E\left[\ln p_i^k - \beta^k X_i | D_i = \varepsilon\right] - \lim_{\varepsilon \downarrow 0} E\left[\ln p_i^k - \beta^k X_i | D_i = \varepsilon\right]$$

γ^k answers the question: how do prices change when one crosses from $D_i = \varepsilon$ to $D_i = -\varepsilon$, where ε is some small number.

We follow Imbens and Lemieux (2007) and estimate γ_k using a local linear regression approach including distance as an additional regressor, interacted with the border dummy:

$$\ln p_i^k = \alpha^k + \gamma^k C_i + \theta^k D_i + \delta^k C_i \cdot D_i + \beta^k X_i + \tilde{\epsilon}_i^k. \tag{10}$$

Importantly, this local linear regression restricts the sample to stores at a distance ε_D from the border, that is $|D_j| < \varepsilon_D$. The optimal distance ε_D is selected using a standard bandwidth selection criterion, based on the cross-validation procedure advocated by Imbens and Lemieux (2007).[31] As for the observable covariates X_i, we measure these by variables that capture the demand characteristics of location i.[32] We consider the number of supermarkets per square km,[33] the population density measured by population per square km, the proportion of people aged 0-19 and aged 65 and up, the proportion of black people, the year the store was opened and household income in US dollars.[34]

The key assumption is that the unobserved characteristics ϵ_i^k do not change discontinuously at the border. Although we cannot test this assumption directly, we do three things to assess its plausibility. First, we examine whether the observable characteristics X_i change discontinuously at the border. If the observable characteristics do not change discontinuously at the border, then this is also likely to be the case for the unobservable characteristics. Moreover, even if observable characteristics are not continuous at the border, this does not

[31] The procedure looks for the minimum value of the cross-validation criterion in 100km increments. The optimal bandwidth ranges from 100km to 700km. For most product-groups week pairs, the optimal bandwidth is either 100km, 350km or 500km. All store level observations beyond this cut-off are effectively discarded.

[32] Holmes (2008), who estimates demand for products sold in Walmart Stores, considers similar variables.

[33] These are establishments in NAICS 445110 (Supermarkets and Other Grocery Stores, but not Convenience Stores)

[34] U.S. data comes from from the US population census and economic census data base. Canadian data comes from from Statistics Canada. There is a difference in the level of disaggregation at which the data is collected because Canadian data is collected at the county level while U.S. data is collected by zip code.

invalidate our design as long as the effect of the covariates X_i on the dependent variable remains the same and we control for these characteristics. In the same spirit, we compare estimates of γ^k with controls for observable characteristics and without controls. Third, we provide estimates of the border effect over time, exploiting the 16% nominal devaluation of the U.S. dollar against the Canadian dollar from 2004 through 2007. Even if unobserved market features are different for U.S. and Canadian stores that are very close to the border, these differences are likely to be fairly stable over time and uncorrelated with the nominal exchange rate.

5.1 Graphical Analysis

We begin by plotting the distribution of the distance of stores in our sample from the U.S.-Canadian border (in kilometers).[35] Figure 9 plots the density of all stores in our sample as a function of the algebraic distance from the border (i.e. distance is negative for Canadian stores and positive for the US stores). Each bin width is 50 kms.

As can be seen, all Canadian stores are located less than 1,000 kms from the border, while many stores in the U.S. are more than 1000 kms from the border. Obviously, the geographical concentration of economic activity in the U.S. is very different from that in Canada, highlighting Gorodnichenko and Tesar (2009)'s caution about estimates that do not take within-country heterogeneity differences into account. Nonetheless, we don't observe any significant discontinuity in store density right at the border. This suggests that the location of stores for this retailer does not appear to be directly influenced by the border. Although this is less of a concern with our approach, since we are only looking at U.S. and Canadian stores that are physically close to each other, we need to recognize that not all border points are the same. From Figure 2, it is clear that many Canadian stores close to the border have no counterpart on the US side. To address both issues, we will also present results with a sample of stores located in Oregon, Washington, and British Columbia (21 Canadian and 41 U.S. stores) where market conditions are likely to be more homogenous and there is an important concentration of stores close to the border. We refer to this sample as

[35] The distance was calculated using the ArcGIS software.

the "West coast sample".

Figure 10 depicts graphically the regression discontinuity for some relevant covariates. Each point is the average value of the relevant variable within 50 km bins. Graphically, for several of these variables no stark discontinuity is apparent. We formally test for this and find that when all stores are included there is some discontinuity at the border for the age variables as well as for the proportion of African Americans. When we restrict attention to our west coast subsample of stores, these discontinuities disappear, but we find some discontinuities for the fraction of seniors as well as median household income. As mentioned above, we will include these variables in our specification when exploring the price and costs gaps.

5.2 Regression Discontinuity Estimates

Figures 11(a)-11(f) plot the kernel density of point estimates obtained by estimating regression (10) by UPC for the first week of 2004 and the 21st week of 2007. For our main specification we use all stores within 500 kms from the border.[36] We also estimated the coefficients using the optimal bandwidth selection criterion proposed in Imbens and Lemieux (2007), with similar results. We do this separately for the price, wholesale cost, and markup for each UPC and for each week, both with controls for the covariates and without the controls. The figures illustrate that the effect of the border varies substantially across products.[37] As can be seen, the border discontinuity in prices is centered around zero in the first week of 2004, but shifts significantly to the right by 2007. The distribution of the border discontinuity in costs also shifts to the right from 2004 to 2007, although the cost discontinuity in the first week of July is centered around a positive number. Thus, it appears that the depreciation of the U.S. dollar over this period increased both the costs and the prices in Canadian stores close to the border relative to US stores on the other side of the border. As for the markups, the border effect on markups shifted slightly to the left from 2004 through 2007, suggesting that the depreciation of the U.S. dollar lowered markups in Canadian stores

[36] We restrict the sample to those UPC's that have a minimum of 10 store observations on both sides of the border.

[37] This finding is consistent with the fact that stores in our sample may not choose their location as a function of the border since for many products, the price gap is positive, but for many others it is negative.

relative to that in U.S. stores. However, a visual inspection of the shift in the distribution of costs and markups suggests that the shift in costs overwhelms the change in markups.

We make two additional points. First, the distributions look very similar when the regression is estimated without (left panel) and with (right panel) covariates. This assuages concerns that omitted variable might result in biased estimates of the border effect. Second, we see the same high pass-through from costs to prices when we extend the sample of stores to those farther away from the border as we do with stores close to the border. Our model suggests that if markets are integrated between the U.S. and Canada, the border effect as estimated using stores at the border should be smaller than that estimated from the larger sample of stores (all else equal). Therefore, the fact that this is not the case suggests that retail markets for the products we consider are almost fully segmented between the U.S. and Canada.

Table 3 reports the summary statistics of the distribution of prices, costs, and markups for week 21 of year 2007 (in the no-covariates case) plotted in Figures 11(a)-11(f). The median net price (imputed cost) gap is 15% (17%) for the full sample. When restricted to the West Coast sub-sample (Panel B) the estimates are 22% (22%). The median absolute net price (imputed cost) gap is 21% (21%) for all stores and 24% (24%) for West coast stores.

Next, we plot in Figures 12(a)-12(d) the median (across UPC's) estimate for price (both net and gross), costs (imputed and wholesale) and mark-ups over time. We also plot the U.S-Canada nominal exchange rate. As can be seen, there is virtually a one-to-one correspondence between movements in the median price and the median cost border effect and the exchange rate. By contrast, the movements in the mark-up are much smaller. In January 2004, the median net price gap was 5 percent lower in Canada relative to the US. By the middle of 2007, the median price gap was 15 percent higher in Canada. Over this time period, the U.S. dollar depreciated by roughly 16 percent relative to the Canadian dollar. Since wholesale costs can be viewed as the most 'traded' component of the retailers costs, the discontinuity in this component of costs is particularly striking. All results hold similarly for the west coast sub-sample.

We take four messages from this evidence. First, there is a great deal of heterogeneity in the "effect" of the border on prices, with both negative and positive price gaps. Second, the fact that the price gaps move almost one to one with costs gaps suggests that the two markets are fully segmented. In that case, our model indicates that price gaps provide a lower bound on the border costs. Since we find significant gaps in both prices and costs, we conclude that the effect of the border is sizeable. Third, the fact that the estimates obtained when comparing adjacent stores across the border are similar to estimates obtained from the entire sample of stores is also suggestive that markets are fully segmented. Fourth, it appears that wholesale markets are highly segmented, even when servicing the same retailer. This is especially striking since the wholesale component is the most tradable component of overall costs.

6 Further Results

6.1 Price indices

We have so far compared products with the same UPC's. Although this has the virtue of comparing identical products in the two countries, the limitation is that the sample of products with common UPC's is a small subset of the available products. We now expand the sample of products by constructing price indices at the store level for each product group and product class. There are 61 product groups in the first week of 2004. At this level of aggregation the match rate across borders is 96%. At a more intermediate level of aggregation, such as 'product class', the match rate is 70% (out of 1165 product classes in the first week of 2004). For details about the construction of the price index refer to Appendix B.

We then use the RD approach to measure the discontinuity in the *percentage change* of the price index as one crosses the border. For this, we consider all stores that are within 200 kms. of the border. The results are reported in Figures 13(a)-13(d). Each panel reports the median discontinuity in the percentage change in the price index across time. The top row presents the median discontinuity for the product groups and the bottom row does the same for the product classes. Superimposed is rate of depreciation of the exchange rate. As is evident the co-movement between the percentage change in price and cost indices, and the rate of depreciation of the exchange rate is very high.

6.2 Store-brand products

A question is whether the cost measures we use are allocative or whether they are accounting costs. Here we focus on products that are sold under the brand of the retail chain to examine this. The idea is that the retail chain arguably controls a larger segment of the supply chain for store branded products, and thus the cost measures are arguably less allocative for these products.[38] To the extent that all production is not done in-house it is still possible that manufacturers might segment markets across borders. However, we expect this segmentation to be less severe than for other products. We investigate this in Figures 14(a)-14(b) by plotting the co-movement between the median RD estimates for store brands and the exchange rate. As is evident the co-movement is much less evident for these goods as compared to the full sample that includes mainly non-store brands (Figures 12(a)-12(d)).

6.3 Intra-national borders

This section compares our evidence on cross-border price costs and mark-up gaps to within country estimates at the Washington-Oregon border, which is a subset of our West coast sample. This serves an important purpose: within country border discontinuities -where transactions costs are presumably lower– provide a natural benchmark for cross border discontinuities. In the language of the treatment effect literature, this internal border serves as placebo.

In Figure 15 we plot the net price of Perrier water regular as a function of the distance from the Washington-Oregon border. Stores in Washington are plotted to the left of the border line ($D < 0$) and stores in Oregon to the right ($D > 0$). Each dot represents the average gross price within a 50 km bin. As is evident, unlike the case of the US-Canada border, there is no evidence of a discontinuity for the Perrier water.

Next, we estimate the RD at the internal border for the products that were matched across borders. Very similar results were obtained when the sample was extended to include

[38] We identify manually 225 store-brand products in our sample of 4,221 matched products.

all UPC's that were traded within U.S. boundaries. Panel C of Table 3 reports the results for the internal borders. There is no evidence of a discontinuity in prices or costs.

In Figures 16(a) and 16(b) we plot the distribution of regression discontinuity estimates by UPC at the Washington-Oregon border for net prices and imputed costs. We find that, in contrast to Figures 11(a)-11(f), the point estimates are almost all concentrated at 0 for every week in our sample.

7 Conclusion

This paper revisits a classic question about the role of international borders in segmenting markets. Our paper improves on the existing literature along three dimensions. Firstly, we use barcode level price and cost data from a single retail chain operating in the US and Canada. Next, we develop a stylized model of price determination along the circle. Finally, we use the model to motivate a regression discontinuity estimate of the border effect.

We find strong evidence of international market segmentation, even for identical goods. The failure of the law of one price that we observe at the UPC level is very similar to the failure observed at a more aggregate level. Therefore the argument that aggregate level evidence arises mainly from a compositional bias is not supported by our results. We also find that most differences in cross border consumer prices arise from differences in an apparently tradable component of costs, and not from systematic mark-up differences.

Since the gains to arbitrage are greater at the whole-sale level, where transacted volumes are much larger than at the retail level, this finding reaffirms the existence of a large border costs. A limitation of our work is that we examine a specific set of goods sold by a large grocery chain. To the extent that the nature of price setting and the costs to arbitrage vary across goods, or across retailers, further work that encompasses a wider range of goods would be very useful.

References

Asplund, Marcus and Richard Friberg, "The law of one price in Scandinavian duty-free stores," *American Economic Review*, 2001, *91* (4), 1072–1083.

Broda, Christian and David Weinstein, "Understanding International Price Differences Using Barcode Data," Technical Report, Chicago GSB October 2007.

Burstein, Ariel and Nir Jaimovich, "Understanding Movements in Aggregate and Product-level Real Exchange rates," Working Paper, UCLA 2008.

Crucini, Mario J. and Mototsugu Shintani, "Persistence in Law-Of-One-Price Deviations: Evidence from Micro-Data," Working Paper 06-W17, Department of Economics, Vanderbilt University December 2006.

_ and _ , "Microeconomic Sources of Real Exchange Rate Variability," Working Paper, Vanderbilt University 2007.

_ , Chris I. Telmer, and Marios Zachariadis, "Understanding European Real Exchange Rates," *American Economic Review*, June 2005, *95* (3), 724–738.

d'Aspremont, C., J Gabszewicz, and J Thisse, "On Hotelling's "Stability in Competition"," *Econometrica*, 1979, *47* (5), 1045–1150.

Engel, Charles and John Rogers, "How Wide is the Border?," *American Economic Review*, 1996, *86*, 1112–1125.

Ghosh, Atish R. and Holger C. Wolf, "Pricing in International Markets: Lessons From The Economist," Working Paper 4806, National Bureau of Economic Research, Inc July 1994.

Goldberg, Pinelopi Koujianou and Frank Verboven, "The Evolution of Price Dispersion in the European Car Market," *Review of Economic Studies*, 2001, pp. 811–48.

_ and _ , "Market Integration and Convergence to the Law of One Price: Evidence from the European Car Market," *Journal of International Economics*, 2005, pp. 49–73.

_ and Michael M. Knetter, "Goods Prices and Exchange Rates: What Have We Learned?," *Journal of Economic Literature*, 1997, *35* (3), 1243–1272.

Gorodnichenko, Yuriy and Linda Tesar, "Border Effect or Country Effect? Seattle May Not Be So Far from Vancouver After All.," *American Economic Journal: Macroeconomics*, January 2009, *1* (1), 21941.

Haskel, Jonathan and Holger Wolf, "The Law of One Price–A Case Study," *Scandinavian Journal of Economics*, December 2001, *103* (4), 545–58.

Hassink, Wolter H.J. and Ronald Schettkat, "On Price-Setting for Identical Products in Markets without Formal Trade Barriers," Discussion Paper 315, Institute for the Study of Labor (IZA) June 2001.

Holmes, Thomas, "The Effect of State Policies on the location of Manufacturing: Evidence from State Borders," *Journal of Political Economy*, August 1998, *106* (4), 667–705.

_ , "The Diffusion of Wal-Mart and Economies of Density," NBER Working Papers 13783 February 2008.

Imbens, Guido and Thomas Lemieux, "Regression Discontinuity Designs: A Guide to Practice," Working Paper 13039, National Bureau of Economic Research, Inc April 2007.

Nakamura, Emi, "Pass-Through in Retail and Wholesale," *American Economic Review*, May 2008, *98* (2), 430–37.

Parsley, David C. and Shang-Jin Wei, "A Prism into the PPP Puzzles: The Micro-Foundations of Big Mac Real Exchange Rates," *Economic Journal*, October 2007, *117* (523), 1336–1356.

Rogoff, Kenneth, "The Purchasing Power Parity Puzzle," *Journal of Economic Literature*, 1996, *34*, 647–668.

Salop, S, "Monopolistic Competition with Outside Goods," *Bell Journal of Economics*, 1979, *10*, 141–156.

8 Tables and Figures

	Number of Unique Products							
	per store-week				per store-pair-week			
	mean	median	10%	90%	mean	median	10%	90%
US	492	497	355	643	272	273	187	365
Canada	414	425	263	533	243	252	146	365
Cross-border pairs	-	-	-	-	164	168	101	225

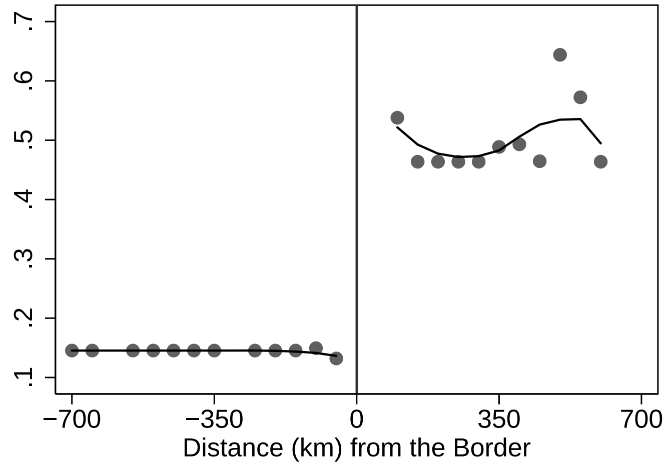
Table 1: Descriptive Statistics. The table reports the mean, median 10th percentile and 90th percentile of the number of unique matched products per store per week, and per store-pair per week.

	Mean	Median	St. Dev.	Mean Absolute	Med. Absolute
	(1)	(2)	(3)	(4)	(5)
	Panel A: Net Prices				
	USA-USA store-pairs (31125)				
Median	0.010	0.000	0.147	0.085	0.037
Average	0.015	0.005	0.145	0.087	0.042
St. Dev.	0.038	0.025	0.034	0.029	0.032
	CAN-CAN store-pairs (2775)				
Median	0.007	0.000	0.055	0.030	0.000
Average	0.010	0.001	0.057	0.030	0.005
St. Dev.	0.025	0.006	0.024	0.020	0.012
	CAN-USA store-pairs (18450)				
Median	0.153	0.118	0.254	0.219	0.146
Average	0.151	0.116	0.255	0.222	0.156
St. Dev.	0.048	0.044	0.030	0.033	0.041
	Panel B: Imputed Costs				
	USA-USA store-pairs (31125)				
Median	0.000	0.000	0.124	0.057	0.008
Average	0.001	0.001	0.126	0.058	0.018
St. Dev.	0.025	0.009	0.038	0.023	0.021
	CAN-CAN store-pairs (2775)				
Median	0.000	0.000	0.122	0.038	0.000
Average	0.000	0.000	0.124	0.038	0.000
St. Dev.	0.013	0.000	0.036	0.011	0.001
	CAN-USA store-pairs (18450)				
Median	0.184	0.144	0.263	0.238	0.178
Average	0.189	0.152	0.267	0.242	0.182
St. Dev.	0.043	0.049	0.035	0.039	0.046

Table 2: Deviations from the Law of One Price for Retail and Whole-sale Prices: Panel A refers to net prices and panel B refers to imputed costs. The table reports within and between-country statistics (the rows) for the mean, median, standard deviation, mean absolute and median absolute (log) price gap within store-pairs (the columns) for the first week of 2005.

	Median	Mean	SD	Frac. sign.	Median Abs.	Mean Abs.	No. of upc's	Median Bandwidth
				Panel A: All Stores				
Net Price	0.15	0.13	0.37	0.70	0.21	0.28	481	10
Imputed Cost	0.17	0.15	0.31	0.80	0.21	0.26	481	10
Imputed mark up	0	-0.02	0.37	0.40	0.14	0.23	481	10
				Panel B: West Coast Stores				
Net Price	0.22	0.26	0.32	0.86	0.24	0.33	212	12
Imputed Cost	0.22	0.20	0.27	0.83	0.24	0.27	212	12
Imputed mark up	0	0.06	0.36	0.44	0.13	0.23	212	12
				Panel C: Washington-Oregon Stores				
Net Price	0	0.01	0.09	0.24	0.01	0.04	370	6
Imputed Cost	0	0	0.06	0.17	0	0.02	370	6
Imputed mark up	0	0	0.10	0.22	0.02	0.04	370	6

Table 3: Regression Discontinuity Estimates - minimizing the cross-validation criterion.



Perrier sparkling natural mineral water, 25fl. oz. (UPC 074780000055). Local linear regression of (log) net price on border dummy B_j, algebraic distance to the border D_j and interaction term. Store distance to the border is positive for the US, negative for Canada. First week of 2004.

Figure 1: Graphical depiction of border discontinuity for Perrier Sparkling Mineral Water

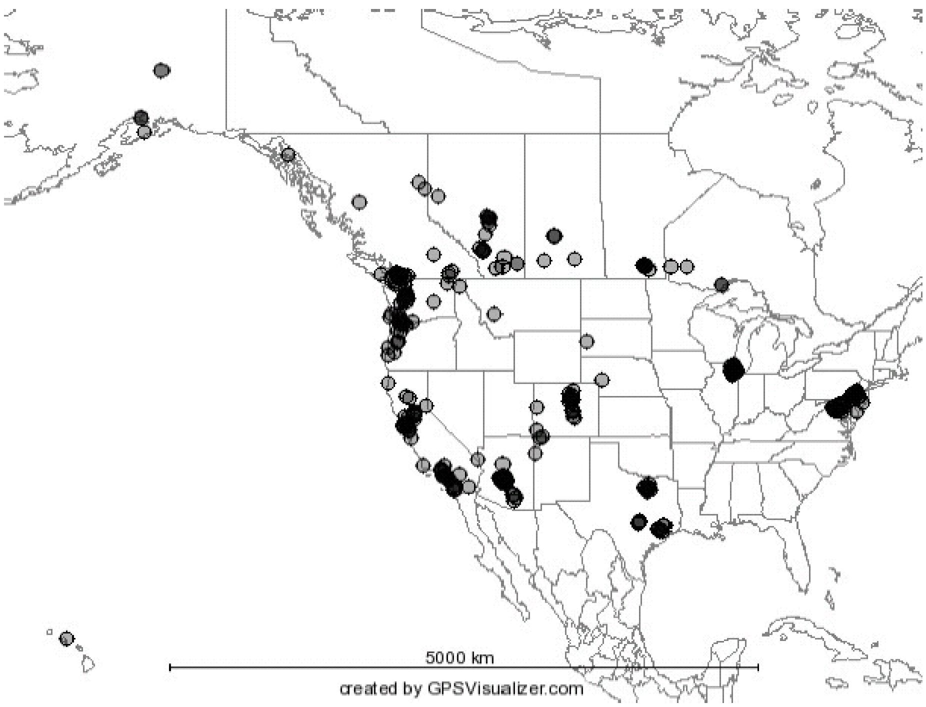


Figure 2: Map of the 325 retail North American stores in our data (250 U.S. and 75 Canada)

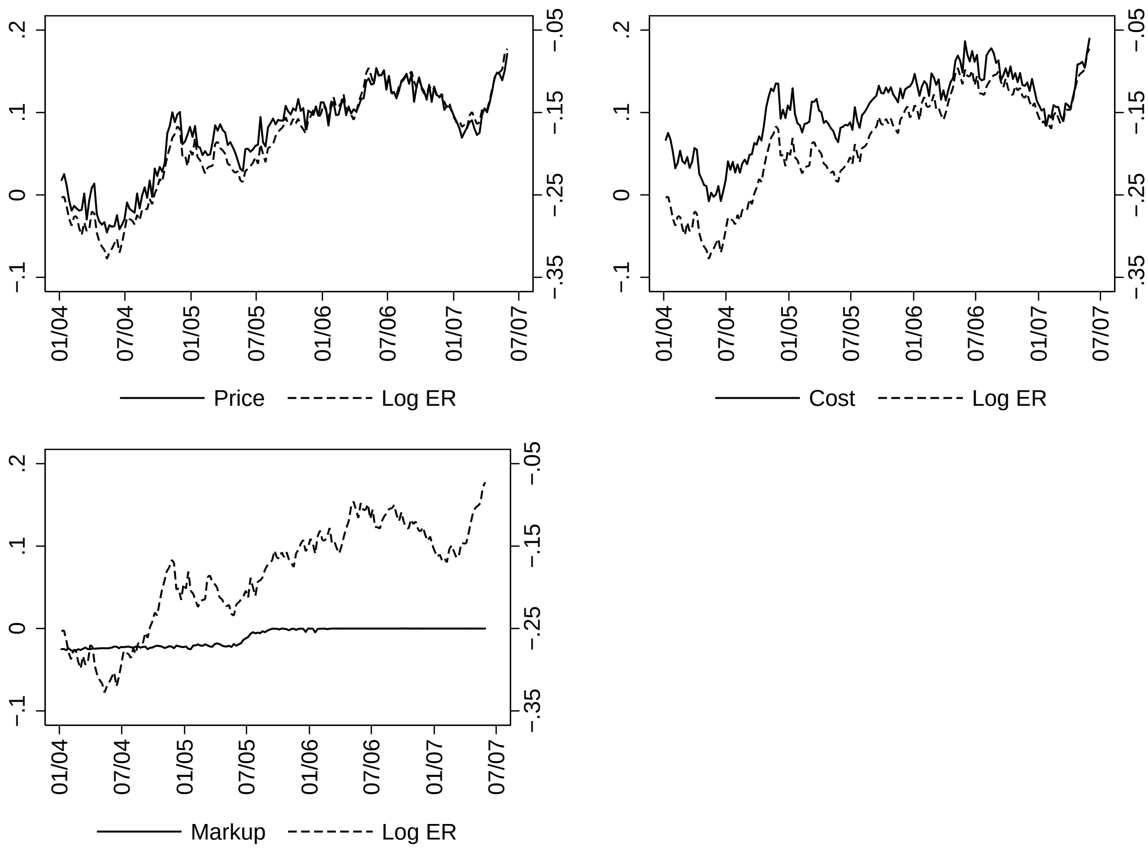
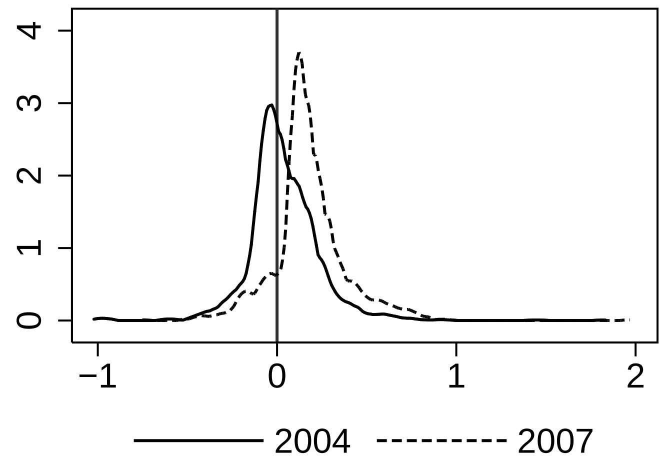
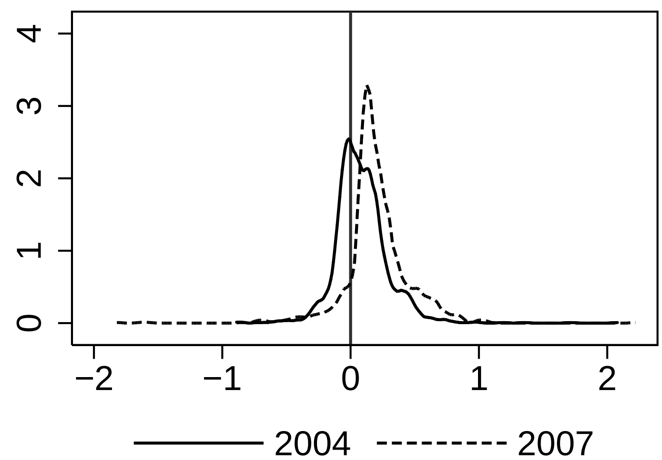


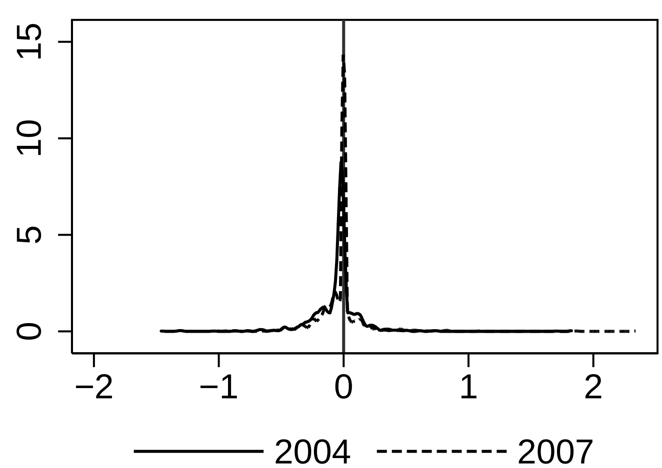
Figure 3: Median net price, imputed cost and markup cross-border gap and exchange rate



(a) (log) net price



(b) (log) imputed cost



(c) (log) markup

Note: 2004 refers to the first week of 2004; 2007 refers to the 21st week of 2007.

Figure 4: The dispersion of cross-border average price, cost, and markup gap

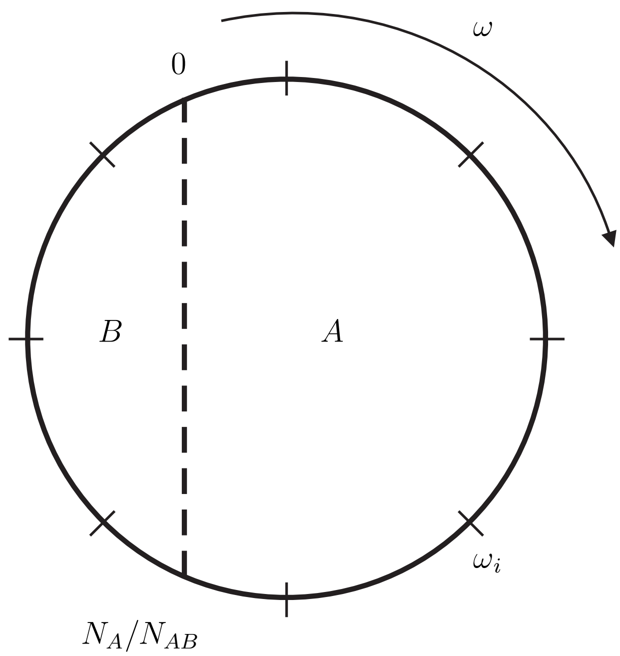
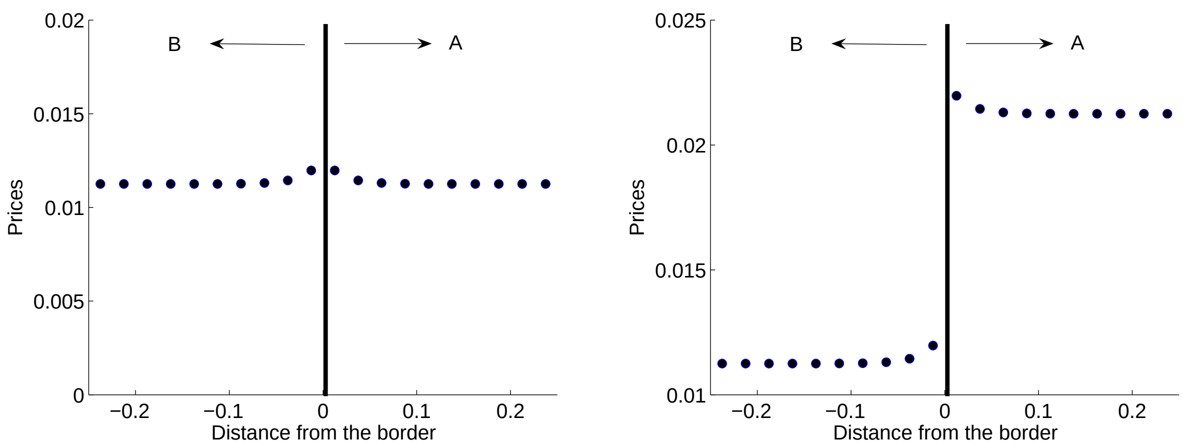
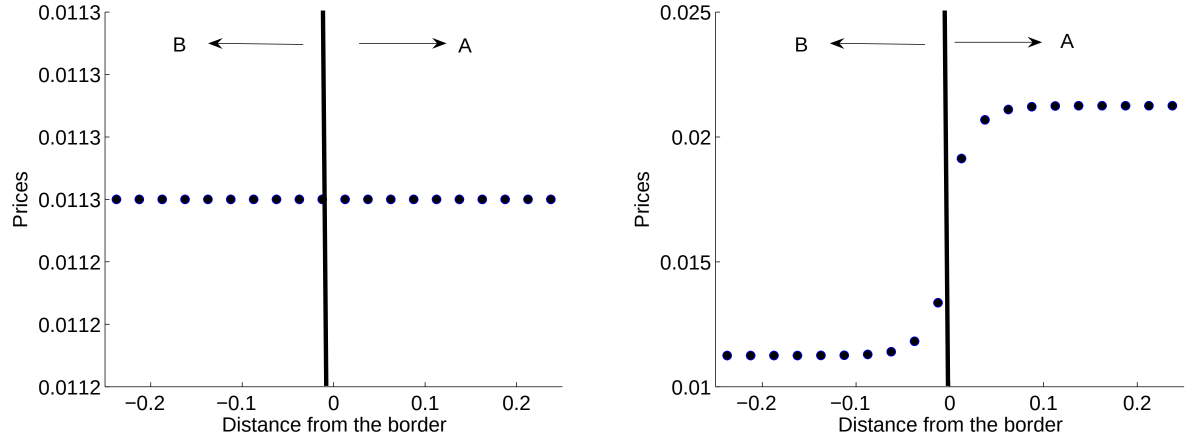


Figure 5: Circular World



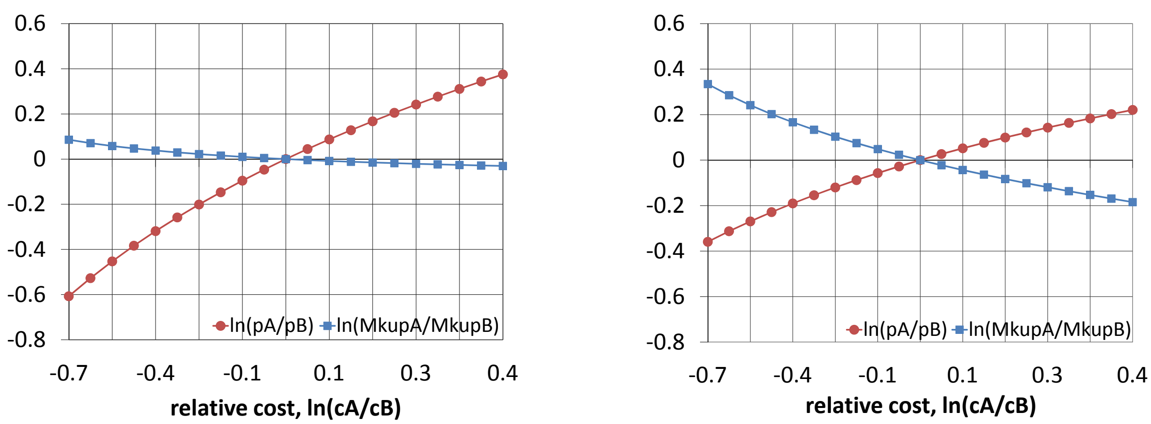
Note: For the left figure the parameters are $N_A = N_B = 20$, $t = 0.05$, $c_A = c_B = 0.01$ and for the figure on the right the parameters are $N_A = N_B = 20$, $t = 0.05$, $c_A = 0.02 > c_B = 0.01$

Figure 6: Price Discontinuity at the Border: Full Segmentation



Note: For the left figure the parameters are $N_A = N_B = 20$, $t = 0.05$, $c_A = c_B = 0.01$, $b = 0$ and for the figure on the right the parameters are $N_A = N_B = 20$, $t = 0.05$, $c_A = 0.02 > c_B = 0.01$, $b = 0$.

Figure 7: Price Discontinuity at the Border: Partial Segmentation

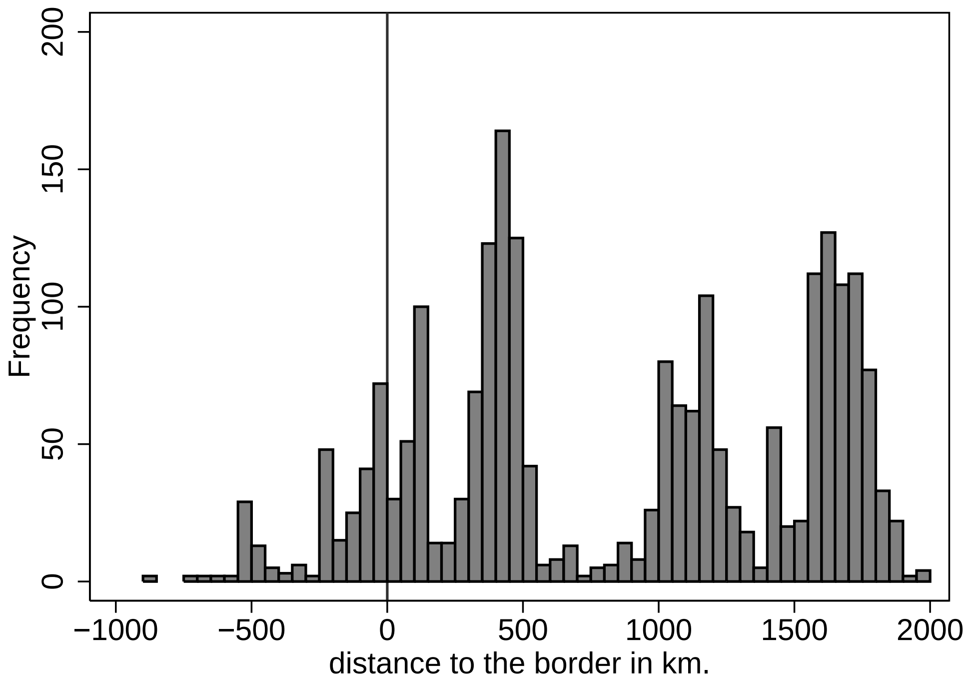


(a) Full Segmentation

(b) Partial Segmentation

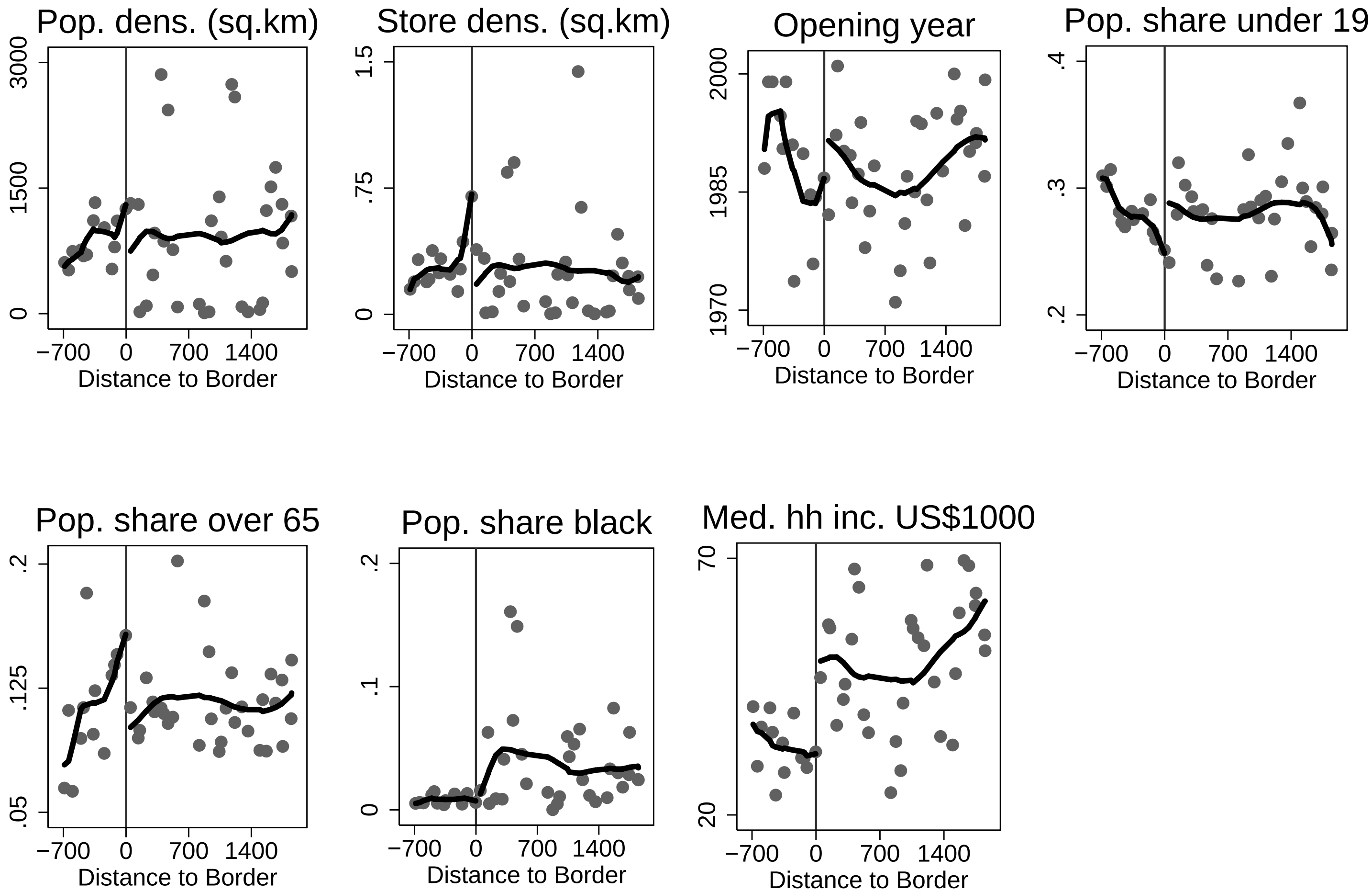
Note: The parameters are $N_A = N_A = 20$ and $t = 0.05$. For panel (a), $|\hat{p}_A - \hat{p}_B| < b$, which is the case of Proposition 2; panel (b) assumes $b = 0$, which is consistent with the case in Proposition 3.

Figure 8: Price and Mark-Up Discontinuities at the Border



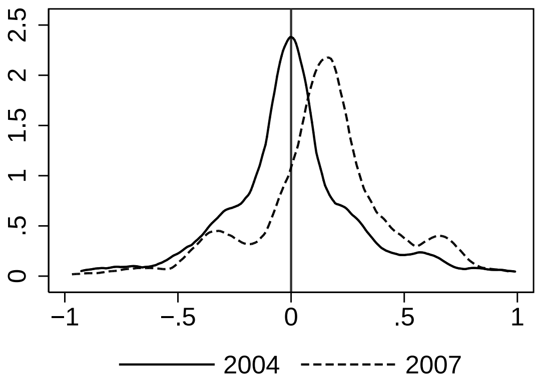
Note: Distance to the border is positive for US stores, negative for Canadian stores.

Figure 9: Distance to the Border

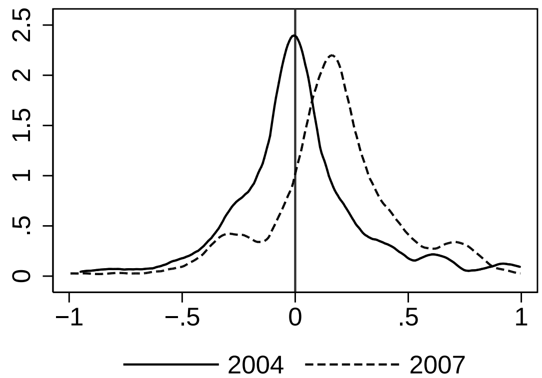


Note: Local linear regression of covariates X_j on border dummy B_j, algebraic distance to the border D_j and interaction term. Store distance to the border is positive for the US, negative for Canada.

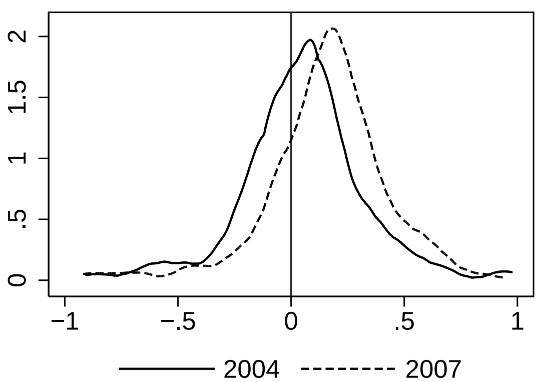
Figure 10: Regression Discontinuity for Covariates



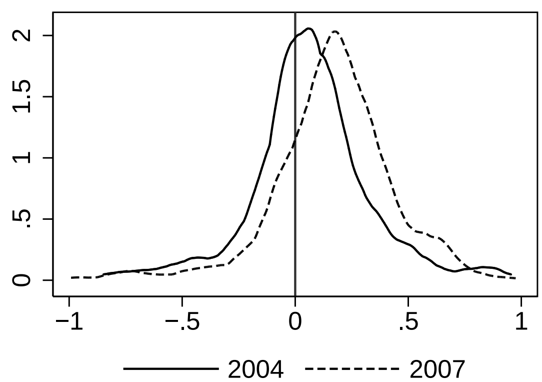
(a) Net price, no covariates



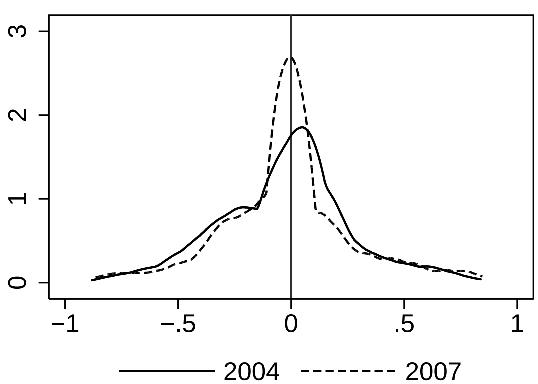
(b) Net price, with covariates



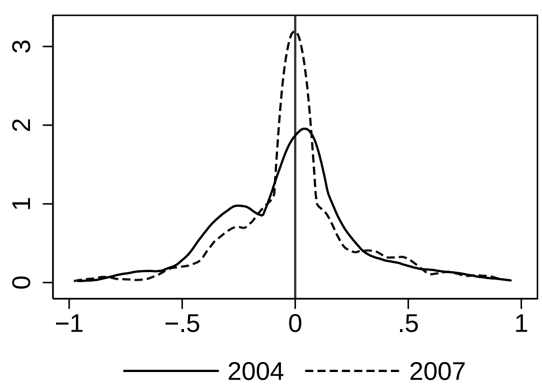
(c) Imputed Cost, no covariates



(d) Imputed Cost, with covariates



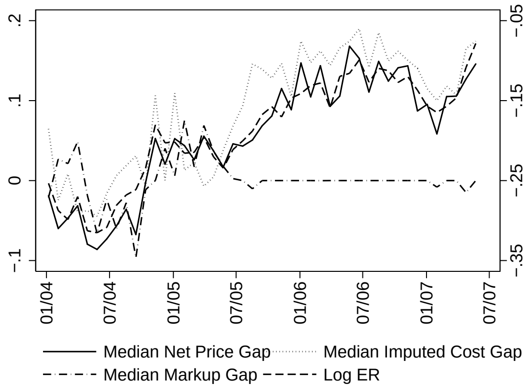
(e) Mark up, no covariates



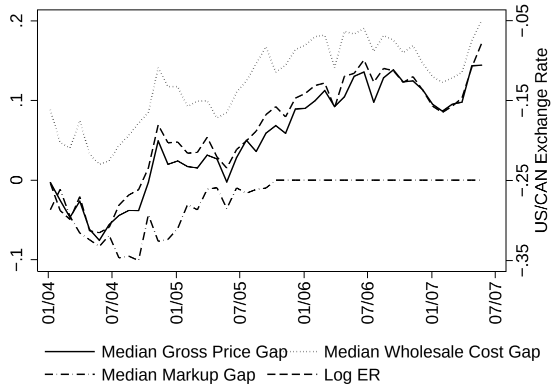
(f) Mark-up, with covariates

Note: 2004 refers to the first week of 2004; 2007 refers to the 21 week of 2007.

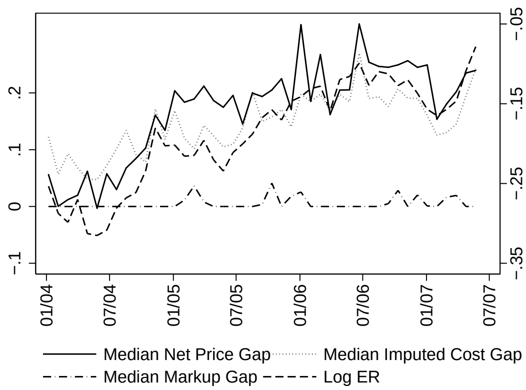
Figure 11: Distribution of RD estimates of Price, Cost and Mark-up Gaps



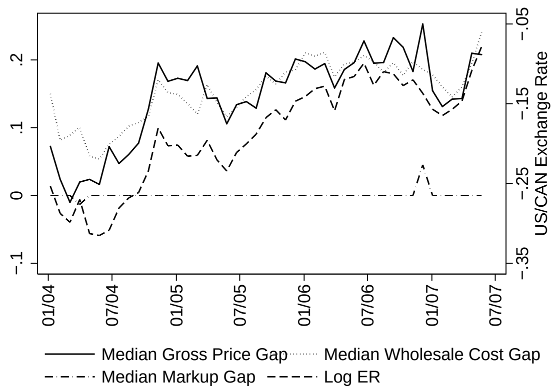
(a) Net, All Stores



(b) Gross, All Stores



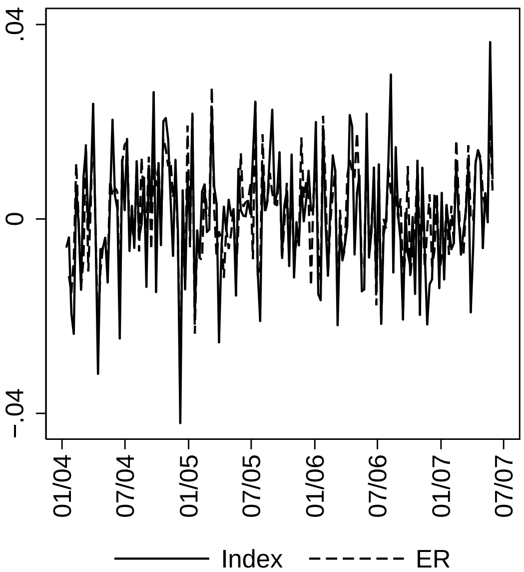
(c) Net, West Coast Stores



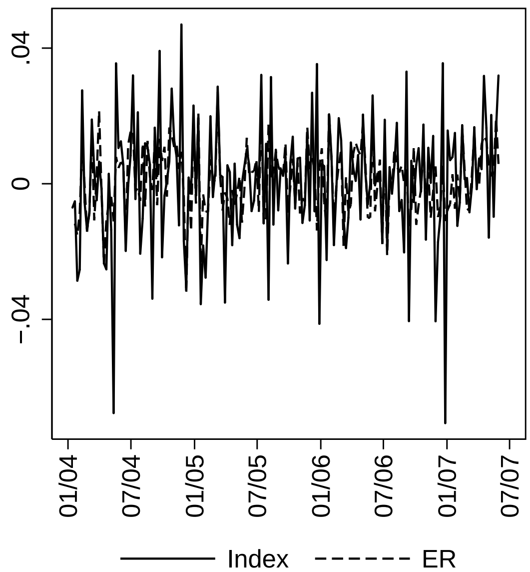
(d) Gross, West Coast Stores

Note: In each panel, the median regression discontinuity (left axis) is plotted alongside the US-Canada (log) nominal exchange rate (right axis). RD includes covariates.

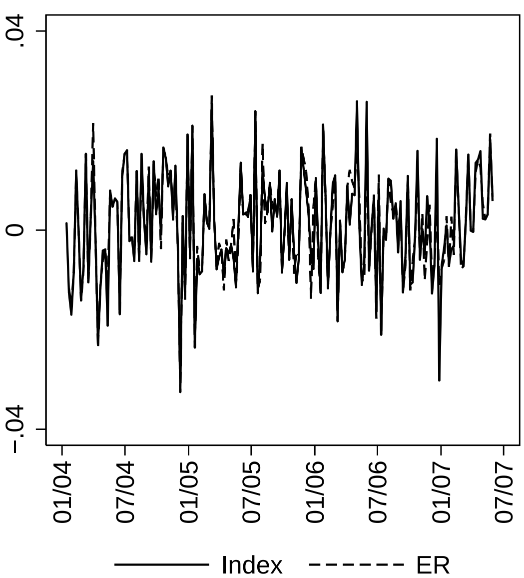
Figure 12: Co-movement in RD estimates of Price and Cost Gaps over time



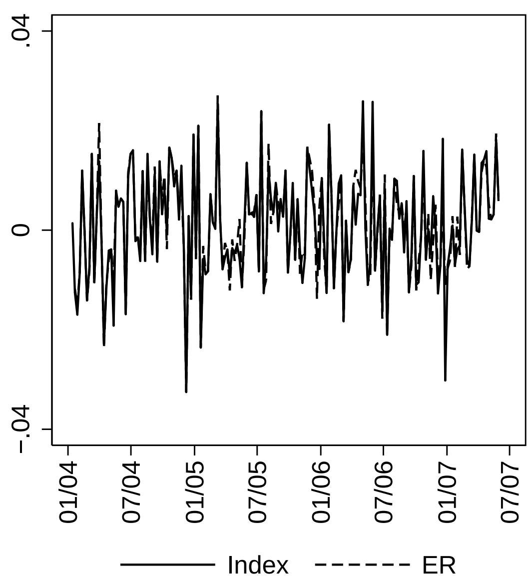
(a) Net Price, Product Group



(b) Imputed cost, Product Group



(c) Net Price, Product class



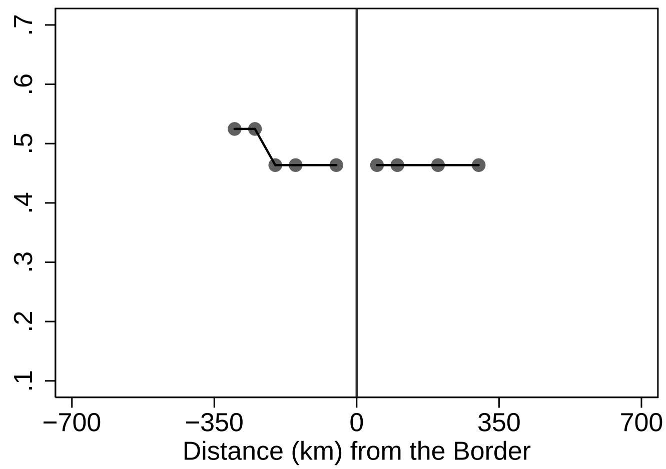
(d) Imputed cost, Product class

Figure 13: Price Indices

(a) Whole-sale cost

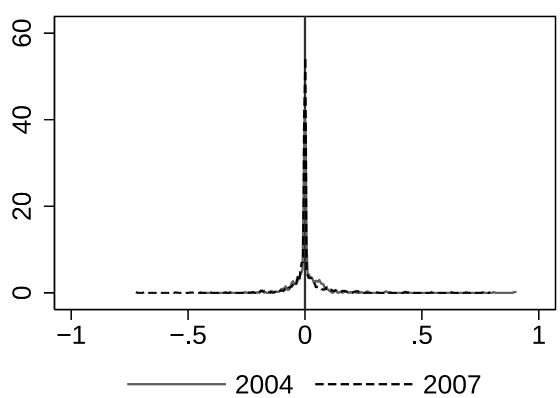
(b) Imputed cost

Figure 14: Retail Chain Brands

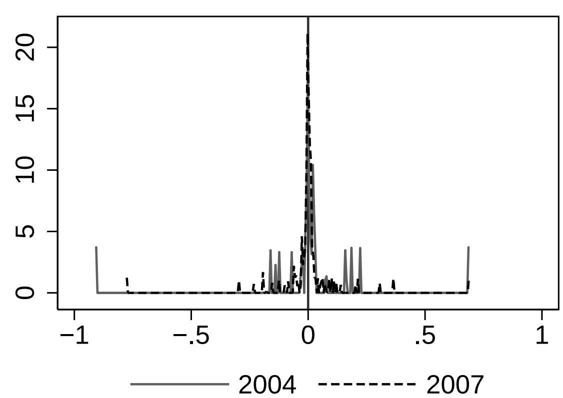


Perrier sparkling natural mineral water, 25fl. oz. (UPC 074780000055). Local linear regression of US (log) net price on Washington-Oregon border dummy B_j, algebraic distance to the border D_j, and interaction term. Store distance to the border is positive for Oregon, negative for Washington.

Figure 15: Graphical depiction of internal border regression discontinuity for Perrier Sparkling Mineral Water



(a) Distribution of Net Price discontinuity



(b) Distribution of Imputed Cost discontinuity

Note: 2004 refers to the first week of 2004 and 2007 refers to the 21st week of 2007.

Figure 16: Intra-national borders regression discontinuity: the Washington-Oregon Border

Appendix

A Derivations for the Circular World model

A.1 Prices charged by interior stores

Without lack of generality, let's consider region A. Given our assumptions in section 4.3, all stores in region A face the same cost c_A. Each interior store maximizes static profits by choosing p_i determined by the first order condition:

$$p_i = \frac{t}{2N_{AB}} + \frac{p_{i+1} + p_{i-1}}{4} + \frac{c_A}{2}, \ i = 2, .., N_A - 1 \tag{11}$$

We solve this system of equations, subject to the boundary condition

$$p_1 = p_{N_A} = \hat{p}_A,$$

In the difference equation (11) all terms are linear in prices (up to the constant term), we can expect a solution in the form of a sum of two exponentials plus a constant. Because of the symmetry between p_1 and p_{N_A}, the sum of the two exponentials should reduce to a hyperbolic cosine centered at $\omega = N_A/2N_{AB}$, or $i = (N_A + 1)/2$. For this reason, we conjecture the following solution[39]

$$p_i = A \cosh \kappa \left(i - \frac{N_A + 1}{2} \right) + B.$$

By plugging this conjecture back into equation (11), we can determine the unknown coefficients A, B,and κ. We obtain

$$\begin{aligned} A \cosh \kappa (i - \frac{N_A + 1}{2}) + B = \frac{t}{2N_{AB}} + \frac{1}{4} \left(A \cosh \kappa (i + 1 - \frac{N_A + 1}{2}) + B \right) \\ + \frac{1}{4} \left(A \cosh \kappa (i - 1 - \frac{N_A + 1}{2}) + B \right) + \frac{c_A}{2} \end{aligned}$$

Using the property that $\cosh(x + y) = \cosh x \cosh y + \sinh x \sinh y$ and after some simplification, it follows that these equations will be satisfied for all i if[40]

$$\begin{aligned} \kappa &= \cosh^{-1} 2 \approx 1.317 \\ B &= c_A + \frac{t}{N_{AB}} \end{aligned}$$

[39] We thank Michal Fabinger for providing us with this conjecture.

[40] $\sinh x = \frac{e^x - e^{-x}}{2}$

The value of A is determined using the boundary condition

$$\hat{p}_A = A \cosh \kappa \left(1 - \frac{N_A + 1}{2}\right) + B$$

This provides,

$$A = \frac{\hat{p}_A - c_A - \frac{t}{N_{AB}}}{\cosh \kappa (\frac{N_A - 1}{2})}.$$

We can summarize the interior solution for stores in region A as

$$p_i = \left(\hat{p}_A - c_A - \frac{t}{N_{AB}}\right) \frac{\cosh \kappa (i - \frac{N_A+1}{2})}{\cosh \kappa \left(\frac{N_A-1}{2}\right)} + c_A + \frac{t}{N_{AB}}, \tag{12}$$

By analogy, the interior solution for country B is

$$p_i = \left(\hat{p}_B - c_B - \frac{t}{N_{AB}}\right) \frac{\cosh \kappa \left(i - \frac{N_B+1}{2} - N_A\right)}{\cosh \kappa \left(\frac{N_B-1}{2}\right)} + c_B + \frac{t}{N_{AB}}. \tag{13}$$

where $\hat{p}_B$ is the price charged by border stores N_{AB} and $N_A + 1$ in region B.

A.2 Prices charged by stores at the border

We use the profit maximization conditions for stores at the border to paste the interior solutions together. As discussed previously, we need to consider several different cases.

A.2.1 Case 1 - The marginal customer is at the border

If $|\hat{p}_A - \hat{p}_B| < b$, the marginal consumer between stores $i = 1$ and $i = N_{AB}$ will be exactly at the border. Similarly for the border between stores N_A and $N_A + 1$. Store 1 will choose $\hat{p}_A$ to maximize

$$\pi_1 \equiv (\hat{p}_A - c_A) \left(\frac{1}{N_{AB}} + \frac{p_2 - \hat{p}_A}{2t}\right),$$

and store N_{AB} will choose $\hat{p}_B$ to maximize

$$\pi_{N_{AB}} \equiv (\hat{p}_B - c_B) \left(\frac{1}{N_{AB}} + \frac{p_{(N_{AB}-1)} - \hat{p}_B}{2t}\right).$$

The corresponding first-order conditions are:

$$\hat{p}_A = \frac{t}{N_{AB}} + \frac{p_2}{2} + \frac{c_A}{2},$$

and

$$\hat{p}_B = \frac{t}{N_{AB}} + \frac{p_{(N_{AB}-1)}}{2} + \frac{c_B}{2}.$$

We can substitute for $p_2$and $p_{N_{AB}-1}$ from eqs. (12) and (13). After some manipulation, we obtain,

$$\hat{p}_A = \frac{\frac{3t}{N_{AB}} + 2c_A - \left(c_A + \frac{t}{N_{AB}}\right)\frac{\cosh\kappa\left(\frac{N_A-3}{2}\right)}{\cosh\kappa\left(\frac{N_A-1}{2}\right)}}{2 - \frac{\cosh\kappa\left(\frac{N_A-3}{2}\right)}{\cosh\kappa\left(\frac{N_A-1}{2}\right)}}\,,\ \hat{p}_B = \frac{\frac{3t}{N_{AB}} + 2c_B - \left(c_B + \frac{t}{N_{AB}}\right)\frac{\cosh\kappa\left(\frac{N_B-3}{2}\right)}{\cosh\kappa\left(\frac{N_B-1}{2}\right)}}{2 - \frac{\cosh\kappa\left(\frac{N_B-3}{2}\right)}{\cosh\kappa\left(\frac{N_B-1}{2}\right)}}$$

With some simplification we arrive at the expressions in (5).

A.3 Case 2a - The marginal customer for the border stores is located in Country A

For this to be the case, we need $\hat{p}_A - \hat{p}_B > b$. The demand for the border stores located near $\omega = 0$, that is $i = 1$ and $i = N_{AB}$, are given by:

$$D_1(\hat{p}_A, \hat{p}_A, p_2) = \frac{1}{N_{AB}} + \frac{p_{(N_{AB}-1)} - 2\hat{p}_A + p_2}{2t} + \frac{b}{2t},$$
$$D_{N_{AB}}(p_{(N_{AB}-1)}, \hat{p}_B, \hat{p}_A) = \frac{1}{N_{AB}} + \frac{p_{(N_{AB}-2)} - 2\hat{p}_B + \hat{p}_A}{2t} - \frac{b}{2t}.$$

The profit maximizing prices are then, for store 1,

$$\hat{p}_A = \frac{t}{2N_{AB}} + \frac{\hat{p}_B}{4} + \frac{p_2}{4} + \frac{b}{4} + \frac{c_A}{2},$$

and for store N_{AB},

$$\hat{p}_B = \frac{t}{2N_{AB}} + \frac{p_{(N_{AB}-1)}}{4} + \frac{\hat{p}_A}{4} - \frac{b}{4} + \frac{c_B}{2}.$$

Substituting for $p_{N_{AB}-1}$ and p_2 using equations (12) and (13), and after some manipulations, we arrive at the expressions in Proposition 6.

$$\hat{p}_A = \frac{(4-\nu_B)\,j_A\ + j_B}{15 - 4\nu_A - 4\nu_B + \nu_A\nu_B}, \qquad \hat{p}_B = \frac{(4-\nu_A)\,j_B\ + j_A}{15 - 4\nu_B - 4\nu_A + \nu_B\nu_A}$$

where

$$\nu_A \equiv \frac{\cosh\kappa(\frac{N_A-3}{2})}{\cosh\kappa(\frac{N_A-1}{2})}, \qquad \nu_B \equiv \frac{\cosh\kappa(\frac{N_B-3}{2})}{\cosh\kappa(\frac{N_B-1}{2})}.$$

and

$$j_A \equiv \frac{3t}{N_{AB}} + 3c_A + b - (c_A + \frac{t}{N_{AB}})\nu_A; \qquad j_B \equiv \frac{3t}{N_{AB}} + 3c_B - b - (c_B + \frac{t}{N_{AB}})\nu_B$$

A.4 Case 2b - The marginal customer for the border stores is located in Country B

This case is symmetric of the previous one when $\hat{p}_B - \hat{p}_A > b$ and is derived analogously.

B Price Index Construction

We calculate the change in the chain-weighted Törnqvist log price index, $\ln P_t^{TQ}(K,i)$, of category K in store i between period $t-1$ and t as

$$\Delta \ln P_t^{TQ}(K,i) \equiv \sum_{k \in K} \ln \left(\frac{p_t(k,j)}{p_{t-1}(k,j)} \right)^{\frac{1}{2}[s_t(k)+s_{t-1}(k)]} \equiv \sum_{k \in K} \omega_t(k) \cdot \Delta \ln p_t(k,j)$$

where the weights $\omega_t(k) = \frac{1}{2}[s_t(k) + s_{t-1}(k)]$ use the expenditure shares of good k as a fraction of total expenditures on category K in week t, i.e.

$$s_t(k) = \frac{\sum_j x_t(k,j)p_t(k,j)}{\sum_{k \in K} \sum_j x_t(k,j)p_t(k,j)} = \frac{\sum_j amt_t(k,j)}{\sum_{k \in K} \sum_j amt_t(k,j)}$$

In summing over j we use all stores in the U.S. and in Canada so that differences in the change in the store level price index arises from differences in the rate of change in prices across stores. However, there are many weeks when a particular UPC is not sold in a particular store, that is we have no recorded price change. In this case we drop the observation for the store missing a price change and re-weight the shares across the UPC's for which price information is available in that store. We construct these price indices for different levels of product classifications: subsubclass, subclass, class, category and group. For the case of net (gross) prices we use the net (gross) expenditure shares. Similarly for the imputed net cost (whole-sale cost) measure we use the net (gross) expenditure shares.

C Data Description

Table 4 describes the distribution of unique UPCs by product groups.

Product Groups	Unique UPCs Freq. (1)	Unique UPCs Percent	Canada Freq. (2)	Canada Percent	United States Freq. (3)	United States Percent	Matched UPCs Freq. (4)	Matched UPCs Percent
Alcoholic Beverages	10,038	8.03	2,268	6.88	8,173	8.3	403	9.55
Baby Food/Diapers/Baby Care	1,220	0.98	384	1.17	930	0.94	94	2.23
Batteries	94	0.08	68	0.21	61	0.06	35	0.83
Books & Magazines	5,361	4.29	3,908	11.86	4,266	4.33	2,505	59.35
Candy, Gum & Mints	4,065	3.25	1,128	3.42	2,967	3.01	29	0.69
Canned Fish & Meat	740	0.59	203	0.62	540	0.55	3	0.07
Canned Fruits	228	0.18	64	0.19	164	0.17		
Canned Vegetables	459	0.37	85	0.26	374	0.38		
Cereal And Breakfast	2,438	1.95	570	1.73	1,875	1.9	7	0.17
Cheese	1,453	1.16	335	1.02	1,130	1.15		
Coffee/Tea/Hot Cocoa...	3,215	2.57	729	2.21	2,606	2.65	120	2.84
Commercial Bread & Baked Goods	4,596	3.68	492	1.49	4,111	4.18	7	0.17
Condiments & Sauces	37	0.03			37	0.04		
Cookies/Crackers & Snacks	2,869	2.29	733	2.22	2,205	2.24	69	1.63
Cough, Cold, Flu, Allergy	15	0.01	1	0	14	0.01		
New Age, Mixers, Bottled Water	4,295	3.43	1,197	3.63	3,135	3.19	36	0.85
Deli/Food Service Items	6,623	5.3	2,313	7.02	4,936	5.01		
Dessert & Baking Mixes	412	0.33	121	0.37	291	0.3		
Detergents & Laundry Needs	1,448	1.16	539	1.64	963	0.98	54	1.28
Diet, Ethnic & Gourmet Foods	3,992	3.19	901	2.73	3,397	3.45	306	7.25
Enhancements	1,086	0.87	279	0.85	825	0.84	18	0.43
Floral	7,360	5.89	1,719	5.22	5,914	6.01		
Flour, Sugar, Corn Meal	122	0.1	26	0.08	96	0.1		
Food Service	1,729	1.38	625	1.9	1,222	1.24		
Fresh Produce	9,985	7.98	2,572	7.8	8,069	8.2		
Frozen Breakfast Items	260	0.21	55	0.17	207	0.21	2	0.05
Frozen Vegetables	895	0.72	139	0.42	757	0.77	1	0.02
Hair Care	1,641	1.31	582	1.77	1,061	1.08	2	0.05
Health Supplements	1,356	1.08	310	0.94	1,064	1.08	18	0.43
Hispanic Products	1,077	0.86	68	0.21	1,013	1.03	4	0.09
Household Cleaners	2,566	2.05	935	2.84	1,790	1.82	159	3.77
Housewares	364	0.29	95	0.29	280	0.28	11	0.26
Ice Cream & Ice	2,713	2.17	544	1.65	2,172	2.21	3	0.07
Fresh Bread & Baked Goods	959	0.77	312	0.95	666	0.7		
Jams, Jellies & Spreads	1,026	0.82	247	0.75	798	0.81	19	0.45
Mayo, Salad Dressings & Toppings	1,268	1.01	249	0.76	1,029	1.05	10	0.24
Meat	5,604	4.48	1,301	3.95	4,370	4.44		
Natural Markets	12	0.01	12	0.04	2	0	2	0.05
Oral Hygiene	978	0.78	303	0.92	682	0.69	7	0.17
Paper, Foil & Plastics	1,378	1.11	322	0.98	1,121	1.14	65	1.54
Pasta & Pasta Sauce	1,963	1.57	362	1.1	1,624	1.65	23	0.54
Pet Food & Pet Needs	2,647	2.12	656	1.99	2,070	2.1	79	1.87
Pickles,Peppers & Relish	849	0.68	147	0.45	709	0.72	7	0.17
Prepared Frozen Foods	3,197	2.56	432	1.31	2,774	2.82	9	0.21
Ready To Eat Prepared Foods	408	0.33	57	0.17	351	0.36		
Refrigerated Dairy	2,841	2.27	786	2.38	2,070	2.1	15	0.36
Refrigerated Foods	1,201	0.96	214	0.65	994	1.01	7	0.17
Refrigerated Juice	435	0.35	105	0.32	331	0.34	1	0.02
Respiratory	537	0.43	219	0.66	319	0.32	1	0.02
Rice & Beans	1,177	0.94	253	0.77	930	0.94	5	0.12
Salt, Seasoning & Spices	1,133	0.91	205	0.62	936	0.95	8	0.19
Salty Snacks	2,367	1.89	579	1.76	1,797	1.83	9	0.21
Seafood	1,901	1.52	311	0.94	1,607	1.63		
Shelf Stable Juices & Drinks	1,267	1.01	383	1.16	887	0.9	3	0.07
Shortening & Cooking Oils	509	0.41	112	0.34	423	0.43	24	0.57
Skin Care	431	0.34	127	0.39	314	0.32	10	0.24
Social Expressions	2,028	1.62			2,028	2.06		
Soft Beverages	707	0.57	167	0.51	541	0.55		
Soups	1,351	1.08	370	1.12	1,011	1.03	30	0.71
Syrups & Pancake/Waffle Mix	291	0.23	65	0.2	227	0.23	1	0.02
Tobacco And Smoking Needs	1,831	1.46	677	2.05	1,154	1.17		
Total	125,048	100	32,961	100	98,430	100	4,221	100

Table 4: Number of Distinct Products by Product Group for both countries, Canada, the United States and the set of uniquely matched products.