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: February 11, 2010

Abstract

To what extent do national borders impose costs that segment markets across countries? We answer this classic question using a data set with product level retail prices and wholesale costs for a large grocery chain operating in the U.S. and Canada. Exploiting comovements between relative prices, costs and the nominal exchange rate within and across borders we show that retail markets are segmented at the border, while domestic markets are integrated. We find that in response to exchange rate changes, relative retail prices between U.S. and Canadian stores are driven entirely by the movements in relative wholesale costs. We propose a regression discontinuity design, using the border as a treatment, to estimate a lower bound on the true size of border costs. The median absolute price discontinuity at the border is large: 24 percent for consumer prices and 24 percent for wholesale costs. On the other hand, within country discontinuity is 0% for prices and costs.

*We are grateful to Paul Bergin, Stefano DellaVigna, Jack Duane, Charles Engel, Edward Glaeser, Penny Goldberg, 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

How large are the costs imposed by international boundaries on the flow of goods across countries? Are these border costs larger than the costs of trade within the same country, and are they large enough to segment markets between countries relative to markets within countries? Some of the classic questions in international economics, ranging from the gains from market integration to the transmission of shocks across borders hinge on the answer to these questions.

There is a large literature that estimates the border cost –that is, the *additional* transaction costs incurred when markets are located in different countries– from price gaps of similar goods across borders. There are, however, two difficult issues that need to be addressed before we can interpret the price gap as the border cost. First, price gaps across borders are also affected by local transaction costs and differences in market conditions, and these factors would also result in cross-border price gaps even if the border cost was zero. A solution proposed by Engel and Rogers (1996) is to control for these other factors by measuring cross-border price gaps relative to price gaps between markets in the same country. However, Gorodnichenko and Tesar (2009) point out that this approach is valid only if heterogeneity in transportation costs and demand are of similar magnitude between countries as within countries, which is unlikely to be the case.

The second issue is that the logic of using price gaps to infer trade costs implicitly assumes that markets remain integrated despite these costs. However, a critical question we want to investigate is whether trade costs are large enough to segment markets and, in particular, whether international markets are more segmented than domestic markets. When markets are segmented, the arbitrage condition that underlies the logic of using price gaps to infer trade costs no longer holds. In that case, the price in each market is only determined by local costs and local market conditions. The price gap may be large or may be small, but does not tell us whether trade costs are large or small. All that we can say with certainty is that the trade cost is higher than the observed price gap.

This paper tackles these two questions – what is the magnitude of the border cost and are these costs large enough to segment markets – with a new approach and with new data. We first address the question of whether international markets are more segmented than domestic markets. Instead of looking at price gaps per se, the key idea we exploit is that the response of prices to cost shocks will depend on the extent of market segmentation. When markets are integrated, the price in a given market will be affected by cost shocks in other markets. When markets are not integrated, prices will only be a function of costs and demand in the same market and will not be affected by cost shocks in other markets.

We implement this idea on a dataset from a large retailer operating in the U.S. and Canada. This dataset provides weekly data at the barcode level on retail prices and wholesale costs for 250 U.S. stores (in 19 states) and 75 Canadian stores (in 5 provinces) for 178 weeks between January 2004 and June 2007.

We report three findings. First, we show that cross-border price gaps are large and exceed within-country price gaps, both within the U.S. and within Canada. We find a similar result for whole-sale costs, which is quite striking since these costs are highly tradable.[1] Unconditionally, movements in cross-border prices are accounted for by both significant movements in cross-border costs and mark-ups. Across products, the median contribution of wholesale costs ranges between 55 percent at the weekly horizon to 77 percent at the quarterly horizon.

However, and this is our second main finding, we show that the variation in cross-border prices between U.S. and Canadian stores in response to changes in the nominal exchange rate is driven entirely by the variation in relative wholesale costs, with no change in the relative markup. This suggests that the well known fact that real exchange rates track the nominal exchange rate is driven entirely by the response of wholesale costs to the nominal exchange rate.

Third, we test whether prices respond differently to shocks in wholesale costs in neigh-

[1] From a consumer's perspective fairly small transaction costs can effectively segment markets. By contrast, at the wholesale level, given the large volumes involved, the gains to arbitraging even small price gaps are large. This is why the evidence on whole-sale costs can be particularly informative.

boring regions depending on whether the neighboring region is located in the same country or in a different country. We find that controlling for the store's own wholesale costs, an increase in wholesale costs in neighboring markets of 1 percent increases the store's retail price by 0.12-0.13 percent when the neighboring store is in the same country. In contrast, prices do not change in response to changes in wholesale costs of neighboring stores when the neighboring store is in a different country. These findings suggest that while domestic markets are likely to be integrated, international retail and wholesale markets are not, for the products we consider. The fact that international markets are not integrated implies that the cross-border price gaps will only provide us with a lower bound estimate of the border cost.

We then estimate a lower bound of the border cost from price data. To address Gorodnichenko and Tesar's (2009) admonition about the dangers of using internal price gaps to control for transaction costs not due to the border, we use a regression discontinuity design to measure the discontinuous change in prices at the border. Our data provides the precise geographic location of the store, which we use to answer the following question: what is the magnitude of the deviation from the law of one price between stores located right across the border from each other? The idea behind the regression discontinuity design is that market conditions are likely to be similar for stores located close to each other. Therefore, any discontinuity in prices we observe at the border will largely be due to the border cost. To illustrate the nature of our evidence, figure 1 plots the (log) average price across stores (in 50-kilometer bins) for 25 ounce bottles of Perrier Sparkling Natural Mineral Water against the individual store's distance from the border. As is evident, there is a clear price discontinuity at the border. We find evidence consistent with the discontinuity evident in figure 1 in our broader sample of products: the median absolute price discontinuity is 24 percent for consumer prices and wholesale costs across our sample of matched products.

This paper builds upon the large body of work measuring the effect of the border on trade costs and market segmentation, particularly the work by Engel and Rogers (1996) and

Gorodnichenko and Tesar (2009) already cited.[2] A recent paper by Broda and Weinstein (2007) follows the approach taken by Engel and Rogers (1996) to estimate the border cost using detailed price data for individual products (UPCs) collected at the consumer level and conclude that there is no border effect. The reason our results differ is because our data comes from a single retailer in all locations, whereas Broda and Weinstein (2007) compare the price at which different consumers purchase a particular good without controlling for the fact that different consumers purchase identical goods from different retail establishments.[3] The literature on pricing-to-market relates to the differential pricing behavior of the same firm for the same product, which is why it helps to have data for the same retailer. The other key differences from the existing literature are that our methodology departs from Engel and Rogers (1996) in that we explicitly test for the effect of border costs on market segmentation and use a regression discontinuity approach to estimate a lower bound on the border cost (given our finding that international markets are segmented). Finally, we have information on the wholesale cost paid by the retailer, which is crucial to our tests of market segmentation.[4]

This work is also related to the literature that decomposes price differences across markets into mark-ups and costs. In the absence of systematic cost data, this decomposition has been limited to a few studies of specific goods such as beer in Goldberg and Hellerstein (2006) and *Big Mac* in Parsley and Wei (2007). In other cases, in the absence of cost data, mark-ups are structurally estimated using price and quantity data such as Goldberg and Verboven (2001, 2005) who study the automobile car market in Europe.

The paper proceeds as follows. Section 3 describes the data. Section 2 describes the theoretical motivation for our empirical specifications. Section 4 presents evidence on price

[2]Engel and Rogers (1996)'s seminal work has been followed up by many authors. Crucini and Shintani (2006) and Crucini et al. (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 McDonald's *duty-free* outlets (Asplund and Friberg 2001).

[3]Broda et al. (2009) shows that price heterogeneity across retailers is attributable in part to differences in shopping experiences and amenities provided by different retailers.

[4]In contemporaneous work, Burstein and Jaimovich (2008), also examine the pattern of wholesale prices in the United States 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.

gaps and an unconditional variance decomposition of these price gaps into costs and markups. Section 5 evaluates the conditional response of price gaps and cost gaps to exchange rate shocks to evaluate market segmentation and Section 6 discusses estimates of border costs using the regression discontinuity approach. Section 7 concludes.

2 Theory

Our goal is to measure the size of cross border transaction costs, both in absolute terms and relative to the cost of arbitraging costs within countries. The border cost, and not price gaps per se, is the object of interest as it is the variable that determines the benefits of market integration and the extent of the transmission of shocks across borders. This section presents a simple model to guide our empirical approach. Consider a world where markets (indexed by i, j, or k) are distributed across space. Consumer h in market i can pay a price p_i to purchase a homogenous good in market i or can pay a (per unit) transaction cost t_{ij}^h to purchase an identical good at price p_j in market j. Notice that we write this transaction cost as specific to consumer h.[5] Different households in a given market may face different transaction costs depending on their demographic characteristics and preferences. For instance, wealthier households can have a higher opportunity cost of time. Further, we write this transaction cost as the sum of two components: $t_{ij}^h = \tilde{t}_{ij}^h + b$. The first term, $\tilde{t}_{ij}^h$, represents the transaction cost for household h were markets i and j to be located in the same country. The second term, b, represents the *additional* transaction costs incurred when markets i and j are located in different countries. This is what we refer to as the 'border cost'.

It should be immediately obvious that estimating b is not an easy task. It requires first estimating the total transaction costs between markets i and j, t_{ij}^h, then subtracting what the transaction costs would have been, if markets i and j had been located in the same country, for households with similar characteristics, $\tilde{t}_{ij}^h$. Both of these tasks are empirically

[5]In the model presented in appendix B, the equilibrium transaction costs t_{ij}^h are a function of demographic characteristics.

difficult.

The first task can be accomplished by making additional assumptions on the structure of markets and competition, to establish an explicit mapping between the distribution of prices across markets $\{p_i\}_i$ and the distribution of total transaction costs $\{t_{ij}^h\}_{ijh}$. For example, in appendix B, we solve for the equilibrium spatial distribution of prices in a modified Salop (1979) model where firms are equally spaced on a circle, consumers are distributed uniformly and face local transaction costs as well as border costs. In general, the shape of the mapping depends on the specific assumptions of the model. However, even without making the strong assumptions necessary to obtain explicit expressions for the equilibrium spatial distribution of prices, simple arbitrage arguments can help us characterize the relationship between transaction costs and retail prices, regardless of the specifics of the model. Two cases need to be considered: when markets are integrated, or fully segmented. We describe these two equilibria and focus on their implications for the effect of transaction costs t_{ij}^h on price gaps.

Integrated Markets: We define markets i and j as integrated if equilibrium prices in these two markets are such that at least one consumer h in one of the markets is indifferent between buying in the market she lives in or paying t_{ij}^h to buy in the other market. Since this marginal consumer is indifferent between the two markets, the price gap $|p_i - p_j|$ must be equal to the transaction cost t_{ij}^h. It follows that when markets are integrated, observing price gaps is equivalent to observing total transaction costs for at least some households. It does not follow, however, that the comparison of price gaps within and across countries provide an estimate of the border cost, b. To see this, suppose that markets i, j and k are integrated, with markets i and j located in the same country and market k located across the border. In that case $|p_i - p_j| = \tilde{t}_{ij}^h$ while $|p_i - p_k| = \tilde{t}_{ik}^l + b$ for some households h and l. Without additional assumptions, b cannot be directly recovered from the observation on price gaps since $\tilde{t}_{ik}^l$ is not observable. An example of one such assumption, often implicitly made in the literature, is that within-country transaction costs are the same regardless of markets: $\tilde{t}_{ij}^h = \tilde{t}_{ik}^l$ for all markets j, k and households h, l. Under this assumption, the border effect can be recovered from the difference between cross border and within-country price gaps. This is problematic on several grounds. First, transaction costs across markets and countries may differ due to differences in the physical environment as well as household characteristics. Second, it requires that markets are integrated within and across-borders, a fact that is often implicitly assumed, and not empirically established.

Segmented Markets: We define market i and j as segmented when t_{ij}^h is large enough relative to the price gap between the two markets such that all consumers in either market are better off purchasing the good in the market where they live. Consider again markets i, j and k with markets i and j located in the same country and market k across the border. Assume now that all three markets are segmented. Then $|p_i - p_j| < \min_h \tilde{t}_{ij}^h$ and $|p_i - p_k| < \min_h \tilde{t}_{ik}^h + b$. When markets are segmented, the price gap is not a function of t_{ij}^h and only provides a lower bound on the distribution of transaction costs. This lower bound is uninformative about the size of t_{ij}^h. Moreover, a comparison of these two lower bounds is uninformative about the size of b even under the additional assumption that $\tilde{t}_{ij}^h = \tilde{t}_{ik}^l$ for all markets j, k and households h, l.

From the previous discussion, it is clear that price gaps alone provide limited information about the degree of market segmentation. To solve this problem, we use information on both prices and costs. The idea is simple. When markets i and j are fully segmented, an increase in relative costs in market i, c_i, will be passed through to prices in market i, but not to prices in market j, since the latter is shielded from market i. It follows that price differences across markets will move closely with cost differences, and markups will not adjust much. On the other hand, when markets are integrated, firms in market i will lose market share if they raise their prices in response to an increase in c_i. In equilibrium, markups will adjust to absorb some of the cost shock. Conversely, firms in market j now face lesser competition from firms in market i and will adjust their markup upwards. The pass-through of relative costs into relative prices will be smaller. This is the basic insight of the pricing-to-market literature pioneered by Dornbusch (1987) and Krugman (1987).[6]

[6] Recent applications of this include Atkeson and Burstein (2008) and Gopinath et al. (2007), among others. Appendix B derives this result formally.

To make this insight explicit within our framework, consider two markets i and j. To fix ideas, assume that $\triangle p = p_i - p_j > 0$, so that all households in market j purchase the good in their local market, while some households in market i may purchase from market j if transaction costs are sufficiently low. Denote $G(t)$ the cumulative distribution of households in market i over transaction costs t. Write also the individual demand for the good $d(p_i; t) \equiv p_i^{-\sigma} h(t)$, where we assume that the price elasticity of demand σ is constant. It is immediate to verify that the price elasticity of aggregate demand when markets are segmented is simply $\varepsilon_s = \sigma$. The corresponding optimal markup is constant, equal to $(1 - 1/\varepsilon_s)^{-1}$: changes in costs c_i are fully reflected in p_i; changes in foreign costs c_j have no effect on local prices. When markets are integrated, the price elasticity of demand in market i becomes $\varepsilon_i(p_i, \triangle p) = \sigma + p_i h(\triangle p) g(\triangle p) / \int_{t \geq \triangle p} h(t) dG(t)$. The second term captures the effect of a change in prices on the mass of i households that decide to purchase the good from market j instead of market i. This term makes the price elasticity of demand in market i responsive both to the local price p_i and the foreign price p_j. In equilibrium, p_i will respond to changes in local costs c_i (with an elasticity smaller than one) and to foreign costs c_j. The pass through of local and foreign costs into local prices reveals the extent of market segmentation.

To accomplish the second task –estimating b– suppose we can observe the same good sold in stores located on opposite side of the border. Now, take the limit as the distance between these stores goes to zero. If market conditions are similar on either side of the border, this implies that the domestic transaction cost term converges to $\tilde{t}_{ii} = 0$. The resulting price gap $|\triangle p|$ directly provides either an estimate of b (when markets are integrated) or a lower bound on b (when markets are segmented). More generally, as long as local transaction costs on both sides of the border become more symmetric as we near the border, the cross border price gap will provide a lower bound on border cost b.[7] This suggests that we look directly at stores located close to the border and motivates the Regression Discontinuity (RD) approach presented in section 6.

There are three points we take away from this discussion. First, price gaps provide infor-

[7]This result is spelled out formally in Appendix B.

mation about the size of transaction cost across markets only when the latter are integrated. The literature on border costs implicitly makes this assumption, which, as we will see, is unwarranted. Second, the key difference between integrated and segmented markets is in the response of price gaps to cost shocks. When markets are integrated, prices in one market respond to changes in costs in the other market, not so when markets are segmented. A related implication is that the pass-through of relative costs into relative prices is high when markets are segmented, and low when markets are integrated. Lastly, one can obtain a direct estimate of (a lower bound of) the border cost b by examining price gaps for stores located close to each other and across the border. These three insights guide our empirical strategy.

3 Data source

We have access to weekly product-level data for 325 grocery stores in the U.S. and Canada (250 stores in 19 U.S. states and 75 stores in 5 Canadian provinces) operated by a single retail chain.[8] Figure 2 plots the location of the stores in our data. Most U.S. 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 United States. The dataset contains information for 125,048 unique products (UPCs) sold in these stores between January 2004 and June 2007 (178 weeks). Most of the products are in the food and beverages categories, housekeeping supplies, books and magazines, and personal care products.[9] This level of disaggregation allows for a very precise identification of products. For instance, in our data, a 25 ounce Perrier Mineral Water with a Lemon Twist and a 25 ounce Perrier Mineral Water with a Lime Twist are two separate items in the soft beverages product group.

The two key pieces of information we use from the data are the price and the marginal

[8]This chain is one of the leading food and drug retailers in the U.S. and Canada and operates directly or through subsidiaries a total of 1,400 stores in the United States and 400 in Canada. The data sharing agreement between this retailer and the research community is managed through the SIEPR-Giannini data center (http://are.berkeley.edu/SGDC/).

[9]Table 8 in appendix D reports a breakdown of UPCs by product categories.

cost of each product (for every store and every week). The retailer reports "gross" and "net" revenues for each product-store-week. Gross revenues refers to revenues computed at the retail list price, while net revenues measure revenues net of rebates, promotions and coupons. We construct corresponding gross and net prices by dividing revenues by quantities.[10]

As for the marginal cost, our data set includes two pieces of information. The retailer reports the "wholesale cost" which is the list price of the product at the wholesale level (vendor cost). The dataset also reports "adjusted gross profits" per unit for each product, defined as the net price minus the sum of wholesale costs and transportation costs plus net rebates from the manufacturer.[11] We subtract these adjusted gross profits from the net retail price to back out the "net cost" of each product. The precise link between the wholesale cost and our imputed net cost is as follows:

$$\begin{aligned}\text{Net cost} &= \text{Wholesale cost} + \text{Freight and Transportation Costs} - \text{Net Rebates} \quad (1)\\ &= \text{Net price - Adjusted gross profit}\end{aligned}$$

It is important to note that neither measure of costs includes local costs (such as labor, rent, and utilities) at the store level. At short horizons, with rent, capital, and labor taken as given, it is natural to interpret the net cost as the full marginal cost of the product faced by the retailer and the net price as the actual price of the product. Unless specified otherwise, our empirical analysis refers to net prices and net costs.[12]

[10] Both the gross and net retail price exclude U.S. sales as well as Canadian federal value-added taxes (VAT) and provincial sales taxes. From a consumer's perspective the relevant price is the price inclusive of sales taxes and VAT. 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 United States and Canada. In general, we found that sales taxes and VAT are higher in British Columbia (13 percent) as compared to Washington State (around 8 percent). To the extent that before-tax prices are higher in Canada than in the United States, as we find for a majority of goods in our sample, this implies that the after-tax price gap between the two countries is larger than what we measure. More importantly, since the VAT and sales tax remained more or less constant over time, they cannot explain the pattern of co-movement with exchange rates that we observe in the data.

[11] Specifically, "adjusted gross profits" is defined as net price minus wholesale cost plus "total allowances." In turn, the documentation provided by the retailer defines "total allowances" as "the sum of shipping allowances, scan allowances, direct-store-delivery case bill back allowances, header flat allowances, late flat allowances, and new item allowances, *minus* the sum of buying allowances, freight allowances, overseas freight, and distress and other allowances."

[12] Our results are qualitatively unchanged if we use instead gross prices and wholesale costs.

Our first task consists in restricting the initial sample of 125,048 unique products to a set of products that appear on both sides of the border in at least one week. This matched set represents the set of goods for which we can evaluate deviations from the law of one price (LOP). It contains 4,221 unique products, or about 3.3 percent of the original dataset.[13] This decline in matched products across the border is an important effect emphasized in Broda and Weinstein (2007) that carries across to our dataset, and underlies the importance of working with unique products.

One concern is that otherwise identical goods have different UPCs because of different labeling requirements in the United States and Canada (for example, language and nutritional information), so that only goods with common labeling would be included in our matched sample. To assess this possibility, we visually inspected the labels of our matched UPCs in a store in the United States (Oakland, CA) and in Canada (Vancouver, BC). We found identical physical characteristics for all matched products, but often different labels in the two countries. Thus, it seems that different labeling does not necessarily imply different UPCs.[14]

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).[15] The distribution of goods across product groups is very skewed, with a median around 11 and a mean of 97.[16] Panel A of table 1 reports information on the number of distinct products (among matched goods) per store-week and per store-pair-week in our data. The average U.S. store in the data carries 493 distinct matched products for which we have data in a

[13]We arrive at this number in the following way. We start with the set of unique UPCs that appear in at least one U.S. 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 digits 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).

[14]It is possible, of course, that many identical products could still have different UPCs in the two countries, which could be a factor behind the low match rate. Matching goods that do not share the same UPC is not currently feasible given the limited product information we have.

[15]"Books and magazines" have a printed sale price that is sticky in the local currency. We find that all our results are roughly unchanged if we exclude this category of goods.

[16]See table 8.

typical week. We find about 272 (251) matched products for a typical within-country store pair in the United States (Canada) in a given week, and 167 for a cross-border store-pair.

Finally, we link the precise geographic coordinates of each store to data from the Canadian and U.S. censuses to measure the characteristics of the local market served by each store.[17] Panel B of table 1 provides some summary statistics for these local market characteristics. The median store in our sample is located in an area with a population density of 810 persons per square kilometer in the U.S., with a density of stores equal to 0.15 stores per square km[18] and with a median household income of US$57,040 in the year 2000. There is significant variation across the stores in our sample, with U.S. population density ranging from 57.2 to 2671 persons per square km, supermarket density ranging from 0.01 to 0.85 stores per square km, and median household income ranging from $34,238 to $82,592.

4 Price Gaps

This section presents three sets of summary statistics of the data. First, we present the distribution of the gap in average prices, costs, and markups between U.S. and Canadian stores. Second, we present the distribution of the gap in price, cost, and markups between all store-pairs, focusing on the difference between store-pairs in the same country and store-pairs located in different countries. Third, we decompose the variance of average price gap between the U.S. and Canada into the variance of the gap in average wholesale costs and the variance of the gap in average markups.

4.1 Cross-border price gaps

We begin by presenting statistics on the deviation in prices, costs, and markups between U.S. and Canadian stores. For every product, we compute the difference in the (log) average

[17] The U.S. data comes from the U.S. population census and economic census data base. The Canadian data comes 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 census subdivision level while U.S. data is collected at the zip code level.

[18] These are establishments in NAICS 445110 (supermarkets and other grocery stores, but not convenience stores) similar to the stores in our data.

price in all U.S. stores from the (log) average price in all Canadian stores. We do the same for costs and mark-ups. Figure 4 plots the resulting distribution of cross-border gaps of the average net price, imputed cost and mark-up for the first week of 2004 (2,242 UPCs) and the twenty-first week of 2007 (2,267 UPCs).[19] A positive value indicates that the average price (and cost or markup) is higher in Canada than in the U.S. Figure 4(a) shows that there is large dispersion of gaps in average prices between the US and Canada across UPCs. For the first week of 2004, 9.9 percent [resp. 4.8 percent] of products are 25 percent more expensive in Canada [resp. the U.S.]. By the twenty first week of 2007, the corresponding numbers are 28.5 percent and 1.9 percent. This shift in the distribution of the average price gap reflects the appreciation of the Canadian dollar relative to the U.S. currency between 2004 and 2007. While the cross-border price gap for any individual UPC is likely to be dominated by idiosyncratic factors, the distribution shifts in line with the exchange rate. This finding is consistent with Crucini and Shintani (2007)'s evidence from more aggregated data.

Figures 4(b) and 4(c) report the corresponding distribution of the cross-border average marginal 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. Between 2004 and 2007, the fraction of products with marginal costs at least 25 percent more expensive in Canada [resp. the U.S.] changes from 14.9 to 34.9 percent [resp. 3.2 to 3.1 percent]. In contrast, the fraction of products with markups at least 25 percent higher in Canada [resp. the U.S.] varies from 2.4 to 3.7 percent [resp. 8.4 to 6.8 percent].

4.2 Price dispersion across stores

We now measure price dispersion across all the stores in our sample, focusing on the price gap between stores located in the same country versus the price gap between stores located in different countries during the first week of 2005. We have a total of 31,125 store-pairs in the US, 2,775 store-pairs in Canada, and 18,450 cross-border store-pairs. The median

[19] This corresponds to the first and next to last week of our sample. There is a significant drop in the number of UPCs in the last week of our sample, which is why we use the next to last week.

number of common UPCs that week is 260 for US store pairs, 242 for Canadian store pairs, and 170 for cross-border pairs. For all the common products in each store pair, we compute the difference in the log price between the two stores. Panel A in Table 2 presents statistics across store pairs on the mean, median and maximum of the absolute price gap for store pairs located in the US (USA-USA), Canada (CAN-CAN), and across the border (CAN-USA). The median price gap across store-pairs is 3.7 percent for U.S. store-pairs, 0 percent for Canada store-pairs, and 14.6 percent for cross-border pairs (col. 2). Since the U.S. store is always treated as the store of reference, this implies that Canadian retail prices were 15 percent higher than U.S. prices in the first week of 2005.[20]

The finding that the dispersion of U.S. prices is larger than the dispersion of Canadian prices is consistent with the evidence in Engel and Rogers (1996) and Broda and Weinstein (2007). However, unlike Engel and Rogers (1996) and Broda and Weinstein (2007), we find that the dispersion in cross-border price gaps vastly exceeds that of within country price gaps. As Gorodnichenko and Tesar (2009) point out the dispersion of prices between U.S. city pairs in Engel and Rogers (1996) is about the same as the dispersion between U.S. and Canadian city pairs. Broda and Weinstein (2007), using barcode level data, also find that deviations from the law of one price are similar within and across countries. However, Broda and Weinstein (2007) measure prices at which different consumers purchase identical goods without controlling for the fact that these goods were purchased from different retailers. In contrast, our data comes from the *same* retailer operating in all locations, which is the relevant unit of observation for evaluating pricing to market and deviations from the law of one price.

Panel B indicates similar results for the median absolute marginal 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-Canadian pairs (0 percent).[21] This finding is surprising given that

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

[21]The corresponding numbers (not reported) for the median (across store-pairs) of the median absolute markup gap (across UPCs) are 6.2 percent within the U.S., 1.3 percent within Canada and 10.5 percent for cross border pairs.

wholesale costs refer to prices of the most tradable component of the retailer's costs.

4.3 Variance Decomposition of Price Gaps

Prices in our sample change very frequently. The median frequency across UPCs is 0.41 for net prices (0.22 for gross prices), implying a median duration of 2.4 (4.5) weeks.[22] Using the decomposition of prices into wholesale costs and markups, we can write the change in the average price gap for product k between the U.S. and Canada at horizon j as:

$$\Delta_j \ln\left(\frac{p^{*k}}{p^k}\right) = \Delta_j \ln\left(\frac{c^{*k}}{c^k}\right) + \Delta_j \ln\left(\frac{\mu^{*k}}{\mu^k}\right) \tag{2}$$

where $\Delta_j x \equiv x - x_{-j}$ is the $j-$period difference operator, p^k denotes the average price of product k in the U.S. while p^{*k} is the average price (in U.S. dollars) in Canada. c^k and μ^k denote respectively the average net cost and average markup in the U.S. (with similar definitions in Canada). Table 3 decomposes the variance of changes in cross-border price gaps (the left hand side of equation (2) into a net cost and markup components, across products for various horizons j (one week to one year).[23]

The table indicates that both wholesale cost gaps and markup gaps contribute to the variability in price gaps. Across products the median contribution of net costs ranges from 55 percent at the weekly horizon to 77 percent at the quarterly horizon. This indicates that –unconditionally– a substantial fraction of the movements in cross border prices is accounted for by relative movements in retail markups (between 23 and 45 percent). This fact is consistent with the within-country evidence documented in Eichenbaum et al. (2008).

[22] We construct the frequency number as follows: we start with the frequency of price adjustment for each UPC-store combination; Next we estimate the average frequency across stores for each UPC. Finally, we report the median (across categories) of the median (within category) frequency.

[23] Our variance decomposition splits evenly the covariance between relative costs and markups into the cost and markup components. Formally, we estimate the share of costs as $cov\left(\Delta_j \ln\left(c^{*k}/c^k\right), \Delta_j \ln\left(p^{*k}/p^k\right)\right)/var\left(\Delta_j \ln\left(p^{*k}/p^k\right)\right)$ with a similar expression for the markup share.

5 Price Gaps and Market Segmentation

The previous section establishes the existence of larger cross-border than within country price and cost differences. However, this fact by itself does not tell us that arbitrage costs are larger across borders than within countries. A first step is to establish whether markets are more or less segmented across the U.S.-Canadian border than markets within the U.S. or Canada.

We evaluate whether retail markets are segmented in two ways. First, if markets are segmented across borders, then relative prices across stores will move closely with relative costs for reasons discussed in Section 2. We examine the relation between relative prices, relative costs and mark-ups conditional on fluctuations in the U.S.-Canadian dollar nominal exchange rate. Fluctuations in the nominal exchange rate are plausibly exogenous to relative market conditions for any single product in our sample. Moreover, they affect differentially stores located on each side of the border, while being common to all stores within the same country. This evidence indicates that both retail and wholesale markets are segmented by the border.

Second, we evaluate how prices charged by a given store comove with that store's cost and with the costs of the same product in nearby stores. When markets are segmented, prices charged by a store should be insensitive to cost shocks to a neighboring store, conditioning on their own cost. When markets are integrated, prices will comove with these cost changes even after conditioning on the store's own cost. This evidence confirm that retail markets are segmented across borders, but integrated within countries.

Overall, our results strongly suggests that the U.S.-Canada border almost perfectly segments the retail and wholesale markets that we examine, while within country retail markets appear partially inegrated.[24]

[24] The evidence that some consumers cross the border to arbitrage price differences (e.g. **?**) does not invalidate our results. Our results simply indicate that the price setting decision by the stores in our sample is not significantly affected by these consumers.

5.1 Median Deviations over Time

We begin by presenting the time variation in the median gap in the average price in U.S. vs. Canadian stores located within 200 km. of the border. The top left part of figure 3 indicates that the median price gap has increased over time, from roughly −5 percent in June 2004 to 15 percent in June 2007. The figure also reports (the dashed line on the right-axis) the (log) U.S./Canadian nominal exchange rate expressed as the U.S. dollar price of the Canadian currency. As is evident, the evolution over time in the median price gap mirrors almost perfectly the evolution of the nominal exchange rate.

The top-right and bottom-left panels perform the same exercise for net costs and the resulting markup. The figure reveals that the movements in the median cross-border cost gap tracks very closely the movements in the nominal exchange rate. This fact is surprising since wholesale costs capture the most tradable component of the retailer's total costs (in particular, excluding local labor and non-traded costs). It is clear that the median price gap and median cost gap move closely together, while relative markups show barely any response to the fluctuations in the exchange rate. This suggests that the well known fact that real exchange rates track the nominal exchange rate is driven entirely by the response of wholesale costs to the nominal exchange rate. This result is robust to the definition of the price (gross versus net) or of the costs (wholesale versus net).[25] According to the theoretical discussion presented in section 2, these results indicate that retail markets are perfectly segmented by the border.[26]

These results differ from the unconditional variance decomposition results presented in section 4 precisely because we are now looking at co-movements with the nominal exchange rate.

[25] It is also robust to restricting the sample to goods that adjust prices frequently.

[26] For a more complete analysis of the nature of market segmentation at the whole-sale level we would need information on costs faced by the wholesaler, which we do not have.

5.2 Conditional variance decomposition

Next, we explore formally the contribution of cross-border cost movements due to nominal exchange rates to variations in cross-border prices. We start by regressing the j-period change in the cross-border (net) price gap $\Delta_j \ln p^{*k}/p^k$ for product k sold in stores located within 200 km from the border on the j-period change in the nominal exchange rate, $\Delta_j \ln S$:

$$\Delta_j \ln p^{*k}/p^k = \alpha^k + \beta^k \Delta_j \ln S + \epsilon_j^k \tag{3}$$

The top panel of table 4 reports the median, 25th and 75th percentile pass-through coefficient β^k, across products, for various horizons between 1 week and 1 year. We find a distribution of pass-throughs narrowly centered on 1 at all horizons.

The second and third parts of the top panel of table 4 report the same regression for the net cost and associated markups. The median pass-through coefficient on costs is also very close to 1, while the pass-through coefficient on markups is always small and close to zero.[27]

Taken together, these results indicate that cross-border wholesale costs account for almost all of the pass-through of exchange rates to cross-border prices. In light of our earlier discussion, we interpret these results as evidence of near perfect segmentation of both retail and wholesale markets between the U.S. and Canada.

5.2.1 Price indexes

So far we have compared products with the same UPCs. This approach has the virtue of comparing identical products in the two countries and avoids the problem, pointed out by Broda and Weinstein (2007), that aggregation averages out idiosyncratic price shocks and creates an artificial border effect through the volatility of common price shocks (such as the nominal exchange rate). One drawback is that the sample of products with common UPCs is a small subset of all the products available in our data set. We now expand the sample of products by constructing price indexes at the store level for each product group and product

[27] The coefficients need not sum to 1 since the median of the average is not the same as the average of the medians.

class. There are 61 product groups and 1165 product classes in the first week of 2004. At these levels of aggregation the match rate across borders is very high: 96 percent for product groups and 70 percent for product classes.[28]

Consistent with the exchange rate pass-through regressions presented in section 5.2, we use all stores within 200km of the border. We calculate separate weekly averages of the percentage change in the price index across stores in each country. We then estimate regression (3) on each product group and product class. As our price indexes are calculated in US dollars, the coefficient on the exchange rate is analagous to the one we calculated before for individual products.

The results are reported in the bottom two panels of table 4. While there is more dispersion in the estimated pass-through coefficient the median remains close to one at all horizons. This provides further evidence of segmentation across the Canada-US border and indicates that our results are not driven by special characteristics of the small set of matched goods.

5.3 Cost-Price Pass-Through

To compare the extent of retail market segmentation within and across countries we estimate cost pass-through regressions. These regressions allow retail prices in one region to depend on costs in the same location as well as costs in neighboring regions. As our theoretical discussion indicates, under segmentation, local product prices should not depend on cost and market conditions in other markets. By contrast, when markets are integrated, local product prices should vary with cost and market conditions in other markets. We find evidence of pass through from neighboring region within country but no evidence of pass through from adjacent regions located in another country.

We implement these regressions for our sample of matched products in the following way. First, we compute the average price and net costs for each product across Washington State stores located within 200km of the U.S.-Canada border (near U.S.). We compute the

[28] For details about the construction of the price index refer to appendix C.

same averages in the next 200km band within the U.S (far U.S.). These stores are located in Washington and Oregon. We replicate the exercise on the Canadian side of the border with stores within 200km of the border in British Columbia (near Canada) and Alberta (far Canada). We then regress for each product the rate of change of the price $\Delta \ln p_i^k$ on the rate of change of its own cost $\Delta \ln c_i^k$ and the rate of change of costs in adjacent regions $\Delta \ln c_j^k$, located on the same side or on opposite side of the border.

$$\Delta \ln p_i^k = \alpha \Delta \ln c_i^k + \beta \Delta \ln c_j^k + \epsilon^k \tag{4}$$

Table 5 reports the results. Across all goods and locations, the median pass-through from own costs to prices is positive and large, ranging between 0.24 and 0.52. The cost pass through from neighboring regions is smaller but still sizeable within countries, especially at shorter horizons. It ranges from 0.12 in the U.S. to 0.15 in Canada for weekly price changes. This indicates significant comovements between prices in one region and costs in an adjacent region within each country. By contrast, the cross-border cost pass through reported in the middle panel is always zero, regardless of the horizon.

While these regressions coefficients should be interpreted simply as establishing some degree of co-movement between costs and prices, and not as the causal impact of changes in costs on prices (unlike the case of the exchange rate shock), they nevertheless confirm our earlier diagnostic that domestic retail markets exhibit less segmentation than cross border markets.

6 A Regression Discontinuity Estimate of the Border Effect

The previous section establishes that markets are segmented across countries. If follows that cross-border price gaps provide a lower bound on *total* cross border transaction costs t_{ij}. This section proposes the use of the regression discontinuity (RD) design, following the discussion in Section 2, to estimate directly a lower bound on the border cost b.[29]

The RD approach allows us to answer the following question: by how much do prices of goods sold in stores located immediately across the border differ?[30] The motivation for the RD approach is twofold. First, equilibrium prices depend on many local factors such as the elasticity of substitution across stores, or demographic characteristics, all of which impact the effective transaction costs for a household and can vary with location. The RD approach controls for all these determinants. Second, by explicitly controlling for distance, we approximate more closely the component of the transaction costs directly associated with the border.

We describe the details of the Regression Discontinuity methodology in appendix A. Formally, we follow Imbens and Lemieux (2007) and use a local linear regression approach including distance to the border as a regressor, interacted with a border dummy:

$$\ln p_i^k = \alpha^k + \gamma^k C_i + \theta^k D_i + \delta^k C_i \cdot D_i + \beta^k X_i + \epsilon_i^k. \tag{5}$$

As before, p_i^k denotes the U.S. dollar price of good k sold in location i. C_i is a border dummy equal to 1 if the store is located in Canada and zero otherwise, D_i denotes the algebraic distance of store i to the border in km. (positive for U.S. stores and negative for Canadian ones) and X_i measures other important observable characteristics of market i. Finally, ϵ_i^k captures unobserved characteristics that are store-and good-specific. The parameter of interest is γ^k. It estimates (a lower bound on) the border effect for good k, expressed as a percent of the price of that good.[31]

Graphically, figure 1 illustrates how the regression discontinuity approach can recover an estimate of the border effect. The figure plots the (log) average price across stores (in 50-

[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 U.S. states.

[31] The local linear regression also restricts the sample to stores within a distance $\bar{D}_k$ from the border. In practice, we set $\bar{D}_k$ to 500km. Imbens and Lemieux (2007) recommend choosing $\bar{D}_k$ using a standard cross validation procedure. For most products the optimal bandwidth is either 100, 350 or 500km. Results are unchanged if we adopt the optimal bandwidth.

kilometer bins) for 25 ounce bottles of Perrier Sparkling Natural Mineral Water against the individual store's distance from the border. As is evident, there is a clear price 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. The discontinuity at the border is interpreted as a lower bound (since markets are segmented) on the pure effect of the border.

As discussed earlier, the covariates X_i capture important demand characteristics that might vary with location.[32] We include population density, density of supermarkets and other grocery stores, the proportion of people aged 0–19 years and aged 65 years and over, the proportion of black people, the year the store was opened, and household income in year 2000 expressed in U.S. dollars. All these variables are described in appendix D. Summary statistics are presented in panel B of table 1.

The key assumption of the RD approach is that the unobserved characteristics ϵ_i^k do not change discontinuously at the border. Although we cannot test this assumption directly, we do two things to assess its plausibility. First we examine the distribution of store's distances to the border to see whether the store's location is discontinuous at the border. A discontinuity would suggest that the store's location is endogenous to the treatment, potentially invalidating our design. Second, we examine whether the observable market characteristics X_i also change discontinuously at the border. If the observable characteristics do not change discontinuously at the border, then one may be more confident that this is the case also for the unobservable characteristics.[33] In the same spirit, we compare estimates of γ^k with and without controls for observable characteristics.

Finally, we will perform the same RD analysis for costs and markups. Since wholesale markets are also segmented at the border, the RD estimates on the cost regressions will also provide a lower bound on the border costs at the wholesale level. This will allow us

[32] Holmes (2008) considers similar variables when estimating the demand for products sold in Walmart Stores.

[33] Moreover, even if observable characteristics are not continuous at the border, this does not invalidate our design, as long as the effect of the covariates X_i on the dependent variable remains the same on each side of the border and we control for these characteristics.

to estimate the extent of market segmentation at the wholesale level and its relationship to market segmentation at the retail level.

6.1 Graphical analysis

We begin by plotting the distribution of each store's distance to the U.S.-Canadian border (in kilometers).[34] Figure 5 plots the density of all stores of the retail chain as a function of the algebraic distance to the border (that is, distance is negative for Canadian stores and positive for U.S. stores). Each bin width is 50 kilometers.

As can be seen, all Canadian stores are located less than 1,000 kilometers from the border, while many stores in the United States are more than 1,000 kilometers from the border. Obviously, the geographic concentration of economic activity in the United States is very different from that in Canada, highlighting Gorodnichenko and Tesar's (2009) caution about estimates that do not take within-country heterogeneity differences into account. Nonetheless, we do not observe any significant discontinuity in store density at the border. This suggests that, for this retailer, the location of stores does not appear to be directly influenced by proximity to the border. From Figure 2, it is also apparent that many Canadian stores close to the border have no counterpart on the U.S. side. This is especially true for Canadian stores in Eastern British Columbia and Alberta. To address this issue, we also present results with a sample of stores located in Oregon and Washington on the U.S. side, and British Columbia in Canada (21 Canadian and 41 U.S. stores) where there is an important concentration of stores close to the border and where market conditions are also likely to be more homogenous. We refer to this group of 62 stores as the "West Coast sample."

Figure 6 depicts graphically the regression discontinuity for the market characteristics included in X_i. Each point is the average value of the relevant variable within 50-kilometer bins. For several of these variables no stark graphical discontinuity is apparent. We formally test for this result and find that 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 the West

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

Coast subsample of stores, these discontinuities disappear, but we find some discontinuities for the fraction of senior citizens as well as for median household income. Further, when we consider all zipcodes and census subdivisions in Canada and the U.S., not just the ones for which we have store data, we cannot find any evindence of discontinuities for any of our covariates. We conclude from this graphical analysis that our RD design is valid, since we can find no clear evidence that market conditions or store locations are systematically affected by the border.

6.2 Regression discontinuity estimates

Figures 7(a)–7(f) plot the kernel density of point estimates obtained by estimating regression (5) by UPC for the first week of 2004 and the 21st week of 2007. For our main specification we use all stores within 500 kilometers from the border.[35] We do this estimation separately for the retail price, net cost, and markup, for each UPC and for each week, both with and without controls for the covariates. Figures 7(a)–7(f) illustrate that the border effect on prices varies substantially across products.[36] 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. Thus, it appears that the depreciation of the U.S. dollar over this period increased both the costs and prices in Canadian stores close to the border relative to U.S. 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 relative to the markups in U.S. stores. However, a visual inspection of the shift in the distribution of costs and markups suggests that the shift in marginal costs overwhelms the change in retail markups.

The covariates do have some explanatory power for price gaps across stores, both within

[35] We also restrict the sample to those UPCs that have a minimum of 10 store observations on both sides of the border.

[36] 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.

and across countries. We calculate the adjusted R^2 from a regression of store prices on our store-level covariates for each UPC in the first month of 2004 and compute the mean and median across UPCs.[37] The adjusted R^2 for prices is in the 10-15 percent range while the adjusted R^2 for costs varies between 8 and 12 percent. This is consistent with our retailer making pricing decisions at a more granular level than the wholesale market - stores that purchase from the same wholesaler at the same cost may nevertheless charge different prices due to different local demand conditions. As for the coefficient on covariates, these are often quite small, as the estimates at the UPC level are quite noisy and dispersed across UPCs. Overall, we find that prices are higher in areas that have higher median incomes, lower population densities, greater competition, fewer senior citizens, and older stores.[38]

Importantly, even though store-level covariates explain a significant share of price differences across stores, they have little effect on our estimated border coefficient. The distributions reported in figures 7(a)–7(f) look very similar when the regression is estimated without (left panel) and with (right panel) covariates. This comparison assuages concerns that an omitted variable might result in biased estimates of the border effect.

Table 6 reports summary statistics for the distribution of prices, costs, and markups for week 21 of year 2007 (without covariates) plotted in Figures 7(a)–7(f). The median price (net cost) treatment effect is 15 percent (17 percent) for the full sample. When restricted to the West Coast subsample (Panel B) the estimates are 22 percent (22 percent). We find no evidence of a border treatment for markups in either sample. Across all weeks, the median absolute price treatment effect varies between 19.6 and 24.2 percent. Recalling that the estimated border effect in week t for product k, $\hat{\gamma}_t^k$, is a lower bound on the true border effect γ^k when markets are segmented, it follows that the true border effect at the retail level is at least as high as 24 percent.[39] Similarly, the median absolute treatment effect for net

[37] These adjusted R^2 are obtained by pooling stores in both countries and partialling out the effect of the border on prices and covariates so that the coefficient estimates are similar to those obtained in our regression discontinuity estimates.

[38] These results can be requested from the authors. We also performed the same regressions using store-level price indexes, relative to a base store. This results in significantly bigger coefficient estimates, as idiosyncratic price differences are averaged out. The signs on the coefficients are similar however.

[39] Formally, if $\hat{\gamma}_t \leq \gamma$, then $\max_t \hat{\gamma}_t \leq \gamma$.

costs varies between 19.9 and 24.8 percent. [40]

As in section 5, we want to compare our evidence on cross-border price, costs and markup border treatments to within-country estimates. We do so by estimating the treatment of the Washington-Oregon border, on our West Coast subsample for the set of matched products.[41] This serves an important purpose: within-country border discontinuities provide a natural benchmark for cross-border discontinuities. In the language of the treatment effect literature, the Washington-Oregon border serves as a placebo. Panel C of table 6 reports the results. We find no evidence of a discontinuity in prices or costs. This is in spite of the fact that some Washington stores are in a different pricing area, as is evident in figure 8. Borders may sometimes form a natural pricing area when they coincide with large differences in competition and demographics, but many pricing areas seem to straddle multiple states or provinces in our sample. This underscores the importance of our RD design - if we simply compared average prices across Washington and Oregon, we might conclude that the state border segments markets and has a big impact. By focusing on the stores closer to the border and including distance as a regressor, we correctly ascertain that these stores share similar local environments and estimate a zero impact of the state border.

Figures 9(a) and 9(b) plot the distribution of treatment effects by UPC at the Washington-Oregon border for prices and net costs. We find that, in contrast to figures 7(a)–7(f), the point estimates are almost all concentrated at 0 for both weeks we consider.

7 Conclusion

This paper revisits a classic question about the role of international borders in segmenting markets. We improve upon the existing literature along three dimensions. First, we use barcode level price and cost data from a single retail chain operating in the United States and Canada. Second, we exploit the comovements between retail prices, wholesale costs and

[40] All these results hold similarly for the West coast sub-sample.

[41] Similar results are obtained when the sample is extended to include all UPCs traded within U.S. boundaries.

nominal exchange rates to determine the extent of market segmentation within and across countries. We find definite evidence for market segmentation at the retail level, even for identical goods. In response to exchange rate shocks, movements in retail prices across the border can be explained almost entirely by changes in wholesale costs, and not by systematic markup differences. By contrast, retail markets within countries appear partially integrated.

Third, we use a regression discontinuity approach to estimate a lower bound on the true border effect, that is, the additional transaction costs resulting from the border. We find that these border costs represents at least 24 percent at the retail level. We estimate the discontinuity at the whole-sale level to range between 19.9 and 24.8 percent. 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 composition bias is not supported by our results.

Quite strikingly, we also find that most differences in cross-border retail prices arise from differences in an apparently tradeable component of costs. A fruitful area for future research is to understand what sustains these large relative price movements in whole-sale costs, even for stores within the same retail chain. Of course, a limitation of our work is that we examine a specific set of goods sold by a single grocery store 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 and retailers would be very useful.

References

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

Atkeson, A. and Burstein, A. (2008). Trade costs, pricing-to-market, and international relative prices. *American Economic Review*, 98(5):1998–2031.

Broda, C., Leibtag, E., and Weinstein, D. (2009). The Role of Prices in Measuring the Poors Living Standards. *Journal of Economic Perspectives*, 23(2):77–97.

Broda, C. and Weinstein, D. (2007). Understanding international price differences using barcode data. Technical report, Chicago GSB.

Burstein, A. and Jaimovich, N. (2008). Understanding movements in aggregate and product-level real exchange rates. Working paper, UCLA.

Crucini, M. J. and Shintani, M. (2006). Persistence in law-of-one-price deviations: Evidence from micro-data. Working Paper 06-W17, Department of Economics, Vanderbilt University.

Crucini, M. J. and Shintani, M. (2007). Microeconomic sources of real exchange rate variability. Working paper, Vanderbilt University.

Crucini, M. J., Telmer, C. I., and Zachariadis, M. (2005). Understanding european real exchange rates. *American Economic Review*, 95(3):724–738.

Dornbusch, R. (1987). Exchange rate and prices. *American Economic Review*, 77(1):93–106.

Eichenbaum, M., Jaimovich, N., and Rebelo, S. (2008). Reference prices and nominal rigidities. Working paper, Stanford.

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

Ghosh, A. R. and Wolf, H. C. (1994). Pricing in international markets: Lessons from the economist. Working Paper 4806, National Bureau of Economic Research, Inc.

Goldberg, P. and Hellerstein, R. (2006). A framework for identifying the sources of local-currency price stability with an empirical application. Yale University and Federal Reserve Bank of New York.

Goldberg, P. and Verboven, F. (2001). The evolution of price dispersion in the european car market. *Review of Economic Studies*, pages 811–48.

Goldberg, P. and Verboven, F. (2005). Market integration and convergence to the law of one price: Evidence from the european car market. *Journal of International Economics*, pages 49–73.

Gopinath, G., Itskhoki, O., and Rigobon, R. (2007). Currency choice and exchange rate pass-through. Harvard University.

Gorodnichenko, Y. and Tesar, L. (2009). Border effect or country effect? seattle may not be so far from vancouver after all. *American Economic Journal: Macroeconomics*, 1(1):21941.

Haskel, J. and Wolf, H. (2001). The law of one price–a case study. *Scandinavian Journal of Economics*, 103(4):545–58.

Hassink, W. H. and Schettkat, R. (2001). On price-setting for identical products in markets without formal trade barriers. Discussion Paper 315, Institute for the Study of Labor (IZA).

Holmes, T. (1998). The effect of state policies on the location of manufacturing: Evidence from state borders. *Journal of Political Economy*, 106(4):667–705.

Holmes, T. (2008). The diffusion of wal-mart and economies of density. NBER Working Papers 13783.

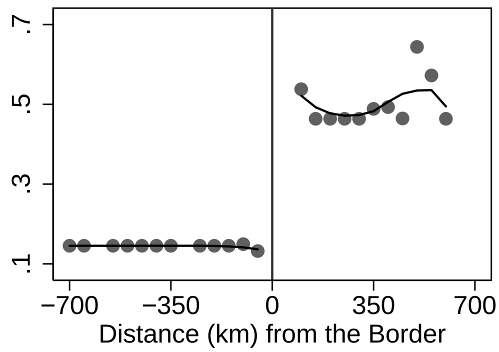
Imbens, G. and Lemieux, T. (2007). Regression discontinuity designs: A guide to practice. Working Paper 13039, National Bureau of Economic Research, Inc.

Krugman, P. (1987). Pricing to market when the exchange rate changes. In Arndt, S. and Richardson, J., editors, *Real Financial Linkages among Open Economies*, pages 49–70. MIT Press, Cambridge.

Parsley, D. and Wei, S. (2007). A prism into the ppp puzzles: The micro-foundations of big mac real exchange rates. *Economic Journal*, 117(523):1336–1356.

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

8 Figures and Tables



Note: Perrier sparkling natural mineral water, 25 ounce. Store distance to the border is positive for the United States, negative for Canada. First week of 2004.

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

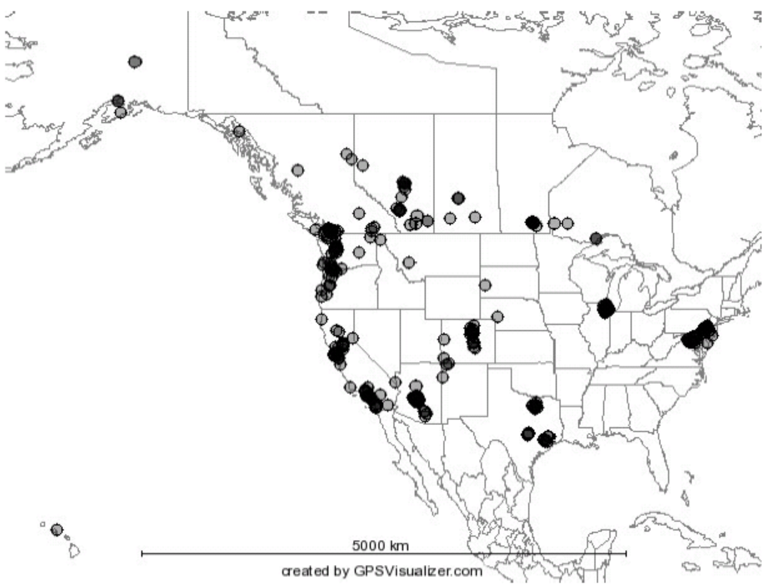


Figure 2: Map of the 325 North American retail stores (250 U.S. and 75 Canadian)

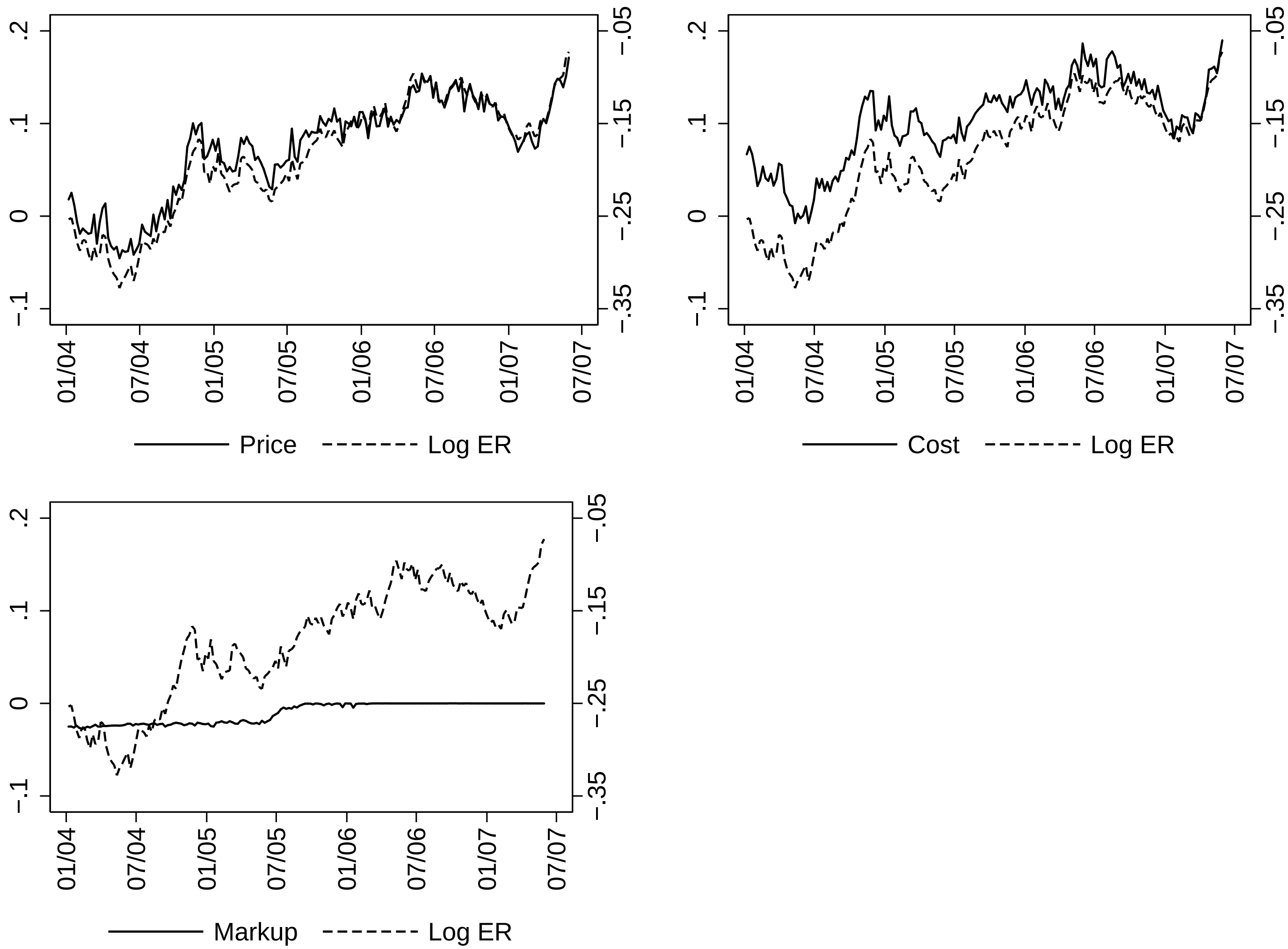
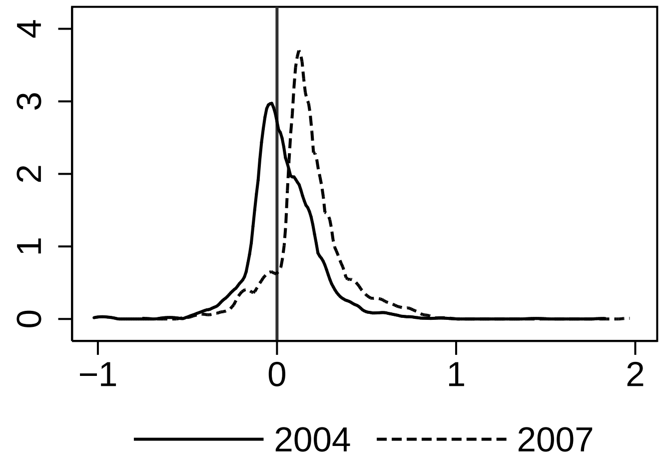
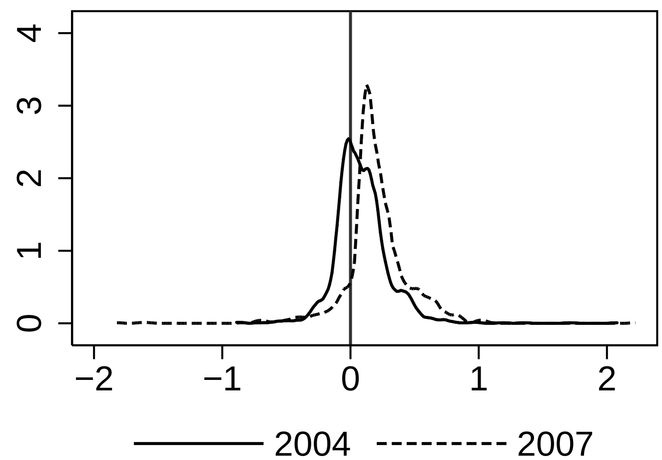


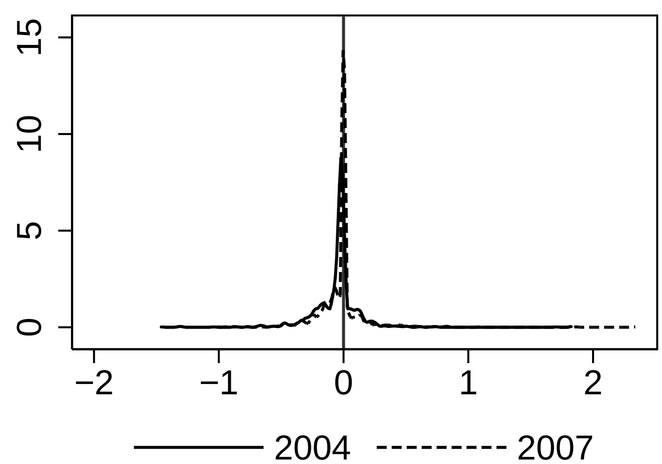
Figure 3: Median net price, imputed cost, and markup cross-border gap and exchange rate



(a) (log) price



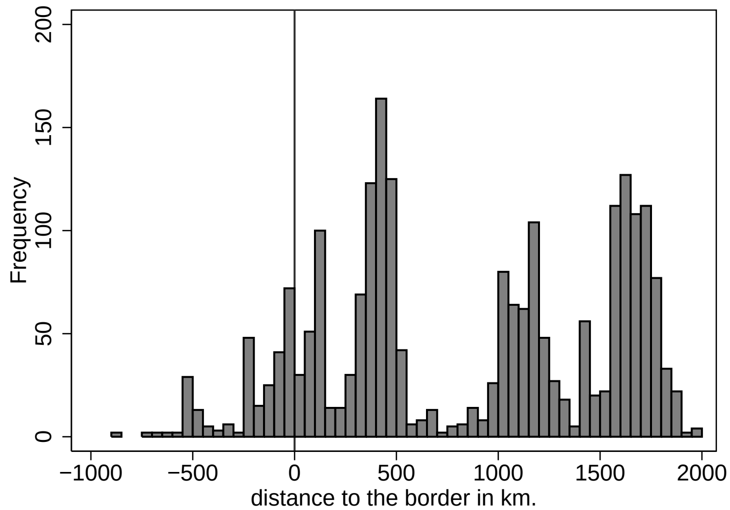
(b) (log) cost



(c) (log) markup

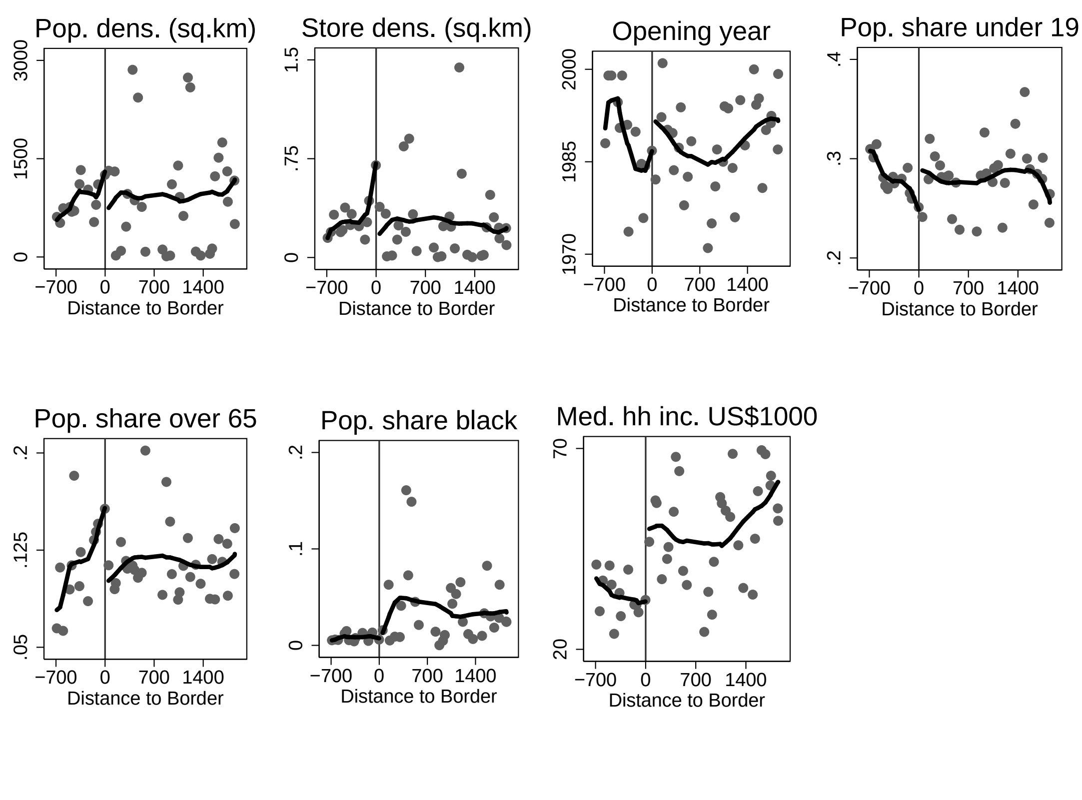
Note: 2004 refers to the first week of 2004; 2007 refers to the 21st week of 2007.

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



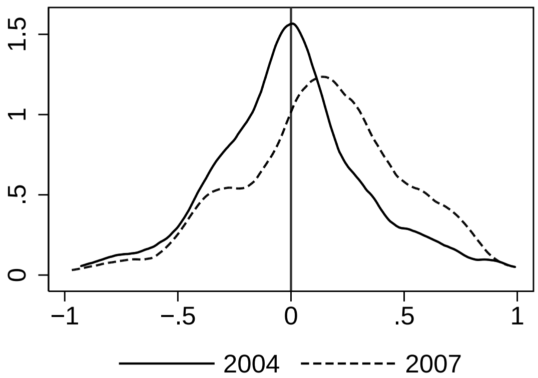
Note: Distance to the border is positive for U.S. stores, negative for Canadian stores.

Figure 5: Distance to the border

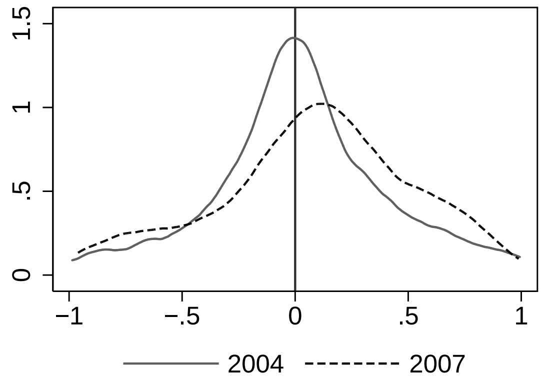


Note: Store distance to the border is positive for the United States, negative for Canada.

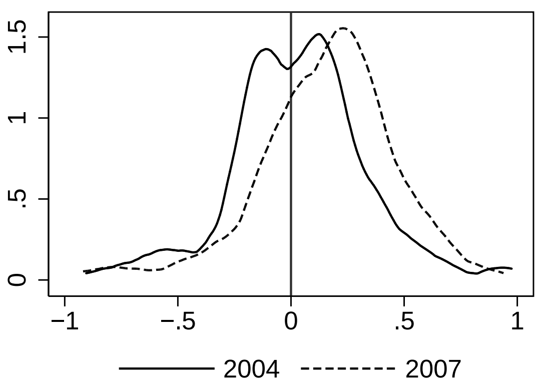
Figure 6: Regression discontinuity for covariates



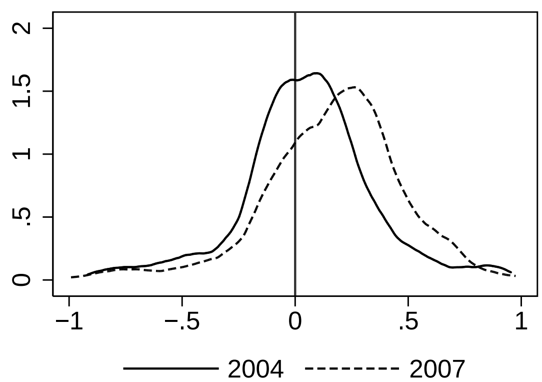
(a) Price, no covariates



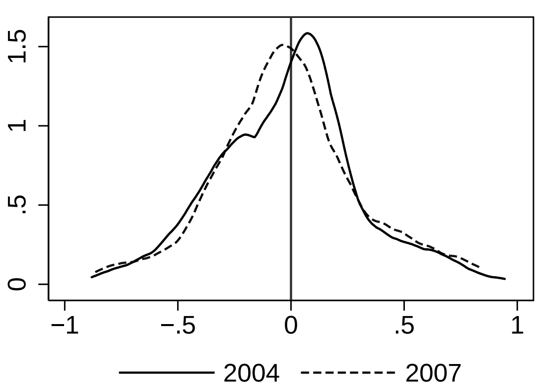
(b) Price, with covariates



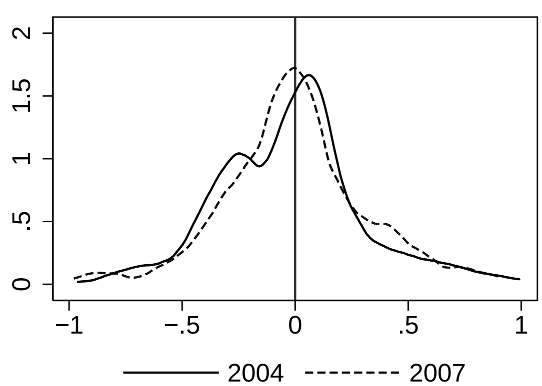
(c) Cost, no covariates



(d) Cost, with covariates



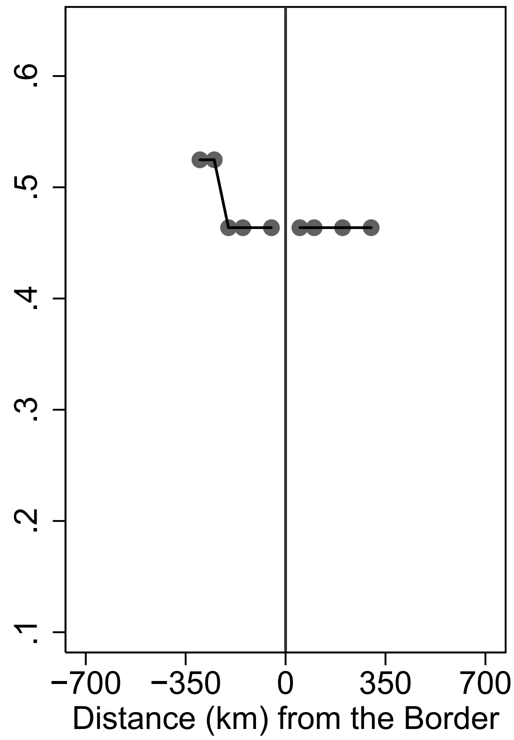
(e) Mark up, no covariates



(f) Mark-up, with covariates

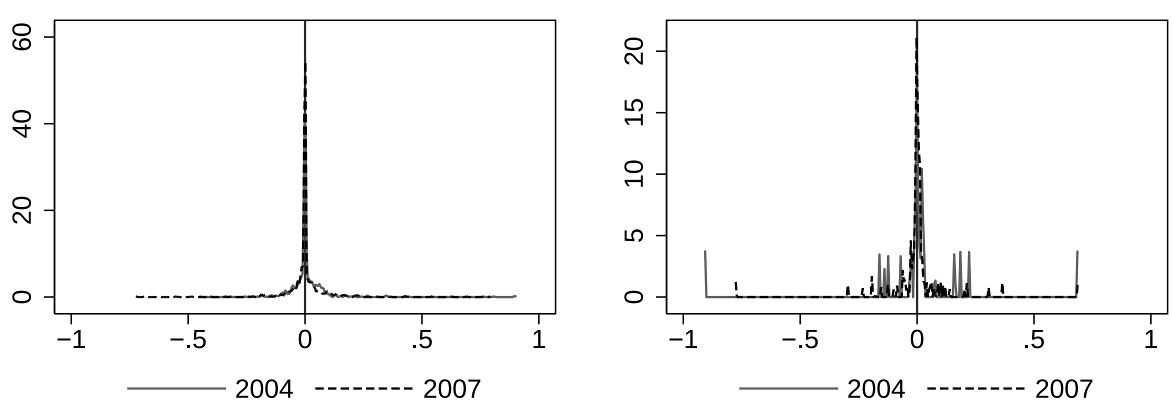
Note: 2004 refers to the first week of 2004; 2007 refers to the 21st week of 2007.

Figure 7: Distribution of regression discontinuity estimates of price, cost and markup gaps



Note: Perrier sparkling natural mineral water, 25 ounce. Store distance to the border is positive for Oregon, negative for Washington.

Figure 8: Graphical depiction of Washington-Oregon border regression discontinuity for Perrier Sparkling Mineral Water



(a) Distribution of price discontinuity

(b) Distribution of cost discontinuity

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

Figure 9: Intra-national borders regression discontinuity: the Washington-Oregon border

	Canada			United States		
	median	10%	90%	median	10%	90%
	Panel A: Number of UPCs sold per store/week					
UPCs per store	7998	5725	9187	10827	8387	12812
Matched UPCs per store	422	261	528	493	354	640
Matched UPCs per store-pair (within country)	251	144	331	272	183	365
Matched UPCs per store-pair (cross country)	167	98	225	167	98	225
	Panel B: Store-level covariates					
Population Density (persons per sq km)	1095.8	97.9	1533.7	810.2	57.2	2671.2
Store Density (stores per square km)	0.2864	0.0354	0.8804	0.1492	0.0105	0.8580
Median Household Income ($2000 US)	31014	28199	41109	57040	34238	82592
Share of population 0-19	0.26	0.22	0.31	0.29	0.19	0.34
Share of population 65 and up	0.13	0.07	0.17	0.10	0.05	0.16
Share of population black	0.01	0.00	0.02	0.03	0.01	0.16
Year store opened	1988	1965	2001	1992	1971	2001
Store selling area (sq feet)	26225	15495	39079	27168	15595	38540
Distance to the border (km)	98	15	503	558	283	1732

Note: The table reports the median, 10th and 90th percentiles across the 75 Canadian stores and 250 US stores in our sample. The variables are either provided to us directly by the retailer or derived by using store addresses to link them to zipcode-level US census data for 2000 and census subdivision level Canadian census data for 2001. Distance is calculated using arcGIS as the great circle distance to the Canada-US border from the geocoded latitude and longitude of the store address.

Table 1: Descriptive statistics

	Mean absolute	Med. absolute	Max absolute
	(1)	(2)	(3)
		Panel A: Net prices	
		USA-USA store-pairs (31125)	
Median	0.085	0.037	0.811
Average	0.087	0.042	0.858
St. Dev.	0.029	0.032	0.312
		CAN-CAN store-pairs (2775)	
Median	0.030	0.000	0.330
Average	0.030	0.005	0.368
St. Dev.	0.020	0.012	0.159
		CAN-USA store-pairs (18450)	
Median	0.219	0.146	1.021
Average	0.222	0.156	1.086
St. Dev.	0.033	0.041	0.303
		Panel B: Net costs	
		USA-USA store-pairs (31125)	
Median	0.057	0.008	0.860
Average	0.058	0.018	0.892
St. Dev.	0.023	0.021	0.402
		CAN-CAN store-pairs (2775)	
Median	0.038	0.000	1.031
Average	0.038	0.000	1.060
St. Dev.	0.011	0.001	0.397
		CAN-USA store-pairs (18450)	
Median	0.238	0.178	1.185
Average	0.242	0.182	1.278
St. Dev.	0.039	0.046	0.194

Note: Panel A refers to net prices and panel B refers to net costs. The table reports within and between-country statistics (the rows) for the mean absolute, median absolute and max absolute (log) price gap within store-pairs (the columns) for the first week of 2005.

Table 2: Deviations from the law of one price for retail and wholesale prices

		Horizon			
		Weekly	Monthly	Quarterly	Annual
Cost share	Median	0.55	0.69	0.77	0.70
	Mean	0.63	0.62	0.66	0.57
Markup share	Median	0.45	0.31	0.23	0.30
	Mean	0.37	0.38	0.34	0.43

Note: The table decomposes the variance of price gap changes at different horizons into a cost and markup component. The results are based on a rolling window over the 178 weeks in our sample.

Table 3: Variance decomposition of cross-border price gaps by product

		Horizon			
		week	month	quarter	year
Panel A: UPC level regression of matched goods					
Price	median	1.00	1.00	1.00	1.00
	25 percentile	0.86	0.84	0.91	0.88
	75 percentile	1.1	1.05	1.07	1.11
Net Cost	median	1.00	1.00	1.00	1.00
	25 percentile	0.97	0.98	0.97	0.93
	75 percentile	1.02	1.02	1.01	1.01
Markup	median	0.00	0.00	0.00	0.00
	25 percentile	-0.18	-0.19	-0.11	-0.15
	75 percentile	0.2	0.1	0.14	0.2
Panel B: Product group price index regression of all goods					
Price	median	0.92	0.9	0.9	1.17
	25 percentile	0.51	0.6	0.51	0.12
	75 percentile	1.38	1.28	1.4	2.03
Net Cost	median	0.83	0.88	0.85	1.18
	25 percentile	0.45	0.52	0.41	-0.17
	75 percentile	1.1	1.29	1.62	2.49
Panel C: Product class price index regression of all goods					
Price	median	1.01	1.03	1.01	1.15
	25 percentile	-0.12	-0.51	-1.29	-1.84
	75 percentile	2.13	2.65	3.32	3.89
Net Cost	median	0.91	0.93	0.99	1
	25 percentile	-0.6	-1.33	-2.17	-2.83
	75 percentile	2.23	3.9	4.11	4.65

Note: The table presents summary statistics (across the matched UPCs in our sample) for the pass-through coefficient. In the top panel, we separately estimate a regression of the change in the log Canada-US price, cost and markup gap on the change in the log Canada-US exchange rate for each UPC. The columns represent the horizon over which we estimate the rolling window regressions. The median, 25th and 75th percentile refer to the distribution of UPCs. The bottom panel uses the change in relative price indexes calculated at the level of product group (61) or product class (1165).

Table 4: Exchange rate pass-through

			Horizon			
		percentile	week	month	quarter	year
Far US	own cost	median	0.35	0.37	0.47	0.45
		25th	0.05	0.05	0.1	0.16
		75th	0.74	0.79	0.89	0.85
	near US cost	median	0.12	0.14	0.05	0.05
		25th	0	-0.01	-0.04	-0.05
		75th	0.38	0.52	0.25	0.19
Near CAN	own cost	median	0.46	0.5	0.52	0.44
		25th	0.22	0.24	0.26	0.22
		75th	0.85	0.92	0.97	0.94
	near US cost	median	0	0	0	0
		25th	-0.05	-0.06	-0.06	-0.08
		75th	0.06	0.08	0.09	0.08
Near CAN	own cost	median	0.26	0.24	0.24	0.24
		25th	0.08	0.05	0.06	0.05
		75th	0.53	0.56	0.54	0.55
	far CAN cost	median	0.15	0.16	0.21	0.17
		25th	0.02	0.01	0.02	0
		75th	0.41	0.42	0.49	0.42

Note: The table presents summary statistics (across the matched UPCs in our sample) for the pass-through coefficient. For each UPC we separately estimate the pass-through from the change in log own costs and log of neighbour's costs into the change in log of own prices. The prices and costs are calculated as weekly averages for the stores in the regions we consider: Near US (WA stores within 200km of the border), Far US (WA and OR stores 200-400km from the border), Near CAN (BC within 200km of the border) and Far CAN (Alberta within 200km of the border). We do this for rolling-windows over horizons from one week to one year. The median, 25th and 75th percentile refer to the distribution of UPCs.

Table 5: Cost pass-through

	Median	Mean	SD	Frac. sign.	Median abs.	Mean abs.	No. of UPCs
				Panel A: All stores			
Price	0.15	0.13	0.37	0.70	0.21	0.28	481
Cost	0.17	0.15	0.31	0.80	0.21	0.26	481
Markup	0	-0.02	0.37	0.40	0.14	0.23	481
				Panel B: West Coast stores			
Price	0.22	0.26	0.32	0.86	0.24	0.33	212
Cost	0.22	0.20	0.27	0.83	0.24	0.27	212
Markup	0	0.06	0.36	0.44	0.13	0.23	212
				Panel C: Washington-Oregon stores			
Price	0	0.01	0.09	0.24	0.01	0.04	370
Cost	0	0	0.06	0.17	0	0.02	370
Markup	0	0	0.10	0.22	0.02	0.04	370

Note: distribution for week 21 of 2007.

Table 6: Regression discontinuity estimates

Appendix

A Regression Discontinuity Methodology

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:

$$\ln p_i^k = \alpha^k + \gamma^k C_i + \beta^k X_i + \tilde{\epsilon}_i^k, \tag{6}$$

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 $\tilde{\epsilon}_i^k$ captures unobserved characteristics that are store-and good-specific. The parameter of interest is γ^k. The inference problem is that the unobserved characteristics may not be independent from the location of store i, that is $E\left[\tilde{\epsilon}_i^k|C_i\right] \neq 0$, which can bias simple border regression estimates.

However, if the unobserved characteristics are a continuous function of the distance between stores and the border, we can control for these characteristics by introducing distance from the border as an additional regressor. Define D_i as the distance (in kilometers) from store i to the border. By convention, stores located in the United States 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:

$$\lim_{\varepsilon \uparrow 0} E\left[\tilde{\epsilon}_i^k|D_i = \varepsilon\right] = \lim_{\varepsilon \downarrow 0} E\left[\tilde{\epsilon}_i^k|D_i = \varepsilon\right].$$

The effect of the border can then be estimated as:

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

In this expression 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:

$$\ln p_i^k = \alpha^k + \gamma^k C_i + \theta^k D_i + \delta^k C_i \cdot D_i + \beta^k X_i + \epsilon_i^k. \tag{7}$$

Importantly, this local linear regression restricts the sample to stores within a distance of $\bar{D}_k$ from the border, that is $|D_i| < \bar{D}_k$. The optimal distance $\bar{D}_k$ can be selected using standard bandwidth selection criterion based on the cross-validation procedure advocated by Imbens and Lemieux (2007).[42] The optimal bandwidth ranges from 100 to 700 kilometers. For most weekly product-group pairs, the optimal bandwidth is either 100, 350, or 500 kilometers. All store-level observations beyond this cut-off are effectively discarded. In

[42] The procedure looks for the minimum value of the cross-validation criterion in 100 kilometers increments.

practice, we choose a bandwidth of 500km.

B Circular world

We present a model that endogenizes the distribution of prices across locations in the presence of border costs. The model is a two-country version of Salop's (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 circumference. A border splits the circle into two countries (country A and country B). The details regarding the derivation of results are available from the authors upon request.

B.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 $i \in \{1, \ldots, N_{AB}\}$, 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.[43]

B.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 prefers to shop in stores that are located close to them. They incur a cost $t \geq 0$ per unit of distance traveled that reflects transportation costs or the individual 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 located in different countries, θ captures the own price elasticity of demand, and t is inversely related to the degree of substitutability across store locations. We assume that ν is large enough so that in equilibrium all consumers purchase one unit of the good.

[43]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 decisions of competitors located in the immediate vicinity than by the 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 like an independent store.

B.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 incurred by the retailer. Note that it will always be the case that c_i is the same for all stores in the same region.

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. Between any pair of stores i and $i-1$, there will be a marginal consumer indifferent between shopping at either store.

B.3.1 Interior stores

Consider an interior store i in country j. That store chooses its price p_i to maximize static profits. 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 :

$$p_i = \left(\hat{p}_A - c_A - \frac{t}{N_{AB}}\right) \cdot \frac{\cosh\left(\kappa(i - \frac{N_A+1}{2})\right)}{\cosh\left(\kappa(\frac{N_A-1}{2})\right)} + c_A + \frac{t}{N_{AB}}, \tag{8}$$

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

$$p_i = \left(\hat{p}_B - c_B - \frac{t}{N_{AB}}\right) \cdot \frac{\cosh\left(\kappa(i - N_A - \frac{N_B+1}{2})\right)}{\cosh(\kappa(\frac{N_B-1}{2}))} + c_B + \frac{t}{N_{AB}}. \tag{9}$$

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.[44]

As equations (8) and (9) 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.

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

B.3.2 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.

The following set of propositions characterizes border prices in these two cases.

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

$$\hat{p}_A = c_A + \frac{t}{N_{AB}} \frac{3-\nu_A}{2-\nu_A}, \qquad \hat{p}_B = c_B + \frac{t}{N_{AB}} \frac{3-\nu_B}{2-\nu_B}, \tag{10}$$

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 border store prices moves one-to-one with the difference in costs, that is, $\partial(\hat{p}_A - \hat{p}_B)/\partial(c_A - c_B) = 1$.

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 costs on the other side of the border, and markets are completely 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

$$\hat{p}_A - \hat{p}_B > b, \tag{11}$$

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

$$\hat{p}_A = \frac{(4-\nu_B)(j_A+b) + (j_B-b)}{(4-\nu_A)(4-\nu_B)-1}, \qquad \hat{p}_B = \frac{(4-\nu_A)(j_B-b) + (j_A+b)}{(4-\nu_A)(4-\nu_B)-1}, \tag{12}$$

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

$$\hat{p}_A = \frac{(4-\nu_B)(j_A - b) + (j_B + b)}{(4-\nu_A)(4-\nu_B) - 1}, \qquad \hat{p}_B = \frac{(4-\nu_A)(j_B + b) + (j_A - b)}{(4-\nu_A)(4-\nu_B) - 1}. \tag{13}$$

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

B.3.3 Discussion

The model presented in the previous section delivers the following insights. First, 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. If countries are completely symmetric this lower bound will be zero, even in the presence of large border costs. Second, stores closest to the border are most sensitive to the border cost. In most of the existing literature, owing to a lack of data, no distinction is made between stores that are close to the border and stores that are far from it. Third, the behavior of relative prices and relative markups is very different in situations of full and partial market segmentation. 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. Lastly, equilibrium prices depend on many factors such as the degree of substitutability across locations (local arbitrage costs), the number of competitors, and the own price elasticity of demand all of which can vary with location, besides the size of the border cost.

C 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, that is

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

In summing over j we use all stores in the United States 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, so we have no recorded price change. In this case we drop the observation for the store that is missing a price change and re-weight the shares across the UPCs for which price information is available in that store. We construct these price indexes 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 (wholesale cost) measure we use the net (gross) expenditure shares.

D Data Description

Variable	U.S. Source	Canadian Source
Population Density	2000 Census	2001 Census
Store Density	2002 US Economic Census	2002 Canadian Business Pattern Survey
Median Household Income	2000 Census	2001 Census
Share of population 0-19	2000 Census	2001 Census
Share of population 65 and up	2000 Census	2001 Census
Share of population black	2000 Census	2001 Census
Year opened	Supplied by retailer	Supplied by retailer

Note: U.S. 2000 census median household income refers to year 1999 in 1999 US dollars. Canadian 2001 census median household income refers to year 2000 in 2000 Canadian dollars, converted into US dollars at the average exchange rate 1.4852 Canadian/US dollar.

Table 7: Data sources for covariates

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 8: Number of distinct products by product group for both countries, Canada and the United States, and the set of uniquely matched products