Preliminary and Incomplete

Price Dispersion: The Role of Borders, Distance and Location*

Mario J. Crucini,† Chris I. Telmer,‡ and Marios Zachariadis§

First version: December 2003
This version: January 2005

Abstract

We use a model of retail price determination to understand the behavior of deviations from the law-of-one-price in a dataset on the prices of 300 different goods and services across 122 major cities in 79 different countries over the years 1990-2000. In our model, deviations from the law-of-one-price arise because retail goods are produced with both a non-traded input, and a traded input which is subject to a shipping cost. We find that several properties of the cross-sectional and time-series distribution of the data are consistent with the model.

*We thank seminar participants at the Vanderbilt University Brown-bag lunch for helpful comments.

†Department of Economics, Vanderbilt University; mario.j.crucini@vanderbilt.edu

‡Carnegie Mellon University; chris.telmer@cmu.edu

§Department of Economics, Louisiana State University; zachariadis@lsu.edu

1 Introduction

The goal of this paper is to understand deviations from the law-of-one-price (LOP). Specifically, if $P_{ij,t}$ is the local-currency cost of good i in location j at date t, we seek to understand

$$q_{ijk,t} = \log \frac{P_{ik,t} e_{jk,t}}{P_{ij,t}} \quad ,$$

where $e_{jk,t}$ is the nominal exchange rate (equal to one if locations j and k share the same currency). We are motivated by two observations. First, while there exists a consensus on a number of empirical regularities about q, there is far less of a consensus on how to interpret them from the perspective of a model. Second, much of what we know about the behavior of q is based on time-series data, not cross-sectional data. This is because most data on P takes the form of index numbers, not absolute prices. Where exceptions exist, the data are confined to a very limited cross-section of goods, i.

With this in mind, we ask if a particular model of deviations from the LOP is consistent with properties of a broad panel dataset on local-currency prices of many different goods and services in many different locations, both within and across countries. What we hope to learn is whether cross-sectional data, in conjunction with an explicit model, changes the economic interpretation of the existing body of empirical facts about q.

Our model presumes that goods in different locations are different because of the tradeability of the inputs required to produce them. A location-specific retail firm combines a non-traded input (*e.g.*, labor) and a traded input to produce a retail good. Retail-good LOP deviations are a simple reflection of input-price LOP deviations. The non-traded good's relative price is assumed to reflect productivity differences in the sense of Balassa (1964) and Samuelson (1964). The traded good's relative price is assumed to reflect a transport cost in the sense of Sercu, Uppal and Van Hulle (1995). Good specificity takes the form of different input shares between the traded and non-traded good. Location specificity takes the form of different transport costs. We model (and measure) transport costs as being log-linear functions of the distance between locations.

Our dataset is from the Economist Intelligence Unit. It consists of annual data, 1990-2000, on local-currency retail prices of roughly 300 different goods and services across 122 major cities in 79 countries. This dataset, while relatively new to economics researchers, has been used previously by Crucini and Shintani (2002), Parsley and Wei (2002), Rogers (2001) and Engel and Rogers (2003).

We find that both the cross-sectional and time-series properties of this dataset are consistent with our model of retail price determination. First, the cross-sectional mean and the cross-sectional variance (across goods, for each bilateral location pair) depend on productivity differences in the manner predicted by the model. That is, there is a 'Balassa-Samuelson effect' in both the first and second cross-sectional moment. Second, the impact of geographical distance (shipping costs in our model) is consistent with the model; distance does not matter for the cross-sectional mean, once we control for productivity differences. We label this the 'averaging-out property' of the shipping cost model; if some goods are imported and others exported, then deviations from LOP will tend to average-out across good. Distance does matter, however, for the cross-sectional variance, as predicted by the model. It also matters for the time-series variance, but only for location-pairs which cross a border. We argue that this is consistent with how nominal exchange rate variability serves to 'trace out' the region of relative prices where the shipping cost model bites.

2 Data

The source of our price data is the annual Economist Intelligence Unit retail outlet price survey. Our data begins in 1990 and ends in 2000, effectively creating (ignoring missing data) a balanced 11-year panel of absolute prices for 220 goods and 84 services across 122 major cities across the globe. The total number of countries is 79 countries, with differences between the number of cities and number of countries reflecting the fact that in 58 of the 79 countries the EIU surveys multiple cities. The country with the most intranational observations is U.S., with 16; the next largest number of intranational observations is 5 (Australia, China and Germany)[1] The same basic data source has recently been used by Crucini and Shintani (2002), Parsley and Wei (2002), Rogers (2001) and Engel and Rogers (2003).

Our basic data unit is $P_{ij,t}$, the price, in units of local currency, of good i in location j at time t. For most of our analysis we transform this data into $q_{ijk,t}$, log deviations from the law-of-one-price (LOP) for each bilateral location-pair:

$$q_{ijk,t} = \log(\frac{P_{ik,t}e_{jk,t}}{P_{ij,t}}) \ ,$$

where $e_{jk,t}$ is the nominal exchange rate between locations j and k, in units of location j, and $e_{jk,t} = 1$ if locations j and k are in the same country.

Figure 1 shows estimates of the density function for $q_{ijk,t}$ for 1990, 1995, 2000; both international city pairs and U.S. city pairs (the graph is quite similar for intranational pairs more broadly). We see what is obvious to anyone who has ever traveled between two locations: the LOP is not very useful for describing the

[1]Specifically, the number of intranational cities are (ordered from the most available cities to the least): United States (16), Germany, China and Australia (5), Canada (4), Saudi Arabia (3) and France, Italy, Russia, Spain, Switzerland, UK, India, Japan, Vietnam, New Zealand (2).

properties of $q_{ijk,t}$. Moreover, as one might expect, price dispersion is considerably higher across international locations. Much of the existing literature has therefore proceeded by describing a number of interesting empirical regularities inherent in the international relative price distribution. The difficulty with this lies in moving from empirical regularities to economic interpretation. Our approach, therefore, will be to (i) briefly list a set of interesting empirical regularities, and then (ii) flesh out their economic interpretation using a simple production function which relates deviations from LOP in retail prices to deviations from LOP in non-traded inputs (such as labor) and transport costs. At each step we compare and contrast the properties that exist within countries and across countries.

Our data on $q_{ijk,t}$ display the following features. In each case, specific details are deferred until the relevant section of the paper. For properties 1 to 5 we average over time work with $q_{ijk} \equiv T^{-1}\sum_{t=1}^{T} q_{ijk,t}$. We use the notation $E_i(q_{ijk} \,|\, jk)$ and $Var_i(q_{ijk} \,|\, jk)$ to denote the cross-sectional mean and variance (across goods for each location-pair jk). Similarly $Var_t(q_{ijk,t} \,|\, ijk)$ denotes the (unconditional) time-series variance for a given good i and location-pair jk.

1. *The Balassa-Samuelson Effect for the Mean.* The cross-sectional mean $E_i(q_{ijk} \,|\, jk)$ — or in words the average relative price of goods in location k in units of goods in location j — is positively related to income in location k less income in location j.

2. *The Balassa-Samuelson Effect for the Variance.* The cross-sectional variance $Var_i(q_{ijk} \,|\, jk)$ increases in the *absolute* income difference between locations j and k.

3. *The Averaging-Out Property.* Once we control for income differences, $E_i(q_{ijk} \,|\, jk)$ is close to zero. That is, for most bilateral location-pairs with similar wealth levels, there tends to be as many overpriced goods as underpriced goods.

4. *Distance Does Not Matter for the Mean.* Intranationally, the average price, $E_i(q_{ijk} \,|\, jk)$, does not depend on the distance between locations j and k. Internationally it does, but not once we control for income/productivity differences.

5. *Distance Matters for Cross-Sectional Dispersion.* $Var_i(q_{ijk} \,|\, jk)$ is increasing in the log of geographical distance between locations j and k. This is true for both intra and international location-pairs. Moreover, the magnitude of the relationship is not affected by the existence of a border.

6. *Distance Does Not Matter For Intranational Time-Series Dispersion.* For intranational city-pairs, $Var_t(q_{ijk,t} \,|\, ijk)$ does not depend on relative distance. This is also true of $Var_t(q_{ijk,t} - q_{ijk,t-1} \,|\, ijk)$, the variance of the *change* in the relative price. The latter was the focus of Engel and Roger's (1996)

study. Their conclusion, based on very different data, was opposite ours. We are still trying to understand the reasons for the differences.

7. *Distance Matters For International Time-Series Dispersion.* If a border separates locations j and k, the variance $Var_t(q_{ijk,t} \,|\, ijk)$ depends positively on the distance between locations j and k. The same applies for the (above) variance in the relative-price change. A 'border effect,' therefore is not separately identified from a distance effect (in our data). There seems to be an interaction effect.

We now demonstrate that each of these empirical regularities is consistent with a retail-good production technology where deviations from LOP are sustained through non-traded intermediate inputs and traded intermediate inputs which are subject to transport costs.

3 Model

This section builds a simple partial equilibrium model retail price determination at the level of individual goods sold in local markets. We assume that trade occurs in intermediate inputs (goods) and retail firms combine local inputs with traded inputs for sale in the local market (allowing for trade in the final products involves a trivial logical extension).

3.1 The Retailer's Problem

In the notation that follows, we drop the time index to conserve notation and reintroduce it when we discuss time series properties. The retail production function is assumed to take the form:

$$Y_{ij} \equiv (N_{ij})^{\alpha_i} (T_{ij})^{1-\alpha_i} .$$

where N_{ij} is a non-traded (i.e. local) input while T_{ij} is an input which is either exported from or imported into location j.

Two examples may help to fix ideas. Suppose Y_{ij} is a men's haircut in Nashville. The traded input, T_{ij}, might be shampoo. The local inputs are the labor of the barber and the rental cost of the barber shop. If Y_{ij} was a PC sold in Gateway country, the traded input would be the PC itself and the local inputs would be sales personnel and the rental cost of the building housing the sales operation. The value of α is expected to be much closer to one for the haircut than the computer.

The cost function is the solution to the following minimization problem at each date:

$$\min_{\{N_{ij}, T_{ij}\}} C_{ij} = P_j^N N_{ij} + P_{ij}^T T_{ij} \tag{1}$$

$$s.t.(N_{ij})^{\alpha_i}(T_{ij})^{1-\alpha_i} \geq Y_{ij} \tag{2}$$

where C_{ij} is the cost of producing good i in location j; P_j^N is the cost of a non-traded input, common to all goods but differing across locations; P_{ij}^T is the price of the traded input into production of retail good i in location j.

We have adopted two standard assumptions. The first is that factor mobility is much higher across sectors within a location than across locations – P_j^N is location-specific, not good-specific. The second assumption is that retailers in all locations produce good i using the same production technology – α_i is good-specific, not location-specific.

3.2 Retail Price Determination

Under constant returns to scale, the cost function takes the form: $Y_{ij} \cdot C_i(\mathbf{P}_{ij}, 1)$ where Y_{ij} is the output level, $\mathbf{P}_{ij} = \left(P_j^N, P_{ij}^T\right)$ and $C_i(\mathbf{P}_{ij}, 1)$ is the unit cost function. Under the assumption of perfect competition, the unit cost is also the retail price:

$$P_{ij} = (P_j^N)^{\alpha_i}(P_{ij}^T)^{(1-\alpha_i)} \quad .$$

To compare this good and location specific price to the price of an identical good sold in another location we work with the bilateral real exchange rate at the level of an individual good (ie. $q_{ijk} = \ln(E_{jk}P_{ij}/P_{ik})$):

$$q_{ijk} = \alpha_i q_{jk}^N + (1-\alpha_i) q_{ijk}^T. \tag{3}$$

As the equation makes clear, this price deviation is a linear combination of analogous deviations in the non-traded and traded input prices. The weights in the linear combination are the shares of non-traded and traded inputs in production, which add up to unity under our assumption of constant returns to scale.

The retailer solves the same problem in each period so it is valid to add time subscripts to Equation (3) to reflect changes are input costs. When we use q_{ijk} it should be understood that we have averaged the good-by-good real exchange rate across time (i.e. $q_{ijk} = T^{-1}\sum_t q_{ijk,t}$) at the outset, thereby eliminating the role of time series variation which is discussed separately.

3.3 The Balassa Samuelson Effect and Trade Costs

Equation (3) amounts to little more than an accounting device, relating input prices to output prices. To push the model further in a structural direction we make two assumptions about the properties of the relative prices on the right-hand-side of equation (3).

The first assumption is that non-traded relative prices reflect productivity differences across locations. This assumption is based on the logic of the Balassa Samuelson (1964) effect. Because we lack reliable data on productivity we use $z_{jk} \equiv \log(y_j/y_k)$ as a proxy for non-traded productivity difference or q^N_{jk}.

The second assumption is that traded intermediate inputs satisfy the Law-of-One-Price up to a trade cost. Moreover, we assume that trade costs are related to distance as follows: $\log(1+\tau^i_{jk}) = \log(D^{\beta_i}_{jk})$ when the source is location k and the destination is location j. Thus the log the deviation for the traded relative price is q^T_{ijk} =$\log\{(1+\tau_{ijk})^{I^i_{jk}}\} = I^i_{jk} log(1+\tau_{ijk}) = I^i_{jk}\beta^i \log D_{jk}$ where I^i_{jk} is an indicator variable for the direction of the trade flow (equal to 1 if goods travel from k to j, and -1 if good go from j to k.

Combining these two assumptions we arrive at the observable implications of a slightly more structured retail model:

$$q_{ijk} = \alpha_i z_{jk} + (1-\alpha_i)\beta_i I^i_{jk} \log D_{jk}. \tag{4}$$

where $0 < \alpha_i < 1$ and $\beta_i > 0$.

We see that for give given bilateral pair, the Balassa-Samuelson effect has a common sign, but the magnitude of the impact varies with the share parameter. Greater distance between locations also increases price dispersion, but the magnitude and the sign depend on specifics related to the good.

4 Properties of the Mean

We begin with an analysis of the cross-sectional mean $E_i(q_{ijk}|jk)$, where it is understood that we have averaged the good-by-good real exchange rate across time (i.e. $q_{ijk} = T^{-1}\sum_t q_{ijk,t}$) at the outset.

Taking a simple average of both sides of the retail pricing equation, we arrive at:

$$q_{jk} = a z_{jk} + b_{jk} \log D_{jk}. \tag{5}$$

where $a = \sum_{i=1}^N \alpha_i$ and $b_{jk} = N^{-1}\sum_{i=1}^N (1-\alpha_i)\beta_i I^i_{jk}$.

4.1 Balassa Samuelson Effects

Ignoring the role of trade costs for the moment, the average deviation reduces to: $q_{jk} = az_{jk}$. Since the parameter a is constant across bilateral pairs by virtue of the common production function, the variance in the mean across jk is simply a scaled version of the variance in income/productivity across locations. Consider a bilateral pair countries with identical (vastly different) levels of income/productivity, we would expect the average Law-of-One-Price deviation to be zero (very large). Another way to visualize this effect is to plot $q_{jk} = az_{jk}$ against income. We expect a strong positive correlation and as Figure 2 amply demonstrates we find one.

4.2 Distance and Trade Costs

Consider, next the implications of the trade cost model and distance. The multiplier on distance is $b_{jk} = N^{-1}\sum_{i=1}^{N}(1-\alpha_i)\beta_i I^i_{jk}$. The dependence on jk is due to the fact that despite the assumption that the production and trade cost parameters are independent of location, trade flows obviously are not.

However, the jk subscripts represent little more than a nuisance because we argue that the coefficient itself is expected to equal zero on a bilateral basis. We refer to this notion as the 'averaging out' property of traded good relative prices. The easiest way to see this is to suppose that the production and trade cost parameters are common across goods and an equal number of goods are imported and exported. Then the expression is literally equal to zero: $(1-\alpha)\beta N^{-1}\left\{\sum_{i=1}^{N/2}(-1)+\sum_{i=(N/2)+1}^{N}1\right\} = 0$. While some interesting asymmetries may give rise to positive or negative coefficient, we expect this property of averaging out of the deviations to prevail for quite general production and trade cost configurations.

4.3 Findings

Figure 3 plots the q_{jk} against bilateral distance. It appears that distance matters internationally, but not internationally. This visual impression is confirmed by a simple pair of regression estimates:

$$|q_{jk}| = (0.0276)0.244 + (0.0031)0.0044\log D_{jk} + \varepsilon_{jk} international \quad (6)$$
$$|q_{jk}| = (0.0433)0.130 - (0.0059)0.0044\log D_{jk} + \varepsilon_{jk} intranational. \quad (7)$$

We use the absolute value of q_{jk} so that the coefficient on distance has a consistent sign across bilateral pairings.

Beginning with the international data we find evidence that distance matters as it should based on the trade cost model. However, the magnitude of the coefficient is small: going from cities that are neighbors to 100 miles apart adds 1% to the price differential, going and additional 2400 miles is needed to add another 0.5%! Controlling for distance, though, the intercept is both highly economically and statistically significant. Thus even after conditioning on distance (a proxy for trade costs), we resoundingly reject the Law-of-One-Price.

The intranational data tells a different story. The distance coefficient is of the wrong sign and statistically insignificant. Using the same interpretation as we did for the international context we reject the Law-of-Price, but unlike the international data distance plays no role at all.

Combining the two results there appears to be a border involving an increase in the unconditional variance in price levels across locations and an increase (from zero) in the impact of geographic distance on price differences.

According to our retail model, though, income/productivity disparities play an independent role. Re-estimating with the absolute value of the log of relative income across bilateral pairs, we have:

$$|q_{jk}| = (.0272)0.2537 - (.0031)0.0022 \log D_{jk} + (.0031)0.0447|z_{jk}| + \varepsilon_{jk} international \quad (8)$$
$$|q_{jk}| = (0.0433)0.130 - (0.0059)0.0044 \log D_{jk} + \varepsilon_{jk} intranational \quad (9)$$

where we have simply repeated our results for the intranational case because we lack data on income levels across regions within countries (assuming that income differences are small across regions within countries, this should not be too problematic, but we will rectify this for the U.S. where we know such data is available).

The absolute deviations of the log of relative income has a positive coefficient as expected. The distance coefficient is no longer not statistically significant as we would expect if the income ratio was effective in picking up the impact of non-traded goods, based on the 'averaging out' property for traded good prices. It is also interesting to note that the reduction in the magnitude of the coefficient on distance coefficient (it actually becomes negative) is consistent with a positive correlation between distance and income differentials.[2]

5 Properties of Cross-Sectional Variance

Next we examine the cross-sectional variance: $Var_i(q_{ijk}|jk)$ which, in the model, is given by:

$$Var_i(q_{ijk}|jk) = a'(z_{jk})^2 + b'(\log D_{jk})^2. \quad (10)$$

$a' = Var(\alpha_i)$ and $b' = var_i \left\{(1-\alpha_i) I^i_{jk} \beta_i\right\}$.

[2]This classic results is derived as follows. Let the true regression be: $y_i = bx_i + cz_i + \varepsilon_i$. If we estimate: $y_i = bx_i + cz_i + \varepsilon_i$ then the difference between the OLS estimate $\widehat{b}$ and the true one is: $\widehat{b} - b = (x'x)^{-1}x'zc$. In our context $c > 0$ and if $x'z > 0$ the OLS estimate is upwardly biased.

5.1 Balassa Samuelson Effects

Taking the Balassa Samuelson effect in isolation of the trade cost give us:

$$Var_i(q_{ijk}|jk) = a'(z_{jk})^2. \tag{11}$$

Thus the variance of Law-of-One-Prices around the mean is increasing in the income/productivity gap. Thus if we select a location pair with very similar incomes there will be very little variation in the size of the deviations across goods. If we take location pairs with vastly different incomes we will find that the Law-of-One_Price deviations are more heterogenous across goods. In other words, the q_{ijk} are expected to be heteroscedastically distributed with a known form of the heteroscedasticity, namely equation (11). Plotting the cross-sectional standard deviation of relative prices for each bilateral pair against the absolute value of the logarithm of relative income gives us the positive association we expect to find. Locations with similar income levels have low price dispersion, as income disparities rise, the deviations become more dispersed across goods for that bilateral pair (see Figure 4).

5.2 Distance and Trade Costs

Obviously we will not have an averaging out property in the second moment of the distribution since the summations in the variance expression involve squared terms. The second term in the $Var_i(q_{ijk}|jk)$ expression simplifies somewhat once we impose the condition that the mean is zero (which is plausible given our earlier analysis of the mean):

$$b' = \sum_{i=1}^{N}(1-\alpha_i)^2(\beta^i)^2$$

where we have exploited the fact that $(I^i_{jk})^2 = 1$. Unfortunately, the coefficient does not reduce conveniently to a simple function of the average value of either the production parameter or the elasticity of trade cost with respect to distance. If either of the parameters α_i or β_i were constant across goods, we would identify the average value of the other parameter up to a scalar (the scalar being the value of the other parameter). If we had traded good prices on the left-hand-side we would be able to estimate the average β_i. In any case, it must be true that $b' > 0$. Thus distance is expected to matter for the second moment properties of the cross-section. Figure 5 plots $Var_i(q_{ijk}|jk)$ against bilateral distance. We see the positive relationship implied by the above algebra.

5.3 Findings

Turning to our regression results, we see that both relative income and distance matter as the theory predicts. We find that the distance coefficient falls in the international case when we add relative income, but remains highly statistically significant. Recall that it became insignificantly different from zero in the analysis of means, which was predicted by the averaging out property. The second moments should be affected by distance and they are. Moreover, the impact of distance is quite similar across the two panels as long as we control for income disparities in the international specification. The coefficient on the distance variable is 0.0012 internationally versus 0.0008 intranationally. Given the standard error on the intranational distance coefficient we probably cannot reject the hypothesis that the two coefficients are equal (had we not controlled for income we might have been able to claim distance matters more internationally).

$$Var_i(q_{ijk} \mid jk) = a + b(\log D_{jk})^2 + cz_{jk}^2 + \text{residuals}$$

	Intercept	$(\log D_{jk})^2$	z_{jk}^2
International Locations	0.1562	0.0016	
	(0.0067)	(0.0001)	
International Locations	0.1497	0.0012	0.0202
	(0.0057)	(0.0001)	(0.0004)
Intranational Locations	0.0416	0.0008	
	(0.0092)	(0.0002)	

Economically, the magnitude of the coefficients is large. If we go from a distance of 100 kilometres to 1000 kilometres, the predicted increase in the cross-sectional standard deviation is $(0.0012(\log 1000)^2 - (\log 100)^2)^{1/2} = 0.178$. This is a significant fraction of the overall cross-sectional standard deviation of 0.52. Similarly, if we consider a location with income 4 times that of another (a relatively small income gap), the predicted increase (vis-a-vis locations with the same income) is $(0.0202(\log 4)^2)^{1/2} = 0.20$, again a significant fraction of the overall level of price dispersion.

6 Properties of the Time-Series Variance

Up to now we have examined how the *cross-sectional* mean and variance — across goods i — of $q_{ijk,t}$ varies across location-pairs jk. That is, we've examined $E_i(q_{ijk,t} \,|\, jk, t)$ and $Var_i(q_{ijk,t} \,|\, jk, t)$. We now turn to the *time-series* variance for each good i, $Var_t(q_{ijk,t} \,|\, i, jk)$ using the stochastic trade cost model of Sercu, Uppal and Van Hulle (1995) and Lee (2003). We also examine the time-series variance of the *change* in $q_{ijk,t}$, $Var_t(q_{ijk,t} - q_{ijk,t-1} \,|\, i, jk)$. The latter was the focus of Engel and Rogers (1996).

Time-series variation in our model can arise propfrom one of four sources: variation in input shares, α_i, non-traded input prices, $P^N_{ij,t}/P^N_{ik,t}$, traded input prices, $P^T_{ij,t}/P^T_{ik,t}$, and nominal exchange rates, $e_{jk,t}$. We ignore the first and second because input shares and relative income are unlikely to vary much at the annual frequency. We focus on the third and fourth; the interaction between nominal exchange rate variability and the trade-cost model of the price of traded inputs.

Re-introducing the time notation into equation (3), we have

$$q_{ijk,t} = \alpha_i q^N_{jk,t} + (1-\alpha_i) q^T_{ijk,t}. \tag{12}$$

In order to emphasize the role of nominal exchange rates, we write this as,

$$q_{ijk,t} = \alpha_i(\log e_{jk,t} + \log \frac{P^N_{ij,t}}{P^N_{ik,t}}) + (1-\alpha_i) q^T_{ijk,t} \text{ , where} \tag{13}$$

$$|q^T_{ijk,t}| = |\log e_{jk,t} + \log \frac{P^T_{ij,t}}{P^T_{ik,t}}| \leq \tau^i_{jk} = \beta_i \log D_{jk} \;\; . \tag{14}$$

This is no more than an alternative way of stating the trade cost model underlying $q^T_{ijk,t}$. Nevertheless, it emphasizes an important link between time-series variation in nominal exchange rates, $e_{jk,t}$, and the local-currency price ratios of traded and non-traded inputs; $e_{jk,t}$ can move independently of $P^N_{ij,t}/P^N_{ik,t}$, but not of $P^T_{ij,t}/P^T_{ik,t}$. Moreover, the latter effect will depend on distance.

We formalize this in the following proposition (yet to be proven and/or tightened-up).

Proposition. Suppose that, for each location-pair jk, we have

1. Initial cross-sectional dispersion (across goods i) at time t in $q^T_{ijk,t}$ such that the *same* fraction, $f > 0$, of the distribution is truncated at $\pm\tau^i_{jk}$
2. For international locations $Var_t(\log e_{jk,t} \,|\, jk)$ is large relative to $Var_t(\log P^T_{ij,t}/P^T_{ik,t} \,|\, i, jk)$, for all goods i
3. Sufficient cross-sectional variance in trade costs $\tau^i_{jk} = \beta_i \log(D_{jk})$

Then,

(i) $Var_i(q^i_{jk,t} \,|\, jk, t) = 0$ is monotonically increasing in the trade cost τ_{jk} and, therefore, in log distance $\log(D_{jk})$.

(ii) $Var_t(q^i_{jk,t}\,|\,jk,i)$ is (somehow) 'less dependent' on τ_{jk} if $Var_t(e_{jk,t}) = 0$

Statement (i) says that, both intra and internationally, we should expect to find that the cross-sectional variance depends on distance. This is what we found in Section 5. Statement (ii) says that we should expect to find stronger distance effects in the international time-series variances than the intranational time-series variances. The intuition is simply that nominal exchange rate variability induces time-series variation in $q^T_{ijk,t}$ which 'traces out' a cross-sectional pattern in trade costs τ^i_{jk}. As an extreme example, suppose that the (initial) cross-sectional variance in intranational LOP deviations depends on distance but that the traded and non-traded input ratios never change. Then, trivially, the time-series variance would be zero and would be independent of distance. In contrast, if nominal exchange rate variability induces variation in $P^T_{ij,t}/P^T_{ik,t}$ across international locations, then distance will matter: locations separated by larger distances leave room for larger movements over time before the arbitrage bounds are hit.

6.1 Empirical Results

Beginning with summary statistics, Figures 6 and 7 show, for each good, the average (across locations) amount of time-series volatility we find in $q_{ijk,t}$. That is, the top graph plots, for each i, $\sum_{jk}^{N_i} Var_t(q_{ijk,t}\,|\,jk,i)^{1/2}/N_i$, where N_i is the number of location-pairs for good i and $Var_t(q_{ijk,t}\,|\,jk,i)$ is the sample variance for good i and location-pair jk.

Figures 6 and 7 show a striking degree of overall variability. Across goods, the average standard deviation for the price levels is 0.29 for international locations and 0.21 for intranational locations. For the price changes, the averages are almost the same: 0.28 and 0.20, respectively. Engel and Rogers (1996), in contrast, used CPI data on bi-monthly changes in prices for Canada-U.S. locations and found average standard deviations on the order of 0.03 (their Table 2). Even if the price deviations follow a random walk, this translates into an annualized standard deviation of just over 0.07. By any measure, then, the EIU data display much more variability.[3]

[3] The two main differences appear to be (i) the inclusion of many more countries above Canada-U.S., (ii) aggregation: the Engel-Rogers data is comprised of 14 aggregate price indices aggregated to the level of, for instance, food, shelter, furnishings, etc.. Regarding (i), when we included only Canada-U.S. locations, we find average standard deviations of roughly 0.23 and 0.21 for levels and changes, respectively. The additional countries, therefore, increase the international volatilities from roughly 0.23 to 0.28, but have very little effect on the intranational volatilities. Regarding (ii), when we aggregate to 12 comparable aggregate indices (using equal weighting, however), we find average volatilities of 0.09 and 0.08 for the inter and intranational, Canada-U.S. differenced data (exactly analogous to Engel-Rogers). At the aggregated price level, therefore, the EIU display time-series volatility which is on the same order of magnitude as the CPI data.

Turning to the effect of distance, Figure 8 plots $Var_t(q_{ijk,t} \,|\, jk, i)^{1/2}$ against log distance, $\log D_{jk}$, for each good and each bilateral location-pair.[4] The graph suggests little, if any, relationship between distance and intranational price volatility, and an increasing relationship for international locations. This is borne out by the following regressions.

$$Var_t(q_{ijk,t} \,|\, i, jk)^{1/2} = a + b \log(D_{jk}) + \text{residuals}$$

	Intercept	Log Distance
Intranational Locations	0.1916	0.0056
	(0.0018)	(0.0007)
International Locations	0.2435	0.0102
	(0.0009)	(0.0002)
International Locations*	0.1508	0.0375
	(0.0015)	(0.0004)

The third regression — the regression involving international locations with an asterisk — uses only international locations separated by 4,400 kilometres or less. The idea is that we don't have intranational data on locations separated by 4,400 km or more. By definition we are not able to identify distance from border effects for these locations. So, we throw away the extreme distance locations.

7 Conclusions

We have shown that a simple model of retail trade is consistent with many of the properties of micro-price dispersion. Aggregate real exchange rates are not a function of geographic distance, consistent with trade costs that 'average out across goods.' Variance in Law-of-One-Price deviations across goods is greater the farther apart are locations, consistent with the model. Income differences play an important quantitative role in both the mean and the variance. Cities with higher incomes tend to have higher price levels. This finding is consistent with a larger literature on absolute PPP. We also find that the variance of LOP deviations is increasing in income differences. Our model assigns the explanation for this to good-specific shares of non-traded inputs into retail good production. For example, since a haircut involves a larger non-traded component to cost than

[4] There are 122 locations and 237 goods, which amounts to 1,749,297 unique good/location-pair observations (122*121*237/2). However, because of our missing data criteria, we use 892,311 observations in total. Our results are not sensitive to loosening-up the missing data criteria and increasing the number of observations substantially.

a retail computer, price dispersion is higher for the former than the latter even when traded inputs bear equal transportation costs.

A natural question that arises in light of our work is the role the nominal exchange rate plays in various facets of our analysis. While we find evidence of a border in every dimension of our work, we are unsure how much of the variation is due to real factors or the lack of nominal price adjustment. One interpretation of the move from fixed to flexible exchange rates is that the real barriers to retail trade (non-traded inputs and trade costs) are made more apparent than during a fixed exchange rate regime. In effect, the nominal variation in exchange rates induces real exchange rates to fluctuate within the bands of arbitrage without necessarily implying a first-order effect on real allocations. We plan to probe this idea more extensively in future work.

References

[1] Balassa, Bela, 1964, "The purchasing power parity doctrine: A reappraisal," *Journal of Political Economy* 72, 584-596.

[2] Crucini, Mario J. Chris I. Telmer and Marios Zachariadis, 2001, "Understanding european real exchange rates," Vanderbilt University Working Paper No. 01-W20.

[3] Engel, Charles and John H. Rogers, 1996, "How wide is the border?" *American Economic Review* 86, 1112-1125.

[4] Engel, Charles and John H. Rogers, 2003, "European product market integration in the wake of EMU," mimeo.

[5] Lee, Inkoo, 2003, "Implications of censoring for absolute and relative deviations from the law-of-one-price," Chapter 2, Vanderbilt University Ph.D. dissertation proposal.

[6] Rogers, John, 2001, "Price level convergence, relative prices, and inflation in Europe," Board of Governors of the Federal Reserve System International Finance Discussion Papers No. 699.

[7] Samuelson, Paul A., 1964, "Theoretical notes on trade problems," *Review of Economics and Statistics* 46, 145-154.

[8] Sercu, Piet., Raman Uppal and C. Van Hulle, 1995, "The exchange rate in the presence of transactions costs: implications for tests of purchasing power parity," *Journal of Finance* 50, 1309-1319.

Figure 1

Distribution of LOP Deviations: $\log q_{ijk,t}$

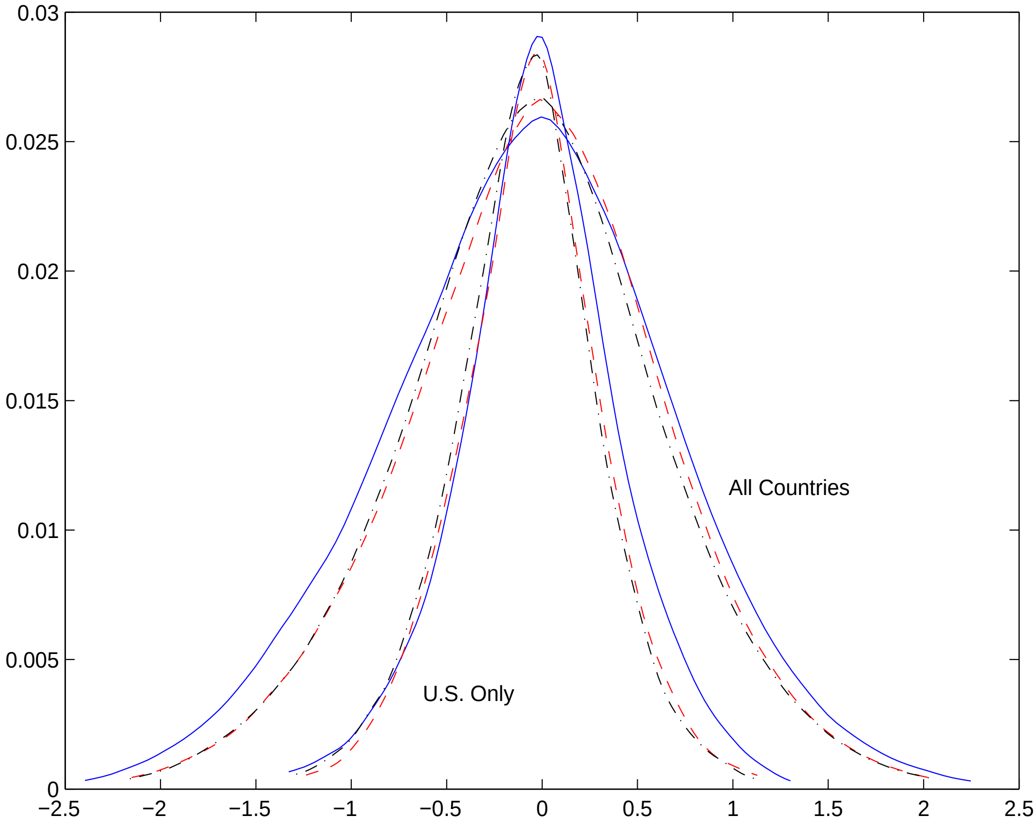


Figure 2

Average Price, q_{jk}, Against Relative Income, z_{jk}

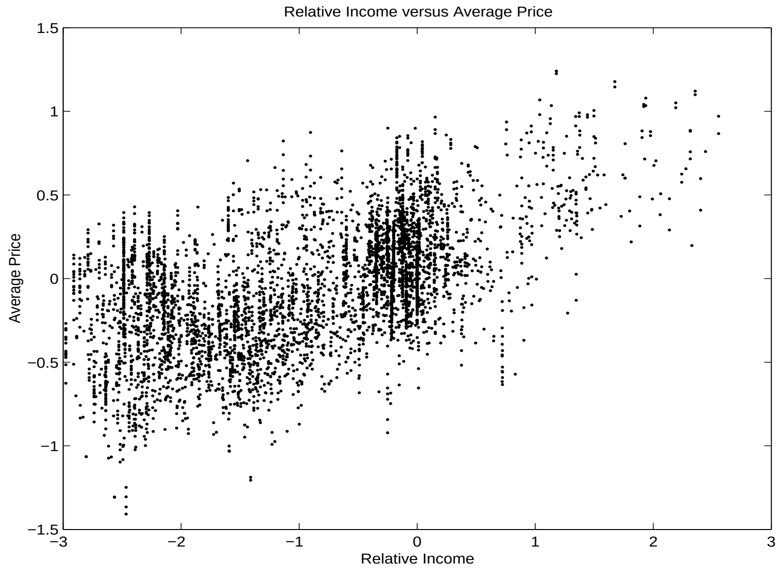


Figure 3

Average Price, q_{jk}, Against Log Distance

(squares are intranational locations, dots are international locations)

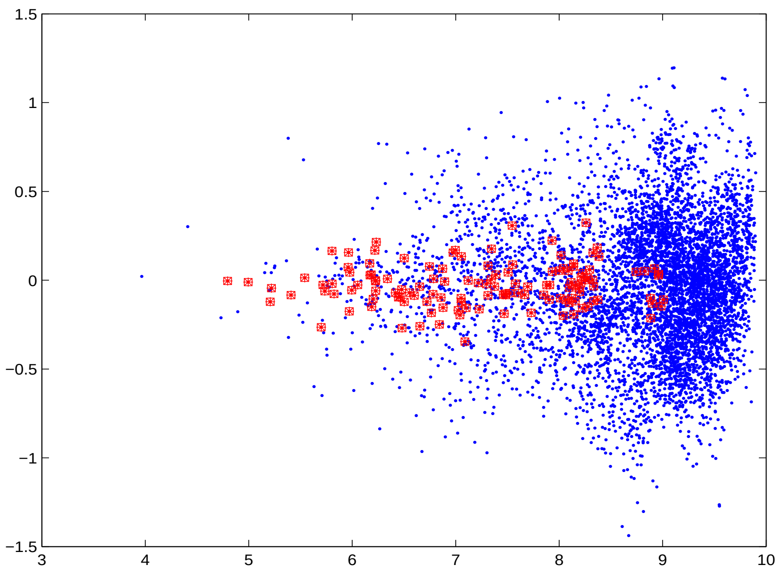


Figure 4

Cross-Sectional Stdev, $Var_i(q_{ijk} \mid jk)^{1/2}$ Against Relative Income, z_{jk}

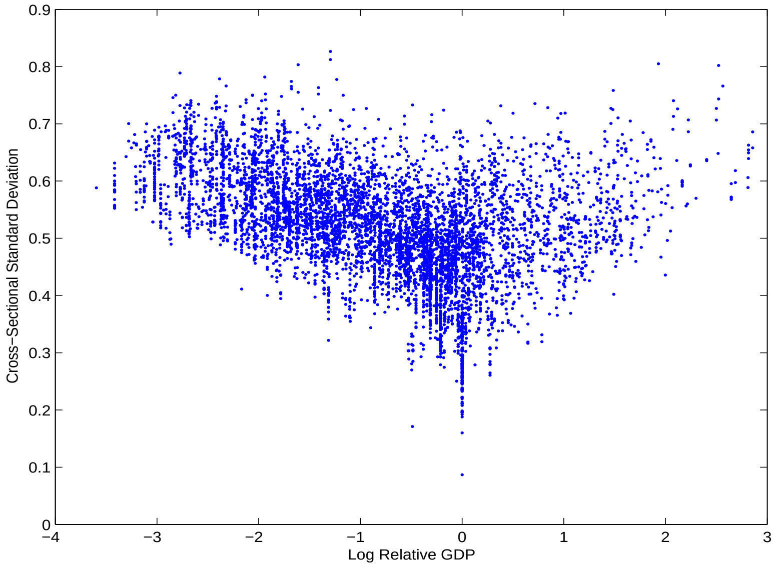


Figure 5

Cross-Sectional Stdev, $Var_i(q_{ijk} \mid jk)^{1/2}$ Against Log Distance

(squares are intranational locations, dots are international locations)

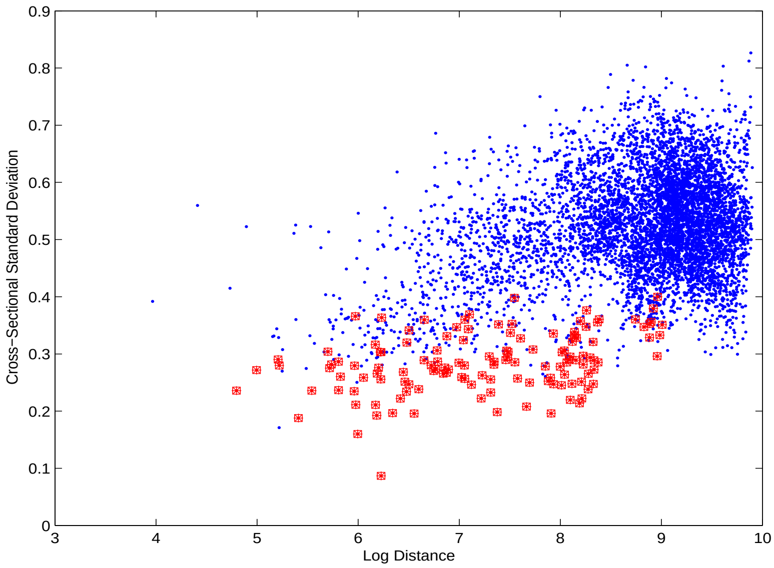


Figure 6

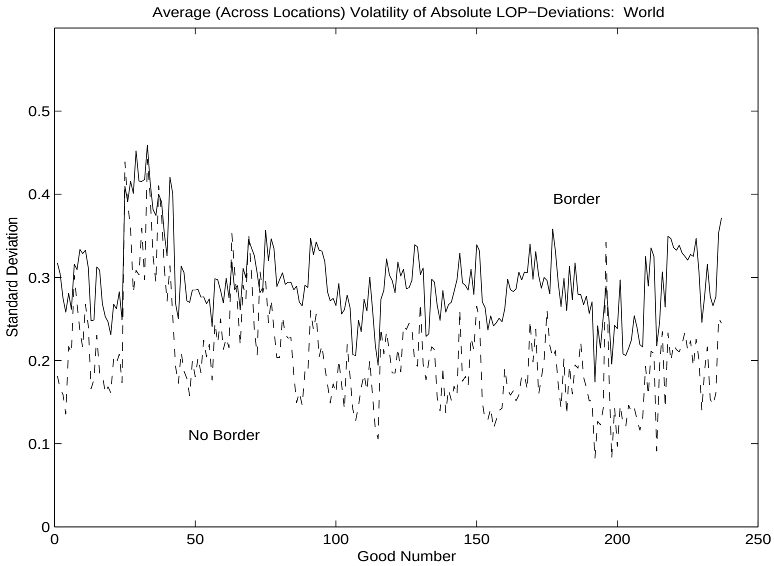


Figure 7

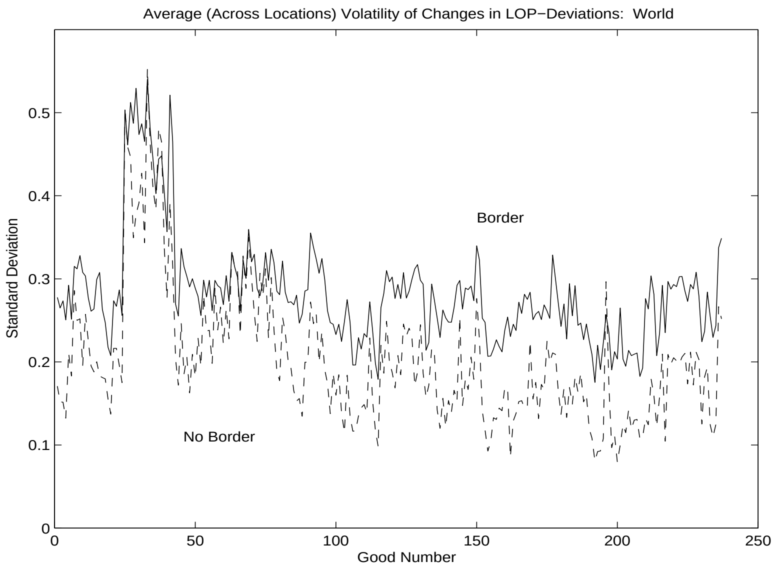
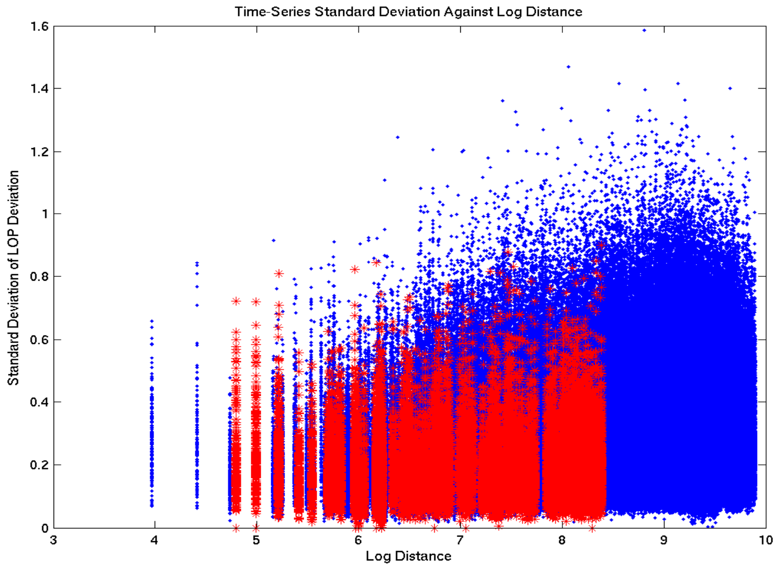


Figure 8



Time-Series Standard Deviation Against Log Distance
Standard Deviation of LOP Deviation
Log Distance
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
3
4
5
6
7
8
9
10