A Spatial Model of National and International Price Dispersion: Theoretical and Empirical Findings

August 2008

Michael A. Anderson*
Washington and Lee University

Kurt C. Schaefer
Calvin College

Stephen L. S. Smith
Gordon College

Abstract We argue that the price dispersion literature ignores markets' spatial character and therefore suffers from missing-variable bias. We estimate spatially-consistent arbitrage models of price dispersion in the United States and Canada using a simple Hotelling spatial-trade model that accounts for price dispersion relative to production location. These results are compared with those from standard arbitrage estimation. The spatial models exhibit significantly improved fit with significant changes in coefficient estimates. Our results suggest that transport costs may be large enough to significantly limit market arbitrage, and that border effects may be significantly misestimated in current models of price dispersion.

JEL classification: E31; F36; F41

Keywords: Spatial Models; Prices; Price Dispersion; Economic Integration

*Corresponding author: (t) 540-463-8971; (e) andersonm@wlu.edu

Acknowledgements Chris Dadak, Adam Hoehn and Parker Sheppard of Washington and Lee University provided able research assistance. Neil Carlson helpfully shared his GIS expertise. Anderson's work was funded in large part by a sabbatical grant from the American Philosophical Society, whose support is gratefully acknowledged. We thank seminar participants at the U.S. International Trade Commission Research Division, the Southern Economics Association meetings, Syracuse University, the Mid-West International Economics Group meetings, and the Federal Reserve Board of Governors for comments on earlier versions of this work. In particular we thank J. David Richardson, John Rogers, Raymond Robertson and John Schindler for their helpful comments. Remaining errors are our responsibility.

A Spatial Model of National and International Price Dispersion: Theoretical and Empirical Findings

I. Introduction

International borders disrupt markets' ability to arbitrage away price differences between localities, and as such add importantly to the variance in prices between locations. By contrast, the distance between locations (a proxy for transport and other costs) is generally not an important contributor to price dispersion. These are among the conclusions of the literature on price dispersion between countries.[1]

These conclusions have been refined as the literature has moved from the use of broad price indices toward micro-level price data, with many recent papers using data on particular items as disparate as men's suits and Big Macs. Whatever the data source, the authors each create a measure of price dispersion using all geographic locations represented in their data. Engel and Rogers (1996) use all combinations of U.S. and Canadian cities' price indices to construct their measure of price dispersion. Crucini, Telmer and Zachariadis (2005a, hereafter referred to as CTZ) likewise use all available European countries' prices to measure price dispersion over time on that continent. Parsley and Wei (2007) use all possible bilateral capital city pairings in the Economist Intelligence Unit (EIU) CityData in decomposing the "Big Mac real exchange rate" into the real exchange rates of its ingredients. These are prominent examples of a common practice: authors obtain a third-party data set and measure price dispersion using every available location.[2]

This approach ignores the geographic nature of production and exchange. It does not control for production location nor distinguish between production and non-production locations.

This concern has been noted by Anderson and van Wincoop (2004, p. 741), who critique Parsley and Wei's (2002) price discrimination measure on the grounds that it misrepresents actual price dispersion in certain spatial cases. But no one has yet demonstrated the theoretical consequences of this error, nor explored its implications for empirical applications.

We argue in this paper that non-spatial models may misestimate key parameters such as the effects of distance and borders on price dispersion. In particular, non-spatial approaches may substantially underestimate the true contribution of distance and transport costs to price variations across locations. After a review of the literature we offer a theoretical spatial model of price dispersion and illustrative empirical evidence, based on price dispersion in the United States and Canada, to support this hypothesis.

But to see this paper's point most simply, suppose that a "topographical map" were drawn of the United States where instead of actual mountains and valleys this map plotted the geographic distribution of some particular good's prices. The "mountains" on this map would be high-price locations and "valleys" would be low-price locations. According to typical price dispersion findings, distance between locations (as a proxy for transport costs) is associated with positive (if small) increases in price dispersion. Therefore, on the map, all else equal, "mountains" would be distant from "valleys." But apart from that there would be no prediction of a spatial dimension to price dispersion. There would be no reason to expect any north-to-south or east-to-west trends in price dispersion; the geographic pattern of price gradients would be random.

Figures 1, 2 and 3 present such maps for carrots, apples and toasters, based on 14 years of price data and 15 U.S. cities. Bands of shading from light to dark show increased prices relative to the low-price location. In each map there is substantial price dispersion, with prices more than doubling from low to high areas. But what is particularly striking is the spatial dimension to dispersion. Prices vary more east-to-west than they do north-to-south. Carrot prices show a nearly constant value along north-south axes, at least until one goes as far east as Michigan and Ohio. For apples the constant-price lines have a diagonal, north-east tilt; apples get more expensive the more south-east one moves from the north-west. For toasters one again finds substantial north-south distances where prices change little.

The point of these maps is not to draw hard conclusions. Nor is there a mystery about the source of price dispersion's spatial character in these cases: prices rise as the distance increases from the point of production. Apples originate in Washington State, carrots in California, and toasters from Los Angeles, the primary port city for imports from China.

But the maps illustrate a key part of our criticism of typical price dispersion models. Such models rest on the assumption that arbitrage occurs uniformly across all location pairs. By ignoring production location, and by assuming that the distance between one pair of locations has the same effect on price dispersion as the distance between any other pair of locations, the models ignore the geographical character of price dispersion.

If our argument is correct, the models used in four related sub-literatures on price dispersion may be misspecified. These literatures, on purchasing-power parity, on border effects, on attempts to measure the gains to international integration, and on measuring price elasticities, are all innocent of any spatial dimension. In this paper we focus on how distance and the border are miscast in non-spatial price-dispersion studies. But our larger point is that the conclusions in all of these related literatures are based upon standard models of price dispersion, and as such we should expect problems to arise including biased coefficients and poorly fitting models.

In what follows we offer a brief literature review focusing on the finding, common in the price-dispersion literature, that distance has small effects on price dispersion. This finding should be puzzling in light of other evidence about the large, persistent effects of distance on trade. In the literature review and theory section we argue that this result is what one might expect when non-spatial models are employed. Finally, in our illustrative regressions, we find that spatial estimation substantially changes distance and border coefficients. Spatial estimates, when compared to non-spatial estimates, also offer dramatically improved goodness of fit measures. In our view, if future empirical work supports our hypothesis, the literatures named above should be revisited with spatially-informed models.

II. Literature Review

The past decade has seen a surge in studies analyzing global price dispersion with detailed micro-level goods price data and location-to-location price comparisons. Many studies focus on documenting and exploring deviations from the law of one price and purchasing-power parity (see Obstfeld and Rogoff 2000, Anderson and Van Wincoop 2004, Parsley and Wei 2007). Other frequent topics include the effects of common currencies and trade agreement membership in diminishing price dispersion (Rogers 2007, Engel and Rogers 2004). Other applications use price dispersion to infer the size of trade costs and border effects (Engel and Rogers 1996, CTZ 2005a, b, Bergin and Glick 2007). In all these cases there is a strong theoretical presumption that, all else equal, a rise in the distance between cities or regions, and the crossing of an international border, should be associated with increased price dispersion. A border variable will measure the aggregate effect of official trade restrictions and the delays and burdens of doing business in another culture and under another legal system, all of which can be expected to disrupt traders' ability to arbitrage away price differences between locations. Distance is also expected to inhibit arbitrage. It is thought to proxy transport costs, as well as the costs of market discovery, both of which add to measured price dispersion.[3]

Empirical results have strongly supported the first proposition regarding borders, but have found little support for the notion that distance adds importantly to price dispersion. In one oft-referenced paper, Engel and Rogers (1996) examine relative price dispersion between pairs of U.S. and Canadian cities. They find that the United States-Canada border adds substantially to price dispersion, famously concluding that crossing the border is equivalent to adding 75,000 miles of distance between pairs of cities. This result reflects the large measured effect of the border, but also the modest measured effect of distance.

The authors' results predict that crossing a border will increase price volatility, relative to that found in the United States, by 37 percent, a very large effect.[4] By contrast, their estimated effect of distance on price variability is quite small. It has the expected positive sign—distance should contribute to higher variability in price differentials—but at 0.00106 it means, for instance, that a 200 percent increase in distance only raises price variability by 0.00212, or about 5 percent of mean price variability. These two broad findings, large border and small distance effects, have been found in a number of other studies.[5]

Interestingly, perhaps because of the seemingly modest effect of distance on price dispersion, several recent studies have dropped distance as an explanatory variable. Engel and Rogers (2004), CTZ (2005a), and Rogers (2007) measure intra-European price dispersion relative to the mean European price and analyze trends in price dispersion without reference to distance between locations. The difference between a location's price and the mean price across locations is a reasonable descriptive statistic for measuring price dispersion. But for understanding the sources of price dispersion, this construction is at odds with the theoretical argument that arbitrage is driven by price differentials between locations rather than by differences from a regional average.

The price arbitrage literature's distance results are paradoxical in light of the (related) literature that applies gravity models to trade flows. A number of authors have examined how borders affect trade volumes, beginning with McCallum's (1995) original finding that Canadian provinces trade 22 times more with each other than with U.S. states after controlling for economic size and distance. There is confirmation here, and in other gravity models of trade-flows, for the finding of a large border effect. But this same literature finds distance effects that do not support the conclusions in the price-dispersion literature. Gravity models consistently yield estimates of the elasticity of trade volume with respect to distance in the range between -1.0 and -2.0. The coefficient invariably has the correct sign and is highly statistically significant.[6] It is important to note just how large these estimates are. At -1.0, toward the low end of recent estimates, two locations 200 miles apart (roughly the distance between London and Paris) will experience twenty times more trade than a trading pair 4000 miles apart (roughly the distance between London and Chicago).

The gravity and price dispersion results suggest a paradox. Consider a distance elasticity of quantity of around -1.0 (plausible on the basis of gravity studies), and a distance elasticity of price of around 0.005 (plausible on the basis of price dispersion studies). The implied "back-of-the-envelope" price elasticity of demand for a typical traded good would therefore have to be -200, which is clearly nonsensical. Which literature gets it right?

III. Theory

We argue that the typical price-dispersion model suffers from misspecification bias due to the exclusion of two relevant variables: the distance of each location from the site at which its imported goods are produced, and production costs at that production site. Moreover, in contrast to the current practice of using all available price comparisons in a given data set, we argue that only certain locations' prices can be meaningfully compared.

Consider the case of a standardized product, produced and exchanged on competitive goods and factor markets. In Figure 4, we illustrate three producers spaced along a one-dimensional world on the horizontal axis. Prices and/or costs are measured on the vertical axis. The first (left-most) producer can manufacture a product at cost C_1. The product can then be shipped elsewhere at constant unit transport cost d.[7] This cost defines the slope of a "trade cone" surrounding the production site—along which the cost of the product when shipped from that location equals the cost of production C_1 (at the "stem" of the cone) plus the total transport cost (d)·(distance). The second (middle) producer has lower production costs (C_2), such that goods produced at location 2 and shipped to location 1 are less expensive (at price P_1, equal to production cost C_2 plus transportation costs) than C_1. Thus no production takes place at location 1, which becomes an importer of this good from location 2.

Location 3's relatively low production costs allow it to produce and compete with location 2. Location 3's market will extend leftward toward location 2, up to the point at which the two locations' trade cones intersect. Call this Location 4. So, in sum, the trade cones associated with locations 2 and 3 supply all locations. Location 1 is "on" trade cone 2 (that is, it is supplied by location 2) and location 4 can be on either cone 2 or cone 3.

Now consider what would happen in this instance if a price-dispersion study regressed price differences against trade-pair distances, using each location-pair's price differential as a unique observation. The distance between location 2 and location 3 is greater than the distance between 2 and 1, yet the price difference between 2 and 3 is smaller. The distance between location 1 and location 4 is large—the second largest in this example—yet the price difference between P_1 and P_4 is smaller than any other price difference. A typical price-dispersion study would conclude that distance has little effect on relative prices, even though distance is (after production costs) *the* major influence on relative prices.

The source of this bias is the omission of a relevant variable: the supply location. The price difference between P_1 and P_4 is very small because locations 1 and 4 are almost equidistant from location 2's supply. Their distance from that stem point has a big influence on their prices, but their distance from each other has no effect on their prices.

The effect on other parameters, like a border coefficient, of regressing all price differences against all location pair distances is difficult to predict *a priori*. If crossing a border is costly over and above distance-related costs, as seems likely, a regression that uses price differences and distances between all city pairs regardless of production location might over- or understate the border's effect on prices, depending upon the exact spatial arrangement of border pairs relative to internal pairs. The exact spatial arrangement of commodity flows would matter, and would likely vary by commodity. In short, a price-dispersion model may be expected to suffer from misspecification bias due to the exclusion of locations' distance from their product sources, exclusion of production costs at that production site, and the inclusion of irrelevant data.[8]

In principle, gravity models are susceptible to the same criticism. But gravity models likely exhibit a smaller misspecification bias in their estimated coefficients, because in gravity equations, quantities (not prices) are the dependent variables. As distance increases, the probability that two locations fall within the same trade cone decreases or, said differently, the probability that they trade nothing increases. Thus distance proxies the otherwise unmeasured spatial effects, while also serving as an indirect measure of per-distance transportation and trade costs of all kinds. This leaves the distance coefficient bearing the burden of several proxies, but they all influence the coefficient in the same direction. This cannot be said of the distance coefficient in arbitrage models. We provide a more formal treatment of this misspecification after introducing some notation in the next section of the paper.

IV. Spatial and Non-Spatial Arbitrage—Reduced Form Equations

In this section we develop the reduced-form equations implicit in the spatial model discussed above. To simplify without sacrificing generality we focus on one type of price-dispersion model, an arbitrage model where the dependent variable is a product price difference between two locations. We continue with the simplest possible case, in which there are no tariffs or trade barriers, identical per-unit-distance shipment costs to all locations, and identical factor costs in each location (so that they contribute nothing to price differentials). Because our empirical work uses city data, we can refer to locations as cities.

Define ΔP_{ij} as the location-pair-specific and good-specific simple difference in prices. Thus if bananas cost \$20/cwt in Miami (P_i in location i) and \$30/cwt in Chicago (P_j in location j), we have a price difference of ΔP_{ij} = \$10 for that good for that location pair.

In our spatial model, price differences between locations arise in two basic ways, corresponding to whether or not the locations are on the same trade cone. Locations on the same trade cone—supplied from the same stem location—pay identical prices net of transport costs: they all pay the stem location's production cost. Therefore price differences that occur in this case reflect differences in locations' distance from the source city. If one location is 100 miles from the stem, and the other is 500 miles from the stem, the extra transport cost associated with traveling 400 miles further is the source of their price differential.

Define C_k to be the production cost at stem location k, and D_i to be the distance between location i and stem k. We define locations' "difference in distance" from the producer/stem as ΔD_{ij}, the distance of consuming city *i* from the producing city (D_i) minus the distance of consuming city *j* from the producing city (D_j). If Miami and Chicago are importers from Honduras (which is, say, 700 miles from Miami and 1500 miles from Chicago), then ΔD_{ij} (= D_i – D_j = 1500 – 700) is 800 miles, which is not necessarily the distance from Miami to Chicago. As an important special case, note that if either city i or city j is the stem city, ΔD_{ij} *is* simply the distance between the two. If the stem is city j, for instance, $D_j = 0$ and $\Delta D_{ij} = D_i - D_j = D_i$.

We can now describe the price difference. As before, d is the unit transport cost from the stem to the consuming location i or j, which we assume to be the same for all destinations on the cone. We therefore have $P_i = C_k + d \cdot D_i$ and $P_j = C_k + d \cdot D_j$, so that $\Delta P_{ij} = P_i - P_j = d \cdot \Delta D_{ij}$. This case is illustrated in Figure 5.

The second type of price difference occurs between locations that are not on the same trade cone. In each location price is the sum of the cost at the stem plus the distance-related transport costs to the consuming location. For location i on stem k we have $P_i = C_k + d \cdot D_i$; for location j on stem m we have $P_j = C_m + d \cdot D_j$. The price difference is $\Delta P_{ij} = (C_k - C_m) + d \cdot \Delta D_{ij}$.

Note that in this case ΔD_{ij} retains its interpretation as the difference in the distances between location i and i's stem and location j and j's stem, though i and j are on different stems. Figure 6 illustrates.[9]

Most generally then, considering cases where locations are on the same or different trade cones, and under free trade with local pure competition, arbitrage should cause prices, local production quantities and trade-flow quantities to adjust so that prices between any two locations differ in the following way:

$$(1) \quad \Delta P_{ij} = \min\{d\cdot\Delta D_{ij}, [(C_k - C_m) + d\cdot\Delta D_{ij}]\},$$

where location *k* is the trade-cone base for location *i* and location *m* is the base for location *j* when i and j are not on the same cone. The price difference equals the first bracketed term when the two locations are in the same trade cone. The second term becomes relevant if the two locations are on separate trade cones, such that no trade flows occur between the two.

Now consider a binary variable that will simplify the reduced-form estimating equation. $NOCONE_{ij}$ takes on a value of one if locations *i* and *j* lie on distinct trade cones. We therefore have the following simple spatial arbitrage regression model:

$$(2) \quad \Delta P_{ij} = \mathbf{d}\cdot\Delta D_{ij} + \mathbf{c}\cdot NOCONE_{ij}\cdot(C_k - C_m) + \varepsilon_{ij},$$

where **d** and **c** are the only regression coefficients to be estimated. Under our initial simplifying assumptions, **d** estimates distance-related trade friction—approximately the same thing as per-unit-distance transit costs—and **c** estimates the effect on price differences of the cost differential when two locations lie on separate trade cones.[10]

At this point the difference between a spatial arbitrage model and a standard arbitrage model in measuring the effects of distance on price differentials comes into clear focus. In a spatial arbitrage model, the required dependent variable is a price differential measured for locations inside a trade cone (i.e. when $NOCONE_{ij} = 0$) relative to their supply source; other comparisons are uninformative. If Boston's rice comes from Louisiana, the Boston-Louisiana price differential—not the Boston-Chicago price differential, nor the Boston-Vancouver differential—is the meaningful measure of trade and distance costs. Standard arbitrage models that look at price differentials between Boston and many other cities both in and outside of Boston's trade cone are combining data that are relevant with data that are irrelevant for understanding the nature of price dispersion.

We can now provide a more formal analysis of the nature of misspecification error in the standard arbitrage model. When *i* and *j* lie on the same trade cone, equation (2) reduces to

(3) $\Delta P_{ij} = \mathbf{d}\cdot(\Delta D_{ij}) + \varepsilon_{ij}$.

If the two consumption sites are on different cones, the equation becomes

(4) $\Delta P_{ij} = \mathbf{d}\cdot(\Delta D_{ij}) + \mathbf{c}(C_k - C_m) + \varepsilon_{ij}$

Equation (3) is a special case of equation (4) in which *k* and *m* are the same place. So let us simply write the model as equation (4), with the understanding that k and m are not distinct places when *i* and *j* lie within a single trade cone. These two cases are illustrated in Figures 5 and 6.

Instead of estimating equation (4), the standard arbitrage model pairs all possible combinations of *destination* cities without regard to the origin of each site's goods, then includes each destination city's production costs, to estimate

$$(5) \quad \Delta P_{ij} = \mathbf{d} \cdot (D_{ij}) + \mathbf{c} \cdot (C_i - C_j) + \zeta_{ij} .$$

Equations (4) and (5) measure distance and marginal costs in completely different ways. Equation (4)'s first term measures the *difference in distance* (ΔD_{ij}) to the production center of locations *i* and *j*; equation (5)'s first term simply measures the distance *between* location *i* and *j* (D_{ij}). The second term in equation (4) measures the difference in production costs at the *origin cities* for goods consumed at *i* and *j;* the parallel term in equation (5) measures production costs *at i and j, even though nothing may be produced there and local production costs may be completely inconsequential to local prices.*

It is not impossible that the estimation of equation (5) would result in unbiased estimates of the coefficients in equation (4), but this requires a very particular spatial arrangement of locations *i, j, k* and *m*. Equation (4) reduces to equation (5) if two conditions are met: all goods originate in the same place so that there is a single trade cone, and all destination cities lie along a one-dimensional straight ray emanating from that single origin and proceeding in a single direction. In other words, all destinations must lie on the same side of a single one-dimensional trade cone. One imagines trade in ancient Egypt originating from a Mediterranean port city, or trade in imports in colonial Gambia, or perhaps some colonial Indian rail lines or pre-modern Chinese river traffic.

In the next more complicated spatial arrangement of trade—a single trade cone, with destinations dispersed in two dimensions around a single origin, such that equation (4) becomes the proper specification—then it is impossible to say with certainty how the incorrect use of equation (5) will bias one's conclusions. This is because irrelevant destination-site production costs are displacing the relevant production-site marginal costs as the second independent variable in the estimations, even though production costs away from the stem of the trade cone do not affect the spatial distribution of prices.

However, we may make some progress under the reasonable assumption that differences in production costs at destination locations within the trade cone are relatively small and/or not systematically related to each location's distance from the stem location. Then the inclusion (or exclusion) of destination production costs as an independent variable will not significantly affect the bias in the distance coefficient. Under this assumption, the distance coefficient will be systematically underestimated because D_{ij} is used as an independent variable rather than ΔD_{ij}. (For a proof, see Proof Appendix.) This misspecification likely also biases the estimation of other independent variables such as the border.

V. Estimation Issues

Proper estimation of a spatial model hinges on identifying production and destination locations on trade cones, by product. In general this would require four types of data: product-level trade-pair commodity-flow data; production data by location; trade-pair distance-from-producer data;

and trade-pair price-difference data. Production location and commodity flow data are necessary to identify the locations and bases of the production cones.[11]

If these high data demands are met, estimating the effect of distance on price dispersion (equation (4)) is still problematic. In contrast to our Figure 4, the costs of transport are not the same from each production location; some locations will have lower cost transport options (coastal cities) while others will face geography that influences the choice of transport mode (e.g. long distances to the nearest city in the data base) and hence per-mile costs. Moreover, there is a likely endogeneity between production location and transport mode.

We solve the problems noted above by careful selection of products and regions, and by simplifying our model. Our data allows two sets of illustrative examples. The first set considers arbitrage within the United States for 12 products; the second set explores arbitrage across U.S. and Canadian cities for a subset of eight products. We use EIU CityData for annual retail prices for all products.[12] In the first case we use U.S. data exclusively in order to eliminate price differences due to tariffs and trade restrictions, currency exchange issues, and border effects. Nine of the commodities—apples, bananas, carrots, lemons, olive oil, oranges, canned peaches, canned pineapples, and fresh orange juice—are food items and thus are not subject to sales tax in most U.S. localities, eliminating another source of price dispersion. The remaining three are standardized manufactured goods: toasters, children's sneakers, and tennis balls. All 12 goods are relatively uniform and precisely defined, minimizing complications that would arise from using broadly-defined product categories. A key benefit from the choice of these products and cities is the relative ease of locating the production location, the "stem city," for each product.

Next, we simplify our model to accurately estimate how distance affects price dispersion. We will use an empirical model where, for each i-j comparison, one of the two cities is the actual production location.[13] This approach, while handicapping our spatial model with a small sample size, allows a clean comparison between a spatial model and a typical arbitrage model.

The dependent variable is the retail price of a product in a “destination city” (a non-producing location inside a trade cone) minus the retail price of the same product at the relevant production location (the base or stem of the cone). As noted above, identifying stems and destinations on particular cones is difficult. We sidestep this problem by choosing products for which U.S. consumption is supplied from a single domestic location (for instance, more than 80 percent of fresh oranges come from the counties immediately surrounding Los Angeles) or, alternatively, for which U.S. consumption is imported through one major port (such as children’s sneakers, produced mostly in China and imported almost exclusively through Los Angeles since the mid-1990s).[14] These are products that plausibly are traded on one “cone.”

The subset of eight goods chosen for the U.S.-Canada estimations are products that are not produced domestically in Canada and are either imported from the United States (lemons and oranges, for instance) or are mostly transshipped through the United States (such as olive oil). In other words, they are products for which the stem city is in the United States, and Canadian destinations cities are likely on the same cone as U.S. destination cities.

Our empirical estimations proceed as follows. First, for every product we run a standard arbitrage model in which price differentials across all possible city pair permutations are included on the left-hand side. This corresponds to estimating equation (5). These results offer a baseline against which the performance of the spatial model can be assessed. Second, we estimate the simplified spatial model which calculates the dependent variable for each product as the retail price differential between destination city j and production location city i, using the distance between the two cities as the distance measure. This corresponds to estimating equation (3).

A specific example helps illustrate. Consider the U.S. markets for lemons in a single year, say 1998. There are 15 U.S. destination cities in the EIU data for which we have retail prices. We know that virtually all lemons come from Los Angeles. For the standard model there would be 105 observations, corresponding to all possible city pair permutations, and "distance" is the literal distance between each city pair. In the spatial model (3), by contrast, Los Angeles is the sole production city (j) while the remaining 14 cities constitute the (i) destinations, for 14 observations. "Distance" is the Los Angeles-to-destination distance.[15]

Bananas, olive oil, toasters and children's sneakers are imported items for which domestic production is negligible. For each product we treat its major port of entry as its production point. These products were chosen because their main port of entry corresponds to—or is geographically close to—a city in the EIU data so it is possible to construct the necessary price differentials for the spatial "difference from stem" calculations.

We follow the convention of estimating arbitrage models in logs; we measure price differentials as the natural log of price ratios.[16] Independent variables include, in addition to logged distance and a border dummy for the U.S.-Canada cases, sales/consumption tax differentials for products and city pairs with such taxes, and wage and rent differentials to capture factor price differences. Finally, all estimations include city and year fixed effects to control for all kinds of local retailing and other costs not reflected in our rough measures of factor costs.[17]

Thus for each good, using pooled annual observations 1990-2003, we estimate the following equation:

(6) $$\ln(\text{priceratio}_{ijt}) = \beta_1 \ln(\text{distance}_{ij}) + \beta_2(\text{border}_{ij}) + \beta_3 \ln(\text{taxratio}_{ijt}) + \beta_4 \ln(\text{wageratio}_{ijt})$$

$$+ \beta_5 \ln(\text{rentratio}_{ijt}) + \sum_i \delta_i \text{Icity}_i + \sum_j \gamma_j \text{Jcity}_j + \sum_t \kappa_t \text{Year}_t + \varepsilon_{ijt}.$$

Destination and production cities are indexed by i and j, respectively, as appropriate for the standard or spatial version of the model. Priceratio_{ijt} is the dollar price in location i divided by the dollar price in location j in period t, and the wage, rent and tax ratios are constructed similarly. Distances are measured in miles on great circles. For standard models this is simply the distance between i and j, while for spatial models it is the distance between the stem and the destination. Thus equation (6) estimates the spatial model, equation (3), the one cone case when distance reflects distance between producer and consumer, and estimates the standard model, equation (5), when distance is calculated in the standard way. The border dummy equals unity when a location pair crosses an international border.

VI. Illustrative Empirical Results

Table 1 displays the selected results from estimating equation (6), on U.S.-only data. The left-most panel presents regression results for the standard model, and the right-most the spatial model. We divide the products between food and manufactures, and for food we further divide by whether the item is produced in the United States or imported.

As expected, the spatial model is estimated on far fewer data points than the standard model, with sample sizes less than 200 for the spatial model, typically over 1000 for the standard. Despite the smaller sample size the spatial model turns in far better goodness-of-fit measures, with R-squared values averaging 69 percent for the food items and 86 percent for the three manufactured goods. By comparison, the standard model returns R-squared values of 24 percent for food and 30 percent for the manufactured goods.

In the theory section we demonstrated that the coefficient on distance is biased downward in standard models; our illustrative empirical results support that claim. Across all the products in Table 1 only two of the distance coefficients in the standard model are statistically distinguishable from zero. In the spatial model nine are significant with the expected positive sign. Moreover, these coefficients are arguably economically significant as well, with an average value of 0.10 and some at 0.20 or higher. Consider a coefficient of 0.12 (the value for canned peaches, and the second smallest of the significant values). Suppose there is a change in distance between city pairs from 1,400 miles to 2,600 miles, a 1,200 mile increase, roughly the increase in distance between the Los Angeles-Houston pair, on the one hand, to the Los Angeles-Boston pair, on the other hand. With an elasticity of 0.12 the distance change suggests a 10 percent increase in the city-pair price difference, an amount that comports with the expectation that transport and other distance-related costs should add meaningfully to price differentials. Two of the estimated spatial distance coefficients are statistically significantly negative. These two puzzling examples could arise from a myriad of uncontrolled-for effects, such as pricing-to-market, mis-assignment of trade cones, or trade routes at variance with great circle distances.

For a subset of the products in Table 1 we were able to determine the production location for the Canadian cities, and estimate border effects between the United States and Canada in addition to distance elasticities. Those results are presented in Table 2. Apart from the smaller number of goods the table has the same structure as Table 1.

Once again the goodness of fit measures improve in the spatial model, by a factor of about 3. Distance measures increase on average from zero in the standard model to 0.10 for all of the food items and 0.03 for the manufactures. The border coefficients change dramatically between the two models. In some cases zero or positive border coefficients (for lemons, olive oil and toasters) switch signs and become negative and large in absolute value. For canned peaches the measured border effect increases from 0.07 to 0.32.

Negative border coefficients are surprising. The underlying absolute price data (summarized in Table 3) suggests that some of our particular products are indeed substantially cheaper in Canada in U.S. dollar terms than in the United States in this period. Therefore, the spatial approach suggests that some price-reducing factors (perhaps marketing customs or distributional networks) dominate what would otherwise be the price-increasing nature of crossing a border, after controlling for other contributors to the retail price. What is striking, however, is the difference between the spatial and standard estimates.

This is a small data set composed of targets of opportunity where we could reasonably claim to know where the production location is for the U.S. and Canadian cities in the EIU data. Moreover, we estimate the simplest possible spatial model where the production location is always part of the location pair. Despite these important caveats, the spatial model delivers substantially improved goodness of fit, dramatic increases in the distance coefficient (as our theoretical model predicts), and changes in the border coefficient. We take this as tantalizing evidence in favor of the spatial approach.

VII. Conclusion

Standard price-dispersion models, by including irrelevant data and mismeasuring distance, are apt to produce biased estimates and poor overall model performance. Spatial models come with daunting data demands, but show promise for returning estimates of border effects and distance that more closely resemble how markets actually arbitrage away profit opportunities.

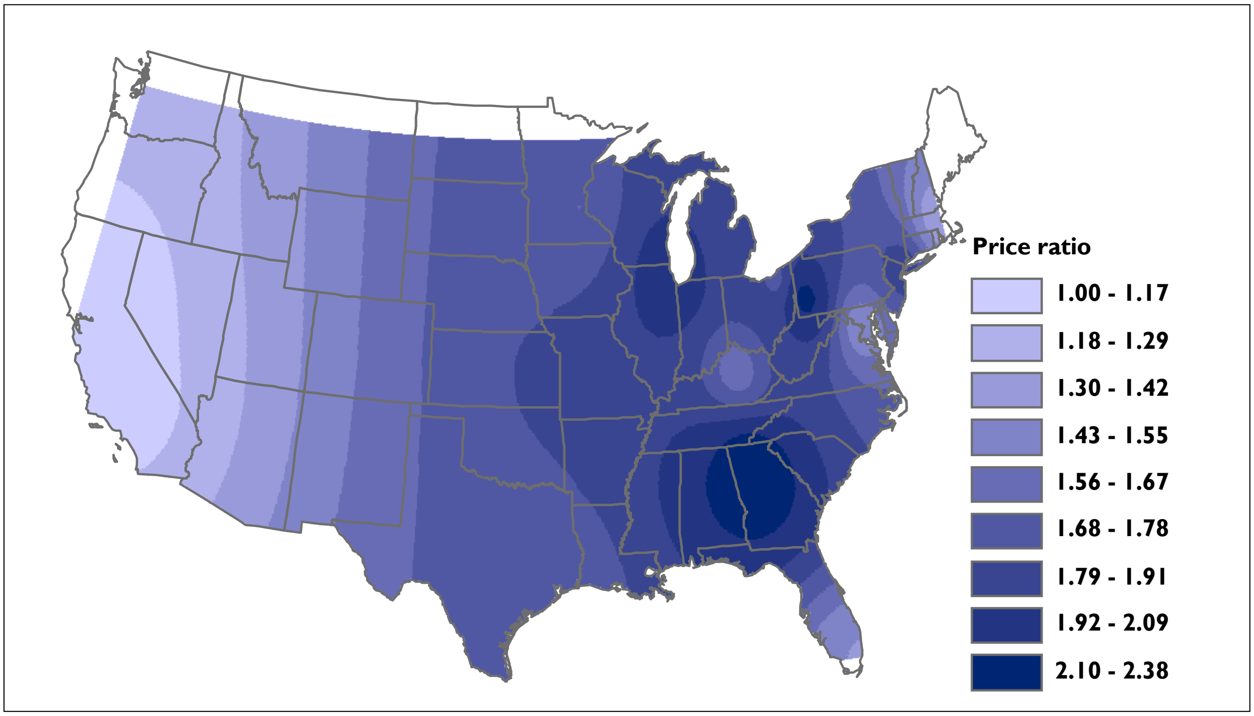
We cannot offer any general conclusions from our very limited data sample, though a couple of provisional observations are in order. In the spatial model (as compared to the standard model) distance is more often economically significant. Border effects are significantly altered and are not always positive (that is, borders do not always add to price dispersion).

In our view, both the spatial theory and these initial empirical results suggest that the conclusions from standard price-dispersion models need to be revisited. The list of areas that could be fruitfully reexamined is quite extensive. We offer the PPP literature, the borders literature, and more narrow papers that calculate price elasticities and measures of non-tariff barriers, as just a few examples.

Distance seems to matter to trade flows, and it should matter to price differences in arbitrage models, too. Nonetheless, as a result of misspecification bias in the standard arbitrage equation, the arbitrage literature is lurching toward the conclusion that distance is irrelevant. A remedy relying on both trade-flow and price data for identification yields a more theoretically-appealing and empirically-satisfying set of results that suggest that distance really does matter in a properly spatially-specified arbitrage model. The next step in this literature should be a large-scale replication of our simple illustrative estimations, using price-dispersion and trade-flow data for an array of clearly-defined product categories.

Figure 1. Absolute Price Dispersion in United States: Carrots, 1990-2003

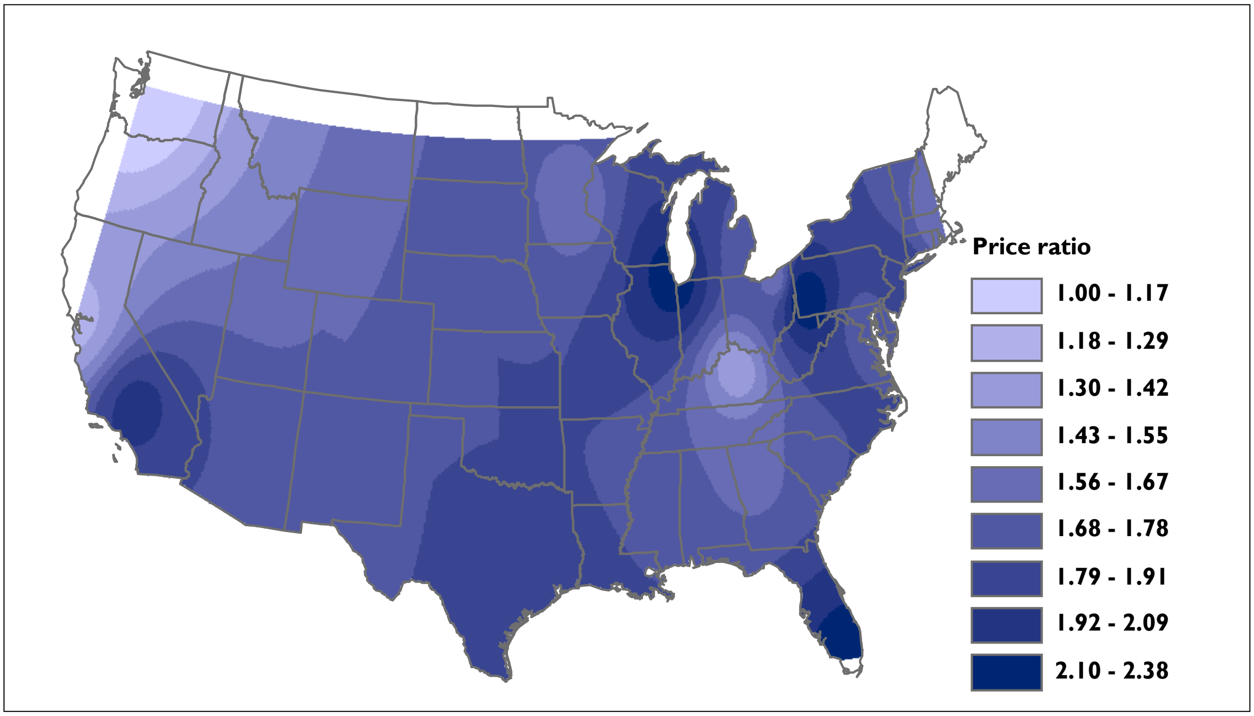
Numeraire: San Francisco average price 1990-2003.



Price ratios are calculated as average absolute retail prices, 1990-2003, relative to the numeraire city, for the 15 EIU cities in the continental United States. GIS software interpolates relative prices between locations to generate the map.

Figure 2. Absolute Price Dispersion in United States: Apples, 1990-2003

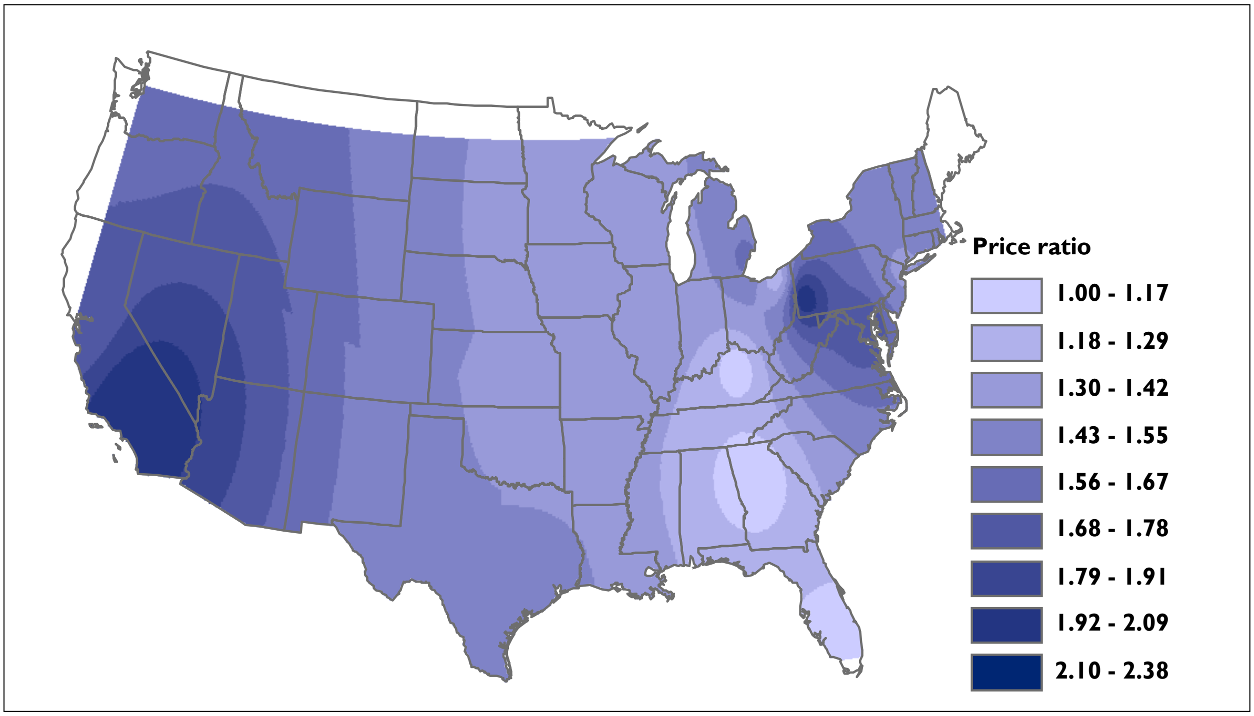
Numeraire: Seattle average price 1990-2003.



Price ratios are calculated as average absolute retail prices, 1990-2003, relative to the numeraire city, for the 15 EIU cities in the continental United States. GIS software interpolates relative prices between locations to generate the map.

Figure 3. Absolute Price Dispersion in United States: Toasters, 1990-2003

Numeraire: Miami average price 1990-2003



Price ratios are calculated as average absolute retail prices, 1990-2003, relative to the numeraire city, for the 15 EIU cities in the continental United States. GIS software interpolates relative prices between locations to generate the map.

Figure 4. A Linear Hotelling Model of Price Dispersion

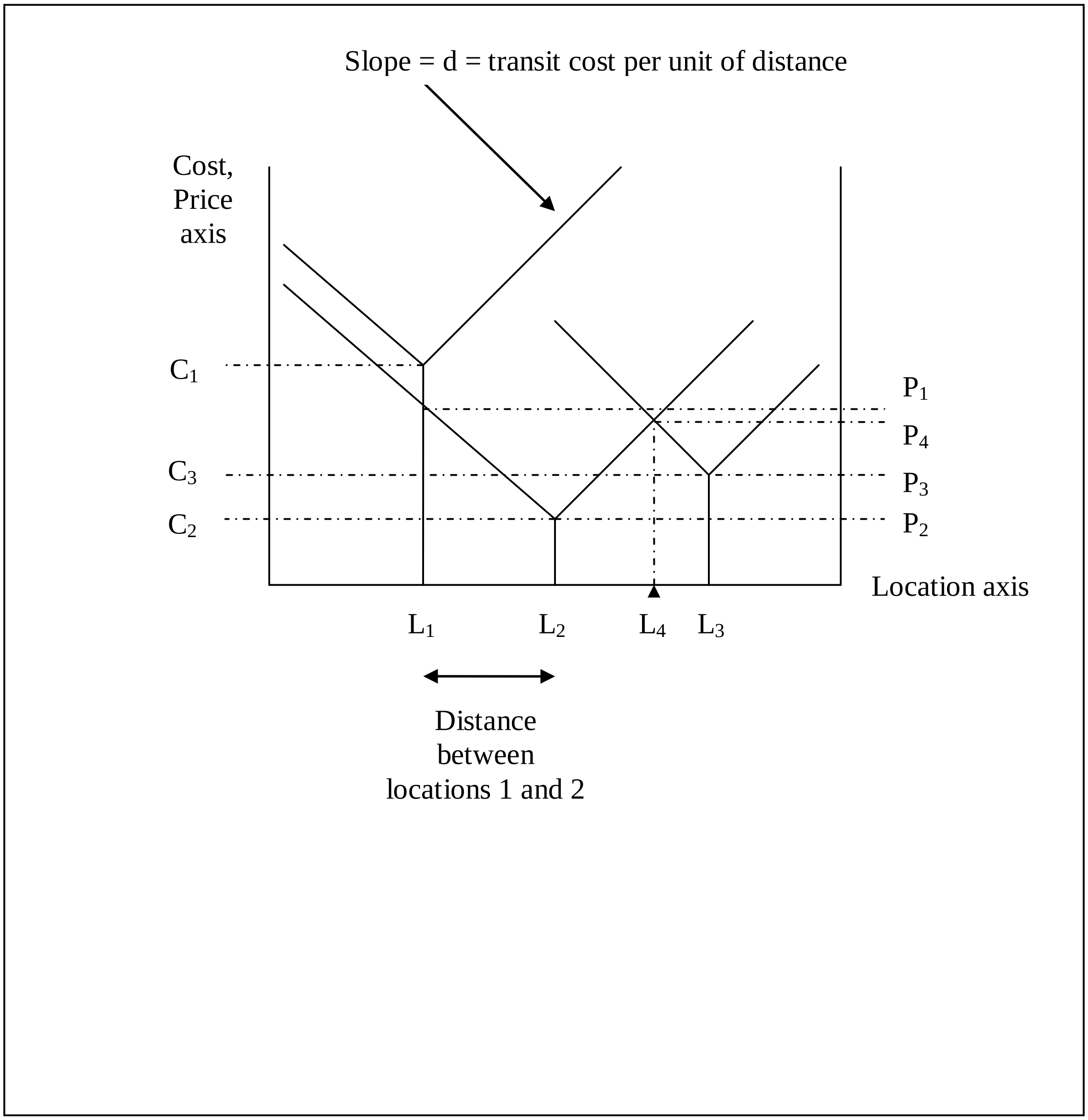


Figure 5. Price and distance when i and j are on the same cone—text equation (3)

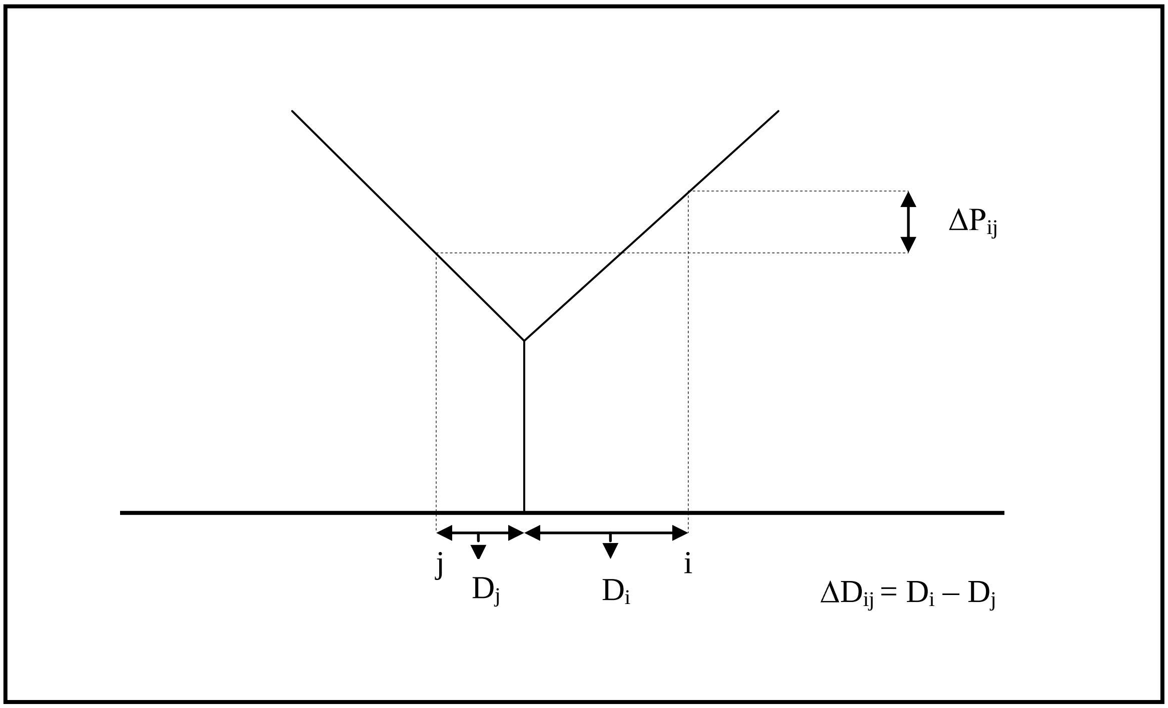


Figure 6. Price and distance when i and j are not on the same cone—text equation (4)

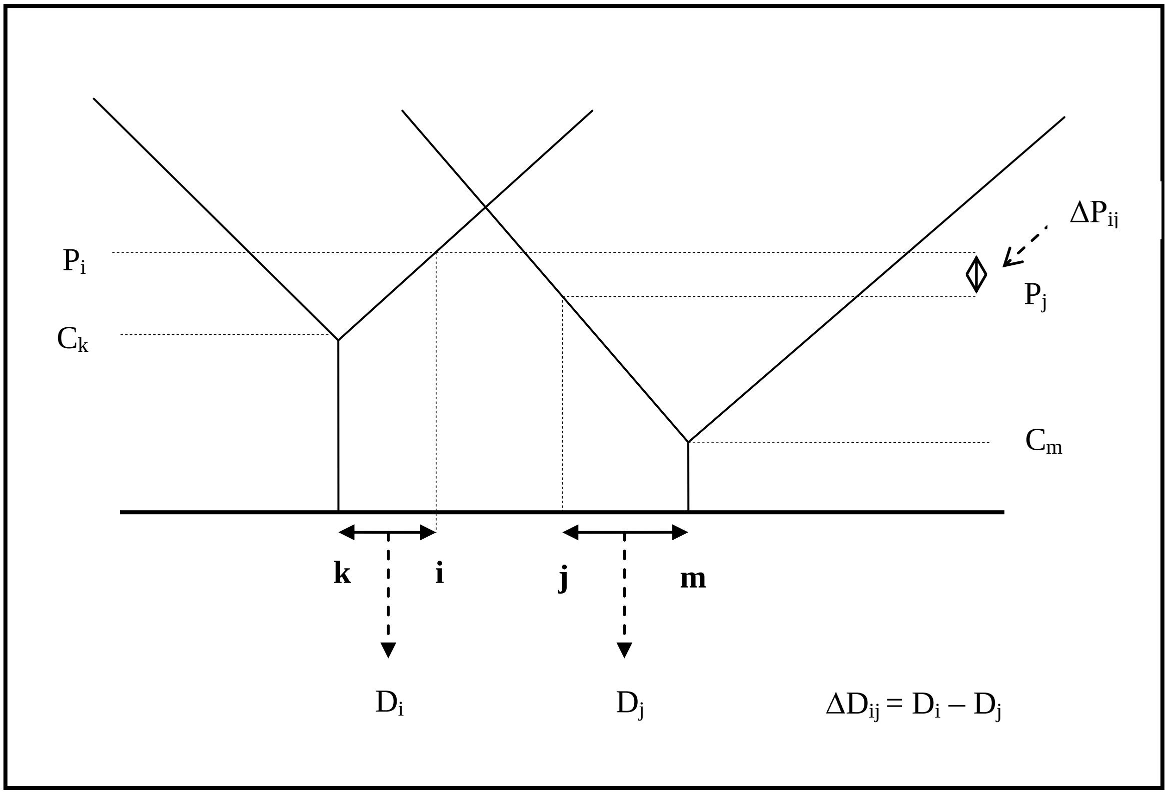


Table 1. Arbitrage Across United States Cities, 1990-2003

Coefficients from estimates of equation (6) in text. Bolded coefficients are significant at 10 percent or better. All regressions include city and year fixed effects, relative wages and rents, and where appropriate relative sales taxes, coefficients not reported.

PRODUCT	"Standard" model			Spatial model			Production location (stem city) and years (if not 1990-2003)†
FOOD	Distance	R^2	n	Distance	R^2	n	
Apples	0.01	0.28	1227	**0.25**	0.69	178	Seattle
Carrots	**0.08**	0.18	1215	**0.31**	0.57	166	Los Angeles
Lemons	**0.05**	0.22	1192	0.10	0.65	176	Los Angeles
Orange Juice	-0.01	0.22	1215	0.01	0.78	177	Miami
Oranges	0.02	0.23	1227	**0.13**	0.74	178	Los Angeles
Peaches	0.01	0.25	1227	**0.12**	0.73	166	San Francisco
Pineapples	0.00	0.28	1227	**-0.06**	0.69	178	Los Angeles
Average	*0.02*	*0.24*		*0.12*	*0.69*		
IMPORTED FOOD							
Bananas	0.01	0.22	1227	**0.09**	0.70	166	New York
Olive Oil	0.00	0.23	1227	**0.10**	0.68	166	New York
Average	*0.00*	*0.23*		*0.10*	*0.69*		
MANUFACTURES							
Children's Sneakers	-0.01	0.32	1227	**0.10**	0.82	94	Los Angeles, 1997+
				0.19	0.91	42	Los Angeles, 2001+
Toasters	0.02	0.26	1227	0.01	0.87	42	Los Angeles, 2001+
				-0.09	0.92	96	Atlanta, 1990-1997
Tennis Balls	-0.01	0.31	915	**0.18**	0.78	130	Los Angeles
Average	*0.00*	*0.30*		*0.08*	*0.86*		

†See Data Appendix for discussion of stem cities and years.

Table 2. Arbitrage Across United States and Canadian Cities, 1990-2003

Coefficients from estimates of equation (6) in text. Bolded coefficients are significant at 10 percent or better. All regressions include city and year fixed effects, relative wages and rents, and where appropriate relative sales taxes, coefficients not reported.

PRODUCT	"Standard" model				Spatial model				Production location (stem city) and years (if not 1990-2003)†
FOOD	Distance	Border	R^2	n	Distance	Border	R^2	n	
Lemons	**0.01**	**0.03**	0.17	2032	0.18	**-0.79**	0.73	110	Los Angeles, 1990-1996
Orange Juice	0.00	0.01	0.19	2063	0.03	0.16	0.69	233	Miami
Oranges	0.01	**0.06**	0.24	2079	**0.18**	0.00	0.78	200	Los Angeles, drop 91, 99
Peaches	**0.01**	**0.07**	0.04	2079	0.02	**0.32**	0.15	218	San Francisco
Average	*0.01*	*0.04*	*0.16*		*0.11*	*-0.08*	*0.58*		
IMPORTED FOOD									
Bananas	0.00	**0.10**	0.24	2079	0.07	-0.09	0.70	218	New York
Olive Oil	0.01	0.04	0.16		0.09	-0.08	0.58	218	New York
Average	*0.00*	*0.15*	*0.29*		*0.09*	*-0.20*	*0.73*		
MANUFACTURES									
Toasters	**0.01**	0.02	0.25	1955	0.11	-0.95	0.82	54	Los Angeles, 2001+
					-0.16	**-0.10**	0.90	128	Atlanta, 1990-1997
Tennis Balls	0.00	0.01	0.29	1411	**0.14**	**-0.49**	0.74	162	Los Angeles
Average	*0.00*	*0.01*	*0.27*		*0.03*	*-0.51*	*0.82*		

†See Data Appendix for discussion of stem cities and years.

Table 3. Product Prices in United States and Canadian Cities

US dollars, average 1990-2003

	Average US Dollar Price	
Product	**US cities**	**Canadian cities**
Lemons	2.30	1.80
Orange Juice	1.56	1.37
Oranges	1.85	1.37
Peaches	1.19	1.35
Bananas	1.09	0.77
Olive Oil	9.31	6.05
Toasters	22.37	26.02
Tennis Balls	7.03	6.88

Source: authors' calculations from EIU CityData.

Data Appendix

1. Data Sources and Construction

Retail prices are from the Economist Intelligence Unit's CityData. We use information from all available U.S. and Canadian cities, with the exception of Honolulu: Atlanta, Boston, Calgary, Chicago, Cleveland, Detroit, Houston, Lexington, Los Angeles, Miami, Minneapolis, Montreal, New York, Pittsburgh, San Francisco, Seattle, Toronto, Vancouver and Washington D. C. All **distances** are calculated on a great circle basis in miles. **Sales taxes** for U.S. locations are from the *Book of the States* 2005 (Lexington, KY: Council of State Government) and earlier issues. Canadian VAT data are from the OECD's *Consumption Tax Trends* (various years). **Rent** values for all locations are in dollars, taken from the EIU CityData series on the rent for unfurnished two bedroom apartments; **wage** data is the EIU CityData series on the hourly cost of domestic cleaning help. Canadian prices are converted into U.S. dollars at the **spot exchange rate** of the date of survey reported by CityData.

2. Spatial Arrangement of U.S. and Canadian Production and Importation

Table 1 Products: Within-U.S. Arbitrage

For the following products, imports during the period under study were a negligible share of U.S. domestic consumption. U.S. consumption was supplied from the specified domestic location(s) as follows:

Apples, fresh: more than 50 percent of the value of U.S. apple consumption is produced in Washington State, for which we use Seattle as the source location. See http://usda.mannlib.cornell.edu/usda/ers/89022/2005/tab-b06.xls and http://www.allaboutapples.com/facts.htm#stateprod.

Carrots, fresh: Over 75 percent of fresh carrot production occurs in California in the counties surrounding Los Angeles, which we use as the source location. See http://www.nass.usda.gov/QuickStats/PullData_US.jsp and http://anrcatalog.ucdavis.edu/pdf/7226.pdf.

Lemons, fresh: Between 80 and 90 percent of lemon production occurs in the counties surrounding Los Angeles, which we use as the source location. See National Agricultural Statistics Service, USDA, tables tabb13 and tabb15, accessed at http://usda.mannlib.cornell.edu/data-sets/specialty/89022/1999/.

Orange juice, fresh: Virtually all fresh orange juice in the United States comes from Florida. See http://usda.mannlib.cornell.edu/usda/current/CitrFrui/CitrFrui-09-21-2006.pdf. We use Miami as the source location.

Oranges, fresh: Over 90 percent of fresh orange production is concentrated in the counties surrounding Los Angeles, which we use as the source location. See, for example, the production data in http://www.ers.usda.gov/publications/fts/jan02/fts296.pdf and USDA data from the National Agricultural Statistics Service at http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1031.

Peaches, canned: Over 90 percent of peach canning takes place in California, the bulk of that in the San Francisco Bay area. We use San Francisco as the source location. See Agricultural Information Center, University of California-Davis, accessed at http://aic.ucdavis.edu/profiles/Peach-2006B.pdf.

Pineapple, canned: Domestically-canned pineapple is produced in Hawaii and shipped to Los Angeles, while imported canned pineapple (mostly from Southeast Asia) is shipped through Los Angeles and New York. We used Los Angeles as the source location. See http://hotdocs.usitc.gov/docs/pubs/701_731/PUB3417.PDF.

The remaining products are almost exclusively imported into the United States with negligible quantities produced domestically. Data on imports and main ports of entry are taken from the U.S. International Trade Commission's data online at http://dataweb.usitc.gov.

Bananas: Main ports are Philadelphia, Los Angeles and Mobile, in that order, which account for more than 70 percent of U.S. imports in this period. We used New York as the source location, as a rough geographic proxy for Philadelphia.

Children's sneakers: Los Angeles and New York are the main ports of entry (in that order), together accounting for more than 50 percent of U.S. "rubber footwear" imports. The share of imports entering through Los Angeles is rising, so we use Los Angeles as the source city for two spatial regressions, for 1997 and 2001 onwards.

Olive oil: Main ports are New York, Los Angeles and San Francisco, in that order, which account for more than 70 percent of U.S. imports in this period. We use New York as the source location.

Tennis balls: Penn balls, manufactured in Arizona, have a greater than 50 percent market share in our period. We use Los Angeles as a rough geographic proxy for the source location.

Toasters: Through 1997, U.S. consumption was predominantly supplied by plants in North Carolina. In the 1998-2000 period supplies from Mexico dominated U.S. consumption, replaced starting in 2001 by surging imports from China shipped via Los Angeles. We estimated two spatial models—one over 1990-7 with Atlanta as the source location, a rough proxy for North Carolina, and the other 2001 onwards with Los Angeles as the source location.

Table 2 Products: U.S.-Canada Arbitrage

We selected a subset of the Table 1 products for which we could determine that (a) domestic Canadian production was minimal and (b) the majority of Canada's consumption was supplied by imports from the United States. To make these calculations we used Statistics Canada data available at http://strategis.ic.gc.ca/sc_mrkti/tdst/engdoc/tr_homep.html. Stem locations and years of estimation are thus the same as described above for the U.S.-only case. In several cases we adjusted the years for estimation, as follows:

Lemons, fresh: 1990-6 only, because from 1997 onwards Mexico and Argentina emerged as strong suppliers of fresh lemons to Canada.

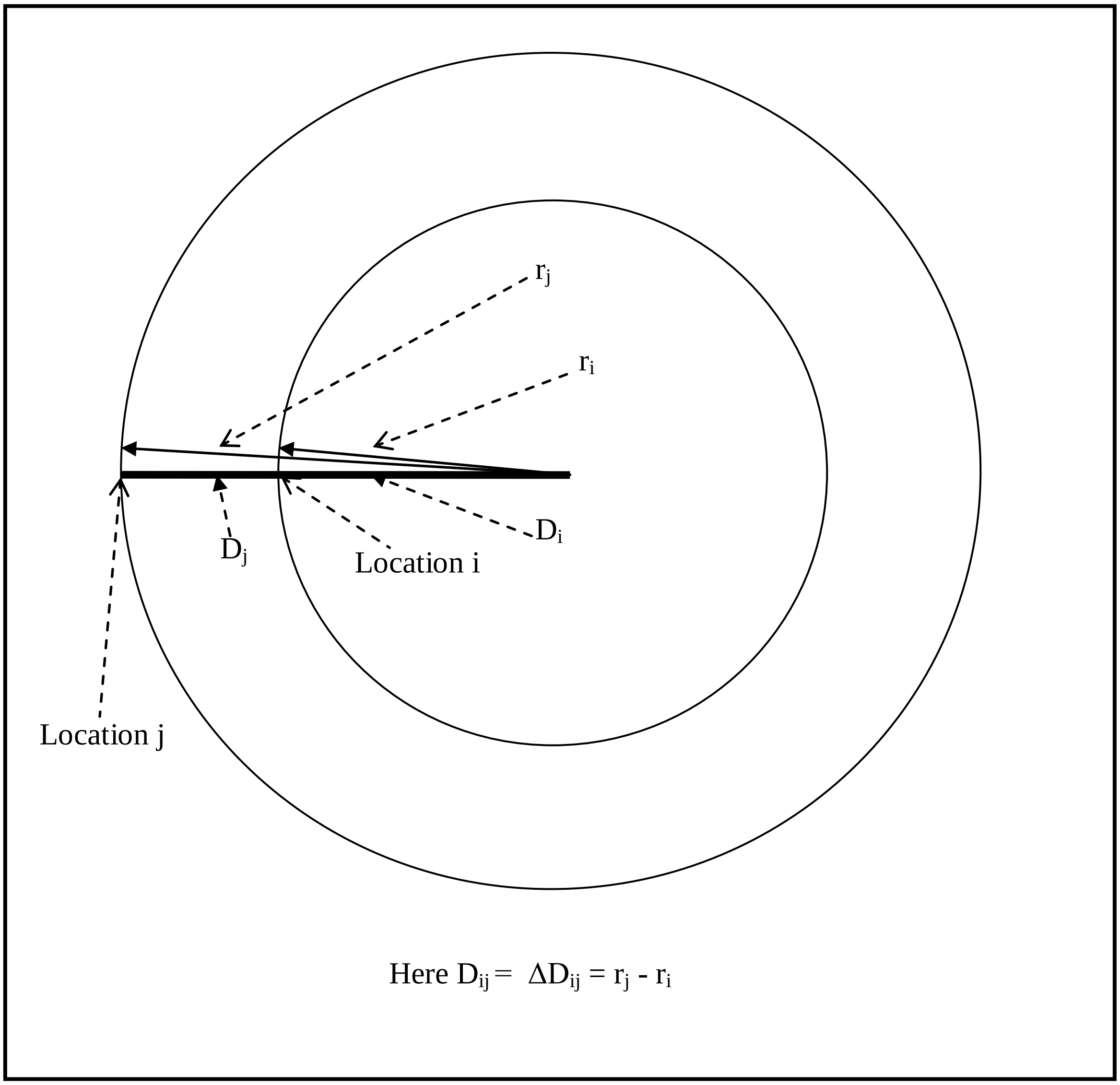
Oranges, fresh: We drop 1991 and 1999 because in those years the U.S. share of Canada's fresh orange market drops below 50 percent. See also the commodity profile prepared by the Agricultural Marketing Resource Center, University of California-Davis, available at http://www.agmrc.org/NR/rdonlyres/E7B0FFE5-3F67-4A4F-A1DA-A9D2AC2594EC/0/califoranges.pdf.

Proof Appendix

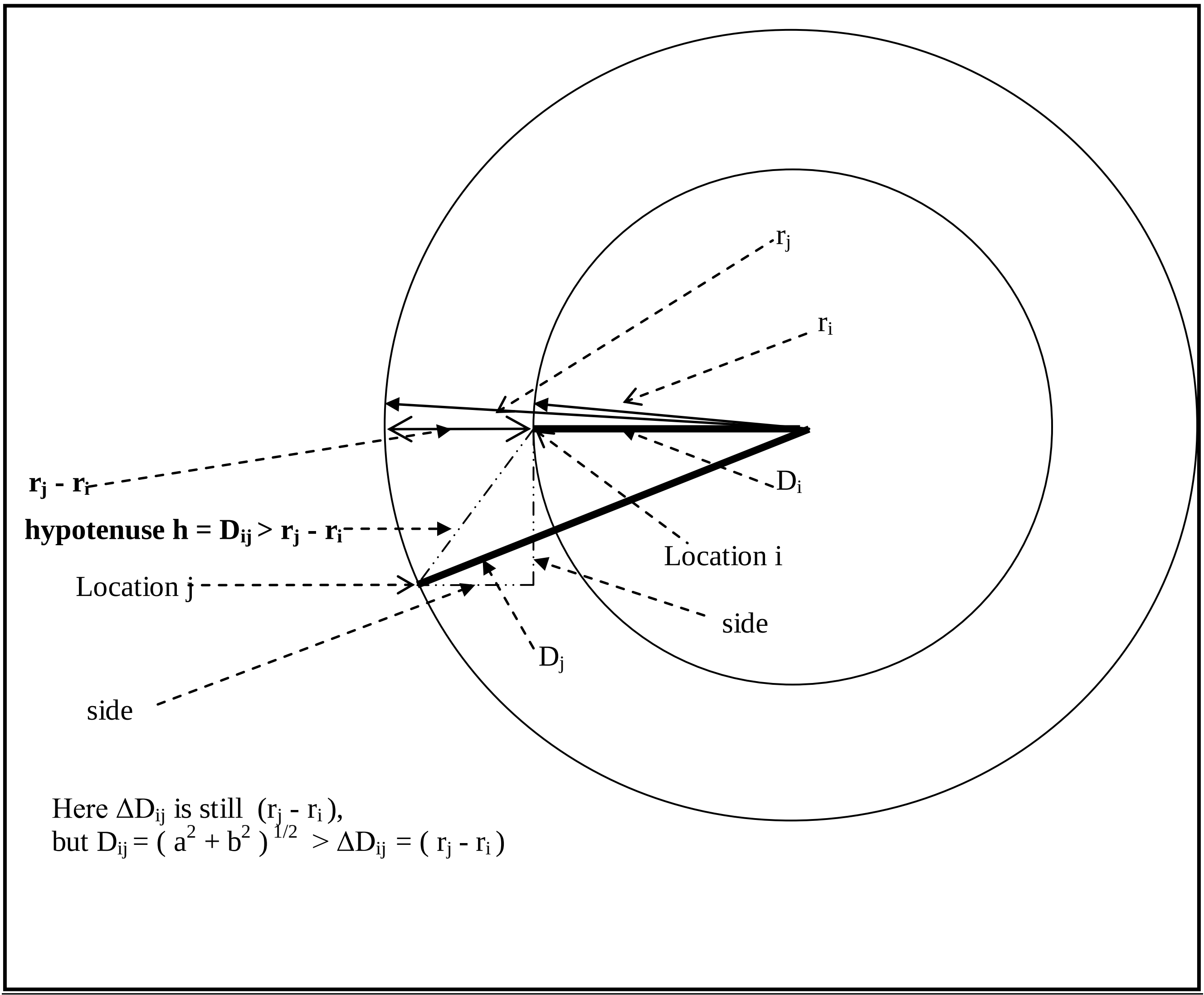
Under the simplifying assumption that differences in production costs at destination locations within the trade cone are relatively small and/or not systematically related to each location's distance from the stem location, the distance coefficient in equation (6) will be systematically underestimated. This is because the (improperly included) distance between a given pair of destinations, D_{ij}, must be at least as large as the difference of their distances from the center of the trade cone, ΔD_{ij}.

D_{ij} (the simple Pythagorean distance between two points) will equal ΔD_{ij} (the difference in the length of the radii from the cone stem to the two locations) when the two locations lie along a single ray emanating outward from the stem (Appendix Figure 1). In all other cases, by simple Pythagorean algebra, D_{ij} exceeds ΔD_{ij} (Appendix Figure 2). Thus a given difference in destination prices is being attributed to an exaggerated measure of distance, reducing the size of the imputed distance coefficient.

Appendix Figure 1. Comparison of D_{ij} and ΔD_{ij} when i and j are collinear



Appendix Figure 2: Comparison of D_{ij} and ΔD_{ij} when i and j are not collinear



REFERENCES

Anderson, James E. and Eric van Wincoop, "Gravity With Gravitas: A Solution to the Border Puzzle," *American Economic Review* 93:1 (2003), 170-192.

_____, "Trade Costs," *Journal of Economic Literature* 42:3 (2004), 691-751.

Anderson, Michael A. and Stephen L. S. Smith, "Borders and Price Dispersion: New Evidence on Persistent Arbitrage Failures," Mid-West International Economics Meetings, Washington University, St. Louis MO (November 2004).

Asplund, Marcus and Richard Friburg, "The Law of One Price in Scandinavian Duty-Free Stores," *American Economic Review* 91:4 (2001), 1072-83.

Bergin, Paul R. and Reuven Glick, "Global Price Dispersion: Are Prices Converging or Diverging?" *Journal of International Money and Finance* 26 (2007), 706-29.

Berthelon, Matias and Caroline Freund, "On the Conservation of Distance in International Trade," *World Bank Policy Research Working Paper 3293* (2004).

Bradford, Scott and Robert Z. Lawrence, *Has Globalization Gone Far Enough? The Costs of Fragmented Markets* (Washington, D.C.: Institute for International Economics, 2004).

Brun, Jean-Francois, Celine Carrere, Patrick Guillaumont and Jaime de Melo, "Has Distance Died? Evidence from a Panel Gravity Model," *World Bank Economic Review* 19:1 (2005), 99-120.

Crucini, Mario J., Chris I. Telmer, and Marios Zachariadis, "Understanding European Real Exchange Rates," *American Economic Review* 95:3 (2005a.), 724-738.

______, "Price Dispersion: The Role of Borders, Distance and Location," processed (2005b).

Engel, Charles, and John H. Rogers, "European Product Market Integration After the Euro," *Economic Policy* 19:39 (2004).

_____, "How Wide is the Border?" *American Economic Review* 86:5 (1996), 1112-25.

_____ and Shing-Yi Wang, "Revisiting the Border: An Assessment of the Law of One Price Using Very Disaggregated Consumer Price Data," Board of Governors of the Federal Reserve System, International Finance Discussion Papers, Number 777 (2003).

Fan, C. Simon, and Xiangdong Wei, "The Law of One Price: Evidence from the Transitional Economy of China," *Review of Economics and Statistics* 88:4 (2006), 682-697.

Feenstra, Robert C., *Advanced International Trade: Theory and Evidence* (Princeton, NJ: Princeton University Press, 2004).

Hufbauer, Clyde; Erika Wada and Tony Warren, *The Benefits of Price Convergence: Speculative Calculations* (Washington, D. C.: Institute for International Economics, 2002).

Obstfeld, Maurice and Kenneth Rogoff, "The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?" National Bureau of Economic Research Working Paper No. 7777 (2000).

Parsley, David C. and Shang-Jin Wei, "A Prism into the PPP Puzzles: The Micro-Foundations of Big Mac Real Exchange Rates," *Economic Journal* 117 (2007), 1336-56.

_____, "Currency Arrangements and Goods Market Integration: A Price-Based Approach" International Monetary Fund, working paper (2002).

Rogers, John H., "Monetary Union, Price Level Convergence, and Inflation: How Close is Europe to the USA?" *Journal of Monetary Economics*, 54:3 (2007), 786-796.

Rogoff, Kenneth, "The Purchasing Power Puzzle," *Journal of Economic Literature*, 32:2 (1996), 647-668.

Rose, Andrew, "Do We Really Know That the WTO Increases Trade?" *The American Economic Review* 94:1 (2004).

Footnotes

[1] We use the phrase "literature on price dispersion" to encompass studies of purchasing-power parity, product-based real exchange rates, and border effects calculated with price data.

[2] Further examples: Fan and Wei (2006) use monthly data on product prices in China, using all available city locations, to test the law-of-one-price. Rogers (2007) uses all EU cities in the EIU data to measure whether economic integration is leading to price convergence like that in the United States. Bergin and Glick (2007) use all city pair combinations in the EIU data in their survey of global price dispersion.

[3] In price dispersion models, distance accounts not only for transport margins, but for all factors that are distance-related. For example, demand shocks could plausibly be correlated geographically, leading to lower price variance among nearby locations.

[4] Engel and Rogers' semi-log regression (Table 3) has a coefficient on the border (for all products) of .0119, which is 37 percent of the mean standard deviation of price volatility for U.S.-U.S. city pairs of .0321 (Table 2).

[5] Anderson and Smith (2004) analyze OECD price dispersion 1990-2003. They find a large border effect—price differentials are about 12 percent higher when city pairs cross an international border—but their estimated elasticity of price differentials with respect to distance is vanishingly small at around 0.001, which is zero in real economic significance. Similarly, Engel, Rogers and Wang (2003), studying U.S.-Canada price dispersion, find a large border effect of 7.3 percent and a small distance elasticity in the 0.000 to 0.006 range (depending upon the type of good). CTZ (2005b) study the variability of price differentials. Their estimates of the distance elasticity range from 0.006 (for intranational pairs, not significantly different from zero) to 0.038 (for international pairs less than 4400 km apart). These are *de minimis* effects, to say the least. Bergin and Glick (2007) analyze global price dispersion 1990-2005. They report a

border effect of 10 percent and an elasticity of mean absolute price differentials with respect to distance of just 0.019. (All the studies cited in this footnote use EIU data.)

[6] In a gravity model the endogenous variable is the trade flow from region i to region j, and standard exogenous variables include the GDPs in both regions and the distance between regions. Dummy variables typically account for policy regimes. Since distance proxies transport costs, it is expected to bear a negative coefficient. For a comprehensive discussion of the theoretical underpinnings of the gravity model and a survey of empirical results see Feenstra (2004). See Berthelon and Freund (2004), Brun *et al* (2005), or Rose (2004) for applications with typically large distance effects.

[7] This is a simplifying assumption that will be modified in the empirical section. A constant transport margin assumes away changes in transport mode (which may be a function of distance) as well as spatial features that affect transport costs. These include geographical features like mountain ranges and rivers, as well as transport features, like the availability and cost of air and rail transport from particular locations.

[8] As noted earlier, by the "price-dispersion literature," we mean all papers that are based upon location-to-location measures of the variance in prices. One example is the interesting discussion of whether the world economy has potential gains from increasing integration (Bradford and Lawrence 2004; Hufbauer, Wada and Warren 2002). Deviations from PPP are used as a measure of potential welfare increases from integration, and Bradford and Lawrence adjust the consumer prices for transport costs. But to do this right the researchers should first determine where production occurs, and then eliminate transport margins based upon distance from production location. By comparing prices for all country pairs, for all of products in their data, they virtually guarantee that they are improperly comparing locations not on the same trade cone. The use of a non-spatial model creates doubts about the conclusions drawn from the study.

[9] As a special case, if the two locations are each themselves stems, we have $\Delta D_{ij} = 0$ and so $\Delta P_{ij} = C_k - C_m$. That is, the price difference boils down to simply the difference in production costs in the two locations. Further, we can relax the assumption that the unit transport costs are the same on different trade cones. Defining d_k (d_m) to be the unit transport cost along cone k (m), if $d_k \neq d_m$, $\Delta P_{ij} = (C_k - C_m) + (d_k D_i - d_m D_j)$.

[10] The parameter **d** could be allowed to vary across subsets of data—across locations separated by different modes of transportation, for example—to test the assumption that trans-shipment costs are identical across trade cones. A regression intercept could incorporate the average tariff and non-tariff barriers that we have temporarily assumed away.

[11] Such data are difficult to construct. For the United States, production location data would come from the Annual Survey of Manufactures (ASM); commodity flow data would come from the Commodity Flow Survey (CFS), the most recent of which is from 1997 and which uses product-based codes different from the ASM's industry codes. Retail prices by location, though tabulated for the construction of the CPI and other price indices, are generally not available from the Department of Labor.

[12] The U.S. cities are Atlanta, Boston, Chicago, Cleveland, Detroit, Houston, Lexington, Los Angeles, Miami, Minneapolis, New York, Pittsburgh, San Francisco, Seattle, and Washington, D.C. Canadian cities are Calgary, Montreal, Toronto and Vancouver.

[13] So, for example, where Los Angeles is the (approximate) production location for carrots, we compare all prices in the sample to the price in Los Angeles The cost is that we lose the ability to compare prices at non-production locations (e.g. Boston's price to New York's price), but we sidestep the problem of having to know anything about how the short Boston-New York distance is traveled relative to the longer Los Angeles-Boston and Los Angeles-New York distances, nor do we have to assume that shorter distances have the same per-mile freight cost as longer

distances. A more detailed empirical investigation could select goods whose transport mode and actual shipping patterns are known, and allow the model to estimate different distance effects for each mode.

[14] We include in the spatial regressions only years for which we could determine the production location. See the Data Appendix for details.

[15] Full estimation of equation (3) would use the price differential between *any* two destination cities on the left-hand-side while using the ΔD_{ij}, the difference between the cities' distances to the production location, on the right-hand-side. Consider lemons again. Price differentials would be calculated across all possible permutation of the 14 destination cities (i.e. excluding the Los Angeles stem city) and the distance measure would be the difference between the cities' distances to Los Angeles. The Boston-Seattle price differential would on the left-hand-side, and on the right-hand-side would be the Boston-Los Angeles distance minus the Seattle-Los Angeles distance (2605 minus 954 equals 1651). This is a preferred spatial model in that, relative to the simpler version we estimate, it has the advantage of a larger sample size. As noted in the text, however, it also has data demands that are beyond the scope of this paper. For the purpose of the illustrative regressions in this paper, we limit our analysis to a simplified spatial model where one member of each city pair is the "stem" or production location. In this case ΔD_{ij} reduces to the distance between the stem and the destination.

[16] The literature includes a variety of ways to measure price dispersion, including differences from geographic averages (Rogers 2007), the variance of goods' prices over time (Engel and Rogers 2006), and the standard deviation of the empirical distribution of the percentage price distribution (Parsley and Wei 2002), among others. Our purpose here is not to decide between measures; we use the log price ratio as a simple way to compare standard and spatial arbitrage

models. Our measure is used by Engel, Rogers, and Wang (2003) and is similar to the measure used by Bergin and Glick (2007).

[17] Including city and time fixed effects and relative factor prices is essential because our theory models producer prices while the EIU CityData is consumer prices.