Documentos de Trabajo 53

Temporal aggregation and convergence to the law of one price: evidence from micro data

Andres Elberg
Universidad Diego Portales

Febrero, 2014

udp FACULTAD DE ECONOMÍA Y EMPRESA

Temporal Aggregation and Convergence to the Law of One Price: Evidence from Micro Data

Andres Elberg*

Feburary 15, 2014

Abstract

Estimates of speed of convergence towards the Law of One Price (LOP) are potentially afflicted by three sources of bias: temporal aggregation; aggregation across goods; and short samples. I empirically assess the importance of temporal aggregation bias (while accounting for the other two sources of bias) using a novel dataset of weekly-sampled retail prices. I find that temporal aggregation can severely bias estimates of persistence in relative prices. Using quarterly aggregated data can overestimate the half-life of deviations from the LOP by a factor exceeding 2. In contrast, I do not find evidence that aggregation across goods biases persistence estimates.

Keywords: Law of One Price, temporal aggregation, micro data

*Universidad Diego Portales, Chile (email: andres.elberg@udp.cl). This paper is based on the first chapter of my Ph.D. dissertation at the University of California, Berkeley. I thank Barry Eichengreen, Yuriy Gorodnichenko, Pierre-Olivier Gourinchas, Michael Jansson Chad Jones and especially Andy Rose and Maury Obstfeld for advice, support and encouragement. David Gelves provided excellent research assistance at an early stage of this project. All remaining errors are my own.

1 Introduction

The Law of One Price (LOP) is a fundamental arbitrage condition in international macroeconomics and finance. In its absolute version, it states that, absent trade barriers, an identical good should sell for the same common-currency price in spatially separated markets.[1] The LOP is important for a number of reasons: It serves as an operational definition of an integrated market (Goldberg and Knetter (1997)) and underlies Purchasing Power Parity (PPP)[2], which is a key building block in a number of workhorse models of international macroeconomics and finance.

In spite of the central role it plays in theoretical models, the LOP has been systematically rejected by the data. An extensive body of empirical work, dating back to Isard (1977) and Richardson (1978), has documented that deviations from the LOP are large, volatile and persistent. Flagrant violations of the LOP have been documented even for highly tradable goods and across cities within a given country (i.e., controlling for exchange rate and tariff fluctuations).

A drawback shared by most, if not all, empirical studies of the LOP is their reliance on inadequate data. Appropriate tests of the LOP require by definition that goods whose prices are being compared be exactly identical across locations. In contrast, prices used by most LOP studies are typically averages at the product-category or sectoral level (Broda and Weinstein (2008)). This type of cross-sectional aggregation can lead both to detecting spurious deviations from the LOP (Broda and Weinstein (2008)) and to overestimating the half-lives of deviations from the LOP (Imbs, Mumtaz, Ravn and Rey (2005)).

[1] The *relative* LOP states that percentage changes in the price of an identical good should be equal across markets.

[2] Purchasing Power Parity (PPP) states that common-currency price *levels* should be equalized across countries or regions.

A less studied but potentially serious problem in estimating persistence of LOP deviations is temporal aggregation.[3] Studies of LOP deviations typically use prices that are observed at lower frequencies than the frequency at which prices are actually generated. In the words of Taylor (2001), "*Typically PPP and LOP have been tested with aggregate data at annual, quarterly, and sometimes monthly frequencies... Sampling the data at low frequencies will never allow one to identify a high-frequency adjustment process*" (p. 474). Furthermore, a number of LOP and PPP studies use temporally averaged prices.[4] Taylor (2001) shows both analytically and using simulation experiments that estimates of persistence based on temporally averaged prices are upwardly biased.

The present paper studies convergence to the LOP using a novel dataset containing retail prices that are not subject, in principle, to either of the above types of aggregation. The data include store-level prices for a set of 40 narrowly defined products sampled weekly from outlets located across 11 urban centers[5] in Mexico. Prices in the sample are surveyed from about 2000 outlets over the 2001-2011 period. The dataset includes both perishable (e.g., "1 kg. of ham San Rafael, 16 percent protein") and non-perishable (e.g., "355 ml. can of Coca-Cola") products which are primarily foodstuffs and health and beauty products.

Two distinctive features of the data are worth emphasizing. First, they include prices that are sampled at the same frequency at which they are generated (assuming, as the evidence suggests, that retail prices for the type of goods included in the data

[3]Temporal aggregation refers to aggregation in which a low frequency time series (e.g., quarterly) is derived from a high frequency time series (e.g., weekly).

[4]Taylor (2001) notes that price levels reported by the International Monetary Fund's (IMF) *International Financial Statistics*, which are in turn based on prices provided by national statistical agencies, are time averages. The use of temporally averaged prices is not limited to those studies using aggregate price indices. Broda and Weinstein (2008) who study deviations from the LOP in the US and Canada using scanner data for narrowly defined products also use temporally (quarterly) averaged prices.

[5]Urban centers correspond to 10 cities and Tlaxcala, the smallest state in Mexico.

are set on a weekly basis).[6] Second, they include prices for exactly identical products across locations. Both of these features are rare in studies examining deviations from the LOP. To the best of my knowledge, this is the first study in the international macro literature that analyzes convergence to the LOP using weekly sampled prices for exactly identical products across locations.

Exploiting the fact that the data are highly disaggregated both on their temporal and cross-sectional dimensions, I examine the severity of different aggregation biases. Specifically, I assess the extent of temporal aggregation bias by replicating estimations of convergence speed using prices averaged at monthly and quarterly frequencies. Similarly, I assess the extent of cross-sectional aggregation by comparing the estimates of persistence based on the average price of a basket of products with the average convergence speed of individual products included in the basket.

The analysis suggests that short-run deviations from the LOP across cities are small compared to those reported in the literature. LOP deviations are about half as large in magnitude as those reported by Parsley and Wei (1996) in their pioneering study of LOP deviations across US cities, and about half as volatile as those reported by Broda and Weinstein (2008) in their study of LOP deviations across US cities and Canadian provinces. Consistent with an inverse relationship between tradability and LOP deviations I find that deviations from the LOP are substantially larger in the case of perishable products.[7] However, unlike most prior literature, I do not find evidence of a strong systematic relationship between geographical distance and LOP deviations. Pooling across all products, the relationship between LOP deviations and geographical distance appears to be weak. It is found to be stronger for nonperishable than for perishable products.

[6] Evidence from the empirical macroeconomics and industrial organization literatures (e.g., Eichenbaum, Jaimovich and Rebelo (2011), Midrigan (2011) suggests that retail prices tend to be set weekly.

[7] This is in line with previous results reported by Parsley and Wei (1996).

The evidence on relative city price dynamics is consistent with fast convergence to the LOP. Panel unit root tests, conducted at the product level, overwhelmingly reject nonstationarity in price gaps. Interestingly, deviations from the LOP appear to converge at a remarkably rapid rate. The average half-life of deviations from the LOP is estimated at 5.6 weeks. Estimated half-lives are substantially –even orders of magnitude– smaller than half-lives previously reported in the literature. For instance, Parsley and Wei (1996) find an average half-life of LOP deviations across US cities in the range of 4-5 quarters and Broda and Weinstein (2008) find an average half-life of LOP deviations between Canadian provinces of 2.9 quarters. Half-life estimates appear to be robust to small sample bias and to measurement error, both of which may induce an underestimation of convergence speed.

Estimated half-lives are of the same order of magnitude as the implied duration of price spells in the micro data. The median implied duration of a price spell equals 6.1 weeks. Hence half-lives of deviations from the LOP appear to be consistent with the extent of price rigidities in the data. This fact is in stark contrast to the difficulty in reconciling large half-lives with relatively high frequencies of price change in (cross-sectionally) aggregate data –a phenomenon dubbed the "PPP-puzzle" by Rogoff (1996).

Turning to the consequences of estimating persistence using aggregate data, the results show that temporally aggregating the data can cause a severe bias in the estimates of persistence in price gaps. Estimating the rate of convergence to the LOP using monthly and quarterly averaged prices causes one to overestimate the half-life of a deviation from the LOP by a factor of 1.9 and 2.6, respectively. The results of Monte Carlo experiments confirm these findings. In contrast, I find no evidence that pooling across individual goods featuring heterogeneous dynamics causes an overestimation of persistence.

This paper is primarily related to a long-standing literature studying the magnitude and persistence of deviations from the LOP. It is primarily related to a strand of the literature studying intra-national convergence to the LOP. The seminal article in this strand of the literature is Parsley and Wei (1996), who study convergence to the LOP across US cities. More recently, Ceglowski (2003), Fan and Wei (2006), Broda and Weinstein (2008) and Crucini and Shintani (2008) provide further evidence on LOP deviations across cities or regions within countries.[8] This paper is also closely related to a strand of the literature that study the consequences of aggregation on LOP and PPP deviations. Imbs et al. (2005) present evidence consistent with aggregation across individual sectoral prices causing estimates of persistence to be upwardly biased. Chen and Engel (2005), Crucini and Shintani (2008) and Broda and Weinstein (2008) do not find evidence of this type of cross-sectional aggregation bias. The seminal article studying the effects of temporal aggregation on estimates of convergence to the LOP and PPP is Taylor (2001). Chambers (2005) provides further theoretical results on this issue.

The paper makes two contributions to the literature. First, it assesses the magnitude of temporal aggregation bias in persistence estimates using actual data and assuming a general autoregressive data generating process. Prior results on temporal aggregation have been obtained either analytically or using simulation and rely on the assumption that price gaps follow an autoregressive process of order one (Taylor (2001), Chambers (2005)). Second, it estimates convergence to the absolute version of the LOP using prices of narrowly defined products observed at the weekly frequency (i.e., prices that are cross-sectionally and temporally disaggregated). Data limitations have constrained previous work to focus on convergence to the relative version of the LOP or

[8] Ceglowski (2003) studies convergence to the LOP for 45 consumer goods across 25 Canadian cities. Fan and Wei (2006) investigate price convergence across 36 cities in mainland China for 93 products and services. Broda and Weinstein (2008) use massive amounts of scanner data to study convergence to the LOP across 10 cities in the US and 6 provinces in Canada. Crucini and Shintani (2008) study convergence to the LOP using a panel of 257 goods prices drawn from 16 major US cities.

to estimate convergence using temporally aggregate prices. In addition, it shows that the speed of convergence in highly disaggregated relative city prices is consistent with the extent of nominal rigidities observed in the micro data.

The remainder of the paper is structured as follows. Section 2 describes the data used in the analysis. Section 3 analyzes short-run deviations from the LOP. Section 4 studies whether the LOP holds in the long run. It presents the results of panel unit root tests and estimates of half-lives of deviations from the LOP. Sections 5 and 6 study the consequences of aggregation. Section 4 studies the effects of estimating convergence using low-frequency data. Section 5 examines the effects of cross-sectional aggregation for the estimates of speed of convergence. Section 7 concludes.

2 Description of the Data

The dataset contains store-level prices for 40 products surveyed across several cities in Mexico over the 2001-2011 period. Products included in the sample are narrowly defined (e.g., "Can of 355 ml. of Coca-Cola") and can be classified in the foodstuffs and healthcare and beauty product groups. The data include both perishable (e.g., fruits and vegetables) and nonperishable (e.g., laundry detergent, instant coffee, soft drinks) products. Table 1 presents a full description of the products included in the dataset.

The data were collected by Mexico's Federal Agency for Consumer Protection (Profeco, for Spanish Procuraduría Federal del Consumidor). Profeco regularly surveys prices for a large number of products and stores across several cities in Mexico. The data are then processed and used to provide consumers with comparative information on the stores charging maximum and minimum prices for a product in a given city.[9]

[9]Profeco publishes the information on stores charging maximum and minimum prices on the following website: http://www.profeco.gob.mx/precios/canasta/default.aspx

The data collection process is performed by Profeco's own personnel who visit stores on a semiweekly to weekly basis. Prices are entered by price collectors in electronic handheld devices and later sent to Profeco's headquarters where they are checked for consistency. The accuracy of the collected data is further checked by price inspectors who visit a sample of stores to ensure there are no errors in the prices collected by price collectors.[10]

The dataset in this paper includes prices surveyed from 1,914 outlets located in 11 urban centers in Mexico. Outlets include large supermarket chains such as Walmart and Carrefour as well as independent stores, open-air-markets, and drugstores. Urban centers include 10 cities[11] and the state of Tlaxcala (the smallest state in Mexico). In what follows I refer to all urban centers as "cities" for expositional convenience. Figure 1 presents a map showing the location of the cities included in the data. Profeco surveys a larger number of products, cities and stores than the ones included in the dataset. A major criterion for selecting the specific set of products, stores and cities included in the data was ensuring that long price trajectories (spanning several years) were available for a given product across several stores and cities in Mexico. Having access to long series of price gaps across cities is crucial for the estimation of convergence to the LOP. Estimates of the speed of convergence towards the LOP is typically performed using an autoregressive specification and it is well known that standard estimates of autoregressive models may be afflicted by a small sample bias. Prices are observed weekly over the period 2002-2011, for a maximum of 526 weeks.

I made several adjustments to the raw data. First, in some instances, the price of a product was surveyed more than once in a given outlet in a given week. In these cases, I chose to keep the last surveyed price in the week. Second, when (at most

[10] I thank Mr. Eliacin Luciano Ocampo for his explanation of Profeco's price collection procedure.

[11] The cities included in the data are the following: Guadalajara, Merida, Mexico City, Monterrey, Morelia, Oaxaca, Puebla, Toluca, Villahermosa and Zacatecas.

two) missing observations were preceded and followed by the same price, I imputed the missing observations using that price. Third, to preserve a base of outlets consistently followed by Profeco, I chose to keep only outlets that remain in the sample for a period of at least one year and outlet/product pairs that have at least one unbroken spell of 4 weeks of non-missing observations. Outlets remain in the sample between 1 and 526 weeks with a median of 474 weeks. A systematic bias due to sample attrition appears to be unlikely because, according to Profeco, the phenomenon of outlets exiting the sample is typically associated to budgetary constraints faced by the federal agency, which are unlikely to be systematically related to the prices set by individual stores. Overall, these adjustments reduced the total number of outlets in the sample from 1914 to 1220.

Fourth, I checked for the presence of outliers by identifying prices that lie outside a band of +/- 3 standard deviations from a 13-week moving average of the price of a product in a given outlet centered in the current week. Only about 0.1 percent of all prices lie outside this bands. I also identified cases in which a price change resulting in an extreme observation is fully reversed the following week. These cases amount to about 0.002 percent of the sample and correspond to 983 observations. I visually inspected these cases and manually corrected prices that appeared to be incorrectly recorded.

Fifth, I corrected for price changes that are transitory in nature. Transitory price changes (coupled with strategic complementarities in pricing) can yield upward biased estimates of the speed of convergence towards the LOP. I define temporary prices using a filter that identifies a temporary price change as one in which the price moves away from the preceding price for a period lasting between one and four weeks before returning to the original price. The filter is two-sided. It identifies both temporary price decreases (i.e., "sales") and temporary price increases (i.e., "spikes"). About 6 percent of all prices in the data correspond to sales and 2 percent correspond to spikes.[12] I replaced temporary prices with the price immediately preceding (or following) the temporary price spell.

Finally, in order to keep a balanced panel, I restrict the analysis to a period of 370 weeks for which prices are available for all 40 products across all 11 cities.

Table 2 presents summary statistics describing the final dataset. It shows the median and standard deviation of prices as well as the frequency of price change for each product.[13] The median price varies between MXN (Mexican pesos) 0.9 (white bread) and MXN 87.4 (ham San Rafael) with a median of MXN 17.4 (about USD 1.56 at the average MXN-USD exchange rate over the period of analysis). Price dispersion across outlets and cities is larger (both in monetary units and relative to the mean) for perishable products than for nonperishable products, as one would expect given that product perishability is likely to partially inhibit the operation of arbitrage. The median frequencies of price change at the product variety level vary between 0.02 and 0.54 with a median of 0.16. This implies that the median duration of a price spell is about 6.1 weeks. Frequencies of price change are in line with evidence reported in the price setting literature. Eichenbaum, Jaimovich and Rebelo (2011), for instance, report that prices at a large US supermarket chain are changed about every two or three weeks. Elberg (2014) uses scanner data for major retail chains in Chile and finds that (posted) prices are changed every 5-6 weeks.

[12] The fraction of sales identified using the price filter should not be taken as an unbiased estimate of the fraction of sales in the population of stores as Profeco aims at collecting regular (as opposed to posted) prices.

[13] The figures in Table 2 are obtained as the median and standard deviation of the median log price where the latter median is taken across products and stores for a given city and week.

3 Short-Run Deviations from the Law of One Price

3.1 Analytical Framework

The absolute version of the law of one price states that, in the absence of trade barriers, the price of an identical good should be the same across locations. Letting p_{ht}^k $(h = i, j)$ denote the price of good k in city h at time t, the absolute LOP states that

$$p_{it}^k = p_{jt}^k$$

The mechanism ensuring that this relationship holds is arbitrage. In a frictionless world, any discrepancy between the price of the good in different locations should be eliminated by the action of traders who buy the good where its price is relatively low and sell it where the price is relatively high. In the case of consumer products, arbitrage can in principle occur both at the wholesale and retail levels. However, retail arbitrage is unlikely to be relevant for LOP deviations. Transportation costs and institutional constraints severely limit the ability of consumers to arbitrage across locations. Arbitrage activity conducive to reducing price gaps across locations is more likely to take place at the wholesale level, where larger volumes of the good are traded.

To see how wholesale price arbitrage works to limit retail price discrepancies between locations consider the following setting. Suppose that good k available in cities i and j is sourced from cities l and l', respectively. All cities are located in the same country. An iceberg cost equal to τ_{nm} $(\tau_{nm} \in (0, 1])$ must be incurred for transporting the good between locations n and m. Letting c_{lt}^k denote the production cost of good k in city l at time t (which could be inclusive of a manufacturer markup) and ξ_i^k denote the retail profit margin in city i, the retail price of good k in city i can be written as follows:

$$p_{it}^{k} = \xi_i^k \, \frac{c_{lt}^{k}}{\tau_{il}}$$

and the ratio between retail prices in the two cities is given by

$$\frac{p_{it}^{k}}{p_{jt}^{k}} = \frac{\xi_i^k \, \tau_{jl} c_{lt}^{k}}{\xi_j^k \, \tau_{il'} c_{l't}^{k}}$$

Assuming that no institutional constraints preclude wholesale arbitrage from taking place, the relative wholesale price between locations i and j should lie within a band defined by transportation costs

$$\tau_{ij} \leq \frac{w_{it}^{k}}{w_{jt}^{k}} \leq \frac{1}{\tau_{ij}}$$

where $w_{ht}^{k} = \frac{c_{ht}^{k}}{\tau_{hl}}$. This, at its turn, implies that the relative retail price can fluctuate within a band which depends on both transportation costs and relative retail profit margins,

$$\frac{\xi_i}{\xi_j} \tau_{ij} \leq \frac{p_{it}^{k}}{p_{jt}^{k}} \leq \frac{\xi_i}{\xi_j} \frac{1}{\tau_{ij}}$$

Thus, in a world characterized by no transportation costs (i.e., $\tau_{ij} = 1$) and where retail profit margins do not vary across locations, wholesale price arbitrage would ensure that retail prices are the same in different locations (i.e., that the absolute LOP at the retail level holds). More generally, under wholesale arbitrage, retail price discrepancies across locations can be attributed to transportation costs (and other trade barriers) and to differential retail profit margins across locations.

3.2 Evidence

Following the international finance literature, I focus on price differentials across cities (as opposed to price differentials across stores).[14] I measure the citywide price of good k as the simple average of store-level prices of good k within a city. According to International Labor Organization (2004), national statistical agencies most commonly use the simple average price across outlets to compute citywide prices. In addition, citywide prices used in the international finance literature are typically simple averages of outlet-level prices within cities (e.g., Parsley and Wei (1996), Crucini, Telmer and Zachariadis (2005), Crucini, Shintani and Tsuruga (2010)).[15]

The choice of the simple average as a citywide price rests on the assumption that prices collected by Mexico's consumer protection agency are representative of the amount of price dispersion observed within cities. It seems reasonable to assume that the set of surveyed outlets in the data is representative of the distribution of prices in a city. This is because Profeco's goal in collecting prices is to provide consumers in a city with useful information on the prices charged on a given product. This presumption is also supported by the fact that all major supermarket chains (e.g., Walmart, Carrefour, Auchan, Aurrera, Comercial Mexicana), together with other types of stores, are included in the sample.

The (log) price differential for good k in week t between cities i and j is defined as

[14] In the international finance literature, locations are typically cities which can be located in the same or in different countries. Although my data include prices at the more disaggregated outlet level, the analysis focuses on citywide price differentials. There are two reasons that justify this choice. First, a major goal of the paper is to provide estimates of the biases incurred when estimating convergence to the LOP using aggregate data. In this sense, it is convenient to follow previous work in analyzing citywide price differentials. Second, some stores in the data are surveyed for relatively short periods of time; this makes store-level data unsuitable for convergence analysis, which, given the high frequency at which prices are observed, requires using long price series.

[15] An exception is Ceglowski (2003). City-average prices are in this case "weighted means of the prices from the individual stores, where the weights represent overall market share of each store" (p. 375).

$$q_{ij,t}^k = \ln\left[\frac{1}{S_i}\sum_{s=1}^{S_i} p_{ist}^k\right] - \ln\left[\frac{1}{S_j}\sum_{s=1}^{S_j} p_{jst}^k\right] \tag{1}$$

where P_{hst}^k $(h = i, j)$ is the price of good k, sold in store s, in city h, in week t and S_h is the total number of stores in city h.[16,17] Table 3 presents summary statistics on the magnitude and dispersion of price differentials across all 55 $(= 11 \times 10/2)$ city-pairs in the sample for both perishable and nonperishable products. Specifically, it shows the median, mean absolute deviation, and standard deviation of the price differential $q_{ij,t}^k$ taken across city-pairs for a given product and week. For each one of these measures, the table presents its mean, median, and standard deviation taken across products and weeks.

Deviations from the LOP in the data appear to be small and exhibit low volatility compared to LOP deviations reported in the literature. The mean price gap between two cities in the sample is 7.5 percent and the average cross-sectional standard deviation of intercity price gaps is 9.2 percent. By way of comparison, Parsley and Wei (1996) report a median price gap of 14.4 percent for perishables and 12.5 percent for nonperishable goods across US cities. Broda and Weinstein (2008) report a median standard deviation of LOP deviations equal to 22.3 percent for US cities and 18.7 percent for Canadian regions; Crucini and Shintani (2008) report a standard deviation of LOP deviations on the order of 60 percent internationally and 25 percent across US cities.

Consistent with the view that the LOP holds better the lower trade costs are,

[16] This notation assumes that the number of stores per city is constant over time. This is only for expositional simplicity. In the data, there is variation in the number of stores surveyed in a city over time.

[17] There are some weeks in which no prices of a product were surveyed in a given city. The fraction of missing observations per city is small, though: it ranges between 0.2 percent and 2.2 percent, with the sole exception of Toluca for which the fraction of missing observations is 6.7 percent. I imputed missing observations using a linear interpolation.

deviations from the LOP are less pronounced in the case of nonperishable products. The mean absolute deviation of price differentials for perishable products is, on average, equal to 10.3 percent, which is about twice the average mean absolute deviation for non-perishable products (5.2 percent). However, unlike prior literature, the data do not support a systematic relationship between LOP deviations and transportation costs as proxied by physical distance.

Prior work has found that LOP deviations are strongly related to the distance separating two cities (e.g., Parsley and Wei (1996), Ceglowski (2003)). Broda and Weinstein (2008) suggest that the strength of the relationship between price gaps and distance may be due to cross-sectional aggregation (i.e., the use of average prices at the product-category level). They find that "distance coefficients are five to ten times larger in aggregate data than in micro data" (p. 2). I use the following standard specification to study the relationship between distance and price gaps:

$$\left|q_{ij}^{k}\right| = \alpha + \beta \log(dist_{ij}) + \xi_k + \varepsilon_{ij} \tag{2}$$

where $\left|q_{ij}^{k}\right|$ is the average absolute log price deviation of product k between cities i and j; $\log(dist_{ij})$ is the natural logarithm of physical distance between city i and city j; ξ_k are product fixed-effects, and ε_{ij} is an error term. Physical distance is measured using the greater circle method.[18] Table 4 presents the results of OLS estimation of equation (2). The results suggest that, as expected, physical distance is positively correlated to LOP deviations. The estimated distance coefficient, using all products in the dataset, equals 0.0024 and is marginally statistically significant (its associated p-value is equal to 0.061). The effect of distance on LOP deviations is small in magnitude compared to previous estimates in the literature. Using similar specifications to (2), Parsley and

[18] The average distance separating two cities in the sample is 379 miles and the largest distance is 885 miles (between Merida and Guadalajara).

Wei (1996) estimate a distance coefficient equal to 0.018 for nonperishables and 0.011 for perishables. Broda and Weinstein (2008) report estimated distance coefficients of 0.0068 and 0.0213 using barcode data for the US and Canada, respectively.

The relationship between physical distance and LOP deviations appears to be stronger for nonperishable products. Estimated distance coefficients are about twice as large for nonperishables than for perishable products (0.0031 vs. 0.0015). In addition, distance is highly statistically significant in the case of nonperishable products but not statistically significant at any conventional level in the case of perishables. The results suggest that transportation costs do not help explain price gaps in the case of less tradable perishable products.

4 The Law of One Price in the Long Run

This section studies convergence to the LOP in the long run. It proceeds in two stages. First, it tests whether shocks to the LOP tend to dissipate over time. Second, finding that in most cases shocks to the LOP tend to disappear over time it estimates the speed of convergence towards the LOP. For interpretational convenience, and following much of the literature, I focus on price gaps relative to a benchmark city –Mexico City. The assumed data generating process for price gaps of a given product k consists of the following SUR system of N (equal to the number of city-pairs) autoregressive processes of order p_i:

$$\Delta q_{i,t}^{k} = \alpha_{ik} + \beta_{ik} q_{i,t-1}^{k} + \sum_{\tau=1}^{p_{ik}-1} \gamma_{ik\tau} \Delta q_{i,t-\tau}^{k} + \varepsilon_{i,t}^{k} \quad , \quad t = 1, ..., T, \ i = 1, ..., N \qquad (3)$$

where $\Delta q_{i,t}^{k}$ is the first-difference of the price differential of product k between city i

and the benchmark city, α_{ik} is a city-pair and product specific constant term and $\varepsilon_{i,t}^{k}$ is an i.i.d. error term with mean zero and constant variance σ_{ε}^{2}. In addition, it is assumed that $Cov\left(\varepsilon_{i,t}^{k}, \varepsilon_{j,t}^{k}\right) = \sigma_{ij}^{k} \in \mathbb{R}$. That is, the model allows for nonzero contemporaneous correlation between shocks to relative prices from different cities. Allowing for cross-sectional correlation in a context in which N is small is especially important in unit root testing. O'Connell (1998) shows that failing to account for cross-sectional correlation might lead to severe size distortions in tests for unit roots. In this formulation (the so-called augmented Dickey-Fuller formulation), the coefficient β_{ik} is equal to the sum of the p_i autoregressive coefficients minus one.[19]

4.1 Panel Unit Root Tests

As is standard in the literature, I assess long-run convergence to the LOP using panel unit root tests. I test for the presence of unit roots in each of the 40 panels of dimensions $N = 10$ and $T = 370$, each one defined for a given product. Here, N stands for the number of city-pairs and T stands for the number of periods, measured in weeks. I use two panel unit root tests whose asymptotics rely on a finite N and $T \to \infty$. The first test is an extension of Choi's (2001) inverse-normal combination test developed by Demetrescu, Hassler and Tarcolea (2006). Demetrescu et al. extend Choi's test to the case in which p-values of individual unit root tests are cross-sectionally correlated. The second test is the multivariate augmented Dickey and Fuller (MADF) test proposed in Sarno and Taylor (1998) and Taylor and Sarno (1998). In both tests, the truncation lag in each equation, p_i, is chosen using the Modified Akaike Information Criterion,

[19] If the original AR(p) model is written as

$$q_{ik,t} = \alpha_{ik} + \rho_{1ik} q_{ik,t-1} + \rho_{2ik} q_{ik,t-2} + \ldots + \rho_{p_i ik} q_{ik,t-p_i} + \varepsilon_{ik,t}$$

then $\beta_{ik} = \rho_{1ik} + \rho_{2ik} + ... + \rho_{p_i ik} - 1$ (see Hamilton 1994 p. 517).

MAIC (Ng and Perron, 2001).[20]

Both types of panel unit root tests overwhelmingly reject the null hypothesis that the processes of price deviations for all city pairs possess a unit root. Table 5 presents the results of performing the inverse-normal combination test. The null hypothesis of a unit root is rejected in 35 out of 40 products at the 5 percent level of significance. The unit root hypothesis cannot be rejected at any conventional significance level in the case of two products: "Pepsi soda" and "Corn Tortilla". Institutional and regulatory constraints may explain non-convergence in the case of these two products. It is well known, for instance, that the price of corn is heavily regulated in Mexico. Using the MADF test, the unit root hypothesis is rejected in all 40 cases.[21]

One feature of panel unit root tests is that they can reject the unit root hypothesis even when a single series within the panel is stationary. To explore whether the rejection of nonstationarity is driven by only a few stationary relative city prices within each panel, I test for the presence of a unit root in each series separately using a less powerful univariate ADF test. As above, the number of lags included in the ADF regression is selected according to the MAIC.[22] I find that the test rejects a unit root at the 5 percent significance level for 62 percent of the total series. This is a relatively high rejection rate compared to the results reported in the literature. For instance, Fan and Wei (2006) reject the unit root hypothesis in 40 percent of the cases in their study of inter-city price differentials in China and Ceglowski (2003) reports an average rejection rate of 45 percent of the inter-city prices in Canada. Thus, the results of the univariate unit root tests tend to confirm the conclusion derived from the panel unit root tests. There is substantial evidence favoring the hypothesis of convergence towards the law of one price.

[20] I limit the search for the optimum number of lags to a maximum of 26 lags.

[21] Results of the MADF test are available from the author upon request.

[22] Similar results are obtained when I use a general-to-specific procedure instead to finding truncation lags.

4.2 Estimates of Speed of Adjustment towards the LOP

Having found supportive evidence for convergence to the LOP, I turn to estimating the speed at which relative prices across cities tend to converge. I focus on the the half-life of a deviation from the LOP, HL, as a scalar measure of the persistence of shocks to the LOP. In autoregressive models of order one, AR(1), the half-life is given by $HL = \ln(0.5)/\ln(\widehat{\rho})$ where $\widehat{\rho}$ is an estimate of the (first-order) autoregressive coefficient. In more general models, as the one estimated in this paper, the previous expression serves only as an approximation of the true half-life because convergence does not necessarily occur at a constant rate. In such cases, the half-life can be estimated as the largest h such that $\widehat{IRF}(h-1) > 0.5$ and $\widehat{IRF}(h) \leq 0.5$, where $\widehat{IRF}(h)$ is the estimated impulse- response function in period h. In what follows, I report the half-lives using the estimated impulse response functions. Estimates of the impulse response functions are obtained from SUR-GLS estimates of system (3).

Figure 2 presents the distribution of estimated half-lives pooling across all products and city-pairs. The mean and median half-lives, pooling across products and city-pairs are equal to 6.5 weeks and 4.3 weeks, respectively. Column 1 in Table 6 presents median half-lives (across city-pairs) at the product level. Half-lives range between 1.8 weeks ("Jitomate Bola de Sinaloa") and 21.9 weeks ("Pepsi soda"). The median and average products present half-lives of 4.2 weeks and 5.6 weeks, respectively. On average, half-lives are only slightly larger in the case of nonperishable products (6 weeks versus 5 weeks).

These results point to an extraordinarily high speed of convergence toward the LOP. By way of comparison, Parsley and Wei (1996) find a half-life of deviations from the LOP between US cities on the order of 4-5 quarters and Broda and Weinstein (2008) estimate a half-life of deviations from the LOP between Canadian provinces of 2.9

quarters. Estimated half-lives are closer to the findings of Fan and Wei (2006), who report a median half-life of 2.4 months for intranational deviations from the LOP in China using monthly data.

It is worth noting that estimated half-lives of deviations from the LOP are of the same order of magnitude as the durations of price spells reported in Section 1. Both the average half-life of deviations from the LOP and the median duration of a price spell are about 6 weeks. Thus, half-lives of deviations from the LOP appear to be consistent with the extent of price rigidities in the data. This fact is in stark contrast to the difficulty in reconciling large half-lives with relatively high frequencies of price change in (cross-sectionally) aggregate data –a phenomenon dubbed the "PPP-puzzle" by Rogoff (1996).

Column 2 in Table 6 presents 95 percent confidence intervals for half-lives. To construct confidence intervals for half-lives I rely on Kilian's (1998) bootstrap-after-bootstrap procedure.[23] Kilian's (1998) procedure improves on the traditional bootstrap estimator by accounting for the small-sample bias and skewness of the small sample distribution of the impulse response function estimator. Pesavento and Rossi (2007) find that the confidence intervals based on the bootstrap-after-bootstrap procedure have good coverage properties as long as the process' largest root is less than one.[24] Confidence intervals are relatively tight. The 95 percent confidence interval for the average perishable (nonperishable) product lies between 3.7 (2.6) weeks and 13.1 (9.1) weeks.

[23]See Kilian (1998) p. 220 for details on the implementation of the bootstrap-after-bootstrap estimator. I use 1,000 and 2,000 bootstrap loops in the first stage and second stage of the procedure, respectively.

[24]Pesavento and Rossi (2007) find that Kilian's (1998) procedure works well in terms of overall coverage as long as the largest root is less than or equal to 0.95 (if the largest root is around 0.97, they find coverage to be still approximately correct as long as there is not much additional serial correlation).

4.3 Robustness Checks

This subsection examines possible biases affecting the persistence estimates presented above. It focuses on two prominent sources of bias: small sample bias and measurement error.

Small Sample Bias. One potential problem with obtaining half-lives using the SUR-GLS estimates from system (3) is that OLS (or GLS) estimates are likely to be biased in small samples (Marriott and Pope (1954), Kendall (1954)). This is because there are lagged dependent variables on the right hand side of the regression equations. While the literature on the LOP and PPP has usually assumed that small sample bias leads to an underestimation of true persistence in the autoregressive process, this is only true in the case of autoregressive processes of order one. In higher order processes, the bias induced by OLS can go in either direction (Stine and Shaman (1988, 1989)).

I follow two approaches to account for small sample bias. One approach uses Kilian's (1998) bootstrap-after-bootstrap estimator. The results from applying the bootstrap-after-bootstrap procedure suggest that the SUR-GLS estimator is not seriously affected by small sample bias (see Figure 3). For most products, half-lives obtained using the bootstrap-after-bootstrap method are larger than the ones obtained from the uncorrected SUR-GLS estimates. However, the differences in half-lives between the two methods are small. The average half-life obtained using the bootstrap-after-bootstrap procedure equals 5.9 weeks, which compares to an average half-life of 5.6 weeks obtained from the original SUR-GLS estimates. The bootstrap-after-bootstrap estimator has, however, been criticized on the grounds that its sampling distribution may be asymmetric and hence, the procedure may not yield a median unbiased estimator (see Pesavento and Rossi (2007)).[25]

[25] An alternative estimator to correct for small sample bias is Andrews and Chen (1994) approximately median unbiased estimator. In addition to being computationally intensive, implementations

As an alternative to the bootstrap-after-bootstrap estimator as a bias correction method, I use a recursive mean adjustment (RMA) procedure (So and Shin (1999)).[26] This method is based on the observation that the small sample bias problem arises in part due to the inclusion of a constant term in the estimated equation. The problem is alleviated if the data is previously demeaned using the recursive mean. Estimating the SUR system (3) using recursively mean adjusted data[27] yields similar half-lives to the ones obtained when no adjustment is introduced. Figure 3 compares the original (unadjusted) estimates of half-lives with estimates obtained using recursively mean adjusted data. Half-lives are even smaller, in most cases, when correcting using the RMA method. The average half-life based on the SUR-GLS-RMA estimator is equal to 4.9 weeks (versus 5.6 weeks in the case in which no correction for small sample bias is made).

The evidence thus suggests that small sample bias has only a negligible (if any) effect on the SUR-GLS estimates of persistence.

Measurement Error. Another source of bias that might affect the results is measurement error. Measurement error –due, for instance, to incorrect recordings made by price collectors– can cause persistence parameters, and hence half-lives, to be underestimated. Suppose that instead of observing the true relative price $q_{i,t}^{k}$ one observes the variable

$$q_{i,t}^{k*} = q_{i,t}^{k} + v_{i,t}^{k} \tag{4}$$

where $v_{i,t}$ is distributed i.i.d. with mean zero and variance σ_v^2. In this case, OLS

of this procedure are usually unstable when the number of lags included in the estimations is large, as is the case with the equations estimated in this paper.

[26] Chen and Engel (2005) is another paper that uses both recursive mean adjustment and Kilian's (1998) bootstrap-after-bootstrap procedure to account for small sample bias.

[27] This procedure involved estimating system (3) without a constant term and with log relative prices, $q_{i,t}^{k}$, expressed as differences from their recursive mean.

estimation of equation

$$q_{i,t}^{k*} = \alpha_i + \sum_{j=1}^{p_{ik}} \rho_{ij} q_{i,t-j}^{k*} + e_{i,t}^{k} \tag{5}$$

leads to inconsistently estimated coefficients as the error term is correlated with the explanatory variables. In the above equation, $e_{i,t}^{k} = -v_{i,t}^{k} + \sum_{j=1}^{p} \rho_{ij} v_{i,t-j}^{k} + \varepsilon_{i,t}^{k}$. In order to assess whether the error term is correlated with the explanatory variables, I perform a Hausman test for endogeneity using the lags $\left\{q_{i,t-p_{ik}-1}^{k}, ..., q_{i,t-2p_{ik}}^{k}\right\}$ as instrumental variables.[28] The results of the Hausman test suggest that measurement error is not a major problem affecting the SUR-GLS persistence estimates. The null hypothesis of consistency is rejected at the 5 percent level in only 12.5 percent of the series.[29]

5 Estimates of Speed of Convergence to the LOP using Lower-Frequency Data

This section turns to studying the consequences of estimating convergence to the LOP using aggregate data. This section focuses on the effects of temporal aggregation; the next section examines potential problems associated with cross-sectional aggregation.

Studies of convergence to the LOP and PPP have typically relied on price data observed at lower frequencies than the frequencies at which prices are generated (Taylor (2001)). While most studies analyzing price gaps in consumer prices use data that is observed quarterly or annually, evidence from studies of price-setting behavior suggests that retail prices are generated at weekly or monthly frequencies (e.g., Eichenbaum, Jaimovich and Rebelo (2011), Midrigan (2011)). Furthermore, the evidence presented

[28] Imbs et al. (2005), among others, use this test to assess the presence of measurement error in their data.

[29] Results of the Hausman test are available from the author upon request.

in Section 1 suggests that prices in the data change, on average, every 6 weeks. In this section, I exploit the fact that my data is available at a weekly frequency to study the consequences of estimating convergence to the LOP using lower frequency data.

The time-series literature recognizes two schemes under which low frequency series can be generated from the original high-frequency ones: Systematic sampling and temporal aggregation.[30] Systematic sampling occurs when observations of a time series are sampled at regular intervals, as when end-of-period prices are observed. Studies of the LOP using systematically sampled data include, among others, Parsley and Wei (1996), Ceglowski (2003) and those analyzing data from the Economist Intelligence Unit's Worldwide Cost of Living Survey such as Rogers (2007) and Crucini and Shintani (2008). Temporal aggregation occurs when the time series available to the researcher are either sums or averages of the original series over a given time interval, as when quarterly or yearly averages are observed. Broda and Weinstein (2008), for instance, study convergence to the LOP using quarterly averaged data. Studies analyzing price indices may also be implicitly using temporally aggregated data as some statistical agencies collect prices at a higher frequency than the one at which the price indexes are constructed and reported (see Taylor (2001) for further details).[31]

There is an extensive literature studying the consequences of systematic sampling and temporal aggregation in the estimation of time series models.[32] Taylor (2001) is the first paper to study the consequences of using low-frequency data to assess convergence to the LOP and PPP. Assuming the data is generated by an AR(1) process, he shows that temporal aggregation biases persistence estimates upward. Furthermore,

[30] "Temporal aggregation" and "systematic sampling" are also known as flow-aggregation and stock-aggregation schemes, respectively.

[31] In Mexico, for instance, the CPI is constructed and reported on a monthly basis but prices are sampled twice a month with food items being sampled as frequently as four times a month (Gagnon (2009)).

[32] Silvestrini and Veredas (2008) provide a survey of the time series literature on temporal aggregation.

the upward bias is increasing in the order of temporal aggregation (e.g., the bias is larger for quarterly averaged prices than for monthly averaged prices). The intuition for inconsistency in the estimates of persistence under temporal aggregation is as follows. Suppose that the original data-generating process for the price gap of a given product, $q_{i,t}$, is an AR(1):

$$q_{i,t} = \rho q_{i.t-1} + \varepsilon_{i,t}, \quad t = 1, ..., T$$

where $\{\varepsilon_t\}$ is an $i.i.d.$ sequence with variance σ_ε^2. The observed price data are non-overlapping K-period averages of the original data. The observed, temporally-aggregated, variable is thus given by

$$Q_{i,s} = (1/K) \sum_{j=1}^{K} q_{i,K(s-1)+j}$$

where $s = 1, ..., T/K$ indexes time units at which prices are actually observed. The temporally aggregated model is an ARMA(1,1):

$$Q_{i,s} = \phi Q_{i,s-1} + u_{i,s}$$

where $u_{i,s} = (1/K) \sum_{j=1}^{K} \sum_{l=1}^{K} \rho^{l-1} \varepsilon_{i,K(s-1)+j-l}$.Thus, as $Q_{i,s-1}$ and $u_{i,s}$ are correlated, the OLS estimator of ϕ is inconsistent.[33]

Consistent estimates of persistence can be obtained in the presence of temporal aggregation using alternative estimators (e.g., instrumental variables or GMM). Chambers (2005), for example, proposes a maximum likelihood estimator to find consistent estimates of half-lives. However, the literature on PPP and the LOP has, for the most part, ignored the problem. The following subsection quantifies the magnitude of the bias introduced by using temporally aggregated prices.

[33] See Chambers (2005) for an analytical derivation of the asymptotic bias.

5.1 Effects of Temporal Aggregation

I assess the magnitude of the temporal aggregation bias by replicating the estimations from Section 3 using temporally aggregated series. Specifically, letting

$$Q_{ik,s} = \ln\left[(1/K)\sum_{j=1}^{K} p_{ik,K(s-1)+j}\right] - \ln\left[(1/K)\sum_{j=1}^{K} p_{mk,K(s-1)+j}\right]$$

denote non-overlapping averages of relative prices ($p_{ik,s}$ and $p_{mk,s}$ are the prices of good k in city i and Mexico City, respectively, in period s) at the (approximately) monthly and quarterly frequencies (i.e., $K = 4$ and $K = 13$), I estimate the system

$$Q_{ik,s} = \alpha_{ik} + \sum_{j=1}^{p_{ik}} \rho_{ikj} Q_{ik,s-j} + u_{ik,s}, \quad s = 1, ..., T/K;\ i = 1, ..., N \tag{6}$$

for each good k.

Estimations using temporally aggregated data show that there is a large upward bias in the estimated half-lives of deviations from the LOP. Figure 4 compares the distributions of half-lives expressed in weeks obtained using temporally disaggregated weekly data with those obtained from monthly and quarterly averaged prices. The distributions of estimated half-lives obtained from aggregate data fall to the right of those obtained from weekly data; the higher the order of aggregation, the larger the shift to the right. Table 7 compares the average and median half-lives estimated from temporally aggregated data with those obtained using weekly data. The mean half-life is 6.46 weeks when estimated using weekly-sampled data. It rises to 9.34 weeks when using monthly data, and to 16.72 weeks when using quarterly aggregate data. Figure 5 compares estimates of half-lives using temporally disaggregated data with estimates using monthly averaged prices at the good level. As can be seen from that figure, in general, half-lives are increasing in the order of aggregation. For 38 out of 40 products the average half-life is larger with monthly relative to weekly aggregation.

5.2 Monte Carlo Experiments

In this subsection, I explore the robustness of the findings on temporal aggregation using Monte Carlo experiments. I use the estimated parameters (including the variance-covariance matrix parameters) for one of the products that bracket the median half-life of deviations from the LOP ("Milk Alpura") to generate simulated series of (log) price differentials. Specifically, I generate 10,000 replications of a panel of dimensions $T = 1300$ (i.e., 25 years of weekly data) and $N = 10$ (i.e., 10 city-pairs) using system (3) as the data-generating process with errors that are multivariate normal. I then obtain temporally aggregated series by calculating non-overlapping averages $Q_s = (1/K) \sum_{j=1}^{K} q_{K(s-1)+j}$, $s = 1, ..., T/K$, for $K = \{1, 4, 13\}$. I compute the half-lives for both disaggregated and aggregate series using the estimated impulse response functions.

Table 8 presents the results of simulations. The mean half-life using disaggregated data (K = 1) is estimated at 5.18 periods (weeks). It rises to 9.58 weeks when data are aggregated every four periods (K = 4) and to 12.79 weeks when the data are aggregated every thirteen periods (K = 13). The bias factor equals 1.85 for monthly aggregation and 2.47 for quarterly aggregation. Thus, the evidence provided by the Monte Carlo experiments reinforces the conclusion that temporal aggregation is a potentially important source of bias in persistence estimates.

6 Cross-Sectional Aggregation

Studies of relative price dynamics have typically relied on data consisting of prices for broadly defined goods or sectors (Broda and Weinstein (2008)). Using price measures that aggregate across several individual goods (or sectors) might pose problems for estimating the speed at which prices converge to parity. Imbs et al. (2005) claim that cross-sectional aggregation across sectors exhibiting heterogeneous dynamics can help to reconcile the high volatility exhibited by real exchange rates with the low speed of convergence towards PPP (the so-called "PPP Puzzle"). They show that when the rates at which relative prices of individual sectors converge to the LOP are heterogeneous, estimations that constrain the parameters capturing convergence speed to be the same across sectors will tend to overestimate the half-life of deviations from PPP.

Broda and Weinstein (2008), on the other hand, find no evidence of the type of bias found by Imbs et al. They find that the estimates of the half-life of deviations from the LOP are similar, whether or not they allow for heterogeneous dynamics across individual goods. Crucini and Shintani (2008) report findings that point in the same direction as those in Broda and Weinstein (2008). In this section, I assess the relevance of this type of bias in the data.

To assess the importance of cross-sectional aggregation bias, I construct an equally weighted price index using 31 products[34] in the sample. The estimated model for each city-pair $i = 1, ...N$ is

$$\Delta\widetilde{q}_{it} = \alpha_i + \beta_i \widetilde{q}_{it-1} + \sum_{j=1}^{p_i-1} \rho_{ij} \Delta\widetilde{q}_{it-j+1} + \epsilon_{it},\ i = 1, ..., N \tag{7}$$

where $\widetilde{q}_{it}$ is the relative aggregate price between city i and the benchmark city. The parameters β_i and $\left\{\rho_{ij}\right\}_{j=1}^{p_i}$ are allowed to vary by city-pair to avoid confounding the effects of aggregation across products with that of aggregation across city-pairs (Choi, Mark and Sul (2006)).

[34]Prices of products for which more than one variety were available (e.g., Pepsi and Coca-Cola) were obtained as a simple average of the product varieties.

Table 9 compares the average half-lives obtained using aggregate relative prices with the average half-life at the city-pair level obtained from weekly data. The half-life of a LOP deviation is larger when using the aggregate price only in 2 out of 10 city-pairs; in those 2 cases, the overestimation of the half-life is rather small. The half-life for the median city-pair estimated using the price index is equal to 4.3 weeks, about 2 weeks smaller than the median half-life obtained using individual prices. Thus, the results suggest that cross-sectional aggregation does not lead to an overestimation of the speed of convergence towards the LOP.

7 Concluding Remarks

This paper studied price differences across cities in Mexico using highly disaggregated price data. Two major results emerge from the analysis. First, the evidence points to a remarkably fast convergence rate towards the LOP. The estimated half-life of 5-6 weeks implies a substantially lower persistence in the process of relative intercity prices than previously reported in the literature. Using similar data for the US, Parsley and Wei (1996) and Broda and Weinstein (2008) find half-lives of deviations from the LOP on the order of 3-5 quarters. It is interesting to note that both the average half-life of deviations from the LOP and the implied duration of price spells are in the order of 6 weeks. Even studies of the LOP that focus on narrowly defined goods (e.g., Broda and Weinstein, 2008) find half-lives of LOP deviations that are orders of magnitude larger than implied durations of individual price spells.

Second, using low frequency data to study relative price dynamics can lead to a large upward bias in persistence estimates. Half-lives of deviations from the LOP derived from monthly and quarterly aggregated data are found to be 1.5 and 2.6 times larger, respectively, than the half-life of a shock to the LOP estimated from weekly data.

The results using actual price data and allowing for higher-order processes for relative prices point in the same direction as those reported by Taylor (2001) using simulations of an AR(1) process. As pointed out above, this problem afflicts studies that analyze price indexes because statistical agencies tend to report average prices; it also affects studies focusing on narrowly defined goods (e.g., Broda and Weinstein, 2008). The message for future work on relative price dynamics should be clear. Using temporally disaggregated data can be crucial for obtaining reliable estimates of half-lives as the bias introduced by aggregation can be large.

In particular, temporal aggregation appears to be a more pernicious source of bias than aggregation across individual goods that exhibit heterogeneous dynamics. In contrast to the findings of Imbs et al. (2005), this paper showed that pooling across goods with varying convergence rates does not affect estimates of persistence in any systematic way. The results, in this respect, are consistent with those reported by Broda and Weinstein (2008) and Crucini and Shintani (2008). An important question left for future research is whether the size of the bias induced by temporal aggregation is similar in other product categories and sectors. Evidence along these lines would be helpful in assessing to what extent temporal aggregation can account for Rogoff's (1996) "PPP puzzle".

References

Andrews, Donald W. K. and Hong-Yuan Chen. 1994. Approximately Median-Unbiased Estimation of Autoregressive Models. Journal of Business & Economic Statistics 12 (2): 187–204 (April).

Broda, Christian and David E. Weinstein. 2008, May. Understanding International Price Differences Using Barcode Data. Working paper 14017, National Bureau of Economic Research.

Ceglowski, Janet. 2003. The Law of One Price: Intra-national Evidence for Canada. Canadian Journal of Economics 36 (2): 373–400.

Chambers, Marcus J. 2005. The Purchasing Power Parity Puzzle, Temporal Aggregation, and Half-life Estimation. Economics Letters 86:193–198.

Chen, Shiu-Sheng and Charles Engel. 2005. Does 'Aggregation Bias' Explain the PPP Puzzle. Pacific Economic Review 10 (1): 49–72.

Choi, In. 2001. Unit Root Tests for Panel Data. Journal of International Money and Finance 20 (2): 249–272 (April).

Choi, Chi-Young, Nelson C. Mark, and Donggyu Sul. 2006. Unbiased Estimation of the Half Life to PPP Convergence in Panel Data. Journal of Money, Credit and Banking 38 (4): 921–938 (June).

Crucini, Mario J. and Mototsugu Shintani. 2008. Persistence in law of one price deviations: Evidence from micro-data. Journal of Monetary Economics 55 (3): 629–644.

Crucini, Mario J., Mototsugu Shintani, and Takayuki Tsuruga. 2010. The Law of One Price Without the Border: The Role of Distance versus Sticky Prices. The Economic Journal 120 (544): 462–480 (May).

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

Demetrescu, Matei, Uwe Hassler, and Adina-Ioanna Tarcolea. 2006. Combining Significance of Correlated Statistics with Application to Panel Data. Oxford Bulletin of Economics and Statistics 68 (5): 647–663 (October).

Eichenbaum, Martin, Nir Jaimovich, and Sergio Rebelo. 2011. Reference Prices, Costs and Nominal Rigidities. American Economic Review 101 (1): 234–262 (February).

Elberg, Andres. 2014. Heterogeneous Price Dynamics, Synchronization, and Retail Chains: Evidence from Scanner Data. Emerging Markets Finance and Trade (forthcoming).

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

Gagnon, Etienne. 2009. Price Setting During Low and High Inflation: Evidence from Mexico. Quarterly Journal of Economics 124 (3): 1221–1263.

Goldberg, Pinelopi K. and Michael Knetter. 1997. Goods Prices and Exchange Rates: What Have We Learned? Journal of Economic Literature XXXV:1243–1272.

Hamilton, James D. 1994. Time Series Analysis. Princeton University Press.

Imbs, Jean, Haroon Mumtaz, Morten O. Ravn, and Hélène Rey. 2005. PPP Strikes Back: Aggregation and the Real Exchange Rate. Quarterly Journal of Economics CXX (1): 1–43 (February).

International Labour Organization, IMF, OECD, World Bank, and the UN. 2004. Consumer Price Index Manual: Theory and Practice. Edited by International Labour Office. Geneva: International Labour Office.

Isard, Peter. 1977. How Far can we Push the "Law of One Price"? American Economic Review 67 (5): 942–948 (December).

Kendall, M. G. 1954. Note on the Bias in the Estimation of Autocorrelation. Biometrika 41:403–404.

Kilian, Lutz. 1998. Small-Sample Confidence Intervals For Impulse-Response Functions. Review of Economics and Statistics 80 (2): 218–230.

Marriott, F. H. C. and J. A. Pope. 1954. Bias in the Estimation of Autocorrelations. Biometrika 41:390–402.

Midrigan, Virgiliu. 2011. Menu Costs, Multiproduct Firms and Aggregate Fluctuations. Econometrica 79 (4): 1139–1180 (July).

Ng, Serena and Pierre Perron. 2001. Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power. Econometrica 69 (6): 1519–1554.

O'Connell, Paul G. J. 1998. The Overvaluation of the Purchasing Power Parity. Journal of International Economics 44:1–19.

Parsley, David C. and Shang-Jin Wei. 1996. Convergence to the Law of One Price Without Trade Barriers or Currency Fluctuations. Quarterly Journal of Economics 111 (4): 1211–1236 (November).

Pesavento, Elena and Barbara Rossi. 2007. Impulse Response Confidence Intervals for Persistent Data: What Have We Learned? Journal of Economic Dynamics and Control 31 (7): 2398–2412.

Richardson, David J. 1978. Some Empirical Evidence on Commodity Arbitrage and the Law of One Price. Journal of International Economics 8 (2): 341–351.

Rogers, John H. 2007. Monetary Union, Price Level Convergence, and Inflation: How Close is Europe to the USA? Journal of Monetary Economics 54 (3): 785–796.

Rogoff, Kenneth. 1996. The Purchasing Power Parity Puzzle. Journal of Economic Literature XXXIV (June): 647–668.

Sarno, Lucio and Mark P. Taylor. 1998. Real Exchange Rates under the Recent Float: Unequivocal Evidence of Mean Reversion. Economics Letters 60 (2): 131–137.

Silvestrini, Andrea and David Veredas. 2008. Temporal Aggregation of Univariate and Multivariate Time Series Models: A Survey. Journal of Economic Surveys 22 (3): 458–497.

So, Beong S. and Dong W. Shin. 1999. Recursive Mean Adjustment in Time-Series Inferences. Statistics and Probability Letters 43 (1): 65–73 (May).

Stine, Robert A. and Paul Shaman. 1988. The Bias of Autoregressive Coefficient Estimators. Journal of the American Statistical Association 83 (403): 842–848.

Stine, Robert A. and Paul Shaman. 1989. A Fixed Point Characterization for Bias of Autoregressive Estimators. Annals of Statistics 17:1275–1284.

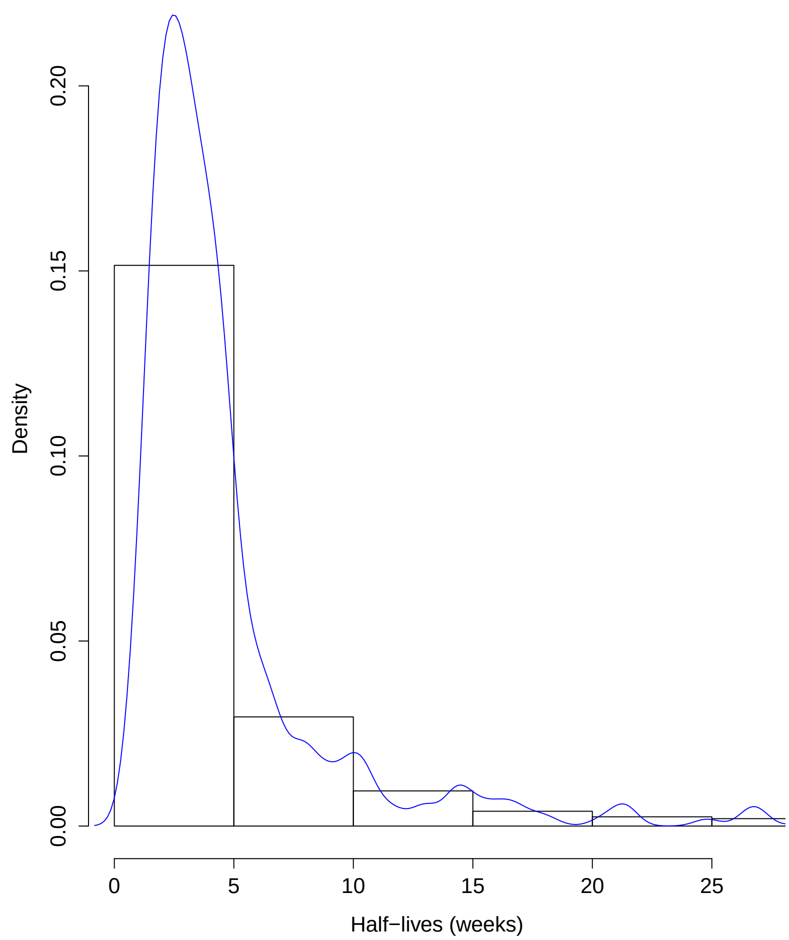
Taylor, Alan M. 2001. Potential Pitfalls for the Purchasing Power Parity Puzzle? Sampling and Specification Biases in Mean-Reversion Tests of the Law of One Price. Econometrica 69:473–98.

Taylor, Mark P. and Lucio Sarno. 1998. The Behavior of Real Exchange Rates During the Post-Bretton Woods Period. Journal of International Economics 46 (2): 218–312.

Teles, Paulo and William W. S. Wei. 1999. The Use of Aggregate Series in Time Series Testing. 1999 NBER/NSF Time Series Conference Proceedings IX & X. Academia Sinica, Taipei, Taiwan, 1–29.

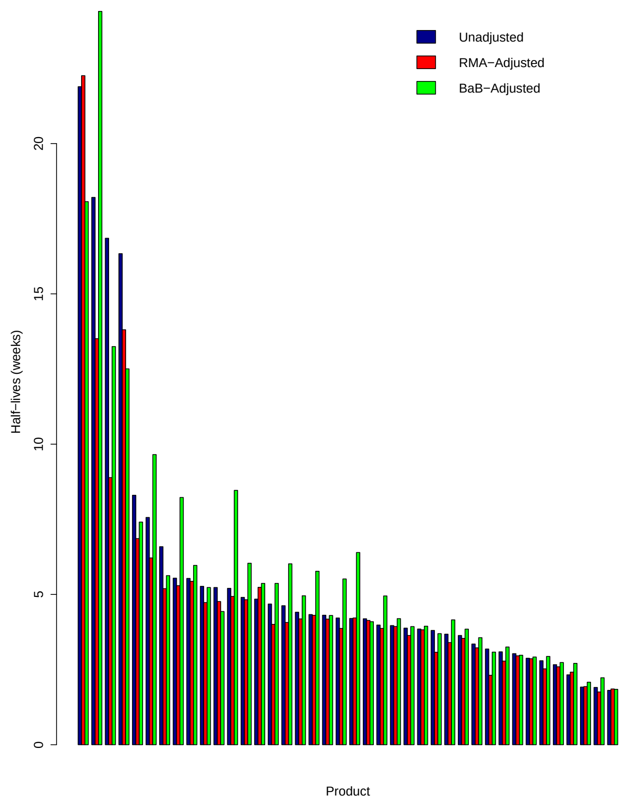
Figure 1: Geographical Location of Cities in the Sample

Figure 2: Distribution of Half-Lives



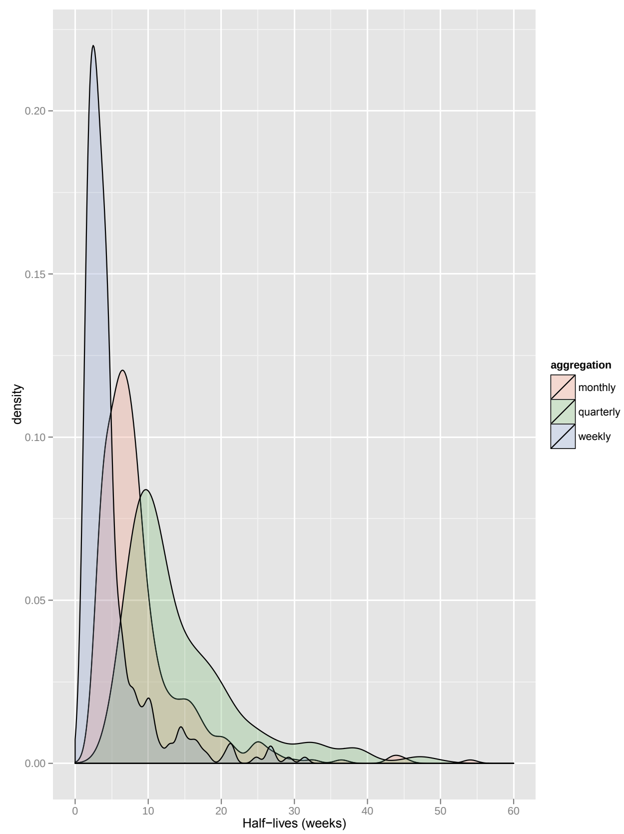
Notes. The figure presents half-lives pooling across all products and city-pairs (using Mexico City as the benchmark city). Half-lives obtained from SUR-GLS estimates of impulse response functions.

Figure 3: Small Sample Bias: Adjusted vs. Unadjusted Estimates of Half-Lives



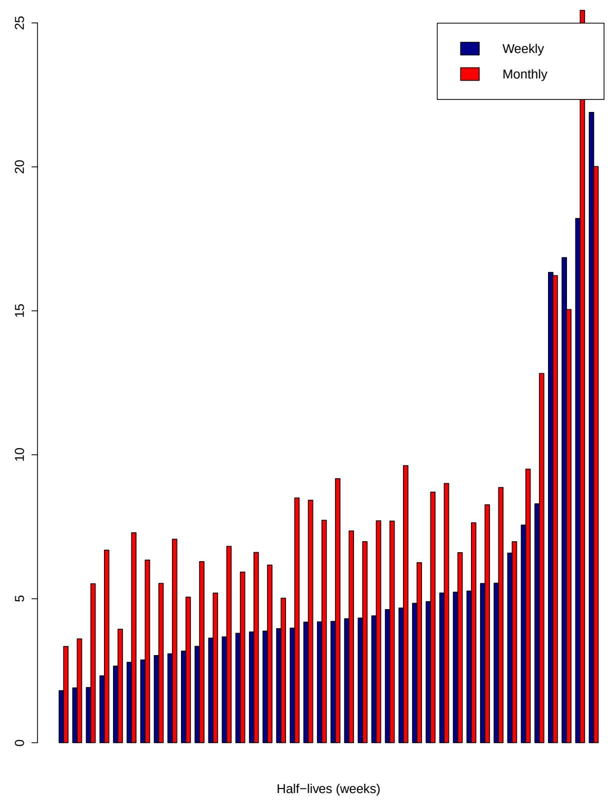
Notes. Unadjusted half-lives obtained from SUR-GLS estimates of impulse response functions. "BaB-Adjusted" denote half-lives corrected for small sample bias using Kilian's (1998) bootstrap-after-bootstrap procedure. "RMA-adjusted" denote half-lives corrected for small sample bias using a recursive mean adjustment procedure.

Figure 4: Estimated Half-Lives under Different Orders of Temporal Aggregation



Notes. The figure presents estimated half-lives obtained from weekly sampled prices, monthly averaged prices, and quarterly averaged prices. Half-lives obtained from SUR-GLS estimates of impulse response functions.

Figure 5: Temporal Aggregation and Half-Lives by Product



Notes. The figure presents estimated half-lives obtained from weekly sampled prices and monthly averaged prices at the product level. Half-lives obtained from SUR-GLS estimates of impulse response functions.

Table 1: Description of Products in the Dataset

Product	Size/Type of package	Perishable	Product	Size/Type of package	Perishable
Cooking oil, Mazola	1 lt. bottle	No	Ham, San Rafael	1 kg.	Yes
Bottled water, Bonafont	1.5 lt. bottle	No	Jitomate Bola de Sinaloa	1 kg.	Yes
Bottled water, Santa Maria	1.5 lt. bottle	No	Jitomate Saladette	1 kg.	Yes
Hass avocado	1 kg.	Yes	Milk (ultrapasteurized), Alpura	1 lt. box	No
Rice, Mexica	1 kg. bag	No	Lentils, Valle Verde	500 gr. bag	No
Rice, Morelos	1 kg. bag	No	Golden apples	1 kg.	Yes
Brown sugar	2 kg. bag	No	White bread	1 piece	Yes
Refined sugar	2 kg. bag	No	White potato	1 kg.	Yes
Bleach, cloralex	950 ml. bottle	No	Chicken leg	1 kg.	Yes
Bleach, clorox	930 ml. bottle	No	Banana macho	1 kg.	Yes
Instant coffee, Nescafe	200 gr. jar	No	Banana Tabasco/Chiapas	1 kg.	Yes
Ground beef	1 kg.	Yes	Whole chicken	1 kg.	Yes
Whilte onion	1 kg.	Yes	Soda, Coca-Cola	355 ml. can	No
Pork chop	1 kg.	Yes	Soda, Pepsi	355 ml. can	No
Deodorant, Lady Speed Stick	45 gr. bar	No	Soup bones	1 kg.	Yes
Deodorant, Obao	65 gr. bar	No	Sardines	425 gr. can	No
Laundry detergent, Foca	1 kg.	No	Napkins, Regio	250 sheets	No
Black beans, Valle Verde	1 kg.	No	Laundry softener, Suavitel	1 lt. bottle	No
Pinto beans, Valle Verde	1 kg.	No	Green tomatoes	1 kg.	Yes
Soap, Tepeyac	400 gr. bar	No	Corn tortilla	1 kg.	Yes

Table 2: Summary Statistics

Nonperishables	Median	SD	freq Δp	Perishables	Median	SD	freq Δp
Cooking Oil, Mazola	21.27	1.80	.08	Hass Avocado	23.00	4.10	.46
Bottled Water, Bonafont	6.30	1.05	.10	Ground Beef	46.50	8.32	.14
Bottled Water, Santa Maria	6.35	.77	.07	White Onions	8.70	2.07	.49
Rice, Mexica	8.35	.76	.08	Pork Chops	59.90	7.73	.12
Rice, Morelos	9.04	.93	.17	Ham, San Rafael	87.40	7.62	.13
Brown Sugar	19.90	1.71	.11	Jitomate Bola de Sinaloa	16.00	3.38	.52
Refined Sugar	20.35	1.06	.09	Jitomate Saladette	14.00	3.56	.54
Bleach, Cloralex	6.34	.81	.12	Golden Apples	22.90	4.06	.35
Bleach, Clorox.	6.80	.61	.08	White Bread	0.90	0.16	.02
Instant Coffee, Nescafe	41.50	1.83	.07	White Potatoes	11.90	2.41	.39
Deodorant, Lady Speed Stick	26.86	2.66	.11	Chicken Leg	26.55	3.32	.09
Deodorant, Obao	17.20	2.03	.09	Banana Macho	11.40	2.11	.21
Laundry Detergent, Foca	13.90	1.16	.11	Banana Tabasco/Chiapas	6.50	1.42	.30
Black Beans, Valle Verde	11.30	1.30	.14	Whole Chicken	19.50	1.89	.12
Pinto Beans, Valle Verde	13.40	1.57	.09	Soup Bones	39.90	3.73	.09
Soap, Tepeyac	5.75	.40	.07	Green Tomatoes	12.40	3.45	.44
Lentils, Valle Verde	8.00	.67	.07	Corn Tortilla	5.80	1.24	.03
Soda, Coca-Cola	5.00	.40	.05				
Soda, Pepsi	4.52	.37	.03				
Sardines, Yavaros	10.35	.82	.06				
Napkins, Regio	13.90	1.67	.09				
Laundry Softener, Suavitel	14.00	1.41	.09				
Milk, Alpura 2000	9.50	.31	.03				
Mean	13.04	1.13	.09		24.31	3.56	.26
Median	10.35	1.05	.09		16.00	3.38	.21
St. dev.	8.63	0.62	.03		22.51	2.33	.18

Notes. The table presents the median, standard deviation (SD) and adjustment frequency of prices for each product in the sample. The median and standard deviation of prices are expressed in local currency (Mexican pesos). The average exchange rate over the period of analysis is approximately 11 MXN per USD.

Table 3: Deviations from the Law of One Price

All Products	Median (1)	MAD (2)	St. dev. (3)
Mean	.0041	.0748	.0916
Median	.0023	.0612	.0734
St. dev.	.0275	.0475	.0620
# of obs	814,000	814,000	814,000
Non Perishables			
Mean	-.0002	.0510	.0615
Median	-.0005	.0473	.0569
St. dev.	.0185	.0223	.0283
# of obs.	468,050	468,050	468,050
Perishables			
Mean	.0100	.1071	.1324
Median	.0089	.0981	.1198
St. dev.	.0355	.0532	.0711
# of obs.	345,950	345,950	345,950

Notes. Columns (1)-(3) present the median, mean absolute deviation (MAD) and standard deviation (St. dev.) of LOP deviations taken over 55 city-pairs for each product and week. Rows present descriptive statistics (mean, median and standard deviation) for each of the above measures taken over products and weeks.

Table 4: LOP Deviations and Transportation Costs

	All Products (1)	**Nonperishables** (2)	**Perishables** (3)
log distance	.0024 (.0013)*	.0031 (.0010)***	.0015 (.0027)
product FE	Yes	Yes	Yes
R-sq	.0014	.0108	.0004
# of obs	2,200	1,265	935

Notes. The dependent variable is the average absolute log price differential of a given product between two cities. All 55 city pairs are included in the estimation. The regressor of interest is the log of distance measured by the greater circle method. Product fixed effects are included in all regressions. Robust standard errors are shown in parenthesis. ***, and *, stand for significant at the 1 percent and 10 percent level, respectively.

Table 5: Results of Panel Unit Root Tests

Non Perishables	Test Statistic	Lags[1]	Perishables	Test Statistic	Lags[1]
Cooking Oil, Mazola	-7.332^{***}	5.5 (8.50)	Hass Avocado	-2.570^{***}	8.0 (8.60)
Bottled Water, Bonafont	-1.589^{*}	9.0 (7.45)	Ground Beef	-4.509^{**}	12.0 (7.63)
Bottled Water, Santa Maria	-1.311^{*}	11.0 (4.24)	White Onions	-6.595^{***}	10.0 (7.63)
Rice, Mexica	-3.291^{***}	17.5 (8.49)	Pork Chops	-1.777^{**}	12.0 (6.11)
Rice, Morelos	-4.818^{***}	15.0 (7.99)	Ham, San Rafael	-7.226^{***}	9.0 (8.28)
Brown Sugar	-5.169^{***}	5.5 (5.27)	Jitomate Bola de Sinaloa	-8.128^{***}	8.5 (11.30)
Refined Sugar	-2.753^{***}	6.0 (7.44)	Jitomate Saladette	-6.648^{***}	13.0 (8.79)
Bleach, Cloralex	-2.276^{**}	9.0 (5.61)	Golden Apples	-6.292^{***}	9.0 (6.30)
Bleach, Clorox.	-1.938^{**}	8.5 (5.59)	White Bread	-2.461^{***}	7.0 (8.00)
Instant Coffee, Nescafe	-2.073^{**}	8.5 (5.17)	White Potatoes	-5.649^{***}	11.0 (6.77)
Deodorant, Lady Speed Stick	-3.477^{***}	8.5 (2.72)	Chicken Leg	-3.655^{***}	7.5 (7.48)
Deodorant, Obao	-1.870^{**}	8.5 (5.62)	Banana Macho	-3.637^{***}	2.5 (3.37)
Laundry Detergent, Foca	-6.552^{***}	8.0 (6.63)	Banana Tabasco/Chiapas	-7.554^{***}	5.0 (4.94)
Black Beans, Valle Verde	-6.934^{***}	5.5 (9.03)	Whole Chicken	-3.154^{***}	7.0 (2.67)
Pinto Beans, Valle Verde	-3.342^{***}	6.0 (3.79)	Soup Bones	-1.716^{**}	8.5 (5.99)
Soap, Tepeyac	-2.800^{***}	7.5 (6.20)	Green Tomatoes	-2.606^{***}	13.0 (7.70)
Lentils, Valle Verde	-9.679^{***}	5.5 (7.71)	Corn Tortilla	-1.239	3.0 (6.60)
Soda, Coca-Cola	-2.633^{***}	2.0 (5.02)			
Soda, Pepsi	-1.191	1.0 (2.26)			
Sardines, Yavaros	-3.372^{***}	5.0 (5.17)			
Napkins, Regio	-1.693^{**}	9.5 (6.04)			
Laundry Softener, Suavitel	-1.331^{*}	11.0 (7.10)			
Milk, Alpura 2000	-3.549^{***}	6.5 (6.24)			

[1] Truncation lags obtained using the Modified Akaike Information Criterion (MAIC). Main entries are the median truncation lag across city pairs (standard deviations in parenthesis).

[2] (*), (**), (***) denote significant at 10 percent, 5 percent and 1 percent level of significance, respectively.

Table 6: Estimates of Convergence Speed

Non Perishables	Median	95% C.I.		Perishables	Median	95% C.I.	
Cooking Oil, Mazola	4.845	[3.422,	9.282]	Hass Avocado	1.919	[1.576,	3.010]
Bottled Water, Bonafont	4.681	[3.387,	10.559]	Ground Beef	3.091	[1.947,	5.964]
Bottled Water, Santa Maria	4.409	[3.035,	8.463]	White Onions	2.664	[1.778,	4.194]
Rice, Mexica	5.231	[2.779,	7.538]	Pork Chops	4.625	[3.487,	10.764]
Rice, Morelos	3.803	[2.515,	5.866]	Ham, San Rafael	3.635	[2.742,	5.617]
Brown Sugar	7.561	[4.827,	16.942]	Jitomate Bola de Sinaloa	1.810	[1.477,	2.543]
Refined Sugar	3.351	[2.698,	7.533]	Jitomate Saladette	1.907	[1.617,	3.048]
Bleach, Cloralex	5.270	[2.938,	9.415]	Golden Apples	3.184	[1.697,	4.386]
Bleach, Clorox.	4.216	[2.944,	11.016]	White Bread	16.850	[5.028,	30.348]
Instant Coffee, Nescafe	3.982	[2.814,	9.920]	White Potatoes	2.326	[1.751,	5.107]
Deodorant, Lady Speed Stick	4.190	[2.699,	7.741]	Chicken Leg	3.679	[2.600,	6.445]
Deodorant, Obao	4.902	[3.616,	12.523]	Banana Macho	8.300	[4.453,	12.522]
Laundry Detergent, Foca	3.880	[2.734,	6.436]	Banana Tabasco/Chiapas	2.795	[0.542,	$+\infty$]
Black Beans, Valle Verde	3.030	[2.155,	5.752]	Whole Chicken	2.878	[1.981,	4.911]
Pinto Beans, Valle Verde	4.309	[3.009,	9.406]	Soup Bones	4.331	[3.134,	17.676]
Soap, Tepeyac	5.542	[4.222,	14.296]	Green Tomatoes	3.850	[2.683,	5.958]
Lentils, Valle Verde	3.965	[2.870,	6.280]	Corn Tortilla	16.337	[3.134,	17.676]
Soda, Coca-Cola	18.209	[8.553,	56.903]				
Soda, Pepsi	21.893	[9.427,	36.049]				
Sardines, Yavaros	6.589	[3.262,	10.080]				
Napkins, Regio	5.203	[3.958,	14.771]				
Laundry Softener, Suavitel	5.533	[3.595,	8.984]				
Milk, Alpura 2000	4.201	[3.048,	14.470]				
Mean	6.035	[3.674,	13.055]	Mean	4.952	[2.606,	9.101]
Median	4.681	[3.035,	9.415]	Median	3.184	[1.981,	5.788]

Notes. Median half-lives obtained from SUR-GLS estimates of impulse response functions. They correspond to the median half-life for a given product across city-pairs. Confidence intervals computed using Kilian's (1998) bootstrap-after-bootstrap procedure.

Table 7: Temporal Aggregation

	Order of aggregation		
	weekly (M=1)	*monthly* (M=4)	*quarterly* (M=13)
Mean	6.46	9.34	16.72
Median	5.26	7.65	14.05
Bias Factor	1.00	1.45	2.56

Notes. The bias factor is computed as the quotient between the average half-life using aggregated data and the average half-life estimated from disaggregated data.

Table 8: Temporal Aggregation: Monte Carlo Experiments

	Order of aggregation		
	weekly (M=1)	*monthly* (M=4)	*quarterly* (M=13)
Mean	5.18	9.58	12.79
Median	4.11	8.19	11.185
Bias Factor	1.00	1.85	2.47

Notes. The bias factor is computed as the quotient between the average half-life using aggregated data and the average half-life estimated from disaggregated data.

Table 9: Cross-Sectional Aggregation

City-pair (1)	Aggregate (2)	Individual (3)	Bias Factor (4)
1	5.19	7.36	0.70
2	3.12	7.11	0.44
3	3.43	4.90	0.70
4	5.45	9.34	0.58
5	2.92	6.82	0.43
6	6.53	5.70	1.15
7	4.26	4.82	0.88
8	4.93	4.75	1.04
9	4.29	7.15	0.60
10	2.47	5.80	0.43

Notes. Column (2) shows the half-life for a given city-pair estimated using an equally weighted price index including 31 products. Column (3) is the average estimated half-life for a given city-pair taken across individual products. Column (4) is computed as the ratio between the figure in Column (2) and the figure in Column (1).