Journal of International Economics 56 (2002) 21–53
www.elsevier.com/locate/econbase

"The bigger they are, the harder they fall": Retail price differences across U.S. cities

Paul G.J. O'Connell[a], Shang-Jin Wei[b,*]

[a]FDO Partners, LLC, 5 Revere Street, Cambridge, MA 02138, USA
[b]Harvard University, Center for International Development, and Brookings Institution, 1775 Massachusetts Ave. N.W., Washington, DC 20036, USA

Received 11 January 1999; received in revised form 28 June 2000; accepted 5 July 2000

Abstract

This paper examines the evidence for nonlinear price behavior in retail goods prices across U.S. cities. First, a simple continuous-time model is used to explore the types of price behavior that can arise in the presence of market frictions. These frictions could be interpreted as transport costs, but we prefer a broader interpretation in which they operate at the level of technology and preferences. Second, we gather price data from 24 U.S. cities on individual goods like orange juice and toothpaste. The empirical analysis reveals that price discrepancies between U.S. cities are stationary and nonlinearly mean-reverting to price parity. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Law of one price; Retail goods prices

JEL classification: F31; C32

1. Introduction

The idea that market frictions are important for understanding deviations from the law of one price (LOP) is not a new one. Obstfeld and Taylor (1997) note that, as far back as 1916, Heckscher argued that transport costs should create some scope for price discrepancies to arise without precipitating goods arbitrage.

*Corresponding author. Tel.: +1-202-797-6023; fax: +1-202-797-6181.
E-mail address: swei@brook.edu (S.-J. Wei).

0022-1996/02/$ – see front matter
PII: S0022-1996(01)00111-8

Notwithstanding its vintage, it is only recently that attention has been devoted to developing the theoretical and empirical implications of this idea. Over the last decade, a number of authors have developed models that tease out the precise implications of market frictions for relative price behavior, and the results display a surprising degree of richness (see, for example, Obstfeld and Rogoff, 2000). On the empirical side, a growing body of work seeks to test the broad implications of these models, including nonlinear price convergence.

In this paper we seek to extend this literature by examining the behavior of retail goods prices across the U.S. We first set out a simple continuous-time model that highlights the relative importance of fixed and proportional market frictions in regulating price behavior. These frictions could be interpreted narrowly as transport costs. However, we prefer a broader interpretation in which they operate at the level of technology and preferences. Differences in demand and supply conditions across locations are apt to create price disparities, but these disparities are regulated by convergence in preferences and technology. Changes in preferences or technology are costly, however. For example, if consumers substitute away from expensive goods, they must research the properties of the substitutes. Technological progress on the supply side incurs up-front fixed costs of R&D. Alternatively, if consumers migrate to regions with lower living costs, they must pay the fixed costs of relocation (see O'Connell, 1997). We dub all such costs 'market frictions.' Our model incorporates both fixed and variable market frictions, and leads to the following conclusions. First, if only variable frictions are present, then the process for relative prices is confined between reflecting barriers that delimit a 'range of no action,' within which there is little incentive to alter technology or to look for substitutes. In these circumstances, when changes do take place, the quantity adjustments are very small, sufficient to prevent the price deviations from growing, but insufficient to shrink them. Second, if there are only fixed costs to altering technology or preferences, the process for relative prices is confined between 'resetting' barriers. These too delimit a band of no action. However, when preferences or technology do change, the changes are sufficient to completely eliminate the price deviation. Thus the process for the deviation is reset to zero. Third, if both fixed and variable costs are present, a hybrid case emerges. There are two bands for price deviations, an inner band within which no changes take place, and an outer band in which some changes to technology and preferences occur. Interestingly, all of these bands are increasing in the variability of relative goods prices.

These types of nonlinear behavior have specific implications for empirical analysis, and this is the second dimension along which we seek to add value. A shortcoming of the extant empirical work is that the data used are typically composite price indices such as the CPI. While these offer good availability and coverage, they provide no information on the absolute size of price discrepancies between locations, only their relative behavior over time. In addition, price indices are subject to aggregation bias, which may mask relevant features of the data. To address these issues, we employ data on disaggregated commodity prices across U.S. cities, yielding a pure measure of deviations from the law of one price. To summarize the results: (a) price discrepancies between U.S. cities are stationary, and indeed often behave in a manner consistent with the absolute law of one price; and (b) there is evidence of nonlinear reversion in the price discrepancies: large price disparities decay faster than small disparities.

At the outset, we should note that the link between the theoretical model and the empirical analysis is intentionally loose. The model generates some stylized predictions about continuous-time relative price behavior between two locations in the presence of market frictions, but it is by no means a complete description of price determination. The empirical section considers discrete-time multi-city data in a panel setting. Accordingly, we let the model stand on its own as one characterization of the effects of fixed and variable market frictions. The empirical section does not seek to test the model, but rather looks for general forms of nonlinear reversion in relative city prices.

The next section categorizes the links that can hold between retail prices in different locations. Then Section 3 sets out our model. In Section 4, we describe our data and empirical strategy. Finally, Section 5 tabulates the results.

2. The determinants of retail goods prices across cities

The goods represented in our dataset are described in Table 1. They run the gamut from basic foodstuffs such as bananas, milk, eggs, steak and potatoes, to services like drycleaning and hairstyling, to nationally-distributed recognized brands such as Winston cigarettes and Levi's jeans. The goods are divided coarsely by the proximity of production to the marketplace. Category A goods are generally not locally-produced, Category B goods may be locally-produced, and Category C goods are always locally-produced.

For the most part, the goods represented here are staples that are regularly purchased, rather than big-ticket items that one might purchase on an infrequent or once-off basis (e.g. cars, white goods, electronics). It is highly unlikely that a consumer would travel to a different location to purchase these items, or pay to have them shipped from another location.[1,2] Therefore, consumer price arbitrage is not compelling as a theory of price interdependence across cities for these goods. In order to tease out the potential links across city prices, we must delve deeper into the process of price determination. Engel and Rogers (1996), who examine

[1] As the retail internet sector has matured, consumers have displayed a willingness to engage in 'internet arbitrage' for some of these products — e.g. aspirin. However, the scope for this was limited during our time sample.

[2] Migration from expensive to cheap locations, a form of consumer arbitrage, is discussed below.

Table 1
Detailed description of price data[a]

Good	Start	Description
		Category A: Not locally-produced
Aspirin	82:2	Bayer; 325 mg tablets, 100 count
Babyfood	75:1	Jar strained vegetables; 4.5 oz.
Bananas	75:1	1 lb
Beer	82:2	Miller Lite or Budweiser; 12 oz., 6 pack
Cheese	82:2	Kraft; Parmesan, grated, canister, 8oz.
Cigarettes	75:1	Winston, king-size, carton
Coffee	75:1	Maxwell House, Hills Brothers or Folgers; 2 lbs, 1 lb, or 13oz.
Cornflakes	79:2	Kellogg's or Post Toasties; 18 oz.
Game	82:2	Monopoly; standard (No. 9) edition
Jeans	82:2	Levi's; straight leg, 501s or 505s
Liquor	75:1	Seagrams 7 Crown or A&B Scotch; 750 ml
Shirt	82:2	Arrow or Van Heusen; white, long sleeve, cotton-poly blend
Orange Juice	75:1	Can, 6 oz. or 12 oz.
Peaches	75:1	Del Monte or Libby's; #2.5 can (29 oz.), halves or slices
Shampoo	82:2	Johnson's or Alberto VO5; bottle, 11 oz. or 15 oz.
Shortening	75:1	Crisco; all vegetable, 3 lb. can
Soda	75:1	Coca-Cola; 1 quart or 2 litre
Tennis	82:2	Wilson or Penn; can, yellow, heavy duty, 3 count
Tissue	75:1	Kleenex; 1 roll, 4 roll or box, 175 count
Toothpaste	82.2	Crest or Colgate; 6 oz. or 7 oz.
Tuna	82.2	Starkist or Chicken of the Sea; in oil, can 6.5 oz.
Underwear	82:2	Package of 3 briefs
Detergent	75:1	Giant Tide, Bold or Cheer; 42 oz. or 49 oz.
Wine	82.2	Paul Masson Chablis or Gallo Sauvignon Blanc or Gallo Chablis Blanc; 750 ml or 1.5 litre
		Category B: May be locally-produced
Bacon	75:1	1 lb package
Bread	75:1	20 oz. or 24 oz.
Eggs	75:1	Grade A, 1 dozen
Minced steak	75:1	1 lb.
Lettuce	75:1	1 head
Margarine	75:1	1 lb.
Milk	75:1	Half-gallon
Potatoes	75:1	White or red, 10 lbs.
Steak	75:1	Round steak or T-bone; USDA choice, 1 lb.
Sugar	79:2	Cane or beet, 4 lbs. or 5 lbs.
Chicken	75:1	Grade A frying, 1 lb.
		Category C: Locally-produced
Fr. chicken	82:2	Kentucky Fried Chicken or Church's; breast and drumstick
McDonalds	82:2	Patty or patty with cheese, pickle, onion, mustard and ketchup
Pizza	82:2	Pizza Hut or Pizza Inn; 12″–13″ crust, regular cheese
App. repair	75:1	Service call for color TV or washing machine; excluding parts
Auto maint.	79:2	Balancing; 1 or 2 front wheels, computer or spin balance
Beauty	82:2	Shampoo, trim and blow-dry; women's visit
Bowling	75:1	Evening price; per line
Dentist	75:1	Office visit; cleaning and inspection, no X-ray or fluoride treatment
Doctor	75:1	Office visit; routine exam of existing patient
Dryclean	75:1	Man's suit; 2-piece
Hospital	75:1	Hospital room; semiprivate cost, per day
Haircut	75:1	No styling; man's
Movie	75:1	First run; indoor evening price

[a] Description of American Chamber of Commerce data as published in *Cost of Living Index*.

disaggregated consumer price indices across 14 U.S. and 19 Canadian cities, adopt the following general framework which is well-suited to our purposes.

Assume that all goods contain both tradable and nontradable components. Even goods that are normally classified as tradable, such as nonperishable commodities, rely on nontradable services such as marketing and distribution to reach the consumer. With Cobb–Douglas technology, price in location i, p_i, is determined by

$$p_i = \beta_i \alpha_i w_i^{\gamma_i} q_i^{(1-\gamma_i)}, \tag{1}$$

where β_i is the markup, α_i is productivity in the final goods sector, w_i is the price of the nontraded intermediate input, γ_i is its share in final output, and q_i is the price of the traded intermediate input. In this schema, q_i is determined largely by global demand and supply for the traded input. w_i, on the other hand, is a function of *local* demand and supply for the intermediate input. α_i and γ_i capture local technology conditions. Finally, β_i captures the mark-up over cost. We consider the influence of each of these in turn.

2.1. Traded intermediate inputs

By nature, traded intermediate inputs can be arbitraged, so the degree to which q_i can vary across locations is limited. However, recalling Heckscher, transport costs can support deviations from the law of one price in these inputs. Assume that the transport costs take the Samuelson 'iceberg form' — if a good is shipped from one location to another, a fraction l melts en route, so that only $(1-l)$ of the good actually arrives. These costs support a 'band of no-arbitrage' — an interval for the relative price q_i/q_j within which arbitrage doesn't pay — of $(1-l) < q_i/q_j < 1/(1-l)$.

A number of authors have developed models of traded goods price behavior that feature this type of no-arbitrage band, including Williams and Wright (1991); Dumas (1992); Uppal (1993); Sercu et al. (1995) and Ohanian and Stockman (1997). These models carry specific implications for the impact of arbitrage activity on relative prices. For instance, if the relative price q_i/q_j reaches the upper threshold generated by the transport costs, $1/(1-l)$, the amount of trade that takes place is sufficient to ensure that it does not rise above this level, but is not so large as to cause it to fall below $1/(1-l)$. In the jargon of continuous time, the thresholds $(1-l)$ and $1/(1-l)$ are 'reflecting barriers' for the relative price process. This means that evidence of arbitrage activity will not be reflected directly in prices, only in quantities. While the barriers render q_i/q_j stationary, it can be difficult to detect this stationarity using standard techniques. This is the basic motivation for much of the recent empirical work on nonlinear price behavior, including Obstfeld and Taylor (1997); Michael et al. (1997); O'Connell (1998b); Taylor and Sarno (1998); Baum et al. (1998) and Taylor (2001).[3]

How relevant are these considerations for the goods in our dataset? From Eq. (1), the importance of traded input prices, and hence transport costs, is inversely related to γ. For the Category A and Category B goods that are not locally-produced, the intermediate traded input is the final good at the wholesale level. The nontraded inputs are confined to local distribution services. For these goods, the relative price q_i/q_j is unlikely to stray outside the band permitted by transport costs. Because γ is relatively low for these goods, this type of nonlinearity can be reflected in the ratio of final prices p_i/p_j. However, cutting the other way is the fact that many of the Category A goods are branded — indeed, it is the very tradeability of these goods that facilitates branding. Branded goods are more likely to be differentiated from potential substitutes, and so the local markups can be more important than input costs in determining final price. We return to this below.

For the Category B and Category C goods that are locally-produced, arbitrage-induced constraints on q_i/q_j are less likely to be reflected in final prices. With local production, nontraded inputs and local technology play the dominant role in determining wholesale prices. That said, if locally-produced goods can be transported, then we might expect some arbitrage to take place at the wholesale level. For example, it is unlikely that the wholesale price of locally-grown potatoes in New England would exceed the wholesale price of imported Californian potatoes in New England for an extended period of time.

Arbitrage at the wholesale level is likely to involve both fixed and variable costs. For example, in order to supply the New England market with Californian potatoes, storage and distribution facilities need to be available, and the installation of such facilities is likely to incur fixed costs. The actual transportation of the goods then incurs variable costs.

2.2. Nontraded intermediate inputs

For many goods in our database, the share of nontraded intermediate inputs is significant. Examples include franchise restaurant meals (e.g. McDonalds), and dental and medical treatment. By definition, nontraded inputs cannot be arbitraged, and so w can vary substantially across locations. This allows for price dispersion that is increasing in γ.

However, the fact that a good has a substantial nontraded component does not imply that its relative price is independent across locations. Factor mobility should ensure that, in the long run, the ratio of nontraded prices w_i/w_j is stationary for many location pairs. This might be particularly relevant in the context of our data,

[3] It is worth noting that the elegant models of Dumas (1992) and Uppal (1993) also render predictions about the behavior of relative prices *within* the band of no arbitrage — thus they may be easier to test empirically.

as factors tend to be more intra-nationally mobile than they are internationally mobile. For example, the nontraded input often takes the form of labor services. It is an established result that one of the determinants of labor migration is the wage ratio w_i/w_j. Consequently, labor mobility should ensure that the wage ratio is stationary. However, as modeled in O'Connell (1997), there are costs associated with migration, and as a result persistent deviations from parity can certainly occur. These migration costs can be both fixed (e.g. actual moving costs) and ongoing (e.g. higher rent in new location). It is the fixed costs of migration that are most relevant in the context of the model we develop below.

Of course, relative nominal wage rates are not the only determinants of migration. Factors such as the cost of living, the quality of living, local services, climate and geography, inter alia, play their part. Not only can these factors drive a permanent wedge between w_i and w_j, changes in these factors can have a permanent effect on w_i/w_j, leading to nonstationarity in relative final goods prices p_i/p_j. We do not hold to the view that all relative nontraded prices are stationary, much less that they all display nonlinear convergence to parity. Rather we view these as viable competing hypotheses which we wish to test.

2.3. Local technology

As already discussed, even in those instances when the prices of intermediate inputs w and q are equalized, final prices can differ across locations if local production technologies are dissimilar. Clearly, this is most relevant for the Category B and C goods in our dataset that are locally-produced.[4]

To what extent can we expect technological convergence to eliminate such price gaps? Some technological differences, such as those stemming from geography or natural resources, are permanent in nature — it is unlikely that the price of locally-produced milk will ever be the same in Wisconsin and Florida.[5] However, other technology gaps can be closed through the transference of technology, or technological innovation in low-productivity regions. For locally-produced but nationally-branded goods, technological convergence is often assured. Franchise restaurant meals (e.g. McDonalds) are typically produced using nationwide or global production technologies. Productivity innovations in one location are rapidly disseminated to other locations. Indeed, this is an essential part of brand maintenance. For non-branded goods, competition can often ensure rapid technological convergence.

Once again, costs must be incurred to bring about this type of price conver-

[4]Of course, productivity in the distribution sector is relevant for those goods that are not locally-produced.

[5]The price of milk was regulated throughout our sample by a federal order program that supported artificially high milk prices to farmers in some regions of the country, while providing low prices for fluid milk produced in the Upper Midwest. We thank the co-editor for drawing this to our attention.

gence. If a producer decides that, to remain competitive, resources must be devoted to R&D, fixed costs are incurred. If instead a new patented technology becomes available to all producers, royalties on the patent represent a per-unit cost (to the producer). These types of fixed and variable costs are apt to generate a band of no arbitrage for final prices that is similar to the transport-cost induced band that applies to traded intermediate goods.

2.4. Price mark-up

It is clear that, for many of the goods in our database, the mark-up set by the producer will play a pivotal role in price determination. This is particularly true of the Category A (non-locally produced) goods. These tend to be nationally-distributed, nationally-priced brands. As a consequence, the supply curve for these goods is relatively similar across locations. The branding of these goods helps to ensure that they are differentiated from potential substitutes, which increases the producer's pricing power. Indeed, the producer may have monopoly pricing power. In these circumstances, prices would be set to maximize profits by equating β_i to $\varepsilon_i/(\varepsilon_i+1)$, where ε is the elasticity of demand. To the extent that producers have price power and demand conditions vary across locations, relative final prices p_i/p_j will depart from parity.

There are a number of points to make here. First, climate and principal industry are likely to give rise to some variation in tastes across locations. For example, aggregate tastes in a southern tourist destination like New Orleans may differ from those in a northern university town like Madison. However, it is not clear that tastes will vary systematically across locations for the staple goods in our dataset like toothpaste, aspirin and breakfast cereal. Moreover, where variations do exist, the ability of consumers to migrate from location to location can help to bring about convergence in preferences. This might be particularly relevant for cities that are quite close to each other, such as New York NY and Newark NJ. Of course, a consumer would be unlikely to move to take advantage of lower prices in just one or two goods: a significant number of the goods in her consumption bundle would have to have lower prices in the destination city.[6]

Second, for many of the goods in our database, the market structure is likely to admit some competition. The extent to which Bayer can price discriminate across consumers is limited by the fact that other aspirin producers will want to compete in the same markets. Competition is likely to be even more important for the goods in Categories B and C. The basic foodstuffs in Category B tend not to be branded, and so consumers have less allegience to specific producers. The non-branded services in Category C are likely to be provided by competing local-producers. And even franchise restaurant brands like McDonalds face stiff competition from

[6]The popularity of web-based cost-of-living calculators (e.g. `homefair.com`) bears witness to the importance of the overall cost of living in migration decisions.

other producers like Burger King. As such, the ability of producers to set the prices of the goods we examine may be limited.

In thinking about market structure, it is important to keep in mind the costs of entry and exit. Suppose that existing producers in a location have sufficient market power to set price above cost. Potential competitors seeking to join the market need to weigh the prospective profits against the costs of entry. Entry costs can be both fixed — for example, the cost of new premises — and flexible (Dixit, 1989). The existence of such costs can support deviations of relative prices from parity. These are the type of market frictions which the general model developed in the next section is designed to capture.

A third consideration is that mark-up is not always set with a view to single-period profit maximization. Retailers often alter mark-ups to control inventory, as with seasonal sales, or to build their customer base. Such alterations can generate large variations in relative prices. The scale and duration of such price variations is determined by supply rather than demand conditions. It is difficult to ascribe the negative serial correlation that sales produce to either producer or consumer arbitrage, and so this is one feature of price behavior that lies outside the theoretical framework considered here. We assess the importance of sales for our data in the empirical section below.

3. A model of arbitrage with transport costs

This discussion in Section 2 highlighted a variety of fixed and flexible market frictions that can lead to deviations from the law of one price. There are the fixed and variable costs of transporting traded intermediate inputs, the fixed costs of labor migration, the fixed and variable costs of technology, and the costs associated with entry and exit in the marketplace. Rather than seek to model each of these in a general model, we present a simple partial equilibrium model which allows us to explore the general implications of fixed and variable market frictions. The model could be generalized to include production, investment and a current account, but the qualitative nature of the predictions would be unchanged. The model draws from the excellent general discussion of optimal control and regulation in Dixit (1993).

3.1. The optimal pattern of trade

We examine the optimal trade strategy in a city that is endowed with two nonstorable commodities, X and Y. The representative agent in this city has the quasilinear utility function

$$U(C^X, C^Y) = C^X - \frac{1}{\gamma}\exp(-\gamma C^Y). \tag{2}$$

Thus the marginal utility of X is fixed at 1. The city is endowed with a nonstochastic supply of X at each instant. This endowment is abundant, so that there will always be some of X consumed.[7] The endowment of Y is stochastic at each instant, with dynamics given by

$$dY = \sigma\, dz, \tag{3}$$

where dz is the increment of a standard Wiener process. Consumption C^Y is equal to the endowment plus net imports of the good M from other cities:

$$C^Y = Y + M. \tag{4}$$

We assume that X and Y are traded in all locations, and that both are priced globally at unity. Moreover the city's trade balance with other cities is always zero.[8] Together these assumptions imply that

$$C^X = X - M. \tag{5}$$

Our ultimate interest is in the process for P, the price of good Y in terms of X in this city. Since the marginal utility of X is one, $P = \exp(-\gamma C^Y)$.

We capture fixed and variable market frictions through the transportation technology available to the economy, which has the following features. First, in order to facilitate imports or exports, trade capacity must be available. We assume that trade capacity is unidirectional — capacity installed for the purposes of exporting cannot be used for imports. The per-unit cost of new capacity is l. Once installed, capacity must be utilized. Capacity may, however, be decommissioned at a cost. To preserve symmetry, we assume that the per-unit decommissioning cost is also l. Second, every time that trade capacity is adjusted, a fixed cost k must be expended. This setup lends itself naturally to a transport cost interpretation, but it also serves as a characterization of entry and exit costs, and the fixed costs of R&D or migration.

In the absence of market frictions, the solution to the representative agent's maximization problem is straightforward. At each instant, the marginal utility from consumption of Y must equal 1, so the agent simply sets M equal to $-Y$. In the presence of market frictions, however, C^Y can deviate from 0. To obtain the solution, it is convenient to work with a utility loss function rather than the level of utility itself. The loss function is defined as

$$L(C) = -\frac{1}{\gamma} + C + \frac{1}{\gamma} e^{-\gamma C} \tag{6}$$

[7] In this sense, the model is partial equilibrium. The endowment of Y is assumed to be insignificant in size relative to the endowment of X.

[8] Implicit here is the assumption that the economy's rate of time preference is equal to the global interest rate. If this were not the case, then there would be no well-defined equilibrium, as marginal utility is not diminishing in X.

where $C \equiv C^Y$ (the Y superscript is dropped hereafter to economize on notation). $L(C)$ captures the relative gain in utility from consuming at 0 instead of consuming at C.[9] It is a convex function that attains a minimum 0 at $C = 0$. The reason for working with this loss function is that it is invariant to the scale of the endowment $Y(t)$. Clearly, *total* utility depends on this endowment, but given quasilinear preferences and the abundance of X the marginal import decision is independent of this endowment, and so nothing is lost by seeking to minimize $L(C)$ rather than to maximize $U(\cdot)$.

Holding M constant, the process for $C \equiv Y + M$ is simply

$$\mathrm{d}C = \sigma\,\mathrm{d}z. \tag{7}$$

Define the value function

$$V(C) = \min_{\{M\}} E_C \int_0^\infty \mathrm{e}^{-\phi t} L(C)\,\mathrm{d}t, \tag{8}$$

where ϕ is the discount rate. The problem is to solve for $V(C)$, taking into account the costs associated with trade. The Hamilton–Bellman–Jacobi equation for this problem is $E[\mathrm{d}V(C)]/\mathrm{d}t + L(C) = \phi V(C)$. Applying Itô's lemma, this can be written as

$$\frac{1}{2}\sigma^2 V_{CC}(C) - \phi V(C) + L(C) = 0. \tag{9}$$

This differential equation has the well-known solution (see, for example, Dixit, 1993)

$$V(C) = A\,\mathrm{e}^{-\alpha C} + B\,\mathrm{e}^{\alpha C} + \frac{e^{-\gamma C}}{\gamma\left(\phi - \frac{1}{2}\sigma^2\gamma^2\right)} + \frac{1}{\phi}\left(C - \frac{1}{\gamma}\right), \tag{10}$$

where $\alpha = \sqrt{2\phi}/\sigma$, and A and B are constants to be determined. The last two terms on the right-hand side comprise the present value of the loss function if net imports are fixed — they are the discounted sum of expected losses $L(C)$. The first two terms represent the change in value that accrues from the ability to control M: the value of the option to import is $A\,\mathrm{e}^{-\alpha C}$, while the value of the option to export is $B\,\mathrm{e}^{\alpha C}$. These quantities are negative as they add to utility and hence subtract from the loss function.

It can be shown (see Dixit, 1993, and the references cited therein) that the optimal trade policy is characterized by four threshold levels of C, $\overline{C}_1 > \overline{C}_2 > \overline{C}_3 > \overline{C}_4$, that have the following features. First, if C rises to $\overline{C}_1$, $(\overline{C}_1 - \overline{C}_2)$ of net export capacity is installed at cost $k + (\overline{C}_1 - \overline{C}_2)l$, and used to export an additional

[9] Consuming at 0 yields $-1/\gamma$ plus Y export revenue, while consuming at C yields $-\exp(-\gamma C)/\gamma$ plus $-M$ export revenue. The difference between these is $L(C)$.

$(\bar{C}_1 - \bar{C}_2)$ of Y to other cities.[10] Second, if C falls to $\bar{C}_4$, $(\bar{C}_3 - \bar{C}_4)$ of net import capacity is installed, and used to import an additional amount $(\bar{C}_3 - \bar{C}_4)$ from other cities. In other words, if C strays too far from its optimum of 0, the representative agent 'resets' it at a value that is closer to the optimum by raising or lowering net imports. To solve for these threshold values, we employ the usual value-matching and smooth-pasting conditions to tie down the constants A and B.

3.2. The behavior of relative prices

For a benchmark case, let the coefficient of absolute risk aversion γ equal 1, the instantaneous variance of the endowment σ equal 0.01, and the rate of discount ϕ equal 0.05. With these fixed, we can examine the behavior of P under varying assumptions.

3.2.1. Costless trade

If there are no market frictions ($k = l = 0$), all four trading thresholds collapse to zero. Trade takes place instantaneously whenever the endowment of C differs from 0. P is therefore fixed at unity.

3.2.2. Infinite costs of trade

If the market frictions are prohibitive, $k \to \infty$ and/or $l \to \infty$. In this case, $\bar{C}_1$, $\bar{C}_2 \to \infty$, $\bar{C}_3$, $\bar{C}_4 \to -\infty$, and no trade takes place. The relative price P will equal $U_C(C) \equiv \exp(-\gamma c)$ at all instants. By Itô's lemma,

$$dP = \frac{1}{2}\sigma^2\gamma^2 P\, dt - \sigma\gamma P\, dz. \tag{11}$$

This implies that $p \equiv \ln(P)$ follows a driftless arithmetic Brownian motion with instantaneous variance $\sigma^2\gamma^2$:

$$dp = \sigma\gamma\, dz. \tag{12}$$

3.2.3. Proportional costs of trade

If all market frictions are proportional (i.e. $k = 0$), then $\bar{C}_1$ coincides with $\bar{C}_2$, and $\bar{C}_3$ coincides with $\bar{C}_4$. In our benchmark case, the solutions for these thresholds are $\bar{C}_1 = 0.283$, and $\bar{C}_4 = -0.256$. The resulting process for consumption shares many of the features of the solution in the Dumas (1992) model. In particular, $\bar{C}_1$ and $\bar{C}_4$ become *reflecting* barriers for the consumption process. If trade takes place, it will involve infinitesimal quantities at these barriers. The process for U_C and hence P will inherit these properties. Thus P will follow the process (11) until such time as C reaches one of the barriers, when sufficient trade will take place to hold P at its barrier level, without driving it back towards parity. In our benchmark case, the reflecting barriers for P are $\exp(-0.283\gamma) = 0.753$ and $\exp(0.256\gamma) = 1.292$.

---

[10] If the city is already importing, then some of its import capacity will be decommissioned.

3.2.4. Fixed costs of trade

If the market frictions are fixed (i.e. $l = 0$), then the middle two thresholds $\overline{C}_2$ and $\overline{C}_3$ coincide. With the benchmark parameters, the threshold solutions are $\overline{C}_1 = 0.688$, $\overline{C}_2 = \overline{C}_3 = 0.019$, and $\overline{C}_4 = -0.600$. The resulting consumption process differs markedly from the case of proportional trade costs. In particular, whenever C hits either $\overline{C}_1$ or $\overline{C}_4$, a discrete amount of trade will take place that is sufficient to bring consumption back to 0.019.[11] The intuition is that, once the fixed cost has been expended, it would be suboptimal to reset consumption to a point that is away from the value-maximizing point close to 0.

The processes for U_C and hence P inherit these resetting features. Thus P will follow the process (11) until such time as C reaches one of the barriers, at which time P is reset to $\exp(-0.019\gamma) = 0.981$. The resetting barriers for P are 0.549 and 1.990.

3.2.5. Fixed and variable costs of trade

Lastly, we consider the behavior of P in the presence of both fixed and proportional market frictions. In the benchmark case, when P hits $\exp(-\gamma\overline{C}_1) = 0.461$, arbitrage moves it to $\exp(-\gamma\overline{C}_2) = 0.890$. Correspondingly, when P hits $\exp(-\gamma\overline{C}_4) = 2.168$, arbitrage resets it to $\exp(-\gamma\overline{C}_3) = 1.070$. The interesting aspect of this solution is that it generates two 'bands' for the deviations from the LOP. Whenever Y hits the outer barriers, it is reset by trade to the inner barriers. This resetting behavior differs from the infinitesimal arbitrage that characterizes models predicated solely on proportional transport costs.

3.3. Implications for testing the LOP

The model developed above has two important implications for empirical analysis of inter-city prices. First, in the presence of market frictions, the stationarity of relative prices may be difficult to detect using conventional tests. Second, by taking advantage of the special structure of price behavior that arises with market frictions, the power to detect stationarity can be increased.

3.3.1. Detecting stationarity in the presence of market frictions

Let the true process for the relative price of a good in two locations, q, be

$$q_t = \begin{cases} b + \epsilon_t & \text{if } q_{t-1} < -a \\ q_{t-1} + \epsilon_t & \text{if } |q_{t-1}| < a \\ b + \epsilon_t & \text{if } q_{t-1} > a \end{cases} \tag{13}$$

[11] Notice that the value-maximizing point is not actually 0. This is because of the asymmetry introduced by the Itô or Jensen's inequality term in the value function.

where $0 < b < a$. This process resembles the one that emerges from our model in the presence of both fixed and variable frictions.[12] The process is globally stationary, but because innovations to the process are i.i.d. for a portion of the time (i.e. whenever $|q_{t-1}| < a$), this stationarity can be difficult to detect using standard techniques. For example, if $a = 4$, $b = 3$ and $\epsilon \sim$ i.i.d.$N(0, 1)$, the power of the Dickey–Fuller test to reject the random walk null with 50 observations on this process is 22 percent at the 5 percent significance level.[13] This can be compared to the power of the Dickey–Fuller test when the true process is $AR(1)$ with the same total variance $V(q_t)$ as the process (13). When $a = 4$, $b = 3$ and $\epsilon \sim$ i.i.d.$N(0, 1)$, $V(q_t) = 3.42$.[14] This is matched by an $AR(1)$ process with a disturbance variance of 1 and a root of 0.84.[15] The power of the Dickey–Fuller test to reject the random walk null under this $AR(1)$ alternative is 37 percent at the 5 percent significance level, 15 percent higher than for process (13).

The fact that standard tests for stationarity have low power under the alternatives generated by market frictions may account for the perennial difficulty of rejecting the unit root null in relative price data.[16]

3.3.2. Increased power to detect stationarity

There is, of course, a positive side to the particular structure that market frictions imply for relative price behavior. It is likely that the power of stationarity tests can be increased by modifying them to take account of the fact that reversion in prices only takes place at certain times. Models of market frictions predict that small price discrepancies will not be arbitraged, but that large ones will. This suggests conditioning reversion on the size of the deviation from the LOP. To illustrate, suppose that only the observations on Δq_t for which $|q_{t-1}|$ exceeds some threshold s are included in the Dickey–Fuller regression. In other words, small deviations from the LOP are excluded from the regression. If the true process for q is (13), then these observations contain little or no information on reversion in q; they just add noise to the estimation. It follows that a more precise estimate of reversion is available from the test, which ought to increase power. This in fact turns out to be the case. If, for example, we choose to look only at the upper

[12] The analogy is not exact. In the continuous time version, p can never stray outside the edge of the outer band a.

[13] This power calculation is carried out by Monte Carlo simulation, using 1000 observations on the distribution of the test statistic under the alternative (13). The Dickey–Fuller regression run here includes an intercept.

[14] $V(p_t)$ is estimated from 1000 simulations of the process (13).

[15] The total variance of an $AR(1)$ process $p_t = \phi p_{t-1} + \epsilon_t$ is $\sigma_\epsilon^2/(1 - \phi^2)$.

[16] Taylor (2001) analyzes this point in depth. 'Standard' tests for relative price stationarity have been carried out by Wei and Parsley (1995); Engel et al. (1996); Frankel and Rose (1996); Papell (1997); Taylor (1996); Papell and Theodoridis (1997a) and O'Connell (1998a). See Froot and Rogoff (1995) and Rogoff (1996) for surveys of the literature on PPP.

quartile of LOP deviations, the power to reject nonstationarity under the alternative (13) is 31 percent.[17]

Recent empirical work has sought to modify standard tests of the LOP and purchasing power parity (PPP) to take account of actual or potential market frictions. Using the same data as is used in this paper, Parsley and Wei (1996) carry out a basic test for nonlinear reversion by adding a higher-order term to the standard Dickey–Fuller regression. Their point estimates suggest that convergence is faster for large initial price differences. However, as shown in O'Connell (1998b), the power of such tests is weak under the unit-root null, and on this basis, newly-developed threshold models are to be preferred. Obstfeld and Taylor (1997) fit such models to detrended real exchange rates for the U.S. sampled from 1973 to 1995. They report evidence that large deviations from the linear trend of the real exchange rate revert to parity quite quickly, while small deviations do not. Taylor and Sarno (1998) fit smooth transition autoregressive (STAR) models to British, German, French and Japanese real exchange rates measured over the 1973–1996 period, and detect significant evidence of nonlinear mean reversion. Michael et al. (1997), employing a substantially similar technique, also detect nonlinear mean reversion using interwar CPI data, and for a 200-year data set of UK and French real CPI exchange rates against the dollar. Similar results are reported by Baum et al. (1998).

A potential problem with all of this work (bar Parsley and Wei, 1996) is that it is based on price indices. This may lead to some aggregation bias. For example, if there are different market frictions associated with each good in the price index, then the limits of the band of no arbitrage for the index will not be clearly defined. In an effort to circumvent this problem, we use detailed price data on individual goods in our empirical analysis, to which we now turn.

4. Empirical analysis

4.1. Data

Our data set is substantially the same as that used in Parsley and Wei (1996). The data is collected from the American Chamber of Commerce Researchers Association publication, *Cost of Living Index* (hereafter, *Index*). Each quarterly issue of *Index* contains comparative average price data for a sample of urban areas, and a cost of living index computed from these data by the Association. In this study we use only the raw price data.

The actual data surveys are conducted by local Chamber of Commerce staff, and responses are voluntary. Explicit instructions and data forms are provided for each

[17] This power calculation is estimated from 5000 simulations of the test statistic under the null, and 1000 simulations under the alternative.

data collector by the association. Some prices are obtained by phone and usually the respondents do not know it is for a survey. Once collected, the data is sent to one of nine different regional coordinators for checking. Finally, the data is sent to Houston where it is transferred to computer and subjected to both computer and visual checks for outliers. Publication occurs approximately five and one half months after the original data are collected.

The sample of cities included in each issue of *Index* varies. At the beginning of our sample period there were 166 cities and 44 items priced. The number of cities steadily increased to 297 by the fourth quarter of 1992. However, each report contains a distinct sample of cities. We choose a sample of 24 cities which appeared in roughly ninety percent of the quarterly surveys. The cities selected can be seen in the first Column of Table 3. Notice major cities like New York, Chicago, Los Angeles, San Francisco, Washington and Boston are excluded from the analysis. There is probably some merit in this as demand and supply conditions are likely to vary with city size.

As already described in Section 2, we choose to examine 48 goods and services. These are selected with three criteria in mind. First, for each commodity we want wide coverage in terms of availability across cities and over time. Second, we want variation in the degree of tradeability of the commodities included in the data set. Finally, we want homogeneity in the definitions of the commodities over time. The definitions of some commodities did change during the sample period, typically as a result of a change in manufacturer packaging. These changes had only small effects on *relative* prices.

Inspecting the data, we find little evidence of transitory changes in mark-up. The exception is the price series for a hamburger sandwich at McDonalds. In a number of cities, notably Indianapolis, the price of these sandwiches is episodically cut to 99 cents. This can represent a price cut of up to 50 per cent, but the change lasts no more than one quarter. These changes aside, the data appear to be 'well-behaved,' in the sense that there are few outliers and the data generating processes appear stable. To quantify this, we measure the fraction of prices that are 15 per cent or more less than the preceding and following quarter's prices. For the goods in Category A, the fraction is 0.5 per cent, for Category B, 1 per cent, and for Category C, 0.25 per cent.[18]

4.2. Construction of relative prices

Mindful of the problem of low power that afflicts many empirical analyses of the LOP, we conduct our analysis in a panel setting. We construct two different sets of relative price panels. The first set groups commodities by type. Thus for

---

[18]We thank the referee and co-editor for their comments on this point. The co-editor suggested the quantitative definition of a 'sale' used here, which is "any time a nominal price falls from one quarter to the next by 15%, and returns to within 3% of its former value in the subsequent quarter."

each of the 48 commodities in the data set, we construct a panel that contains all the price series that are available for that commodity across the country. The second set of panels groups commodities by location. In this case, the panels contain all the price series that are available for a given city. Within these two types of panels, we further subdivide by the Categories A, B and C identified in Table 1.

The absolute prices that are included in each panel must be converted to relative prices in order to test the LOP. This requires choosing a numeraire for each panel. For some of the tests run, the choice of numeraire will prove immaterial, but for others it will make an important difference. Accordingly we choose the numeraire with a eye to its economic meaning. For the panels which group commodities by type, we choose the *average price across all cities* as the numeraire. So for example, each series in the panel of aspirin prices is calculated as

$$q_{\text{Aspirin},jt} = p_{\text{Aspirin},jt} - \frac{1}{M}\sum_{j=1}^{M} p_{\text{Aspirin},jt}, \tag{14}$$

where $p_{\text{Aspirin},jt}$ is the price of aspirin in city j at time t, and M is the total number of cities for which aspirin price series are available. One of the $q_{\text{Aspirin},j}$ series is redundant, and we arbitrarily choose this to be the series for Louisville, Kentucky in all of the panels. For the panels which groups commodities by location, we choose prices in New Orleans as the numeraires. This is because the data for New Orleans are relatively complete. So, for example, each series in the Houston panel is constructed as

$$q_{i,\text{Houston},t} = p_{i,\text{Houston},t} - p_{i,\text{New Orleans},t}. \tag{15}$$

Here i indexes each of the commodities. Note that because we will use GLS to estimate our parameters, these numeraire choices are relatively innocuous. GLS naturally reduces the sensitivity of estimates to the numeraire; indeed, in the case of unit-root testing, GLS often renders the estimates invariant to the numeraire. In all panels, we delete series that have fewer than 43 quarterly observations. This yields balanced panels, which simplifies the empirical analysis substantially. After deleting these series, 20 cities remain in the location panels.

4.3. Empirical tests

Three types of tests are carried out on each panel of price data. First, each is tested for stationarity by means of a GLS panel unit root test. Second, a threshold autoregression (*TAR*) model is fitted to the panels. This model allows for a discrete change in the strength of reversion once relative prices reach a certain departure from parity. Finally, an exponential smooth threshold autoregression (*ESTAR*) model is estimated for each panel. Favored in recent work on nonlinear price reversion, this model allows for a smooth increase in the rate of reversion as departures from price parity grow.

In order to properly size our tests, it is essential to control for both contemporaneous and serial correlation in the data. O'Connell (1998a) shows that the failure to control for contemporaneous correlation in testing for reversion in relative prices can lead to very serious size distortions, while Papell (1997) demonstrates the importance of accounting for serial correlation. To gauge the serial correlation present in our data, we follow Taylor and Sarno (1998) in examining the sample partial autocorrelation functions of the relative price series.[19] For the Category A goods, none of the partial autocorrelations for lags four and higher are significant, while for the Category B and C goods, there is some evidence that the fourth lag is important. Taking into account that there may be seasonal effects at the quarterly frequency, we choose to allow for serial correlation of order four in our analysis.[20]

The data-generating process (DGP) for all three models can be characterized as follows

$$\Delta(q_{ijt} - \mu_{ij}) = \sum_{k=1}^{4} \phi_k(q_{ij,t-k} - \mu_{ij}) + F(q_{ij,t-d} - \mu_{ij}) \sum_{k=1}^{4} \phi_k^{*}(q_{ij,t-k} - \mu_{ij}) + \epsilon_{ijt}, \quad t = 1, \ldots, T, \tag{16}$$

where $\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, and the type of panel (good or location) is selected by holding either i or j constant. The ϕ parameters are assumed to be the same across all relative prices in a panel. Not only does this greatly increase statistical power, it seems like a natural constraint to impose, given that the factors that influence the member series are likely to be similar.[21] Moreover, this restriction guarantees that the unit root test is 'numeraire invariant.' That is to say, the same estimates of the ϕ coefficients will be obtained no matter which price series is chosen as the numeraire in Eq. (14) (see O'Connell, 1998a).

The key element of the DGP is the transition function $F(q_{ij,t-d} - \mu_{ij})$, which dictates how the behavior of relative prices changes as price disparities grow. The 'delay parameter' d determines the interval over which the transition takes place. Under the random walk model, $F(q_{ij,t-d} - \mu_{ij}) \equiv 0$, and the DGP takes the form of a modified panel Dickey–Fuller specification, with lagged levels rather than lagged changes of the dependent variable included in the regression. In this case,

[19] Standard information criteria tend to penalize high-order lags too much. On the other hand, likelihood ratio tests based on estimation of the multivariate data-generating process under the null can attribute too much significance to high-order lags when the null is false. The partial autocorrelation function appears to offer a reasonable balance as a data-dependent lag selection procedure.

[20] Partial autocorrelation results are available from the authors on request.

[21] The variances and partial autocorrelations of the underlying relative price series display relative homogeneity. Moreover, as shown in the tables of results to follow, the mean departures from absolute price parity are relatively small. Accordingly, the symmetry restriction does not seem unjustified.

the null hypothesis is that the root of the process is unity, or $\Sigma_{k=1}^{4}\phi_k = 1$, and the alternative is that the root is less than one. For both the *TAR* and *ESTAR* models, $F(q_{ij,t-d} - \mu_{ij})$ is a symmetric convex function, centered on μ_{ij}. For the *TAR* specification,

$$F(q_{ij,t-d} - \mu_{ij}) = \begin{cases} 0 & \text{if } |q_{ij,t-d} - \mu_{ij}| < c \\ 1 & \text{otherwise} \end{cases} \tag{17}$$

c is the distance from equilibrium at which the transition function 'switches on,' causing the rate of reversion to undergo a discrete change. In the *ESTAR* formulation,

$$F(q_{ij,t-d} - \mu_{ij}) = 1 - \exp[-\theta(q_{ij,t-d} - \mu_{ij})^2]. \tag{18}$$

In this case, the transition function changes smoothly from 0 when $q_{ij,t-d}$ is at the equilibrium μ_{ij} to 1 as $(q_{ij,t-d} - \mu_{ij}) \to \infty$. The parameter θ determines the speed at which the transition takes place.

The inclusion of the mean μ_{ij} merits separate discussion. One of the distinct advantages of our data is that it affords measures of absolute rather than relative price differences. As such, we might consider testing for stationarity without the inclusion of an intercept. A problem that arises, however, is that the small-sample distribution of ϕ is affected by the initial value of the relative price, q_{ij0}. We could control for this by simulation, initializing the bootstrap samples at the observed initial prices. However, this would render our tests conditional on the particular initial values of our data. Accordingly, we allow for an intercept in each relative price series. The test for stationarity is still based on ϕ. Conditional on this, it will be informative to examine whether the μ_{ij} terms differ from zero.

The ϕ coefficients are estimated in all cases by maximum likelihood. A consistent estimate of the covariance matrix $\hat{\mathbf{\Sigma}}$ is obtained from ordinary least squares residuals, and initial estimates of ϕ, ϕ^* and the transition function F are then calculated by minimizing the sum of squares

$$SSE = \sum_{r=1}^{N}\sum_{s=1}^{N} \hat{\sigma}_{rs}\epsilon'_{rt}\epsilon_{st}, \tag{19}$$

where $\hat{\sigma}_{rs}$ is the r,sth element of $\hat{\mathbf{\Sigma}}^{-1}$. Iterating these steps to convergence produces the maximum likelihood estimates. The transition parameter c in the *TAR* model renders the likelihood function nondifferentiable, and so conditional on a set of candidate estimates for the other parameters, it is estimated by grid-search. The delay parameter d is also estimated by a simple grid search over {1, 2}.

For the unit root test, in the absence of contemporaneous correlation, the sampling distribution for the root of the process would be well-approximated by that tabulated in Levin and Lin (1992). However, with nonzero correlations clearly present in the data, critical values must be obtained by simulation. Toward this end, a simulation DGP is constructed by modifying the fitted *VAR*(4) process.

Table 2
Estimated models for goods not locally-produced (Category A) grouped by commodity type[a]

Good	Random walk model				*TAR* model				*ESTAR* model			
	T	N	$\overline{\lvert\mu\rvert}$	Φ	c	Φ	Φ^*	ΔAIC	θ	Φ	Φ^*	ΔAIC
Aspirin	43	24	0.067	0.611	0.12	0.735	−0.024	−21.4	61.35	0.69	0.05	−19.3
			11	(0.016)		(0.067)	(0.039)		(45.02)	(0.04)	(0.05)	
Babyfood	72	23	0.067	0.791	0.11	0.766	0.088	−21.0	76.95	0.96	−0.20	−32.3
			12	(0.035)		(0.040)	(0.038)		(39.56)	(0.04)	(0.05)	
Bananas	72	23	0.070	0.410	0.1	0.502	−0.156	−6.6	52.51	0.51	−0.14	−3.3
			14	(0.000)		(0.068)	(0.060)		(71.01)	(0.06)	(0.07)	
Beer	43	24	0.071	0.694	0.06	0.581	−0.022	−30.3	76.35	0.66	−0.02	−9.0
			17	(0.274)		(0.079)	(0.036)		(95.46)	(0.04)	(0.04)	
Cheese	43	22	0.063	0.740	0.09	0.562	0.068	−64.0	64.88	0.59	−0.01	−33.3
			13	(0.275)		(0.053)	(0.040)		(76.95)	(0.04)	(0.08)	
Cigarettes	72	23	0.051	0.796	0.12	0.848	−0.440	−3.1	99.00	0.82	0.09	−9.9
			19	(0.000)		(0.020)	(0.289)		(96.16)	(0.02)	(0.05)	
Coffee	72	22	0.048	0.687	0.12	0.652	−0.007	−22.8	78.46	0.69	−0.05	−13.7
			8	(0.004)		(0.044)	(0.037)		(62.71)	(0.04)	(0.05)	
Cornflakes	55	21	0.055	0.521	0.12	0.632	−0.158	−16.5	85.55	0.61	0.02	−18.9
			17	(0.000)		(0.054)	(0.102)		(64.15)	(0.05)	(0.09)	
Game	43	24	0.087	0.463	0.06	0.809	−0.298	−35.9	41.23	0.64	−0.07	−9.3
			20	(0.007)		(0.080)	(0.054)		(42.71)	(0.04)	(0.05)	
Jeans	43	24	0.064	0.461	0.07	0.624	−0.238	−35.7	49.61	0.62	−0.16	−14.4
			18	(0.008)		(0.084)	(0.068)		(53.68)	(0.06)	(0.09)	
Liquor	72	22	0.062	0.773	0.07	0.805	0.028	−9.2	61.40	0.79	0.06	−5.6
			14	(0.010)		(0.028)	(0.025)		(81.86)	(0.02)	(0.04)	
Shirt	43	24	0.070	0.357	0.05	0.574	−0.141	−45.1	43.11	0.61	−0.02	−25.3
			17	(0.000)		(0.109)	(0.077)		(40.07)	(0.05)	(0.07)	
Orange Juice	72	20	0.049	0.498	0.07	0.743	−0.068	−7.3	73.48	0.80	−0.14	−10.2
			14	(0.000)		(0.059)	(0.053)		(61.80)	(0.06)	(0.06)	
Peaches	72	23	0.060	0.583	0.06	0.773	−0.021	−30.7	73.61	0.71	0.04	−26.1
			13	(0.000)		(0.047)	(0.043)		(41.51)	(0.04)	(0.05)	

Shampoo	43	24	0.067	0.446	0.06	0.593	−0.255	−34.4	73.99	0.47	−0.06	−16.8
			14	(0.001)		(0.114)	(0.088)		(63.15)	(0.06)	(0.07)	
Shortening	72	23	0.056	0.723	0.12	0.758	−0.083	−9.7	79.34	0.81	−0.12	−11.4
			12	(0.013)		(0.036)	(0.034)		(66.11)	(0.04)	(0.04)	
Soda	72	22	0.072	0.705	0.09	0.544	−0.014	−12.5	49.83	0.64	−0.06	−9.1
			7	(0.023)		(0.070)	(0.052)		(44.78)	(0.05)	(0.05)	
Tennis	43	24	0.152	0.580	0.1	0.898	−0.021	−19.9	39.06	0.71	−0.09	−10.4
			19	(0.010)		(0.033)	(0.039)		(34.58)	(0.04)	(0.04)	
Tissue	72	22	0.050	0.563	0.06	0.608	0.021	−16.5	70.71	0.67	−0.09	−12.1
			12	(0.000)		(0.064)	(0.061)		(52.46)	(0.04)	(0.04)	
Toothpaste	43	24	0.073	0.523	0.12	0.443	0.194	−59.5	75.96	0.54	0.14	−16.1
			19	(0.006)		(0.054)	(0.029)		(63.51)	(0.05)	(0.05)	
Tuna	43	23	0.075	0.610	0.05	0.911	−0.438	−18.2	42.51	0.57	−0.20	−9.0
			14	(0.098)		(0.152)	(0.148)		(54.62)	(0.06)	(0.11)	
Underwear	43	24	0.115	0.649	0.05	0.418	0.045	−18.9	41.92	0.64	−0.08	−4.4
			16	(0.100)		(0.153)	(0.128)		(70.42)	(0.06)	(0.06)	
Detergent	72	21	0.094	0.732	0.12	0.644	0.079	−3.0	45.84	0.69	−0.03	−3.4
			16	(0.021)		(0.041)	(0.038)		(68.42)	(0.04)	(0.06)	
Wine	43	24	0.122	0.131	0.11	0.625	−0.186	−29.1	48.25	0.55	−0.12	−19.2
			23	(0.000)		(0.065)	(0.046)		(32.48)	(0.05)	(0.06)	

[a] Estimated parameters for the random walk, *TAR* and *ESTAR* models. $\overline{|\mu|}$ is the mean absolute value of the estimated equilibrium for each relative price series in the panel, under the liner model. As the data measure absolute price differences, our prior is that $\overline{|\mu|}=0$. The integer number below each estimate of $\overline{|\mu|}$ reports the number of estimated μ values that are statistically different from zero in the panel. For each model, Φ is shorthand for $\Sigma_{k=1}^{4}\phi_k$, and similarly $\Phi^* \equiv \Sigma_{k=1}^{4}\phi_k^*$. c is the estimated transition parameter for the *TAR* model, and θ is the estimated transition parameter for the *ESTAR* model. All estimates are obtained by maximum likelihood. For the random walk model, *P*-values for the random walk null are shown below Φ, while for the other models the parenthetical numbers are standard errors. As the models are not nested, the Akaike Information Criterion (*AIC*) is used to compare their overall goodness of fit. The columns labeled ΔAIC show, for the *TAR* and *ESTAR* models, the change in the *AIC* relative to the linear *AR*(4) model. Negative values in these columns indicate improved goodness-of-fit relative to the linear case.

Specifically, $(1 - \Sigma_{k=1}^{4} \hat{\phi}_k)/2$ is added to both $\hat{\phi}_1$ and $\hat{\phi}_2$, thereby forcing the root of the process to unity. The resultant data-generating process is used to generate 5000 panels of relative prices under the null, and the sampling distribution of $\Sigma_{k=1}^{4} \phi_k$ is then obtained by running *FGLS* on each simulated panel. For the *TAR* and *ESTAR* models, conditional on stationarity, standard inference procedures are aysmptotically valid. In order to contrast the overall fits of each model, we use the Akaike Information Criterion (*AIC*).

5. Empirical results

The empirical results are presented in six tables. Tables 2 and 3 look at the Category A goods, Tables 4 and 5 at Category B, and Tables 6 and 7 at Category C. In order to economize on notation, we use the shorthand $\Phi = \Sigma_{k=1}^{4} \phi_k$ and $\Phi^* = \Sigma_{k=1}^{4} \phi_k^*$.

5.1. Results for goods not locally-produced (Category A)

The results for the linear model (Columns 4) leave one with the overall impression that the goods in this category are stationary. For 20 of the 24 goods in Table 2, the random walk null is rejected at at better than the 5 per cent level, and it is rejected at the 10 per cent level for a further two goods. The estimated roots are quite low, averaging 0.60, suggesting that under linear reversion, deviations from parity decay very rapidly indeed, with half-lives under 2 quarters. This can be contrasted with the commonplace assumption that international deviations from parity have half-lives on the order of 4–5 years. The non-rejections are for beer and cheese. It is noteworthy that all four of the panels that produce nonrejections at the 5 per cent level have only 43 time series observations, and hence offer lower statistical power. The beer and cheese panels are also interesting in that they display the third and fourth lowest absolute price variation (the lowest variation is in cigarette prices, followed by liquor prices). One consistent interpretation of the nonrejections is that price disparities for these goods have simply not been large enough to cause convergence of the relative price series.

Accepting the alternative hypothesis of stationarity, the question arises as to whether convergence is to zero, as the theory of absolute price parity would predict. Approximately 40 per cent of the estimated equilibrium prices for all series (Column 3) are statistically indistinguishable from zero. The mean absolute deviation of the equilibria from zero is typically between 5 and 10 per cent, though it rises to 15 per cent for tennis balls. This suggests that price discrepancies of modest size exist across the cities in our sample; prices, while stationary, do not converge absolutely.

Conditional on stationarity, the random walk test specification can be interpreted as a linear mean-reversion model. How well does this model perform? Are there

Table 3
Estimated models for goods not locally-produced (Category A) grouped by location[a]

Good	Random walk model				*TAR* model				*ESTAR* model			
	T	N	$\overline{\lvert\mu\rvert}$	Φ	c	Φ	Φ^*	ΔAIC	θ	Φ	Φ^*	ΔAIC
Mobile, AL	72	9	0.117	0.565	0.08	0.921	−0.417	−30.6	43.55	0.83	−0.39	−24.1
			9	(0.000)		(0.086)	(0.104)		(23.95)	(0.09)	(0.13)	
Blythe, CA	72	8	0.085	0.714	0.11	0.811	−0.164	−7.1	30.90	1.21	−0.72	−31.3
			4	(0.001)		(0.086)	(0.101)		(13.73)	(0.11)	(0.15)	
Indio, CA	72	8	0.085	0.740	0.05	0.739	0.064	−8.7	26.93	1.27	−0.45	−14.8
			2	(0.006)		(0.221)	(0.224)		(13.58)	(0.12)	(0.13)	
Denver, CO	72	9	0.094	0.696	0.09	0.765	−0.022	−7.4	47.43	0.84	−0.07	−17.1
			4	(0.000)		(0.093)	(0.100)		(30.04)	(0.09)	(0.12)	
Indianap., IN	72	9	0.063	0.633	0.1	0.948	−0.265	−18.3	34.35	1.17	−0.67	−34.4
			4	(0.000)		(0.083)	(0.108)		(14.55)	(0.09)	(0.14)	
C. Rapids, IA	72	8	0.047	0.566	0.07	0.934	−0.369	−19.6	55.54	0.82	−0.28	−16.3
			3	(0.000)		(0.108)	(0.125)		(37.18)	(0.10)	(0.15)	
Lex., KY	72	9	0.093	0.701	0.11	0.820	−0.057	−11.8	29.43	0.92	−0.27	−23.2
			4	(0.000)		(0.069)	(0.089)		(17.76)	(0.08)	(0.13)	
Louisville, KY	72	9	0.067	0.641	0.06	0.903	−0.302	−8.2	32.40	0.67	0.02	−7.6
			3	(0.000)		(0.117)	(0.129)		(29.13)	(0.08)	(0.15)	
St. Louis, MO	72	9	0.050	0.651	0.12	0.862	−0.198	−9.7	44.29	1.16	−0.58	−22.1
			3	(0.000)		(0.069)	(0.096)		(19.45)	(0.10)	(0.13)	
Hastings, NE	72	9	0.084	0.667	0.12	0.803	−0.066	−12.8	33.42	0.99	−0.54	−24.5
			6	(0.000)		(0.057)	(0.094)		(20.25)	(0.09)	(0.14)	
Omaha, NE	72	8	0.069	0.550	0.1	0.653	−0.153	−8.0	52.43	0.72	−0.27	−11.9
			4	(0.000)		(0.087)	(0.112)		(38.38)	(0.10)	(0.15)	
Rap. City, SD	72	9	0.149	0.606	0.08	0.921	−0.343	−10.9	38.09	0.88	−0.34	−10.5
			7	(0.000)		(0.106)	(0.112)		(26.30)	(0.10)	(0.12)	
Vermill., SD	72	9	0.094	0.555	0.06	1.015	−0.472	−9.6	47.76	0.97	−0.58	−16.3
			7	(0.000)		(0.119)	(0.134)		(29.43)	(0.11)	(0.15)	
Ch'nooga, TN	72	9	0.132	0.700	0.11	0.922	−0.340	−12.0	28.65	1.10	−0.71	−34.0
			6	(0.000)		(0.071)	(0.087)		(12.97)	(0.09)	(0.13)	
El Paso, TX	72	8	0.075	0.707	0.08	0.983	−0.240	−11.4	36.56	1.16	−0.57	−26.3
			3	(0.001)		(0.096)	(0.103)		(17.53)	(0.09)	(0.10)	
Houston, TX	72	9	0.055	0.664	0.06	0.969	−0.351	−6.7	37.66	0.81	−0.14	−10.5
			2	(0.000)		(0.104)	(0.119)		(30.90)	(0.08)	(0.10)	
Lubbox, TX	72	8	0.076	0.733	0.07	1.006	−0.262	−13.9	45.08	0.90	−0.18	−19.8
			2	(0.001)		(0.102)	(0.110)		(26.99)	(0.09)	(0.13)	
S. L. City, UT	72	9	0.075	0.608	0.05	0.843	−0.190	−12.2	29.40	1.03	−0.46	−38.4
			3	(0.000)		(0.160)	(0.163)		(12.67)	(0.08)	(0.12)	
Appleton, WI	72	9	0.087	0.736	0.09	0.844	−0.034	−9.7	34.71	1.19	−0.56	−32.1
			5	(0.000)		(0.086)	(0.097)		(15.41)	(0.09)	(0.11)	
Casper, WY	72	9	0.139	0.564	0.06	0.397	0.287	−4.6	26.80	1.20	−0.76	−23.6
			7	(0.000)		(0.119)	(0.126)		(10.02)	(0.08)	(0.12)	

[a] Estimated parameters for the random walk, *TAR* and *ESTAR* models. $\overline{|\mu|}$ is the mean absolute value of the estimated equilibrium for each relative price series in the panel, under the liner model. As the data measure absolute price differences, our prior is that $|\mu| = 0$. The integer number below each estimate of $|\mu|$ reports the number of estimated μ values that are statistically different from zero in the panel. For each model, Φ is shorthand for $\Sigma_{k=1}^{4}\phi_k$, and similarly $\Phi^* \equiv \Sigma_{k=1}^{4}\phi_k^*$. c is the estimated transition parameter for the *TAR* model, and θ is the estimated transition parameter for the *ESTAR* model. All estimates are obtained by maximum likelihood. For the random walk model, P-values for the random walk null are shown below Φ, while for the other models the parenthetical numbers are standard errors. As the models are not nested, the Akaike Information Criterion (*AIC*) is used to compare their overall goodness of fit. The columns labeled ΔAIC show, for the *TAR* and *ESTAR* models, the change in the *AIC* relative to the linear *AR*(4) model. Negative values in these columns indicate improved goodness-of-fit relative to the linear case.

Table 4
Estimated models for goods that may be locally-produced (Category B) grouped by commodity[a]

Good	Random walk model				*TAR* model				*ESTAR* model			
	T	N	$\overline{\lvert\mu\rvert}$	Φ	c	Φ	Φ^*	ΔAIC	θ	Φ	Φ^*	ΔAIC
Bacon	67	24	0.054	0.412	0.06	0.666	−0.267	−16.7	59.11	0.59	−0.26	−24.8
			16	(0.000)		(0.093)	(0.083)		(30.12)	(0.07)	(0.07)	
Bread	72	22	0.078	0.469	0.1	0.506	0.117	−8.1	28.28	0.57	0.13	−5.6
			13	(0.000)		(0.069)	(0.062)		(33.69)	(0.05)	(0.06)	
Eggs	72	23	0.107	0.781	0.05	0.829	−0.077	−15.1	59.32	0.75	0.16	−45.0
			10	(0.076)		(0.091)	(0.073)		(22.30)	(0.05)	(0.05)	
Minced steak	72	23	0.061	0.596	0.06	0.629	0.008	−20.5	43.97	0.59	0.14	−9.2
			12	(0.000)		(0.078)	(0.074)		(39.29)	(0.05)	(0.07)	
Lettuce	72	23	0.099	0.321	0.07	0.338	−0.020	−1.8	33.19	0.36	−0.02	1.2
			17	(0.000)		(0.102)	(0.092)		(62.03)	(0.05)	(0.06)	
Margarine	72	23	0.081	0.592	0.08	0.570	0.008	−16.8	47.42	0.81	−0.25	−34.0
			14	(0.000)		(0.070)	(0.056)		(21.01)	(0.06)	(0.06)	
Milk	72	23	0.071	0.745	0.1	0.719	0.127	−15.4	50.41	0.70	0.22	−27.6
			18	(0.015)		(0.037)	(0.033)		(35.39)	(0.03)	(0.09)	
Potatoes	72	23	0.099	0.392	0.1	0.247	0.181	−19.3	39.48	0.33	0.14	−9.6
			13	(0.000)		(0.083)	(0.072)		(37.57)	(0.07)	(0.08)	
Steak	72	22	0.058	0.625	0.09	0.592	−0.027	−14.1	40.95	0.63	0.00	−12.3
			10	(0.000)		(0.063)	(0.049)		(36.01)	(0.04)	(0.06)	
Sugar	55	20	0.062	0.540	0.1	0.596	−0.163	−8.3	60.91	0.60	−0.05	−3.9
			16	(0.000)		(0.059)	(0.063)		(61.68)	(0.05)	(0.08)	
Chicken	72	23	0.070	0.490	0.08	0.641	−0.092	−13.2	58.02	0.61	−0.05	−5.9
			16	(0.000)		(0.082)	(0.073)		(57.75)	(0.06)	(0.08)	

[a] Estimated parameters for the random walk, *TAR* and *ESTAR* models. $\overline{|\mu|}$ is the mean absolute value of the estimated equilibrium for each relative price series in the panel, under the liner model. As the data measure absolute price differences, our prior is that $\overline{|\mu|} = 0$. The integer number below each estimate of $\overline{|\mu|}$ reports the number of estimated μ values that are statistically different from zero in the panel. For each model, Φ is shorthand for $\Sigma_{k=1}^{4}\phi_k$, and similarly $\Phi^* \equiv \Sigma_{k=1}^{4}\phi_k^*$. c is the estimated transition parameter for the *TAR* model, and θ is the estimated transition parameter for the *ESTAR* model. All estimates are obtained by maximum likelihood. For the random walk model, $p-$values for the random walk null are shown below Φ, while for the other models the parenthetical numbers are standard errors. As the models are not nested, the Akaike Information Criterion (*AIC*) is used to compare their overall goodness of fit. The columns labeled ΔAIC show, for the *TAR* and *ESTAR* models, the change in the *AIC* relative to the linear *AR*(4) model. Negative values in these columns indicate improved goodness-of-fit relative to the linear case.

nonlinearities in the data that it fails to capture? The *TAR* and *ESTAR* fits shed light on these questions. It turns out that the nonlinear models dominate the linear model in every instance. Both models involve the estimation of an additional five parameters for each panel, but as a consequence the likelihood function, or equivalently the determinant of the inverse of the residual covariance matrix, increases, and the net result is that *AIC* falls. Columns 8 and 12 show the decline in the *AIC* for each model relative to the linear fit.

Turning first to the *TAR* model, the estimated thresholds (c) are in the range of 5

Table 5
Estimated models for goods that may be locally-produced (Category B) grouped by location[a]

Good	Random walk model				*TAR* model				*ESTAR* model			
	T	N	$\overline{\lvert\mu\rvert}$	Φ	c	Φ	Φ^*	ΔAIC	θ	Φ	Φ^*	ΔAIC
Mobile, AL	72	7	0.073	0.238	0.05	0.813	−0.610	−3.5	48.52	0.52	−0.49	−2.8
			6	(0.000)		(0.352)	(0.360)		(55.31)	(0.19)	(0.24)	
Blythe, CA	72	7	0.179	0.743	0.12	0.821	−0.073	1.1	29.47	1.24	−0.40	−20.9
			3	(0.003)		(0.084)	(0.100)		(14.90)	(0.12)	(0.13)	
Indio, CA	72	7	0.178	0.745	0.05	0.130	0.657	−4.9	26.92	1.17	−0.42	−10.8
			3	(0.006)		(0.271)	(0.276)		(16.61)	(0.12)	(0.13)	
Denver, CO	72	7	0.056	0.536	0.11	1.058	−0.451	−17.7	28.28	1.12	−0.85	−27.7
			2	(0.000)		(0.097)	(0.117)		(15.69)	(0.15)	(0.20)	
Indianap., IN	72	7	0.058	0.421	0.11	0.680	−0.184	−8.8	40.27	0.78	−0.38	−14.6
			3	(0.000)		(0.125)	(0.150)		(25.00)	(0.17)	(0.22)	
C. Rapids, IA	72	7	0.068	0.549	0.11	0.834	−0.205	−10.6	44.70	1.01	−0.56	−20.9
			4	(0.000)		(0.107)	(0.123)		(23.09)	(0.16)	(0.19)	
Lex., KY	72	7	0.102	0.588	0.1	0.827	−0.304	−2.9	41.21	0.93	−0.51	−15.9
			4	(0.000)		(0.120)	(0.136)		(29.23)	(0.14)	(0.18)	
Louisville, KY	72	7	0.093	0.491	0.05	0.748	−0.238	−1.0	26.35	0.82	−0.50	−5.7
			4	(0.000)		(0.281)	(0.295)		(22.12)	(0.14)	(0.19)	
St. Louis, MO	72	7	0.063	0.655	0.12	0.725	−0.065	−2.6	37.95	0.82	−0.08	−19.5
			1	(0.000)		(0.107)	(0.124)		(20.17)	(0.12)	(0.16)	
Hastings, NE	72	7	0.088	0.409	0.06	0.648	−0.181	−5.8	29.54	0.88	−0.89	−18.3
			5	(0.000)		(0.216)	(0.229)		(17.85)	(0.14)	(0.28)	
Omaha, NE	72	7	0.101	0.482	0.11	0.786	−0.312	−4.6	30.20	1.03	−0.91	−22.7
			6	(0.000)		(0.103)	(0.141)		(15.87)	(0.15)	(0.23)	
Rap. City, SD	72	7	0.090	0.346	0.09	0.808	−0.523	−10.1	23.22	0.80	−0.74	−12.6
			4	(0.000)		(0.159)	(0.177)		(17.04)	(0.16)	(0.20)	
Vermill., SD	72	7	0.065	0.579	0.05	0.132	0.520	−6.1	44.14	0.90	−0.50	−4.2
			3	(0.000)		(0.331)	(0.335)		(44.81)	(0.18)	(0.20)	
Ch'nooga. TN	72	7	0.049	0.503	0.1	0.546	0.024	−1.8	29.21	0.44	0.17	−1.6
			2	(0.000)		(0.154)	(0.168)		(38.28)	(0.14)	(0.20)	
El Paso, TX	72	7	0.114	0.574	0.07	0.460	0.155	−0.6	25.45	0.58	0.08	−0.3
			5	(0.000)		(0.180)	(0.181)		(35.89)	(0.10)	(0.14)	
Houston, TX	72	7	0.057	0.563	0.06	0.701	−0.104	−6.6	43.21	0.81	−0.24	−18.0
			1	(0.000)		(0.229)	(0.233)		(29.04)	(0.13)	(0.17)	
Lubbox, TX	72	7	0.075	0.449	0.12	0.640	−0.055	−11.4	32.44	0.73	−0.20	−10.2
			4	(0.000)		(0.123)	(0.144)		(22.18)	(0.15)	(0.21)	
S. L. City, UT	72	7	0.141	0.642	0.07	0.563	0.103	−0.4	22.81	0.66	0.05	−1.5
			4	(0.000)		(0.216)	(0.222)		(28.05)	(0.11)	(0.17)	
Appleton, WI	72	7	0.077	0.545	0.11	0.852	−0.346	−8.2	46.35	0.96	−0.49	−17.1
			5	(0.000)		(0.097)	(0.128)		(30.42)	(0.14)	(0.18)	
Casper, WY	72	7	0.088	0.467	0.11	0.826	−0.287	−17.7	37.11	1.10	−1.01	−30.6
			4	(0.000)		(0.109)	(0.124)		(17.15)	(0.17)	(0.20)	

[a] Estimated parameters for the random walk, *TAR* and *ESTAR* models. $\overline{\lvert\mu\rvert}$ is the mean absolute value of the estimated equilibrium for each relative price series in the panel, under the liner model. As the data measure absolute price differences, our prior is that $\overline{\lvert\mu\rvert} = 0$. The integer number below each estimate of $\overline{\lvert\mu\rvert}$ reports the number of estimated μ values that are statistically different from zero in the panel. For each model, Φ is shorthand for $\Sigma_{k=1}^{4}\phi_k$, and similarly $\Phi^* \equiv \Sigma_{k=1}^{4}\phi_k^*$. c is the estimated transition parameter for the *TAR* model, and θ is the estimated transition parameter for the *ESTAR* model. All estimates are obtained by maximum likelihood. For the random walk model, *P*-values for the random walk null are shown below Φ, while for the other models the parenthetical numbers are standard errors. As the models are not nested, the Akaike Information Criterion (*AIC*) is used to compare their overall goodness of fit. The columns labeled ΔAIC show, for the *TAR* and *ESTAR* models, the change in the *AIC* relative to the linear *AR*(4) model. Negative values in these columns indicate improved goodness-of-fit relative to the linear case.

Table 6
Estimated models for goods that are locally-produced (Category C) grouped by commodity[a]

Good	Random walk model				*TAR* model				*ESTAR* model			
	T	N	$\overline{\|\mu\|}$	Φ	c	Φ	Φ^*	ΔAIC	θ	Φ	Φ^*	ΔAIC
Fr. Chicken	43	24	0.055	0.525	0.06	0.768	−0.227	−19.9	40.71	0.71	−0.14	−15.4
			14	(0.036)		(0.084)	(0.063)		(36.29)	(0.04)	(0.06)	
McDonalds	43	24	0.031	0.503	0.07	0.353	0.045	−72.9	155.35	0.50	−0.07	−23.0
			18	(0.002)		(0.068)	(0.089)		(117.28)	(0.05)	(0.07)	
Pizza	43	24	0.108	0.709	0.08	0.597	−0.086	−53.6	104.44	0.72	−0.04	−27.9
			20	(0.104)		(0.070)	(0.019)		(96.73)	(0.03)	(0.02)	
App. Repair	72	23	0.096	0.721	0.07	0.890	−0.078	−5.3	44.16	0.82	0.00	−13.8
			14	(0.001)		(0.046)	(0.045)		(40.36)	(0.03)	(0.03)	
Auto maint.	55	22	0.117	0.739	0.1	0.656	0.068	−32.4	53.12	0.66	0.05	−31.8
			13	(0.130)		(0.046)	(0.028)		(36.61)	(0.03)	(0.04)	
Beauty	43	24	0.148	0.564	0.11	0.835	−0.135	−27.8	41.89	0.80	−0.06	−25.2
			22	(0.011)		(0.032)	(0.049)		(43.32)	(0.03)	(0.04)	
Bowling	72	22	0.127	0.730	0.1	0.893	−0.012	−28.3	56.50	0.82	0.10	−40.0
			16	(0.196)		(0.027)	(0.023)		(30.13)	(0.03)	(0.03)	
Dentist	72	23	0.114	0.715	0.11	0.807	−0.074	−7.5	27.82	0.85	−0.15	−18.1
			18	(0.000)		(0.037)	(0.040)		(18.72)	(0.04)	(0.04)	
Doctor	72	23	0.129	0.789	0.05	0.823	0.011	−7.9	35.46	0.91	−0.12	−16.1
			17	(0.013)		(0.065)	(0.063)		(24.22)	(0.03)	(0.04)	
Dryclean	72	23	0.130	0.762	0.12	0.833	−0.071	−24.2	35.88	0.84	−0.07	−31.2
			19	(0.001)		(0.025)	(0.019)		(23.50)	(0.02)	(0.02)	
Hospital	72	23	0.162	0.890	0.07	0.870	0.061	−41.8	24.30	1.04	−0.17	−76.2
			15	(0.766)		(0.024)	(0.021)		(11.89)	(0.02)	(0.03)	
Haircut	72	23	0.132	0.706	0.11	0.761	0.031	−12.1	31.50	0.78	0.01	−6.2
			20	(0.004)		(0.029)	(0.018)		(29.44)	(0.02)	(0.03)	
Movie	72	23	0.096	0.605	0.09	0.872	−0.047	−10.7	51.27	0.84	−0.03	−5.5
			19	(0.000)		(0.026)	(0.025)		(79.98)	(0.03)	(0.03)	

[a] Estimated parameters for the random walk, *TAR* and *ESTAR* models. $\overline{|\mu|}$ is the mean absolute value of the estimated equilibrium for each relative price series in the panel, under the liner model. As the data measure absolute price differences, our prior is that $|\mu| = 0$. The integer number below each estimate of $|\mu|$ reports the number of estimated μ values that are statistically different from zero in the panel. For each model, Φ is shorthand for $\Sigma_{k=1}^{4}\phi_k$, and similarly $\Phi^* \equiv \Sigma_{k=1}^{4}\phi_k^*$. c is the estimated transition parameter for the *TAR* model, and θ is the estimated transition parameter for the *ESTAR* model. All estimates are obtained by maximum likelihood. For the random walk model, *P*-values for the random walk null are shown below Φ, while for the other models the parenthetical numbers are standard errors. As the models are not nested, the Akaike Information Criterion (*AIC*) is used to compare their overall goodness of fit. The columns labeled ΔAIC show, for the *TAR* and *ESTAR* models, the change in the *AIC* relative to the linear *AR*(4) model. Negative values in these columns indicate improved goodness-of-fit relative to the linear case.

to 12 per cent. Inside these thresholds, the root of the DGP is Φ, while outside the root is $\Phi + \Phi^*$. So the sign of Φ^* indicates whether reversion is stronger or weaker outside the band. Seventeen of the 24 estimates of Φ^* are negative, indicating stronger reversion for large price gaps. Nine of these are statistically less than zero at conventional significance levels, while 3 are statistically positive.

Table 7
Estimated models for goods that are locally-produced (Category C) grouped by location[a]

Good	Random walk model				*TAR* model				*ESTAR* model			
	T	N	$\overline{\lvert\mu\rvert}$	Φ	c	Φ	Φ^*	ΔAIC	θ	Φ	Φ^*	ΔAIC
Mobile, AL	72	8	0.103	0.808	0.06	1.043	−0.156	−8.3	30.66	1.22	−0.53	−38.2
			4	(0.022)		(0.121)	(0.122)		(12.19)	(0.09)	(0.09)	
Blythe, CA	72	7	0.239	0.702	0.08	1.037	−0.339	−4.3	15.97	1.25	−0.68	−19.4
			7	(0.000)		(0.130)	(0.140)		(8.90)	(0.09)	(0.15)	
Indio, CA	72	7	0.267	0.731	0.1	0.740	0.056	−3.2	22.70	0.79	−0.01	−12.6
			6	(0.000)		(0.090)	(0.100)		(18.73)	(0.08)	(0.13)	
Denver, CO	72	8	0.134	0.773	0.05	0.711	0.037	−5.5	25.04	1.27	−0.87	−36.7
			4	(0.006)		(0.175)	(0.178)		(10.18)	(0.09)	(0.14)	
Indianap., IN	72	8	0.072	0.810	0.12	0.751	0.143	−16.3	47.96	1.01	−0.06	−43.8
			3	(0.005)		(0.063)	(0.074)		(18.67)	(0.07)	(0.09)	
C. Rapids, IA	72	7	0.053	0.859	0.05	0.774	0.120	−4.8	43.71	1.00	−0.07	−27.6
			0	(0.037)		(0.213)	(0.215)		(24.88)	(0.10)	(0.10)	
Lex., KY	72	8	0.082	0.866	0.12	0.734	0.194	−7.8	46.49	1.23	−0.32	−27.3
			2	(0.139)		(0.061)	(0.066)		(20.22)	(0.12)	(0.12)	
Louisville, KY	72	8	0.070	0.833	0.11	0.863	−0.094	−6.3	37.89	1.05	−0.29	−17.2
			1	(0.034)		(0.060)	(0.073)		(22.59)	(0.08)	(0.10)	
St. Louis, MO	72	8	0.064	0.835	0.12	0.858	−0.005	−0.7	35.85	1.08	−0.36	−11.4
			2	(0.009)		(0.057)	(0.061)		(22.91)	(0.09)	(0.09)	
Hastings, NE	72	8	0.124	0.803	0.11	0.894	−0.201	−16.1	40.80	1.05	−0.51	−33.2
			5	(0.005)		(0.056)	(0.070)		(18.52)	(0.08)	(0.10)	
Omaha, NE	72	8	0.094	0.675	0.11	0.880	−0.195	−6.4	45.38	0.98	−0.33	−12.8
			5	(0.000)		(0.067)	(0.084)		(31.10)	(0.09)	(0.12)	
Rap. City, SD	72	8	0.091	0.828	0.07	1.045	−0.253	−12.0	41.21	1.22	−0.43	−19.9
			2	(0.036)		(0.128)	(0.131)		(18.96)	(0.11)	(0.12)	
Vermill., SD	72	8	0.141	0.808	0.08	0.838	−0.018	−11.7	19.90	0.87	−0.09	−4.6
			6	(0.010)		(0.099)	(0.100)		(25.63)	(0.07)	(0.10)	
Ch'nooga. TN	72	7	0.115	0.789	0.09	0.885	−0.082	−8.3	40.36	1.06	−0.25	−31.2
			3	(0.003)		(0.083)	(0.089)		(21.51)	(0.08)	(0.09)	
El Paso, TX	72	7	0.095	0.843	0.12	0.961	−0.107	−7.8	48.32	0.76	0.19	−10.1
			1	(0.031)		(0.050)	(0.059)		(35.17)	(0.09)	(0.09)	
Houston, TX	72	8	0.162	0.829	0.05	0.754	0.076	−11.2	23.98	1.16	−0.44	−15.8
			7	(0.009)		(0.141)	(0.144)		(15.81)	(0.07)	(0.09)	
Lubbox, TX	72	7	0.090	0.765	0.12	0.837	−0.059	−4.5	27.41	0.98	−0.12	−20.8
			2	(0.000)		(0.067)	(0.085)		(16.14)	(0.08)	(0.09)	
S. L. City, UT	72	8	0.103	0.777	0.05	0.817	−0.066	−0.3	40.95	0.91	−0.22	−2.5
			3	(0.002)		(0.146)	(0.148)		(44.00)	(0.09)	(0.09)	
Appleton, WI	72	8	0.145	0.857	0.12	1.006	−0.250	−12.5	35.27	1.08	−0.36	−13.4
			4	(0.050)		(0.053)	(0.062)		(20.38)	(0.08)	(0.09)	
Casper, WY	72	8	0.109	0.694	0.08	1.014	−0.324	−2.2	41.44	1.10	−0.39	−7.7
			5	(0.000)		(0.117)	(0.127)		(25.21)	(0.10)	(0.10)	

[a] Estimated parameters for the random walk, *TAR* and *ESTAR* models. $\overline{\lvert\mu\rvert}$ is the mean absolute value of the estimated equilibrium for each relative price series in the panel, under the liner model. As the data measure absolute price differences, our prior is that $\lvert\mu\rvert = 0$. The integer number below each estimate of $\lvert\mu\rvert$ reports the number of estimated μ values that are statistically different from zero in the panel. For each model, Φ is shorthand for $\Sigma_{k=1}^{4}\phi_k$, and similarly $\Phi^* \equiv \Sigma_{k=1}^{4}\phi_k^*$. c is the estimated transition parameter for the *TAR* model, and θ is the estimated transition parameter for the *ESTAR* model. All estimates are obtained by maximum likelihood. For the random walk model, P-values for the random walk null are shown below Φ, while for the other models the parenthetical numbers are standard errors. As the models are not nested, the Akaike Information Criterion (*AIC*) is used to compare their overall goodness of fit. The columns labeled ΔAIC show, for the *TAR* and *ESTAR* models, the change in the *AIC* relative to the linear *AR*(4) model. Negative values in these columns indicate improved goodness-of-fit relative to the linear case.

The positive estimates are obtained for the babyfood, detergent and toothpaste panels. These three goods are unremarkable in terms of their price volatility — each has a standard deviation of absolute price levels in the 8–9 per cent range — so the finding of weaker reversion outside the threshold is something of a puzzle.

As for the *ESTAR* model, a similar story emerges, though with some twists. The transition parameter θ is in every case positive. To give the estimates meaning, Fig. 1 plots the estimated transition functions $F(\cdot)$ for the *TAR* (dotted line) and *ESTAR* (solid line) models for Category A goods. The results for the two models are broadly consistent. As the standard errors in Column 9 make clear, the θ coefficients are estimated imprecisely, indicating that the model has difficulty identifying the exact form of the nonlinearity. Eighteen of the 24 estimates of Φ^* are negative, suggesting once again that reversion is stronger for larger price gaps. Nine of these are statistically less than zero at conventional size, while 2 are statistically positive.

Comparing the two models, there are some discrepancies. Babyfood prices appear to be highly nonlinear under the *ESTAR* formulation, but not under the *TAR* model, and the converse is true for cigarettes. These differences stem from

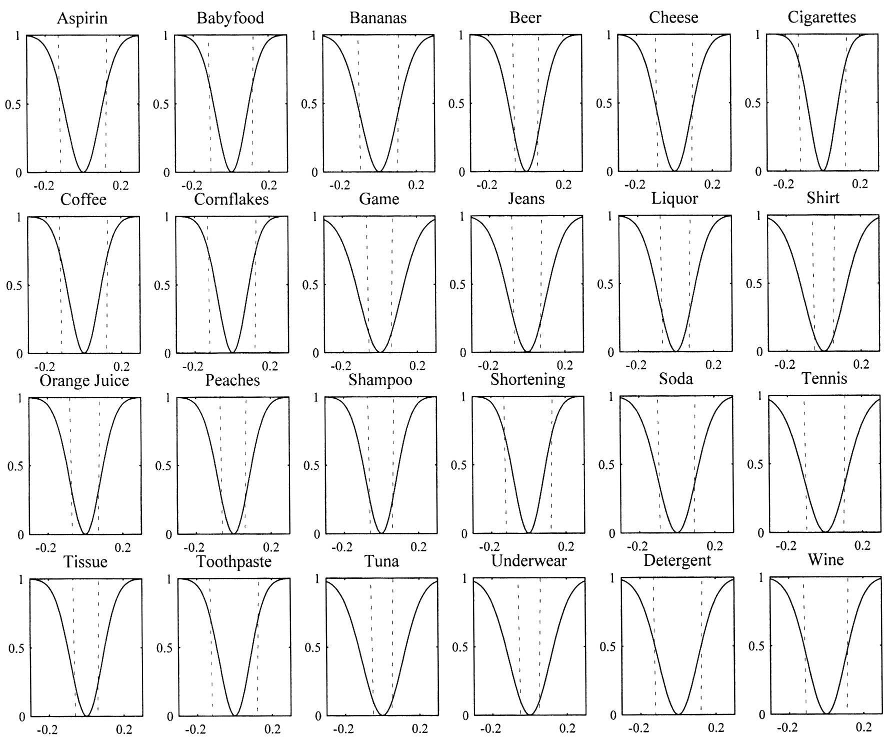

Fig. 1. *TAR* (· · ·) and *ESTAR* (—) transition functions for goods not locally produced (Category A).

the difference in the form of the transition function $F(\cdot)$. Choosing the preferred specification for each panel, 10 of the goods have statistically negative estimates of Φ^*, while only 2 (toothpaste and cigarettes) have statistically positive estimates of Φ^*.

Table 3 reports results in the same format for Category A goods grouped by location. Grouped this way, the data reveal stronger evidence of nonlinear reversion. All panels appear to be stationary, and the nonlinear models outperform the linear specification on the basis of the *AIC*. For 11 of the cities, the *TAR* estimate of Φ^* is negative and statistically significant; only for one city (Casper, WY) is the estimate positive and statistically significant. Sixteen of the *ESTAR* estimates of Φ^* are negative and significant, while none are positive.

Overall, then, the prices of Category A goods appear to converge to equilibria that are at or close to parity, and this convergence is nonlinear, with larger deviations from equilibrium producing more rapid convergence.

5.2. Results for goods that may be locally-produced (Category B)

The goods in this category appear to be stationary. Column 4 of Table 4 indicates that all the price processes are statistically distinguishable from the random walk at 8 per cent size or better. Once again, the roots are relatively low, averaging 0.54. As with the Category A goods, approximately 40 per cent of the estimated equilibrium prices are also statistically indistinguishable from zero. Moreover, the average absolute deviation of the equilibria from zero is within 10 per cent. Only for egg prices is there evidence of somewhat slower price convergence and relatively higher equilibrium price disparities. Inspecting the data it is clear that it is the price of eggs in California that gives rise to this differential behavior — the price of a dozen eggs in Blythe, Indio and Palm Springs doubled from roughly \$1 to \$2 from 1989 to 1992 while the price of eggs elsewhere in the country remained stable.

The goods in this category are staples like bread and milk. Accordingly, it is perhaps not too surprising that, given their commodity-like nature, their prices seem to conform well to the LOP. As for nonlinear convergence, however, the evidence in this category is much less compelling. For the *TAR* model, 2 of the estimates of Φ^* are statistically less than zero, and 3 are statistically greater than zero. For the *ESTAR* model, 2 estimates are negative and 4 are statistically positive. As a general rule, these goods display much higher price volatility than the goods in Category A: the average standard deviation of price levels in each panel is 11.5 per cent, while for changes the volatility is 12.3 per cent. These numbers can be contrasted with 8.7 per cent and 8.4 per cent for the Category A goods.

Table 5 reveals that, when grouped by location, these goods continue to exhibit stationarity — in all cases, the *P*-values in Column 4 indicate strong rejection of the unit root null. There is somewhat more evidence that the reversion is nonlinear.

Eight of the estimated roots for the *TAR* model are statistically less than 0, and 14 of the roots from the *ESTAR* model are statistically less than 0.

5.3. Results for goods locally-produced (Category C)

As might be expected, there is less evidence against the random walk null in this category, which includes many services such as dry cleaning and dentistry. Column 4 of Table 6 reveals that, for 4 of the 13 goods, the random walk null cannot be rejected. The non-rejections are for automobile maintenance, bowling, and hospital services. For hospital services the failure to reject is particularly acute. In addition, only 20 per cent of the estimated equilibria for the stationary series are statistically indistinguishable from zero, and the average absolute deviation of the estimated equilibria from zero for the stationary panels is in excess of 10 per cent.

Having said this, for the goods that do appear to be stationary in this panel, prices seem to be relatively well-behaved. The estimated roots under the linear model average 0.65. Of the 9 goods that are stationary, 6 exhibit evidence of nonlinear price convergence according to the *TAR* model, and 4 do so according to the *ESTAR* model.[22] Neither of the nonlinear models is dominant: for appliance repair, dentistry, doctor services and dry cleaning the *AIC* selects the *ESTAR* model, while for the remaining stationary series the *TAR* model is preferred. The average volatilities of price changes and price levels are lower than for Category B, at 9.0 and 6.9 per cent respectively.

When grouped by location (Table 7), the prices appear to be stationary across locations, though the *P*-values are higher than was the case for the other groupings and the random walk null cannot be rejected for Lexington, KY. There appears to be substantial evidence of nonlinear convergence to equilibrium. Selecting the preferred model for each city (excluding Lexington), 13 of the estimated Φ^* parameters are statistically less than 0.

6. Conclusion

In this paper we have sought to develop the literature on nonlinear commodity price behavior along two dimensions. First, we set out a simple continuous-time framework to help understand how general market frictions might affect relative price behavior. Second, we employ a detailed data set on U.S. goods prices to examine the pattern of reversion exhibited by deviations from the LOP. The data,

[22] Despite the episodic repricing of McDonald's sandwiches to 99 cents, this good does not exhibit evidence of nonlinear price convergence. This provides some comfort that our results are not driven by once-off 'sale' adjustments to mark-up.

measured in 24 cities over the period 1975:1–1992:4 afford a 'purer' measure of deviations from price parity than is possible with aggregate price indices. The empirical results can be distilled to the following main findings.

1. There is every indication that relative prices across U.S. cities are stationary, with some exceptions in the services area, notably health care. The estimated roots of the data generating processes indicate that reversion to equilibrium is quite rapid, and is significantly more rapid than has been found for relative international prices.
2. For approximately 40 per cent of non-locally produced, branded goods in Category A and the basic staples in Category B, reversion is to zero, which is to say the LOP holds absolutely. For the remainder, sustained price gaps on the order of 5 to 10 per cent can and do persist.
3. The volatility of price levels and price changes is almost 50 per cent higher for the staples in Category B than it is for either Category A or Category C goods.
4. Grouping the prices by commodity type, there is persuasive evidence that large price disparities mean revert more rapidly than small price disparities for the goods in Category A. For Category B goods, the evidence of nonlinear reversion is much weaker. In Category C, there are important differences in price behavior across the goods. Four of the 13 members of this category exhibit nonstationarity price behavior. For the remaining goods, there appears to be moderate to strong evidence of nonlinear reversion to the mean.
5. Grouping the prices by location in general leads to much stronger results. The prices in almost all locations appear to mean-revert, and for some three-fourths of the panels, the strength of reversion rises with the distance from parity.

These results can be seen as complementary to the body of recent work that has found evidence of nonlinear reversion in relative international prices.

Acknowledgements

We are grateful to John Campbell, Charles Engel, Jon Faust, Ken Froot, John Leahy, Alan Taylor, Aaron Tornell and two anonymous referees for helpful comments. All errors are our own.

References

Baum, C.F., Caglayan, M., Barkoulas, J.T., 1998. Nonlinear adjustment to purchasing power parity in the post-Bretton Woods era. Unpublished, Boston College.

Dixit, A., 1989, Hysteresis, import penetration, and exchange rate pass−through, Quarterly Journal of Economics, May, 104(2), 205–228.

Dixit, A.K., 1993. The Art of Smooth Pasting. Harwood Academic Publishers, Chur, Switzerland.

Dumas, B., 1992. Dynamic equilibrium and the real exchange rate in a spatially separated world. Review of Financial Studies 5, 153–180.

Engel, C., Rogers, J.H., 1996. How wide is the border. American Economic Review 86, 1112–1125.

Engel, C., Hendrickson, M., Rogers, J.H., 1996. Intra-national, intra-continental and intra-planetary PPP. Journal of Japanese and International Economics 11, 480–501.

Frankel, J., Rose, A., 1996. A panel project on purchasing power parity: Mean reversion within and between countries. Journal of International Economics 40, 209–225.

Froot, K., Rogoff, K., 1995. Perspectives on PPP and long-run real exchange rates. In: Grossman, G., Rogoff, K. (Eds.). Handbook of International Economics, Vol. 3. North-Holland, Amsterdam.

Levin, A., Lin, C.-F., 1992. Unit root tests in panel data: asymptotic and finite sample properties, UCSD Discussion Paper 92–93.

Michael, P., Nobay, A.R., Peel, D.A., 1997. Transactions costs and nonlinear adjustment in real exchange rates: An empirical investigation. Journal of Political Economy 105, 862–879.

Obstfeld, M., Rogoff, K., 2000. The six major puzzles in international macroeconomics: Is there a common cause? In: Bernanke, B.S., Rogoff, K. (Eds.), NBER Macroeconomics Annual 2000, forthcoming.

Obstfeld, M., Taylor, A.M., 1997. Nonlinear aspects of goods-market arbitrage and adjustment: Heckscher's commodity points revisited. Journal of Japanese and International Economics 11, 441–479.

O'Connell, P.G.J., 1997. Migration under uncertainty: 'Try your luck' or 'Wait and see'. Journal of Regional Science 37, 331–347.

O'Connell, P.G.J., 1998a. The overvaluation of purchasing power parity. Journal of International Economics 44, 1–19.

O'Connell, P.G.J., 1998b. Market frictions and real exchange rates. Journal of International Money and Finance 17, 71–95.

Ohanian, L.E., Stockman, A.C., 1997. Arbitrage costs and exchange rates. Discussion Paper, Department of Economics, University of Rochester. Presented at the CEPR-PIES Conference on 'Market Integration and Real Exchange Rates,' Georgetown University, Washington DC, May 2–3, 1997.

Papell, D.H., 1997. Searching for stationarity: Purchasing power parity under the current float. Journal of International Economics 43, 313–332.

Papell, D.H., Theodoridis, H., 1997. Increasing evidence of purchasing power parity over the current float. Unpublished (Mar.), University of Houston.

Parsley, D., Wei, S., 1996. Convergence to the law of one price without trade barriers or currency fluctuations. Quarterly Journal of Economics 108, 1211–1236.

Rogoff, K., 1996. The purchasing power parity puzzle. Journal of Economic Literature 34, 647–668.

Sercu, P., Uppal, R., Van Hulle, C., 1995. The exchange rate in the presence of transactions costs: Implications for tests of purchasing power parity. Journal of Finance 50, 1309–1319.

Taylor, A.M., 1996. International capital mobility in history: Purchasing power parity in the long run. Prepared for the NBER conferences on 'The Determination of Exchange Rates,' Cambridge, MA, May 10–11, 1996.

Taylor, A.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, forthcoming.

Taylor, M.P., Sarno, L., 1998. Linear and nonlinear mean reversion in real exchange rates: Reconciling the empirical evidence. Unpublished (Oct.), Oxford University.

Uppal, R., 1993. A general equilibrium model of international portfolio choice. Journal of Finance 48, 529–553.

Wei, S., Parsley, D., 1995. Purchasing power *dis*-parity during the floating rate period: Exchange rate volatility, trade barriers, and other culprits. NBER Working Paper No. 5032.
Williams, J.C., Wright, B.D., 1991. Storage and Commodity Markets. Cambridge University Press, Cambridge, MA.