Threshold Cointegration and Nonlinear Adjustment to the Law of One Price

Ming Chien Lo and Eric Zivot
Department of Economics, University of Washington

Preliminary draft: November 7, 1999. Please do not quote.

Abstract

Previous studies investigating threshold behavior in real exchange rate and price difference data have used rather ad hoc statistical methods and have focused on univariate threshold models for relative prices. We utilize a general multivariate threshold cointegration model and develop a systematic testing and estimation strategy for this model building on the work of Balke and Fomby (1997), Tsay (1998) and Hansen (1999). Using Monte Carlo experiments, we systematically compare the use of univariate and multivariate techniques for testing threshold cointegration, estimating various threshold models and specification testing. We apply our methodology to a large data set of U.S. disaggregated CPI data. We find evidence of threshold cointegration mainly for tradable goods. However, the type of threshold nonlinearity we find generally does not support the transactions cost view of commodity arbitrage.

Acknowledgement 1 *Paper presented at the Cowles Foundation Conference "New Developments in Time Series Econometrics," New Haven, October 23-24, 1999, and at the Erasmus University Workshop "Nonlinear Modeling of Macroeconomic Relations," Rotterdam, September 17-18, 1999. The authors thank Dick van Dijk for helpful comments and suggestions and Charles Engel for generously providing the data. Research support from the Tinbergen Institute, Erasmus University Rotterdam, is gratefully acknowledged.*

Threshold Cointegration and Nonlinear Adjustment to the Law of One Price

1 Introduction

Several recent papers have presented evidence revealing threshold type nonlinearity in real exchange rate data and have shown that threshold nonlinearity is a possible explanation for the apparent unit root behavior of real exchange rates. For example, Michael, Nobay and Peel (1997) and Sarantis (1999) apply smooth transition autoregressive (STAR) models and reject linearity for many bilateral and effective real exchange rates for a group of industrial countries. Using threshold autoregressive (TAR) models, O'Connell (1998) finds evidence of threshold nonlinearity in narrow panels of European CPI real exchange rates and Obstfeld and Taylor (1997) reject linearity for a large number of within country disaggregated price differential series as well as aggregated between country real exchange rate series. O'Connell and Wei (1997) and Parsley and Wei (1996) find that nonlinearities exist in U.S. disaggregated price data for a panel of location pairs. All of these works are more or less motivated by the idea that arbitrage is the force that eliminates any purchasing power parity (PPP) or law of one price (LOP) deviation only when it is profitable, i.e. the price difference (denominated by one currency) is large enough to offset the per unit cost of transporting the goods between the two locations.

The empirical TAR models utilized by the above authors to investigate nonlinearity in adjustments to PPP or to the LOP are univariate versions of the bivariate threshold cointegration models described by Balke and Fomby (1997). That is, the authors focus on the threshold behavior of the univariate cointegrating residual implied by the LOP or by PPP (equal to log price difference) and do not investigate threshold behavior in the broader bivariate model for log prices. Analysis of threshold behavior in the bivariate model allows one to uncover potential nonlinearities and asymmetries in the adjustment of individual prices and provides more information regarding the dynamics of the data. In addition, multivariate procedures for testing threshold cointegration that utilize the full structure of the model should have higher power, provided the model is true, than univariate procedures that ignore the restrictions imposed by the multivariate structure. Most empirical work has also largely ignored specification testing of the imposed TAR models. Specification testing is particularly important in threshold analysis of the LOP because the transactions cost theory that motivates the empirical specification of the TAR model imposes strong testable restrictions on the model.

In this paper we utilize a bivariate threshold vector error correction model of cointegration for log price differences and develop a systematic testing, estimation and specification strategy for investigating nonlinear adjustment to the LOP. Our strategy builds on the recent work of Balke and Fomby (1997), Tsay (1998) and Hansen (1999). Using Monte Carlo experiments, we compare the use of univariate and multivariate techniques for testing cointegration in the presence of threshold nonlinearity and for estimating the parameters of various threshold cointegration models. We also evaluate several model specification schemes, based on nested hypothesis tests, to see if restricted threshold models can be identified from unrestricted models. The results from our Monte Carlo experiments suggest that multivariate tests for cointegration have much higher power than univariate tests, multivariate and univariate tests for linearity have similar power, and it is difficult to detect certain types of threshold models, especially in small samples. We then apply our methodology for analyzing threshold cointegration to a wide range of U.S. disaggregated price data for 41 goods and service categories and 29 cities. We find evidence of threshold cointegration mainly for pairs of "tradable" goods prices. However, the type of threshold nonlinearity we find generally does not support the strong restrictions implied by the transactions cost view of commodity arbitrage.

The paper is organized as follows. First we provide some motivation for the use of threshold cointegration models in the analysis of PPP and the LOP. Next, we present the general bivariate threshold cointegration model and discuss the types of restricted models used in most empirical work. Then, we extend Balke and Fomby's (1997) methodology for testing threshold cointegration to a multivariate setting and we evaluate this extension using a small Monte Carlo experiment. Following the Monte Carlo study, we apply our methodology to a large data set of disaggregated prices to investigate the evidence for threshold nonlinearity in the adjustment to the law of one price. Technical details and derivations are gathered in the appendix.

2 The Threshold Cointegration and the Law of One Price

2.1 Motivation and Literature Review

Threshold Cointegration and Nonlinear Adjustment to the Law of One Price

There is a large literature on testing whether PPP holds in the long run using univariate and multivariate linear dynamic models. Froot and Rogoff (1996) offers an exhaustive review. Essentially, the approach taken involves testing to see if the real exchange rate is $I(0)$ and, hence, a mean reverting process. There is a consensus view that real exchange rate data indeed has mean reverting dynamics, yet a deviation from equilibrium is found to be very persistent and the estimated half-life of a shock is typically about four years. Obstfeld and Taylor (1997) question the simplicity of this premise and they revisit a notion raised by Hechscher (1916) who suggests that international transaction costs may play a role in explaining the deviation from PPP. The recent theoretical works by Dumas (1992), Uppal (1995) and Sercu, Uppal and Van Hulle (1995) lend support for this idea. O'Connell and Wei (1997) demonstrate the following simplified version of the effects of transactions costs on the real exchange rate.

Assume the only transaction cost is for transportation and is of the "iceberg" form. For goods purchased in location i at P_i and sold in location j at P_j, the per-unit revenue is $(1-\tau)P_j$ where τ denotes the proportional loss of value from transportation, like the melting of an iceberg, and $0 < \tau < 1$. That is to say, instead of defining transportation costs as the product of how many units of goods are traded and how many miles they are transported, this functional form takes it as a devaluation proportional to the gross revenue for each unit, namely the price. The greater the distance between two locations, the higher τ is. Thus, arbitrage from i to j is profitable if and only if $(1-\tau)P_j - P_i > 0$ or $1-\tau > P_i/P_j$. Conversely, for a unit of goods from j to i, positive profits implies $(1-\tau)P_i - P_j > 0$ or

$P_i/P_j > 1/(1-\tau)$. Putting the *opposite* of these two conditions together, there will be no arbitrage from either direction when $1/(1-\tau) > P_i/P_j > 1-\tau$, or, equivalently when taking logarithms, $-\ln(1-\tau) > \ln(P_i) - \ln(P_j) > \ln(1-\tau)$. The band surrounding relative prices wherein arbitrage is not profitable is therefore $[-\ln(1-\tau), \ln(1-\tau)]$. As arbitrage is profitable only outside of the bands, we expect to observe mean-reverting behavior of the log price difference only when its absolute value exceeds $\ln(1-\tau)$. Within the bands, the log price difference can exhibit random walk type non-stationary behavior.

Empirically, two types of simple threshold cointegration models, popularized by Balke and Fomby (1997), have been applied to investigate nonlinear adjustment to PPP or the LOP suggested by the transactions cost theory described above. The first type of model is the symmetric three regime BAND-TAR model

$$\Delta z_t = \begin{cases} \phi(z_{t-1} - c) + \eta_t, \text{ if } z_{t-1} > c \\ \eta_t, \text{ if } -c \leq z_{t-1} \leq c \\ \phi(z_{t-1} + c) + \eta_t, \text{ if } z_{t-1} < -c \end{cases} \tag{1}$$

where z_t denotes the log price differential at time t, η_t is an *i.i.d.* error term, $[-c, c]$ represents the symmetric transactions cost band wherein arbitrage is not profitable, and ϕ is a speed of adjustment parameter satisfying $-2 < \phi < 0$. In this model, if arbitrage is not profitable so that $|z_{t-1}| \leq c$ then z_t follows a random walk since there are no economic forces pushing prices together. In contrast, if arbitrage is profitable such that $|z_{t-1}| > c$ then z_t follows a stationary AR(1) process with mean equal to $\pm\, c$ depending on whether $z_{t-1} > c$ or $z_{t-1} < c$. Notice that the forces of arbitrage in this model push relative prices only to the edge of the transactions cost band. If the relative price is just outside the band, say z_{t-1} slightly larger than c, then the expected adjustment of relative prices, $\phi(z_{t-1} - c)$, is small for a given ϕ.

The second type of model is the symmetric three regime equilibrium (EQ) TAR model

$$\Delta z_t = \begin{cases} \phi_1 z_{t-1} + \eta_t, \text{ if } z_{t-1} > c \\ \phi_0 z_{t-1} + \eta_t, \text{ if } -c \leq z_{t-1} \leq c \\ \phi_1 z_{t-1} + \eta_t, \text{ if } z_{t-1} < -c \end{cases} \tag{2}$$

where it is expected that $\phi_0 \approx 0$ and $\phi_1 < \phi_0$ so that large deviations from parity should be less persistent than small deviations. In contrast to the BAND-TAR model, in the EQ-TAR model, if $|z_{t-1}| > c$, the relative price reverts to parity and not to the edge of the band. As a consequence, the magnitude of adjustment in response to an arbitrage opportunity is greater in the EQ-TAR than in the BAND-TAR. In other words, for given values of ϕ and c, z_t in the EQ-TAR model is less persistent than the BAND-TAR model[1] .

Pippenger and Goering (1993), Balke and Fomby (1997), O'Connell (1998), Enders and Granger (1998) and Berben and van Dijk (1999) show that if data are generated by TAR models like (1) and (2) then standard unit root tests can have very low power, which may explain the commonly found high degree of persistence in real exchange rates, and this result has motivated interest in the empirical relevance of threshold models for real exchange rates and price differences.

The empirical evidence in support of the transactions cost view for real exchange rates based on estimates of models like (1) and (2) are mixed. Obstfeld and Taylor (1997) estimate

[1] As noted by Obstfeld and Taylor (1997), if the transactions cost theory is correct then the width of the band $[-c, c]$ in the EQ and BAND TAR models should be positively related to measures of economic distance, transportation costs and measures of trade barriers. Also, the magnitude of the adjustment coefficient ϕ should be negatively related to these measures.

symmetric EQ-TAR and BAND-TAR models using monthly disaggregated and aggregated CPIs for 32 city and country locations (with the U.S. as the home country) over the period 1980 through 1994 and find evidence in support of threshold nonlinearity based on the transaction costs theory. Their estimates of half-lives of shocks to relative prices range from 8 to 12 months compared to those from a linear AR(1) model that range from 17 to 40 months[2] . They also find that the widths of their estimated transactions costs bands are positively related to measures of economic distance and exchange rate volatility as suggested by theory. Berben and Van Dijk (1998) re-evaluate the results of Obstfeld and Taylor (1997) using more sophisticated tests for unit roots in the presence of threshold nonlinearity as well as a more general BAND-TAR model that allows for asymmetry in the speed of adjustment and in the threshold values. Despite the fact that they find threshold-type nonlinearity, they claim that the asymmetry of the estimated parameters of the model, which is statistically significant, leads to nonsensical conclusions that contradicts the transaction costs view and they conclude that one should not attribute goods arbitrage as the factor that causes the nonlinearity in the data. O'Connell (1998) performs unit root tests and estimates EQ-TAR models on various panels of real exchange rates based on a different set of price indexes for industrial countries than used by Obstfeld and Taylor. O'Connell's results for broad panels of real exchange rates generally do not support the transaction costs view as he often fails to reject unit roots using panel unit root tests. Moreover, his panel estimates of EQ-TAR models like (2) indicate that large deviations from PPP can be more persistent than small deviations from PPP.

[2] The half-life of a shock from a stationary AR(1) with autoregressive coefficient ρ is given by $\ln(0.5)/\ln(\rho)$.

The evidence for the transactions cost view using disaggregated price data is more favorable. There are several cross sectional panel studies that show how price differences of individual commodities and services between two locations are affected by transaction costs. By regressing some volatility measure of price differences on selected exogenous variables, Engel and Rogers (1996,1998a,1998b), with no intention of testing for commodity arbitrage, show that the distance between two locations is a significant factor in explaining market segmentation for many goods within the U.S. borders. However, time series studies on price differences are rare. Parsley and Wei (1996) perform panel unit root test on monthly price data of 51 goods and services for a group of 48 cities with a sample period from 1975 to 1992. For log price differences of both "non-perishable/tradable" and "perishable" goods, they are able to reject unit roots in the data which provides support for the LOP. They also test for nonlinearity, in an ad hoc manner, and their results are supportive of the view that convergence of prices is faster when the initial price difference is wider. O'Connell and Wei (1997) are the first to test for nonlinearity on disaggregated price data using threshold models. They use the same data as Parsley and Wei (1996) consisting of monthly observations of individual prices on 48 goods and services from 24 cities. Two panels of data are formed: one is the price difference of a good in each city versus the all-cities average and the other is the price difference of each good in a city versus that of a benchmark city. Using panel unit root tests they generally do not reject unit roots. However, their panel estimates of Band-TAR and EQ-TAR models are highly supportive of the transactions costs view in that large deviations from the LOP appear to be mean reverting.

Although there is substantial evidence for threshold nonlinearity in the adjustment to PPP and the LOP, most of the empirical work has been rather *ad hoc*. For the most part, EQ-TAR and BAND-TAR models have been imposed on the data and no specification testing has been done to see if these models are actually appropriate. The transactions cost view imposes restrictions of symmetry of thresholds and symmetry of adjustment parameters and, apart from the analysis in Berben and van Dijk (1998), these restrictions generally have not been tested. In addition, all of the analysis has been based on univariate threshold models for log price differences. There has been no analysis of the threshold behavior of multivariate systems of log prices. Clearly there is a need for a more systematic and rigorous investigation of threshold nonlinearity. In the following sections we outline a general strategy for investigating threshold cointegration in bivariate systems of log prices.

2.2 A General Model of Threshold Cointegration

We start our analysis by considering a general three regime bivariate threshold VAR (TVAR) model for log prices similar to the one discussed in Tsay (1998). Let p_{it} be the log price of a good in location $i, (i = 1, 2)$, at time t and define $\mathbf{p}_t = (p_{1t}, p_{2t})'$. The bivariate threshold VAR (TVAR) model for $\mathbf{p}_t$ with lag length k, threshold variable z_t and delay d is given by

$$\mathbf{p}_t = \boldsymbol{\alpha}^{(j)} + \boldsymbol{\Phi}_1^{(j)}\mathbf{p}_{t-1} + \boldsymbol{\Phi}_2^{(j)}\mathbf{p}_{t-2} + \cdots \boldsymbol{\Phi}_k^{(j)}\mathbf{p}_{t-k} + \boldsymbol{\varepsilon}_t^{(j)}, \text{ if } c^{(j-1)} \leq z_{t-d} \leq c^{(j)} \qquad (3)$$

where $t = 1, \ldots, T$, $j = 1, 2, 3, -\infty = c^{(0)} < c^{(1)} < c^{(2)} < c^{(3)} = \infty$, and $\boldsymbol{\varepsilon}_t^{(j)}$ is a serially uncorrelated error term with mean zero and covariance matrix $\boldsymbol{\Sigma}^{(j)}$. The threshold variable z_t is assumed to be stationary and have a continuous distribution. Additionally, the transition variable z_t is assumed to be known whereas the delay variable d, lag length k and the threshold values $c^{(1)}$ and $c^{(2)}$ are potentially unknown. Typically, the delay parameter d is assumed to be less than or equal to the lag length k.

The TVAR model can be rearranged as

$$\Delta \mathbf{p}_t = \boldsymbol{\alpha}^{(j)} + \mathbf{\Pi}^{(j)}\mathbf{p}_{t-1} + \sum_{i=1}^{k-1} \mathbf{\Psi}_i^{(j)} \Delta \mathbf{p}_{t-i} + \boldsymbol{\varepsilon}_t^{(j)}, \text{ if } c^{(j-1)} \leq z_{t-d} \leq c^{(j)} \tag{4}$$

where $\mathbf{\Pi}^{(j)} = \sum_{i=1}^{k} \mathbf{\Phi}_i^{(j)} - \mathbf{I}_2$ and $\mathbf{\Psi}_i^{(j)} = -\sum_{l=i+1}^{k} \mathbf{\Phi}_l^{(j)}$. If, within each regime j, $\mathbf{p}_t$ is $I(1)$ and cointegrated with common cointegrating vector $\boldsymbol{\beta}' = (1, -\beta_2)$ then $rank(\mathbf{\Pi}^{(j)}) = 1$ and

$$\mathbf{\Pi}^{(j)} = \boldsymbol{\gamma}^{(j)}\boldsymbol{\beta}' = \begin{pmatrix} \gamma_1^{(j)} \\ \gamma_2^{(j)} \end{pmatrix} (1, -\beta_2)$$

The threshold vector error correction model (TVECM) representation is then

$$\Delta \mathbf{p}_t = \boldsymbol{\alpha}^{(j)} + \boldsymbol{\gamma}^{(j)}\boldsymbol{\beta}'\mathbf{p}_{t-1} + \sum_{i=1}^{k-1} \mathbf{\Psi}_i^{(j)} \Delta \mathbf{p}_{t-i} + \boldsymbol{\varepsilon}_t^{(j)}, \text{ if } c^{(j-1)} \leq z_{t-d} \leq c^{(j)} \tag{5}$$

The TVECM specifies that the adjustment toward the long-run equilibrium relationship $\boldsymbol{\beta}'\mathbf{p}_t$ is regime specific. For ease of exposition, in the following we focus on the TVAR with $k = 1$ and $d = 1$ which implies a TVECM with $k = 0$. This simple model is also the one used most in empirical applications to date.

In the TVECM with $k = 0$, it is straightforward to show that the cointegrating residual $\boldsymbol{\beta}'\mathbf{p}_t$ has the regime specific AR(1) or threshold autoregressive (TAR) representation

$$\boldsymbol{\beta}'\mathbf{p}_t = \delta^{(j)} + \rho^{(j)}\boldsymbol{\beta}'\mathbf{p}_{t-1} + \eta_t^{(j)}, \tag{6}$$

with

$$\rho^{(j)} = 1 + \boldsymbol{\beta}'\boldsymbol{\gamma}^{(j)} = 1 + \gamma_1^{(j)} - \beta_2\gamma_2^{(j)} \tag{7}$$

where $\delta^{(j)} = \boldsymbol{\beta}'\boldsymbol{\alpha}^{(j)}$and $\eta_t^{(j)} = \boldsymbol{\beta}'\boldsymbol{\varepsilon}_t^{(j)}$. Hence, $\boldsymbol{\beta}'\mathbf{p}_t$ is stable within each regime if $\left|\rho^{(j)}\right| = |1+\gamma_1^{(j)} - \beta_2\gamma_2^{(j)}| < 1$. Provided this stability condition holds the regime specific mean of the cointegrating residual is

$$\mu^{(j)} = \frac{\boldsymbol{\beta}'\boldsymbol{\alpha}^{(j)}}{\boldsymbol{\beta}'\boldsymbol{\gamma}^{(j)}} = \frac{\delta^{(j)}}{1-\rho^{(j)}}. \tag{8}$$

To eliminate drift in $\mathbf{p}_t$ the regime specific constants can be restricted to the to the error correction term if

$$\boldsymbol{\alpha}^{(j)} = -\mu^{(j)}\boldsymbol{\gamma}^{(j)}$$

where $\mu^{(j)}$ is given by (8). The TVECM then becomes

$$\Delta\mathbf{p}_t = \boldsymbol{\gamma}^{(j)}(\boldsymbol{\beta}'\mathbf{p}_{t-1} - \mu^{(j)}) + \boldsymbol{\varepsilon}_t^{(j)},\ \text{if } c^{(j-1)} \leq z_{t-1} \leq c^{(j)} \tag{9}$$

and $\mu^{(j)}$ is interpreted as the regime specific mean of the cointegrating relation $\boldsymbol{\beta}'\mathbf{p}_t$.

Since the LOP indicates that $p_{1t} - p_{2t}$ should be $I(0)$, we focus on the TVECM for $\mathbf{p}_t$ where the cointegrating vector is known to take the value $\boldsymbol{\beta} = (1,-1)'$, the threshold variable z_t is equal to the residual from the cointegrating relationship, $z_{t-1} = \boldsymbol{\beta}'\mathbf{p}_{t-1} = p_{1t} - p_{2t}$, and the errors in each regime have a common covariance matrix $\boldsymbol{\Sigma}$:

$$\Delta\mathbf{p}_t = \boldsymbol{\alpha}^{(j)} + \boldsymbol{\gamma}^{(j)}z_{t-1} + \boldsymbol{\varepsilon}_t,\ \text{if } c^{(j-1)} \leq z_{t-1} \leq c^{(j)} \tag{10}$$

2.2.1 The BAND-TVECM

A special case of the restricted constant TVECM model (9) occurs when $\mathbf{p}_t$ in the middle regime is not cointegrated but is $I(1)$ without drift. In this case $\boldsymbol{\gamma}^{(2)} = 0$, $\boldsymbol{\alpha}^{(2)} = 0$ and (9) becomes the *BAND-TVECM*

$$\Delta \mathbf{p}_t = \begin{cases} \boldsymbol{\gamma}^{(3)}(z_{t-1} - \mu^{(3)}) + \boldsymbol{\varepsilon}_t, \text{ if } z_{t-1} > c^{(2)} \\ \boldsymbol{\varepsilon}_t, \text{ if } c^{(1)} \leq z_{t-1} \leq c^{(2)} \\ \boldsymbol{\gamma}^{(1)}(z_{t-1} - \mu^{(1)}) + \boldsymbol{\varepsilon}_t, \text{ if } z_{t-1} < c^{(1)} \end{cases} \tag{11}$$

A sufficient condition for the stability of (11) is that the cointegrating residual z_t be stable in the outer regimes[3] . From (6), stability in the outer regimes requires $|\rho^{(j)}| = |1+\gamma_1^{(j)} - \gamma_2^{(j)}| < 1$ for $j = 1, 3$. Notice that in the middle regime $\boldsymbol{\gamma}^{(2)} = 0$ and $\boldsymbol{\alpha}^{(2)} = 0$ which implies that $z_t = z_{t-1} + \eta_t$. The BAND-TVECM has the following interpretation. If z_{t-1} is within the band $(c^{(1)}, c^{(2)})$ then $\mathbf{p}_t$ follows a random walk without drift; if $z_{t-1} > c^{(2)}$ then z_t reverts to the regime specific mean $\mu^{(3)}$ with adjustment coefficient $\rho^{(3)}$ and $\Delta \mathbf{p}_t$ adjusts with speed of adjustment vector $\boldsymbol{\gamma}^{(2)}$; if $z_{t-1} < c^{(1)}$, then z_t reverts to the regime specific mean $\mu^{(1)}$ with adjustment coefficient $\rho^{(1)}$ and $\Delta \mathbf{p}_t$ adjusts with speed of adjustment vector $\boldsymbol{\gamma}^{(1)}$.

It is important to emphasize that the speeds of adjustment of prices in the outer bands can be different for each element of $\mathbf{p}_t$. To see this more clearly, the model for Δp_{1t} is

$$\Delta p_{1t} = \begin{cases} \gamma_1^{(3)}(z_{t-1} - \mu^{(3)}) + \varepsilon_{1t}, \text{ if } z_{t-1} > c^{(2)} \\ \varepsilon_{1t}, \text{ if } c^{(1)} \leq z_{t-1} \leq c^{(2)} \\ \gamma_1^{(1)}(z_{t-1} - \mu^{(1)}) + \varepsilon_{1t}, \text{ if } z_{t-1} < c^{(1)} \end{cases} \tag{12}$$

[3] Chan, Petrucelli, Tong and Woolford (1985) give stability conditions for the three regime TAR model. We note that the locations of the regime specific means $\mu^{(1)}$ and $\mu^{(3)}$ also influence the stability of the model. Sensible models in the present context should have $|\mu^{(3)}| \leq |c^{(2)}|$ and $|\mu^{(1)}| \leq |c^{(1)}|$. Tsay (1998) notes that necessary and sufficient conditions for stationarity in general multivariate threshold models are largely unknown.

and the model for Δp_{2t} is

$$\Delta p_{2t} = \begin{cases} \gamma_2^{(3)}(z_{t-1} - \mu^{(3)}) + \varepsilon_{2t}, \text{ if } z_{t-1} > c^{(2)} \\ \varepsilon_{2t}, \text{ if } c^{(1)} \leq z_{t-1} \leq c^{(2)} \\ \gamma_2^{(1)}(z_{t-1} - \mu^{(1)}) + \varepsilon_{2t}, \text{ if } z_{t-1} < c^{(1)} \end{cases} \tag{13}$$

In general, $\gamma_1^{(3)} \neq \gamma_2^{(3)}$ and $\gamma_1^{(1)} \neq \gamma_2^{(1)}$ although we expect to find $\gamma_1^{(j)} \leq 0$ and $\gamma_2^{(j)} \geq 0$ so that both prices "error correct" toward parity when the LOP deviation is large. An interesting special case occurs if $\gamma_i^{(j)} = 0$ (for some $i = 1, 2$ and $j = 1, 3$) since this restriction implies that p_{it} is not responding to possible arbitrage opportunities. This could happen, for instance, if prices are sticky in some locations relative to others.

Several special cases of the BAND-TVECM (11) are of interest. The *continuous* model results when the regime specific means of the cointegrating residual z_t are equal to the neighboring threshold values; i,e., $\mu^{(3)} = c^{(2)}$ and $\mu^{(1)} = c^{(1)}$. The *symmetric threshold* model occurs when $c^{(2)} = -c^{(1)} = c$. The *EQ-TVECM* arises if $\mu^{(3)} = \mu^{(1)} = 0$.

2.2.2 The BAND-TAR Model

The BAND-TVECM (11) implies the following three-regime *BAND-TAR* model for z_t

$$\Delta z_t = \begin{cases} \phi^{(3)}(z_{t-1} - \mu^{(3)}) + \eta_t, \text{ if } z_{t-1} > c^{(2)} \\ \eta_t, \text{ if } c^{(1)} \leq z_{t-1} \leq c^{(2)} \\ \phi^{(1)}(z_{t-1} - \mu^{(1)}) + \eta_t, \text{ if } z_{t-1} < c^{(1)} \end{cases} \tag{14}$$

where $\phi^{(j)} = \rho^{(j)} - 1$. This model is more general than (1), which has been used in most empirical work, as it allows for asymmetric thresholds and adjustment parameters as well as regime specific means that are different from the neighboring thresholds.

Several special cases of the BAND-TAR model (14), analogous to the special cases of the BAND-TVECM, are of interest. A *continuous* model has $\mu^{(3)} = c^{(2)}$ and $\mu^{(1)} = c^{(1)}$; a *symmetric threshold* model has $c^{(2)} = -c^{(1)} = c$; a *symmetric adjustment* model has $\rho^{(3)} = \rho^{(1)} = \rho$; and an EQ-TAR model results when $\mu^{(3)} = \mu^{(1)} = 0$. Notice that the symmetric BAND-TAR model (1), used most often in empirical work, results from a continuous, symmetric threshold and symmetric adjustment BAND-TAR model[4] .

3 Testing for Threshold Cointegration

Balke and Fomby (1997) discuss some of the general problems associated with testing for threshold cointegration. They note that testing the null hypothesis of no cointegration against the alternative hypothesis of threshold cointegration is complicated by the combination of unit root asymptotics and the presence of nuisance parameters that are only present under the alternative hypothesis. Additionally, to construct tests with high power for a specific type of TVECM the specific form of the threshold model under the alternative generally needs to be specified and estimated and this can be difficult since there are many possible types of threshold models. Based on the outcome of a small set of Monte Carlo experiments, Balke and Fomby suggest the following practical strategy. First test the null hypothesis of no cointegration against the alternative of (non-threshold) cointegration. Next, if the hypothesis of no cointegration is rejected then test for threshold nonlinearity. If linearity is rejected, a third step, not investigated by Balke and Fomby, is also necessary. This is the

[4] Symmetric adjustment in the Band-TAR model does not imply symmetric adjustment in the BAND-TVECM since the restriction $\rho^{(3)} = \rho^{(1)}$ occurs whenever $\gamma_1^{(3)} - \gamma_2^{(3)} = \gamma_1^{(1)} - \gamma_2^{(1)}$.

specification and estimation of the threshold model.

We consider several modifications of the Balke-Fomby procedure for testing threshold cointegration. In their implementation of the two step strategy, Balke and Fomby focus on the properties of the univariate series $z_t = \boldsymbol{\beta}'\mathbf{p}_t$ and not on the properties of the multivariate process driving $\mathbf{p}_t$. That is, in the first step they investigated the power properties of various residual-based tests for no-cointegration on z_t in cases where $\boldsymbol{\beta}$ is known and unknown and in the second step they considered power properties of several univariate tests for linearity of z_t. We adopt a similar two step strategy as Balke and Fomby but focus instead on multivariate estimation and testing procedures. The idea is that if the bivariate TVECM is the appropriate model then multivariate procedures that utilize the structure of the model should have higher power than univariate procedures that ignore the restrictions imposed by the multivariate structure. Accordingly, in the first step we consider the power of multivariate tests of the hypothesis of no-cointegration and in the second step we consider the power of multivariate tests of linearity. In testing for no-cointegration, Balke and Fomby did not consider tests that made use of the threshold nature of the alternative. We also add to their analysis by considering some recently developed tests for unit roots that are designed to have power against threshold alternatives. Finally, we add a third step to the analysis consisting of a specification analysis of the form of the threshold model based on nested hypothesis tests within a general unrestricted threshold model. The following sub-sections describe our additions to the Balke-Fomby methodology.

3.1 Testing for No Cointegration

Given that we assume the cointegrating vector is known and constant across regimes under threshold cointegration, we avoid many complications associated with estimating general multivariate threshold cointegration models[5] . Still, the power of tests for no cointegration against threshold cointegration depend on how the test is constructed. We investigate tests of the hypothesis of no cointegration against the alternative of linear cointegration and against the specific alternative of threshold cointegration.

3.1.1 Testing no cointegration against linear cointegration

Given that the cointegrating vector $\boldsymbol{\beta}$ is known, standard univariate unit root tests on the cointegrating residual $z_t = \boldsymbol{\beta}'\mathbf{p}_{t-1}$ can be used to test the no cointegration null hypothesis. The results of Pippenger and Goering (1993) and Balke and Fomby (1997) show that standard unit root tests can have low power in EQ-TAR and Band-TAR models if the autoregressive coefficients in the outer regimes are close to 1 and/or the width of the band relative to the variance of the errors is large. However, to date no one has considered multivariate tests that are based on a known cointegrating vector. Since we assume that under the alternative of cointegration $\mathbf{p}_t$ has a VECM structure we consider Horvath and Watson's (1995) multivariate test for no-cointegration. The VECM for $\mathbf{p}_t$ ignoring threshold effects is

$$\Delta\mathbf{p}_t = \boldsymbol{\alpha} + \boldsymbol{\gamma} z_{t-1} + \sum_{i=1}^{k-1} \boldsymbol{\Phi}_i \Delta\mathbf{p}_{t-i} + \boldsymbol{\varepsilon}_t$$

and under the null hypothesis of no cointegration $\boldsymbol{\gamma} = \mathbf{0}$. Horvath and Watson's test statistic is the standard seeming unrelated regressions (SUR) Wald statistic for testing $\boldsymbol{\gamma} = \mathbf{0}$ and is

[5] González and Gonzalo (1999) discuss testing for threshold cointegration in a general framework where the cointegrating vector is unknown.

given by

$$HW = \hat{\gamma}' \widehat{var}(\hat{\gamma})^{-1} \hat{\gamma} \tag{15}$$

where $\hat{\gamma}$ denotes the (equation by equation) OLS estimate of γ and $\widehat{var}(\hat{\gamma})$ is the OLS estimate of the covariance matrix of $\hat{\gamma}$. Under the null of no cointegration the limiting distribution of HW is a function of bivariate Brownian motions and Horvath and Watson provide tables with the appropriate critical values.

Horvath and Watson (1995) show that their test can have much better power against cointegrated alternatives than the univariate ADF unit root test on z_t especially when the correlation between the errors in ε_t is strong. Additionally, Zivot (1999) shows that the Horvath-Watson test generally has higher power than the ADF test when the dynamics of the data invalidate the common factor restrictions imposed by the ADF test. Given that the Horvath-Watson test generally performs better than the ADF test against linear cointegration alternatives it is conjectured that it will have higher power than the ADF test against threshold cointegration alternatives as well.

3.1.2 Testing No Cointegration Against Threshold Cointegration

Recently Gonzáles and Gonzalo (1997), Caner and Hansen (1998), Enders and Granger (1998), and Berben and van Dijk (1998, 1999) have addressed the issue of testing for a unit root in a univariate autoregressive model against the alternative of a stationary TAR model and they develop unit root tests that can have higher power than tests that ignore the specific nature of the threshold alternative. It appears then that these methods can be applied to the cointegrating residual $z_t = \boldsymbol{\beta}'\mathbf{p}_t$. However, we must be careful because some of these new tests require assumptions about the nature of the transition variable, z_{t-1}, that are not satisfied in the present context. In particular, the tests of Gonzáles and Gonzalo (1997) and Caner and Hansen (1998) require the threshold variable to be stationary under the unit root null. Since we use the cointegrating residual as the transition variable the tests of Gonzáles and Gonzalo (1997) and Caner and Hansen (1998) are not applicable. The tests of Enders and Granger (1998) and Berben and van Dijk (1998, 1999) do not require the transition variable to be stationary and so are applicable in the present context.

Enders and Granger (1998), hereafter EG, consider testing for a unit root in the two regime TAR model

$$\Delta z_t = \begin{cases} \phi^{(2)}(z_{t-1} - c) + \eta_t, \text{ if } z_{t-1} > c \\ \phi^{(1)}(z_{t-1} - c) + \eta_t, \text{ if } z_{t-1} \leq c \end{cases} \tag{16}$$

which is like the band-TAR model (14) with $\mu^{(3)} = \mu^{(1)} = c$ so that the middle regime vanishes. They estimate the threshold value c using the sample mean of z_{t-1} and test the null hypothesis that $\phi^{(1)} = \phi^{(2)} = 0$ using the standard F-statistic from the regression

$$\Delta z_t = \phi^{(1)}(z_{t-1} - \widehat{c})I(z_{t-1} \leq \widehat{c}) + \phi^{(2)}(z_{t-1} - \widehat{c})(1 - I(z_{t-1} \leq \widehat{c})) + \sum_{j=1}^{k-1} \psi_j \Delta z_{t-j} + \eta_t, \tag{17}$$

where $\widehat{c}$ is the mean of z_{t-1}. Under the null of a unit root, the distribution of the F-statistic is a function of Brownian motion and EG provide the appropriate critical values. EG find, rather surprisingly, that their F-test actually has lower power than the ADF t-test that ignores the threshold nature of the two regime alternative but may have higher power than the ADF test in three regime models with asymmetric thresholds and dynamics. Hence, the EG test may prove useful in the present context.

Berben and van Dijk (1999), hereafter BVD, point out that the low power of the EG test relative to the ADF test is likely due to the fact that the test makes use of a biased estimate of the threshold parameter under the alternative hypothesis. They develop a more powerful test that uses a consistent estimate of the threshold under the alternative. Their test is based on the standard F-statistic for testing the null hypothesis $\phi^{(1)} = \phi^{(2)} = 0$ in (17) where c is estimated from (16) by sequential conditional least squares[6] . BVD show that their $F-$statistic can have much higher power than EG's $F-$statistic for the two regime model (16) if the dynamics are highly asymmetric. It remains to be seen how BVD's $F-$statistic performs for three regime BAND-TAR models. Even though the BVD test is designed for two-regime TAR alternatives, the test should also have power against three-regime TAR alternatives. To see why, if, for example, a three-regime TAR model is the true model then the results of Bai (1997) show that the least squares estimate of the threshold on the misspecified two regime model will be consistent for one of the thresholds. Hence, one of the estimated autoregressive coefficients in (17) should be less than zero and this will give the test power.

3.2 Testing Linearity

Once it has been determined that $\mathbf{p}_t$ is cointegrated with known cointegrating vector $\boldsymbol{\beta}$, the next step is to determine if the dynamics in the cointegrating relationship is linear or exhibits threshold nonlinearity. Several univariate and multivariate tests for linearity that have power against threshold nonlinear alternatives have been proposed and we briefly review them here.

[6] BVD derive the asymptotic distribution of the $F-$statistic based on the assumption of a "drifting threshold" in terms of the maximum and minimum values of z_{t-1} in the sample and show that it is a function of standard Brownian motion and parameters $\tau_1, \tau_2 \in (0,1)$ that are determined from the data. They tabulate critical values for various values of τ_1 and τ_2 and provide an algorithm for computing approximate $p-$values.

Balke and Fomby test for linearity in the univariate cointegrating residual $z_t = \boldsymbol{\beta}'\mathbf{p}_t$ by testing for structural breaks in an arranged autoregression for z_t. An arranged autoregression for z_t orders the data according to the value of the threshold variable, here z_{t-1},instead of by time. The rearrangement of the data does not alter the dynamic relationship between z_t and it lags and is useful for detecting threshold nonlinearity since the existence of a threshold in the time ordered data translates into a structural change in the rearranged data. Using arranged autoregressions for z_t, Balke and Fomby consider Tsay's (1989) nonparametric test for structural change based on recursive residuals as well as sup-Wald type tests for one and two breaks. Tsay's test has the attractive property that it is independent of the form of threshold nonlinearity and its limiting distribution is independent of nuisance parameters. The distributions of the sup-Wald type tests, however, are complicated by the fact that the estimated break dates are not identified under the null hypothesis of linearity and need to be simulated or bootstrapped on a case by case basis. Based on a small Monte Carlo study, Balke and Fomby find that the Tsay test and the Sup-Wald test for one break have similar power against the symmetric three regime EQ-TAR and BAND-TAR models and that the sup-Wald test for two breaks has the best power, although it tends to be size distorted.

Tsay (1998) generalizes his univariate test for threshold nonlinearity based on arranged autoregressions to multivariate models and shows that it is also valid for cointegrated processes. Hence it is of interest to see how his multivariate test performs relative to his univariate test in the present context. To implement his test in the present context we need to consider an arranged multivariate regression for the VECM. Details of this test are complicated and are described in the appendix.

Hansen (1997,1999) describes another method for testing the null hypothesis of linearity versus the alternative of a TAR(m) model, where m denotes the number of regimes, based on nested hypothesis tests. To illustrate, consider the TAR(3) model for $z_t = \boldsymbol{\beta}'\mathbf{p}_t$ in (6). A linear autoregressive, or TAR(1), model results under the restrictions that $\delta^{(j)} = \delta$ and $\rho^{(j)} = \gamma$, $\forall\ j$. Hansen's linearity test is a test of the null hypothesis of TAR(1) against the alternative of TAR(m) for some $m > 1$ using sup-F type (sup-Wald) tests of the form

$$F_{1m} = T\left(\frac{S_1 - S_m}{S_m}\right)$$

where S_1 and S_m denote the sum of squared residuals from the estimation of a TAR(1) model and a TAR(m) model, respectively. A drawback of this procedure is that the asymptotic distributions of the sup-F tests are influenced by the unidentified threshold parameters under the null hypothesis of linearity, the so-called Davies problem (see Davies (1977, 1987)), and simulation techniques must be used to evaluate these distributions on a case-by-case basis. Hansen shows that a simple bootstrap procedure can be used to compute p-values for various linearity tests and we use this procedure in our analysis. A useful by product of this testing strategy is the estimation of the parameters of the TAR(m) models. Hansen's linearity testing procedure has the apparent advantage over Tsay's nonparametric procedures since it is based on the specific form of the threshold model under the alternative. Hansen's testing procedure has not been investigated against BAND-TAR models and we evaluate it here[7] .

[7] If the true model is the BAND-TAR model one would like a direct test of the linear autoregressive model against the BAND-TAR(m). However, the stationary TAR(1) is not nested within the general BAND-TAR(m), except in very special circumstances, and so Hansen's LR-type tests based on nested models are not appropriate.

We note that Hansen's method for testing linearity in univariate TAR models based on nested hypothesis tests can be easily extended to test linearity in multivariate TVECMs. We simply test the null hypothesis of a linear VECM against the alternative of a TVECM(m) for some $m > 1$. The most convenient test statistic to use in this case is the sup-LR statistic (which is equivalent to the sup-Wald)

$$LR_{1m} = T(\ln(|\widehat{\mathbf{\Sigma}}|) - \ln(|\widehat{\mathbf{\Sigma}}_m(\widehat{\mathbf{c}}, \widehat{d})|))$$

where $\widehat{\mathbf{\Sigma}}$ and $\widehat{\mathbf{\Sigma}}_m(\widehat{\mathbf{c}}, \widehat{d})$ denote the estimated residual covariance matrices from the linear VECM and m-regime TVECM, respectively. Using the arguments in Hansen (1997), the distribution of the sup-LR statistic will be non-standard. We use Hansen's bootstrap procedure to compute p-values for various linearity tests based on LR_{1m}.

3.3 Model Specification

Given that we reject no cointegration and linearity, how do we determine which kind of threshold model is appropriate for the data? How many regimes are in the model? In a three regime model are the threshold values symmetric? Is a continuous threshold model specification appropriate? Is an EQ-TAR model or EQ-TVECM more appropriate than a BAND-TAR model or BAND-TVECM? These are the kinds of questions that need answering in an empirical analysis. Two approaches have generally been taken to determine the appropriate threshold specification of a model. The first approach, advocated by Tong (1990), Clements and Herzog (1998), and Tsay (1998), uses a model selection criterion like AIC to determine the best specification from the data. The second approach, recently reviewed by Hansen (1999), uses a sequential testing procedure based on nested models. We follow Hansen and consider nested hypotheses tests based on unrestricted estimation of TAR models and TVECMs.

The transactions cost theory of commodity arbitrage reviewed in section 2 implies a three regime symmetric threshold and symmetric adjustment BAND-TAR model for z_t as well as a three regime symmetric threshold and symmetric adjustment BAND-TVECM for $\mathbf{p}_t$. The symmetric BAND-TAR model is nested within an unrestricted TAR(3) model and the symmetric BAND-TVECM is nested within an unrestricted TVECM(3). This nested structure allows for a systematic specification analysis.

Consider first the determination of the number of regimes. Given that linearity is rejected in favor of threshold nonlinearity, to determine if a TAR(3) model for z_t is appropriate we can follow Hansen (1999) and test of the null of a TAR(2) model against the alternative of a TAR(3) model using the F-statistic

$$F_{23} = T\left(\frac{S_2 - S_3}{S_3}\right).$$

where S_2 and S_3 denote the sum of squared residuals from the estimation of an unrestricted TAR(2) model and an unrestricted TAR(3) model, respectively. Similarly, to determine if a TVECM(3) for $\mathbf{p}_t$ is appropriate we can test the null of a TVECM(2) against the alternative of a TVECM(3) using the LR statistic

$$LR_{23} = T(\ln(|\widehat{\mathbf{\Sigma}}_2(\widehat{c}, \widehat{d})|) - \ln(|\widehat{\mathbf{\Sigma}}_3(\widehat{\mathbf{c}}, \widehat{d})|)).$$

where $\widehat{\mathbf{\Sigma}}_2(\widehat{c}, \widehat{d})$ and $\widehat{\mathbf{\Sigma}}_3(\widehat{\mathbf{c}}, \widehat{d})$ denote the estimated residual covariance matrices from the unrestricted VECM(2) and TVECM(3), respectively. As with the linearity tests discussed previously, the asymptotic distributions of F_{23} and LR_{23} are nonstandard and bootstrap methods can be used to compute approximate $p-$values.

Next, consider specification tests for the BAND-TAR model and the BAND-TVECM. Chan (1993) shows that the threshold estimates from estimating unrestricted (stable) TAR models are superconsistent (converge at rate T) and that the remaining parameters are asymptotically normally distributed, with the usual formulas for covariance matrices, and independent of the threshold estimates. Tsay (1998) gives analogous results for the parameters of stable or cointegrated TVAR models. The superconsistency of the thresholds means that the estimated thresholds can be treated as the true thresholds for inference purposes regarding the remaining parameters. Hence, Wald tests of restrictions on the parameters (excluding the thresholds) can be computed in the usual way and these tests have asymptotic chi-square distributions. LR tests can also be computed but are more costly than Wald tests because they require the estimation of restricted threshold models[8] . Inference regarding the thresholds, however, is problematic since the limiting distribution of the thresholds from unrestricted estimation is nonstandard and generally depends on the nusiance parameters and the data, see Hansen (1997).

4 Monte Carlo Results

In this section we use Monte Carlo methods to compare the performance of the univariate and multivariate procedures to analyze a bivariate threshold cointegration model. We

[8] The estimation of restricted threshold models requires considerable programming costs and hence the computation of LR tests is not trivial. However, the finite sample behavior of LR tests may be better than the finite sample behavior of Wald tests.

compare the performance of tests for no cointegration against the alternatives of linear cointegration and threshold cointegration, we compare the performances of tests designed to capture threshold nonlinearity, we compare different estimation strategies for various kinds of threshold models and we evaluate specification tests based on nested hypothesis tests. Unless noted otherwise, all experiments are based on 1000 replications.

4.1 Design of the experiments

The design for our set of Monte Carlo experiments is the one used by Balke and Fomby (1997)[9] . They specify a cointegrated system for $\mathbf{p}_t = (p_{1t,}p_{2t})'$ as

$$\begin{aligned} p_{1t} - 3p_{2t} &= z_t, \\ p_{1t} - 2p_{2t} &= B_t \\ B_t &= B_{t-1} + u_t \end{aligned}$$

where z_t follows either a continuous symmetric EQ-TAR model as in (2), with $\phi_0 = 0$, or a continuous symmetric BAND-TAR model as in (1). In both cases, η_t and u_t are i.i.d. $N(0,1)$ random variables, $\phi = \rho - 1 = -0.6$, $c = 3, 5$ and 10, and $T = 100, 250$ and 500. The implied continuous and symmetric BAND-TVECM is given by (12) and (13) with $\gamma_1^{(j)} = -1.8$ and $\gamma_2^{(j)} = -0.6$ $(j = 1, 3)$, and the correlation between the errors in the BAND-TVECM is $corr(\varepsilon_{1t}, \varepsilon_{2t}) = 0.98$. This implied parameterization of the BAND-TVECM is a bit odd since the speed of adjustment coefficients are both negative and the errors are very highly correlated.

[9] We use the Balke-Fomby design mainly to check our code and to compare our multivariate methods with the univariate methods used by Balke and Fomby.

4.2 Results for Balke-Fomby Design

Table 1 summarizes the results for the tests of no-cointegration for the EQ and BAND models. The powers of all tests are higher for the EQ specification than for the BAND specification and the power is higher for smaller values of the thresholds. For the tests that ignore the threshold nature of the alternative, the multivariate Horvath-Watson (HW) test has about twice the power as the univariate ADF tests for both the EQ and BAND models. However, the powers of the EG and BVD tests, which take account of the threshold nature of the alternative, have higher power than the HW test. In particular, the BVD test has excellent power for moderate sample sizes.

Table 2 gives the results for the linearity tests. As with the tests for no cointegration, the linearity tests have higher power for the EQ specification. For the BAND specification, all of the tests have low power for small sample sizes and the power decreases with the magnitude of the thresholds. Somewhat surprisingly, the univariate tests generally have higher power than the multivariate tests even for large sample sizes. This could be due, however, to the restrictive nature of the data generating process. The univariate F_{12} and F_{13} statistics have slightly higher power than the univariate Tsay statistic, with F_{12} having higher power than F_{13}, and the multivariate LR_{12} and LR_{13} statistics have higher power than the multivariate Tsay statistic.

Tables 3 and 4 presents Monte Carlo means and standard deviations for the estimated parameters of the BAND specification obtained from unrestricted univariate TAR models and multivariate TVECMs. The estimated univariate model is (6) and the estimated multivariate model is (10). Estimation of the models is performed using sequential conditional least squares as outlined by Hansen (1999) and explained in the appendix. First, consider the estimated thresholds. The univariate and multivariate estimates of the thresholds both capture the symmetry of the true thresholds and are numerically very close. They are both biased towards zero, with the bias increasing with the magnitude of the thresholds, and the bias slowly decreases with the sample size. The apparent slow convergence of the thresholds is surprising given the result from Chan (1993) that the thresholds are super-consistent. Figure 1 illustrates the asymmetric distribution of the threshold estimates for the band specification with $c = 3$ and $T = 250$. The histogram of the thresholds clearly shows that there is considerable uncertainty in the estimates for moderate sample sizes. Next, consider the estimates of the regime specific means. For both univariate and multivariate models, the estimated means downward biased but are closer on average to the true thresholds, which are equal to the true means in the continuous specification used here, than the estimated thresholds. However, there is substantially more uncertainty in the estimated means than in the estimated thresholds. Finally, consider the estimates of the speed of adjustment parameters. The estimates in the outer regimes are very similar, capture the symmetry of the true model, are closest to the true values for $c = 3$ and tend to be biased towards nonstationary values as the threshold band increases. The estimates in the middle regime are generally biased toward stationary values but show considerably more sampling uncertainty than the outer regime estimates. Overall, all of the estimates look reasonable on average but there is considerable sampling uncertainty even for fairly large sample sizes.

Table 5 presents results for the estimated parameters of the restricted BAND-TAR model and the restricted BAND-TVECM as well as Wald and LR tests for the restrictions imposed by these models. The Wald and LR tests based on the TAR model (6) are for the null hypothesis of a band specification with symmetric adjustment with equal regime specific means: $\rho^{(1)} = \rho^{(3)}, \delta^{(1)} = -\delta^{(3)}, \rho^{(2)} = 1, \delta^{(2)} = 0$. The tests based on the TVECM (5) are for the restrictions $\gamma_1^{(1)} = \gamma_1^{(3)}, \gamma_2^{(1)} = \gamma_2^{(3)}, \alpha_1^{(1)} = -\alpha_1^{(3)}, \alpha_2^{(1)} = -\alpha_2^{(3)}, \boldsymbol{\gamma}^{(2)} = 0, \boldsymbol{\alpha}^{(2)} = 0$. The restricted estimates of the thresholds are much closer to the true values than the unrestricted estimates and the sampling uncertainty is also much smaller. This result is striking given Chan and Tsay's (1998) result that the threshold estimates from a continuous TAR model converge at a slower rate than the threshold estimates from an unrestricted (discontinuous) TAR model. The restricted estimates of the common autoregressive coefficient in the outer regimes is slightly downward biased for small values of the threshold and slightly upward biased for large values. The 5% Wald and LR tests of the restrictions imposed by the continuous, symmetric threshold models are substantially size-distorted even for large sample sizes. The large size distortions indicate that the asymptotic normal distributions of the estimated parameters are not good approximations to the finite sample distributions and that specification tests based on nested hypothesis tests are essentially useless.

5 Application to U.S. Disaggregated Price Data

5.1 Data Description

We use U.S. Bureau of Labor Statistics monthly price indexes from December, 1986 through June, 1996. These consist of 43 categories of goods for 29 cities, although due to missing data in some series we restrict our analysis to 41 categories. Details of the data can be found in Appendix B. The data are the same as that used recently by Engel and Rogers (1998). Compared to the individual goods price data used by Parsley and Wei (1996) and O'Connell and Wei (1997), our data are less specific and more aggregated but cover a larger geographical dispersion. An advantage of using price index data is that the price index for a particular good is less likely to be affected by the marketing behavior of one or a few manufacturers or wholesalers which can distort the effect of arbitrage forces on prices.

We consider bivariate systems of log prices of common goods in different locations relative to a benchmark city. Following O'Connell and Wei (1997) we choose New Orleans as our benchmark city. Let $p^i_{k,t}$ denote the log price of good k in location i (excluding the benchmark location) at month t and let $p^{NO}_{k,t}$ denote the log price of good k in New Orleans, our benchmark city. With 41 goods and 28 cities (excluding the benchmark) we have 1,148 bivariate systems of log prices. Under the assumption that log prices are $I(1)$, the LOP indicates that, for all i and k considered, $(p^i_{k,t}, p^{NO}_{k,t})'$ is cointegrated with cointegrating vector $(1, -1)'$. Accordingly, for each system of log prices we also construct the log price differential (cointegrating residual) defined as $z^i_{k,t} = p^i_{k,t} - p^{NO}_{k,t}$. We note that our analysis of LOP deviations differs from that of O'Connell and Wei (1997) in that we do not group similar goods, or all goods from a particular city, into panels. Our view is that panel models impose too many restrictions on price dynamics that are likely to be violated in the data.

5.2 Testing Threshold Cointegration

In this sub-section, we perform some preliminary analysis on the data to illustrate some properties and stylized facts. As the first step in our analysis of the LOP, we test the null hypothesis of no cointegration for bivariate systems of log price data. We then investigate a measure of persistence in the deviation from the law of one price by fitting linear AR(1) models and computing half-lives. Next, we summarize the adjustment behavior of log price differences by fitting linear VECMs. Finally, we test for linearity.

Table 6 gives the results of various tests for no-cointegration. For each test, we count the number of bivariate price systems in each category of the 28 goods and services for which we reject the null hypothesis of no cointegration at the 10% level. We would expect the LOP to hold and, hence, cointegration to be found for systems of prices of tradable homogenous goods in the absence of government price controls. ADF tests on the LOP deviation $z^i_{k,t}$, however, generally indicate the presence of unit roots. Some exceptions are tradable goods such as Fresh Fruits and Vegetables, Fuels, Men's and Boy's Apparel and Motor Fuels. More suggestive results emerge from the application of the multivariate Horvath-Watson test. Cointegration is generally found in categories with relatively homogenous products (Meats, Fresh Fruits and Vegetables, Eggs, Fuels, Fish and Seafood, Poultry, Motor Fuels) and is generally not found in categories that contain relatively heterogeneous goods and services (Used Cars), categories of goods subject to severe federal or state government interventions such as taxation and sale regulations (Alcoholic Beverages, Tobacco and Smoking Products), and domestically oriented goods and services (Household Maintenance and Repairs, Personal and Educational Expenses, Residential Rent, Automobile Maintenance and Repairs). The results from the EG and BVD tests are almost the same as those from the HW test and indicate that a few more tradable goods categories exhibit cointegration (Cereals and Bakery Products and Processed Fruits and Vegetables). However, it is puzzling that the EG and BVD tests find cointegration in the nontradable category Other Renter's Costs and Residential Rent.

Table 7 summarizes the results from estimating AR(1) models to the log price differentials. Consistent with the outcomes of the HW, EG and BVD no cointegration tests, the categories that, on average, have highly persistent LOP deviations are goods and services that are relatively heterogeneous, goods subject to severe federal or state government interventions and domestically oriented goods. Other Apparel Commodities and Infants' and Toddlers' Apparels are anomalies in this group. Price differences of relatively homogenous products tend to have stronger mean-reverting behavior with half-lives less than 3 months. Interestingly, the standard deviations of the autoregressive coefficients in each category of this group are much higher than in the other groups.

To get an indication of the potential asymmetry of the dynamics in price adjustment we estimate simple linear VECMs for the price pairs that appear to be cointegrated based on the outcomes of the no cointegration tests. We generally find evidence for asymmetry in the adjustment of goods prices to the LOP between certain locations. For example, Table 8 summarizes the estimated speed of adjustment parameters for the VECM using the

New York-New Orleans price data. The estimated coefficients for γ^{NO} are relatively large and positive (from 0.1 to 0.4), and the estimates for γ^{NY} are negative and close to zero. Apparently, prices in New Orleans tend to "catch up" to prices in New York in response to LOP deviations while the prices in New York are unaffected.

Given that cointegration is found in some categories, we proceed to test for linearity. The results based on 10% tests are reported in Table 9. Both the univariate and multivariate Tsay's tests find more evidence of nonlinearity than the Hansen-type F and LR tests. This result may be due to the fact that the former two tests are non-parametric and may be subjected to less model mis-specification error. Also, the rejections of linearity must be interpreted with care because the tests are designed to test for linearity within a stationary null hypothesis. If the data have unit root or near unit root behavior the size of the tests may be distorted. Nevertheless, all four tests fail to find the pattern of persistence in LOP deviations in Table 7 and are not very informative.

5.3 Estimation Results for Threshold Models

Based on the results of the no-cointegration tests in the previous section and our finding that specification tests are likely severely size distorted, we use the continuous and symmetric versions of the TAR model (14) and TVECM in (11) to characterize the nonlinear dynamics for categories that contains threshold-cointegrated prices. We estimate models only for the city pairs of the categories that appear to be threshold-cointegrated, at the 10% level, based on the outcome of the BVD test, which we find to be the most powerful test for detecting Band-TAR type cointegration . By estimating these models, we attempt to answer the following questions: (i) Can we detect faster convergence to the LOP when threshold nonlinearity is observed? (ii) In the presence of threshold-type nonlinearity, is there asymmetry in price adjustments between two cities?

Because the "bands of no arbitrage" in these models are symmetric, it is reasonable to demean the data so that zero becomes the equilibrium level or the global unconditional mean of the price differences. However, unlike Obstfeld and Taylor (1997), we find no a priori reason to detrend the data. Table 10 reports the estimated coefficients from the univariate and multivariate threshold models. We summarize the results with the median estimates within each category. The improvement in the estimated speed of price adjustment for the univariate model is astonishing. The medians of the coefficient estimates of the linear error correction model are very close to zero. That of the TAR model is significantly smaller. This seems to be a feature of the TAR model where adjustment is forced towards either the bands or the equilibrium. Judging from the half-lifes computed, the persistence almost disappears. This seems to be a feature of the TAR model where adjustment is forced towards either the bands or the equilibrium. Asymmetry in dynamics is also apparent in the estimates from the TVECM. Notice that, in general, prices in New Orleans tend to catch up to the prices of the other cities much faster than the latter prices tend to decrease. The coefficients of price adjustment for other cities range from -0.3542 to 0.0297. Yet, for example, the coefficients for New Orleans can be as high as 0.6483 (Meats) and 0.7534 (Entertainment Services).

The values of the symmetric thresholds estimated from the TVECM are graphed in Figure 2. If the notion of commodity arbitrage is true, everything else being constant, we should expect the shape of the bar charts to have similar patterns, indicating that market frictions against price adjustment between two cities depends on some characteristics of the city pair; for instance, distance[10] . From Figure 2, we see that no apparent pattern is found in the distribution of the thresholds. For example, Anchorage Alaska (the first bar in all figures) is probably the furthest from New Orleans; in fact, both people and goods have to pass through the U.S.-Canadian border from New Orleans to anywhere in Alaska. We would expect, even if prices converge between the two cities, market friction should be always higher than most city pairs. This only holds, however, for a few commodities such as Men's and Boy's Apparel, Textile and House Furnishings and Women's and Girl's Apparel. The thresholds estimated for Anchorage-New Orleans in other categories, especially the food categories, are rather small. Hence, the estimated thresholds appear not to be consistent with the commodity arbitrage theory and we cannot interpret the nonlinearity found is due to the distance between cities.

These results imply two possibilities. First, distance is not the only crucial factor for market segmentation. Among some of the figures, one feature is interesting: for more homogenous goods, such as Fish and Seafood, Meats, Motor Fuel and Poultry, the variation among the threshold in each of these categories is small; but for more heterogeneous goods, such as Footwear, Men's and Boy's Apparel and Women's and Girl's Apparel, the variation is larger and there are more outliers in it. Therefore, in order to test the transaction costs theory in future work, we may want to investigate individual goods markets as the charac-

[10]For international data, we would also expect formal and informal trade barriers and exchange rate volatility will play a role, but this should not be the case in our domestic price data.

teristics of a particular commodity may create some distortions in estimation of threshold models. Second, the strongly restricted versions of threshold models we estimate are potentially misspecified. Relaxing the assumptions of symmetric bands, symmetric dynamics and the continuous property of the models may give us more informative estimates. Of course, there are also many choices of other nonlinear models, such as STAR models, that may be more appropriate.

6 Conclusion

This paper assesses a battery of cointegration and linearity tests that are useful for testing for threshold cointegration in bivariate systems of price data. We find that tests of no-cointegration that incorporate the threshold nature of the alternative in general have higher power than tests that ignore the form of the alternative model. Comparison within the group of linearity tests suggests that multivariate tests have similar power as univariate tests. We also find that estimates from unrestricted univariate TAR models and multivariate TVECMs behave similarly and that convergence to asymptotic normal distributions for the estimated parameters is very slow for data generated from BAND-TVECMs. Moreover, we find simple specification tests based on Wald and LR statistics to be almost useless for moderate to large sample sizes. For our empirical investigation, we investigate the evidence for threshold-type nonlinearity in a large data set of U.S. disaggregated goods prices. We find threshold-type nonlinearity mostly in goods that are tradable and relatively homogenous. In addition, we find that prices may adjust at different speed for different cities. Nevertheless, the threshold estimates are not consistent with the motivation of commodity arbitrage. We suggest that, in order to address to issue of nonlinearity with this particular motivation in mind, we may want to investigate individual markets or adopt a different type of threshold model.

7 Appendix A

Here we describe the technical aspects of some of the testing and estimation procedures used in the paper.

7.1 Estimation of Multivariate TVECM

The estimation of univariate TAR models is outlined in Tong (1983) and is recently reviewed in Hansen (1999). The estimation of continuous univariate TAR models is discussed in Chan and Tsay (1998) and Berben and van Dijk (1998, 1999). The main estimation technique is sequential conditional least squares. The estimation of multivariate TVAR models involves a similar strategy to that used for univariate models and is discussed in Tsay (1998). Since the estimation of a TVECM with a known cointegrating vector has not been discussed very much in the literature we briefly review it here. Our strategy is to combine Hansen's treatment of the estimation of two and three regime TAR models with Tsay's treatment of the estimation of multivariate TVAR models.

Consider the unrestricted two regime bivariate TVECM with $k-1$ lags (10) written as the multivariate regression

$$\Delta \mathbf{p}_t' = \begin{cases} \mathbf{x}_{t-1}'\mathbf{\Theta}^{(2)} + \boldsymbol{\varepsilon}_t', & \text{if } z_{t-d} > c \\ \mathbf{x}_{t-1}'\mathbf{\Theta}^{(1)} + \boldsymbol{\varepsilon}_t', & \text{if } z_{t-d} \le c \end{cases}$$

where $\mathbf{x}'_{t-1} = (1, z_{t-1}, \Delta\mathbf{p}'_{t-1}, \ldots, \Delta\mathbf{p}'_{t-k+1})$, $z_{t-1} = \boldsymbol{\beta}'\mathbf{p}_{t-1}$, $\boldsymbol{\beta}$ is known and $\boldsymbol{\Theta}^{(j)}$ is a $(k+1)\times 2$ matrix. Defining $I_t(c,d) = I(z_{t-d} < c)$, where $I(\cdot)$ is the indicator function such that $I(A) = 1$ if A is true and 0 otherwise, the above model may be rewritten as the multivariate regression model

$$\Delta\mathbf{p}'_t = \mathbf{x}'_{t-1}\boldsymbol{\Theta}^{(1)}I_t(c,d) + \mathbf{x}'_{t-1}\boldsymbol{\Theta}^{(2)}(1 - I_t(c,d)) + \boldsymbol{\varepsilon}'_t \tag{18}$$

The delay parameter d is integer valued and is assumed to be less than some upper bound $\overline{d}$. Since the threshold value c only arises through the indicator function $I(z_{t-d} < c)$ there is no loss in restricting the possible values for c to the observed values of z_{t-d}. Also, for practical matters it is necessary to restrict the threshold c so that each regime contains a minimal number of observations. Let T_i denote the number of sample observations in regime i and let T denote the sample size. Hansen (1999) suggests constraining the thresholds so that as $T \to \infty, T_i/T \geq \tau$ for some $\tau \in (0,1)$. Hansen suggests setting $\tau = 0.1$.

The model (18) may be estimated by sequential multivariate least squares in two steps. In the first step, conditional on (c,d), the parameters $(\boldsymbol{\Theta}^{(1)}, \boldsymbol{\Theta}^{(2)})$ may be estimated by multivariate least squares giving the residual sum of squares

$$S_2(c,d) = trace\left(\widehat{\boldsymbol{\Sigma}}_2(c,d)\right)$$

where $\widehat{\boldsymbol{\Sigma}}_2(c,d)$ denotes the multivariate least squares estimate of $\boldsymbol{\Sigma} = var(\boldsymbol{\varepsilon}_t)$ conditional on (c,d) for the two-regime model. In the second step, the least squares estimates of (c,d) are obtained as

$$(\widehat{c}, \widehat{d}) = \arg\min_{c,d} S_2(c,d)$$

The requirement that $T_1 \geq T\tau$ restricts the search over c to values of z_{t-d} that lie between the τ^{th} and $(1-\tau^{th})$ quantiles of z_{t-d}. The final estimates of $\mathbf{\Theta}^{(j)}$ are given by $\widehat{\mathbf{\Theta}}^{(j)} = \widehat{\mathbf{\Theta}}^{(j)}(\widehat{c}, \widehat{d})$ and the estimate of the residual covariance matrix is given by $\widehat{\Sigma}_2(\widehat{c}, \widehat{d})$

Under mild regularity conditions, Tsay (1998) shows that the sequential conditional multivariate least squares estimates of $(\mathbf{\Theta}^{(1)}, \mathbf{\Theta}^{(2)}, c, d)$ are strongly consistent and the least squares estimates of $\mathbf{\Theta}^{(1)}$ and $\mathbf{\Theta}^{(2)}$ are asymptotically normally distributed independent of c and d. Also, the estimates of c and d converge at rate T.

Next, consider a general three regime TVECM

$$\Delta p_t' = \begin{cases} \mathbf{x}_{t-1}'\mathbf{\Theta}^{(3)} + \boldsymbol{\varepsilon}_t', \text{ if } z_{t-d} > c^{(2)} \\ \mathbf{x}_{t-1}'\mathbf{\Theta}^{(2)} + \boldsymbol{\varepsilon}_t', \text{ if } c^{(1)} \leq z_{t-d} \leq c^{(2)} \\ \mathbf{x}_{t-1}'\mathbf{\Theta}^{(1)} + \boldsymbol{\varepsilon}_t', \text{ if } z_{t-d} < c^{(1)} \end{cases} \tag{19}$$

This model can be compactly expressed as

$$\Delta \mathbf{p}_t' = \mathbf{\Theta}^{(1)}\mathbf{x}_{t-1}' I_t^{(1)}(\mathbf{c}, d) + \mathbf{\Theta}^{(2)}\mathbf{x}_{t-1}' I^{(2)}(\mathbf{c}, d) + \mathbf{\Theta}^{(3)}\mathbf{x}_{t-1}' I_t^{(3)}(\mathbf{c}, d) + \boldsymbol{\varepsilon}_t' \tag{20}$$

where $\mathbf{c} = (c^{(1)}, c^{(2)})'$ and $I_t^{(j)}(\gamma, d) = I(c^{(j-1)} < z_{t-1} < c^{(j)})$. Conditional on $(\mathbf{c}, d)$, the parameters $(\mathbf{\Theta}^{(1)}, \mathbf{\Theta}^{(2)}, \mathbf{\Theta}^{(3)})$ may be estimated by multivariate least squares giving the residual sum of squares $S_3(\mathbf{c}, d)$ and the estimates of $\mathbf{c}$ and d may be found by minimizing $S_3(\mathbf{c}, d)$ using a three dimensional grid search.

The above method is computationally burdensome since if N points are evaluated at each value of $c^{(1)}$ and $c^{(2)}$ then the grid search over $(\mathbf{c}, d)$ in the second step involves $\overline{d} \cdot N^2$ regressions, which can be time consuming if N is reasonably large. Hansen (1999) suggests a computational short-cut that is related to the sequential estimation of multiple breakpoints proposed by Bai (1997). First, estimate the (misspecified) two regime model (18) giving least squares estimates $(\widehat{d}, \widehat{c}^1)$. The results of Bai (1997) indicate estimate $\widehat{d}$ will be consistent for d and $\widehat{c}^1$ will be consistent for one of the threshold pairs $(c^{(1)}, c^{(2)})$. Next, estimate $\mathbf{c} = (c^{(1)}, c^{(2)})$ by least squares on (19) imposing $d = \widehat{d}$ and that one element of $\mathbf{c}$ equals $\widehat{c}^1$. The resulting estimate $\widehat{c}^2$ will be consistent for the remaining element of the pair $(c^{(1)}, c^{(2)})$[11]. This sequential procedure to estimate the three regime model offers considerable computational savings over a three dimensional grid search and allows us to conduct an extensive Monte Carlo analysis[12] .

To estimate the multivariate TVECM (9) where the constant is restricted to the error correction term the above procedure needs to be modified since multivariate least squares is no longer efficient, although it is consistent, due to the cross equation restrictions. Since $\mathbf{c}$ and d are super consistent in the unrestricted model (19), a simple two-step estimation procedure can be used. In the first step, $\mathbf{c}$ and d are estimated from the unrestricted model (19). In the second step, the restricted constant three-regime model with the first step estimates of $\mathbf{c}$ and d imposed is estimated using Zellner's seeming unrelated regression (SUR) technique

$$(\widehat{\boldsymbol{\Theta}}^{(1)}_{SUR}(\widehat{\mathbf{c}}, \widehat{d}), \widehat{\boldsymbol{\Theta}}^{(2)}_{SUR}(\widehat{\mathbf{c}}, \widehat{d}), \widehat{\boldsymbol{\Theta}}^{(3)}_{SUR}(\widehat{\mathbf{c}}, \widehat{d})) = \underset{\boldsymbol{\Theta}^{(1)}, \boldsymbol{\Theta}^{(2)}, \boldsymbol{\Theta}^{(2)}}{\arg\min} \; \frac{1}{2} \log\left(|\boldsymbol{\Sigma}_{SUR,3}(\widehat{\mathbf{c}}, \widehat{d})| \right)$$

where $\boldsymbol{\Sigma}_{SUR,3}(\widehat{\mathbf{c}}, \widehat{d}) = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\varepsilon}_t(\widehat{\mathbf{c}}, \widehat{d})\boldsymbol{\varepsilon}_t(\widehat{\mathbf{c}}, \widehat{d})'$ and the parameters corresponding to the intercepts in $\boldsymbol{\Theta}^{(j)}$ are given by $-\mu^{(j)}\gamma_1^{(j)}$ and $-\mu^{(j)}\gamma_2^{(j)}$.

[11] Bai (1997) shows that the sequential procedure gives consistent estimates. Additionally, he shows that a simple iteration of the method yields asymptotically efficient estimates as well.

[12] In the estimation, the restriction that all three regimes have at least $T\tau$ needs to be imposed. Additionally, in the second and third-stage searches the requirement that at least $T\tau$ observations lie in the regime where $c^{(1)} \leq z_{t-d} \leq c^{(2)}$ (or $c^{(2)} \leq z_{t-d} \leq c^{(1)}$ if $c^{(2)} < c^{(1)}$) needs to be imposed.

7.2 Tsay's Multivariate Test for Threshold Nonlinearity

Tsay (1998) discusses a multivariate generalization of his univariate test for threshold nonlinearity. To implement his test in the present context we need to consider an arranged multivariate regression for the VECM

$$\Delta \mathbf{p}_t' = \mathbf{x}_t' \boldsymbol{\Theta} + \boldsymbol{\varepsilon}_t', \ t = h+1, \ldots, T$$

where $h = \max(k, d)$, $\mathbf{x}_{t-1}' = (1, z_{t-1}, \Delta \mathbf{p}_{t-1}', \ldots, \Delta \mathbf{p}_{t-k+1}')$, $z_{t-1} = \boldsymbol{\beta}' \mathbf{p}_{t-1}$, $\boldsymbol{\beta}$ is known and $\boldsymbol{\Theta}$ is a $(k+1) \times 2$ matrix with first row giving the coefficients for Δp_{1t} and second row giving the coefficients for Δp_{2t}. The threshold variable z_{t-d} assumes values in $S = \{z_1, \ldots, z_{T-d}\}$. Consider the order statistics for S and denote the ith smallest element of S by $z_{(i)}$ and let $t(i)$ denote the time index of $z_{(i)}$. Then the arranged multivariate regression is

$$\Delta \mathbf{p}_{t(i)+d}' = \mathbf{x}_{t(i)+d}' \boldsymbol{\Theta} + \boldsymbol{\varepsilon}_{t(i)+d}', \ i = 1, \ldots, T-h \tag{21}$$

Let $\widehat{\Theta}_m$ denote the multivariate least squares estimate of $\boldsymbol{\Theta}$ from (21) using data from $i = 1, \ldots, m$. Define

$$\widehat{\boldsymbol{\varepsilon}}_{t(m+1)+d} = \Delta \mathbf{p}_{t(m+1)+d} - \widehat{\boldsymbol{\Theta}}_m' \mathbf{x}_{t(m+1)+d}$$

and

$$\widehat{\boldsymbol{\xi}}_{t(m+1)+d} = \frac{\widehat{\boldsymbol{\varepsilon}}_{t(m+1)+d}}{\left[1 + \mathbf{x}_{t(m+1)+d}' \mathbf{V}_m \mathbf{x}_{t(m+1)+d}\right]^{\frac{1}{2}}},$$

where $\mathbf{V}_m = \left(\sum_{i=1}^m \mathbf{x}_{t(m+1)+d} \mathbf{x}_{t(m+1)+d}'\right)^{-1}$, as the predictive residual and the standardized predictive residual of the arranged regression computed using recursive multivariate least squares. Next, consider the multivariate regression

$$\widehat{\boldsymbol{\xi}}_{t(l)+d}' = \mathbf{x}_{t(l)+d}' \boldsymbol{\Psi} + \mathbf{w}_{t(l)+d}', \ l = m_0 + 1, \ldots, T-h, \tag{22}$$

where m_0 denotes the starting point of the recursive least squares estimation. If there is no threshold nonlinearity then $\boldsymbol{\Psi}$ should be zero in (22). Tsay's test statistic for testing $H_0 : \boldsymbol{\Psi} = \mathbf{0}$ versus $H_a : \boldsymbol{\Psi} \neq \mathbf{0}$ in (22) is given by

$$C(d) = (T - h - m_0 - (2 \cdot (k-1) + 1)) \cdot \{\ln(\det(\mathbf{S}_0)) - \ln(\det(\mathbf{S}_1))\} \tag{23}$$

where

$$\mathbf{S}_0 = \frac{1}{T - h - m_0} \sum_{l=m_0+1}^{T-d} \widehat{\boldsymbol{\xi}}_{t(l)+d} \widehat{\boldsymbol{\xi}}'_{t(l)+d},$$
$$\mathbf{S}_1 = \frac{1}{T - h - m_0} \sum_{l=m_0+1}^{T-d} \widehat{\mathbf{w}}_{t(l)+d} \widehat{\mathbf{w}}'_{t(l)+d}$$

and $\widehat{w}_t$ is the least squares residual from (22). Under the null of linearity $C(d)$ is asymptotically distributed as a chi-square random variable with $2 \cdot (k-1) + 1$ degrees of freedom.

The null hypothesis $\Psi = 0$ includes a zero intercept for all predictive residuals. Tsay (1998) remarks that due to finite sample bias, in some applications one may wish to exclude the intercept terms from the nonlinearity test in (23). In this case S_0 should be mean corrected and the resulting test has an asymptotic chi-squared distribution with $2 \cdot (k-1)$ degrees of freedom.

References

[1] Balke, N.S., and T.B. Fomby (1997), "Threshold cointegration," *International Economic Review*, 38, 627-645.

[2] Bai, J. (1997), "Estimating multiple breaks one at a time," *Econometric Theory* 13, 315-352.

[3] Berben, R.-P. and D. van Dijk (1998), "Nonlinear adjustment to purchasing power Parity: Evidence on the Effects of Transactions Costs," Working Paper, Tinbergen Institute, Erasmus University Rotterdam.

[4] Berben, R.-P. and D. van Dijk (1999) "Unit root tests and asymmetric adjustment: a reassessment," Working Paper, Tinbergen Institute, Erasmus University of Rotterdam

[5] Caner, M. and B.E. Hansen (1998)," Threshold autoregressions with a unit root," Working Paper, Department of Economics, University of Wisconsin, Madison.

[6] Chan, K.S. (1993), "Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model," *Annals of Statistics* 21, 520-533.

[7] Chan, K.S., and R. Tsay (1998), "Limiting properties of the least squares estimator of a continuous threshold autoregressive model," *Biometrika* 85, 413-426.

[8] Chan, K.S., J.D. Petrucelli, H.Tong and SW. Woolford (1985), "A Multiple Threshold AR(1) Model," *Journal of Applied Probability* 22, 267-279.

[9] Dumas, B. (1992), "Dynamics equilibrium and the real exchange rate in a spatially separated world," *The Review of Financial Studies* 5, 153-180.

[10] Clements, M.P. and H.-M. Krolzig (1998), "A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP," *Econometrics Journal* 1, C47-C75.

[11] Davies, R.B. (1977), "Hypothesis testing when a nuisance parameter is present only under the alternative," *Biometrika* 64, 247-254.

[12] Davies, R.B. (1987), "Hypothesis testing when a nuisance parameter is present only under the alternative," *Biometrika* 74, 33-43.

[13] Enders, W. and C.W.J. Granger (1998), "Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates," *Journal of Business of Economic Statistics* 16, 304-311.

[14] Engel, C. and J.H. Rogers (1996), "How wide is the border," *American Economic Review* 86, 1112-1125.

[15] Engel, C. and J.H. Rogers (1998a), "Relative price volatility: what role does the border play?," Working Paper, Department of Economics, University of Washington.

[16] Engel, C. and J.H. Rogers (1998b), "Violating the law of one price: should we make a federal case out of it?," Working Paper, Department of Economics, University of Washington.

[17] Frankel, J. A. (1986), "International capital mobility and crowding-out in the U.S. economy: imperfect integration of financial markets or of goods markets?," in *How open is the U.S. Economy*, eds R.W. Hafer, Lexington Book, Lexington.

[18] Froot, K., and K. Rogoff (1995), "Perspectives on PPP and long-run real exchange rate," in G. Grossman and K. Rogoff eds., *Handbook of International Economics*. Amsterdam: North-Holland.

[19] González, M. and J. Gonzalo (1997), "Threshold unit root models," Working Paper No. 97-50, Universidad Carlos III de Madrid.

[20] González, M. and J. Gonzalo (1999), "Inference and estimation in threshold cointegration models," manuscript, Universidad Carlos III de Madrid.

[21] Hansen, B.E. (1997), "Inference in TAR models," *Studies in Nonlinear Dynamics and Econometrics* 1(2), 1-14.

[22] Hansen, B.E. (1999), "Testing linearity," forthcoming in *Journal of Economic Surveys*.

[23] Hechscher, E.F. (1916), "Växelkursens Grundval vid Pappersmyntfot," *Ekonomisk Tidskrift* 18, 309-312.

[24] Horvath, M.T.K. and M.W. Watson (1995), "Testing for cointegration when some of the cointegrating vectors are prespecified," *Econometric Theory* 11, 984-1014.

[25] Michael, P., R.A. Nobay, and D.A. Peel (1997), "Transactions costs and nonlinear adjustment in real exchange rates: an empirical investigation," *Journal of Political Economy* 105, 862-879.

[26] Obstfeld, M. and A.M. Taylor (1997), "Nonlinear aspects of goods-market arbitrage and adjustment: Heckshер's commodity points revisited," *Journal of the Japanese and International Economies* 11, 441-479.

[27] O'Connell, P.G.J. (1998), "Market frictions and real exchange rates," *Journal of International Money and Finance* 17, 71-95.

[28] O'Connell, P.G.J. and S.-J.Wei (1997), "'The bigger they are, the harder they fall' How price differences across U.S. cities are arbitraged," Working Paper, National Bureau of Economic Research, No. 6089.

[29] Parsley, D.C. and S.-J.Wei (1996), "Convergence to the law of one price without trade barriers or currency fluctuations," *Quarterly Journal of Economics* 111, 1211-1236.

[30] Pippenger, M.K. and G.E. Goering (1993), "A Note on the empirical power of unit root tests under threshold processes," *Oxford Bulletin of Economics and Statistics* 55, 473-481.

[31] Sarantis, N. (1999), "Modeling non-linearities in real effective exchange rates," *Journal of International Money and Finance* 18, 27-45.

[32] Sercu, P., R. Uppal, and C. Van Hulle (1995), "The exchange rate in the presence of transaction costs: implications for tests of purchasing power parity," *Journal of Finance* 10, 1309-1319.

[33] Tong, H. (1990), *Nonlinear time series: A Dynamical Systems Approach*, Oxford University Press, Oxford, U.K.

[34] Tsay, R. (1989), "Testing and modeling threshold autoregressive processes," *Journal of the American Statistical Association*, 84, 231-240.

[35] Tsay, R. (1998), "Testing and modeling multivariate threshold models," *Journal of the American Statistical Association*, 93(443), 1188-1202.

[36] Zivot, E. (1999), "The power of single equation tests for cointegration when the cointegrating vector is pre-specified," forthcoming in *Econometric Theory*.

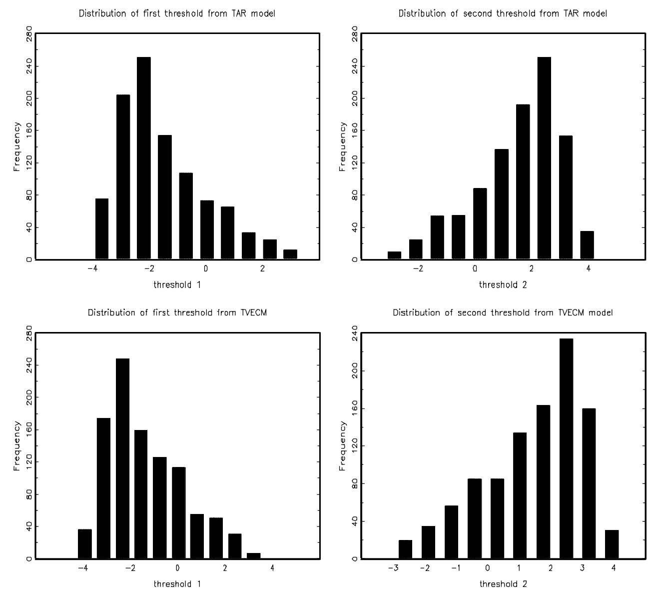
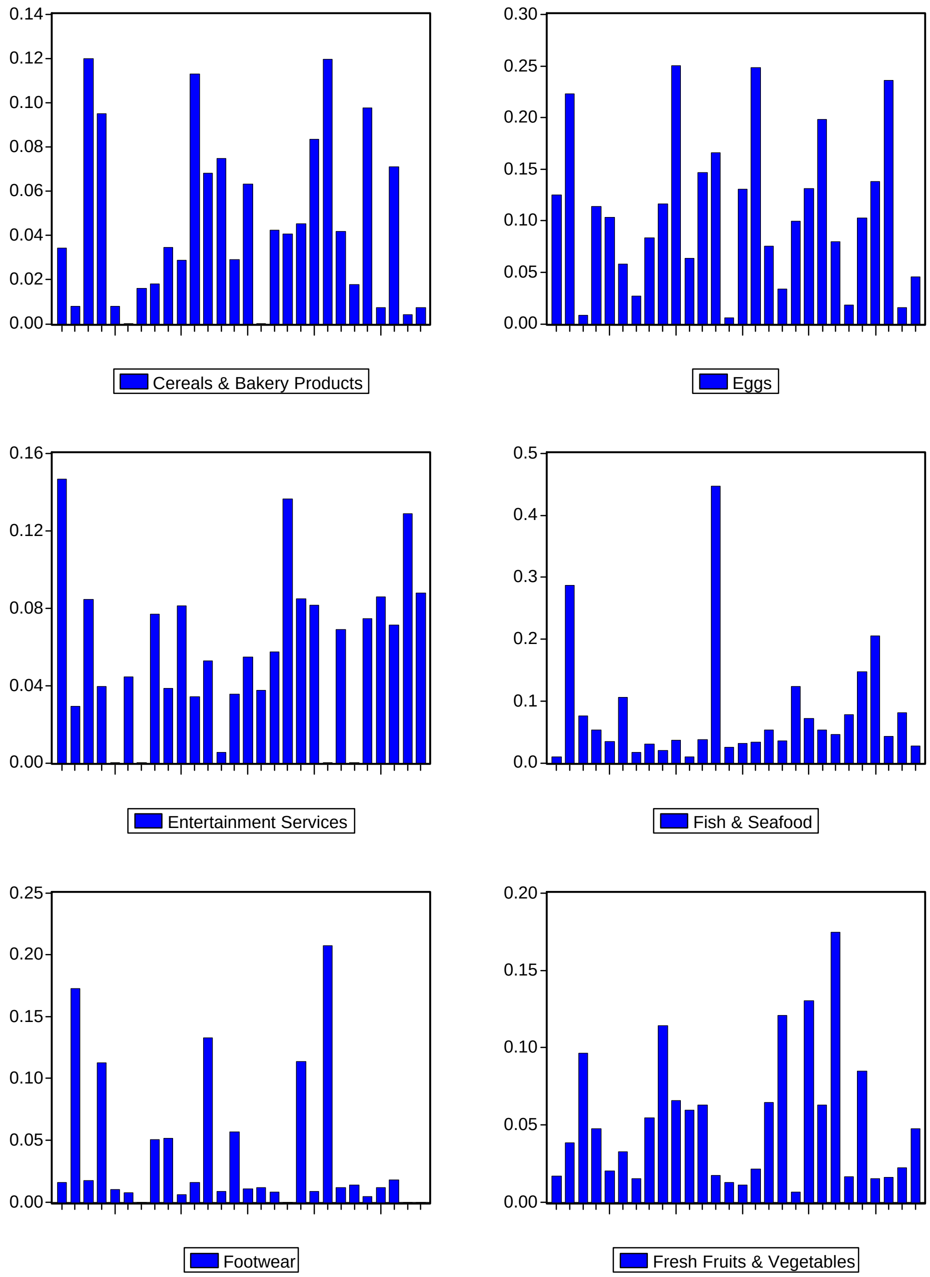


Figure 1: Distribution of estimated thresholds from unrestricted TAR(3) and TVECM(3) with $c = 3$ and $T = 250$.

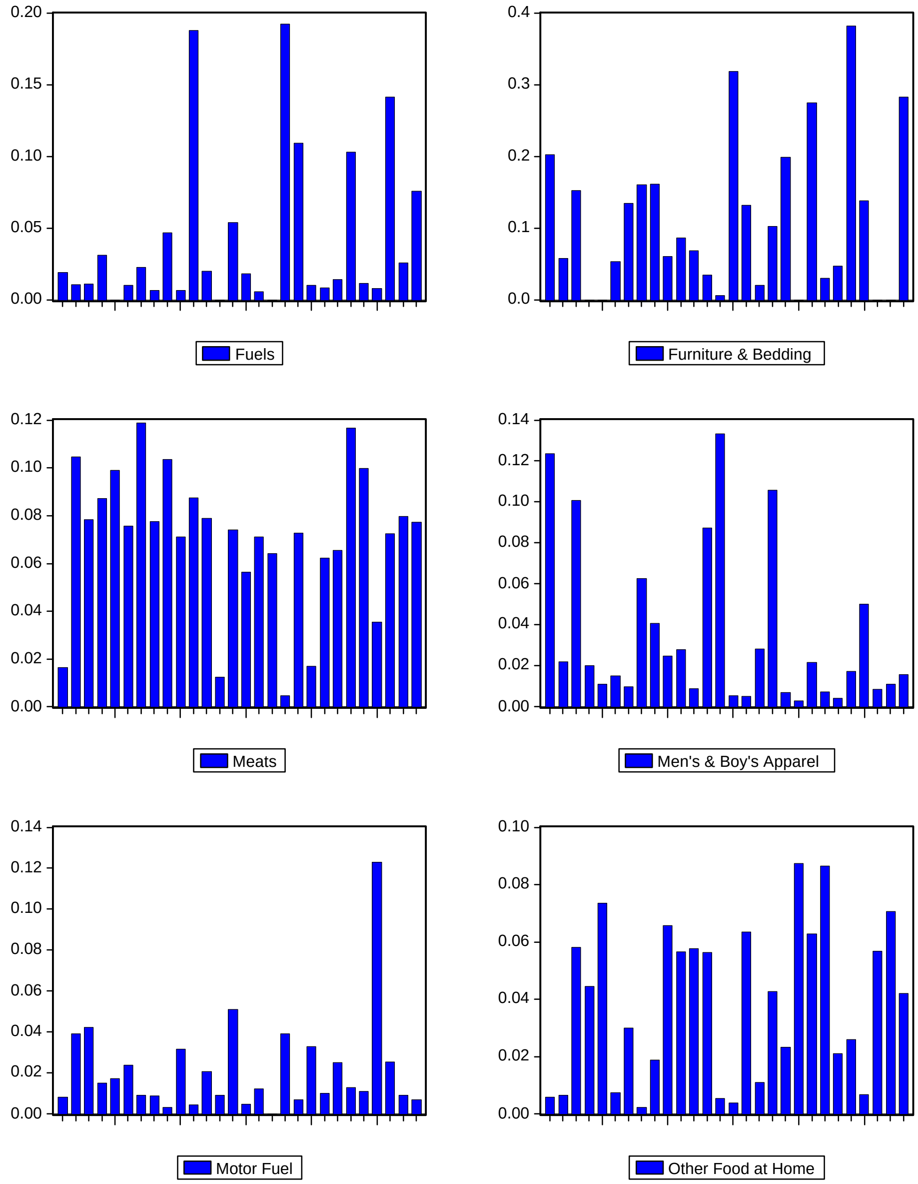
Appendix B: Details of the Data

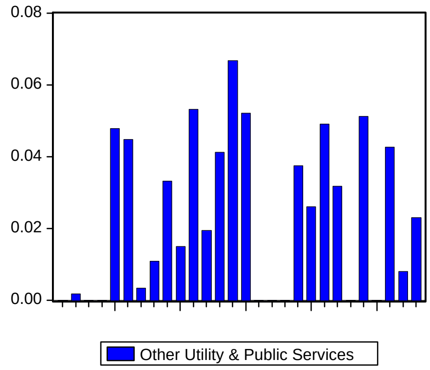
Categories	Cities	Abbv.
1. Alcoholic Beverages	1. Anchorage	AN
2. Apparel Services	2. Atlanta	AT
3. Appliances, incl. electronic equipment	3. Baltimore	BT
4. Automobile Maintenance & Repairs	4. Boston	BO
5. Cereals and Bakery Products	5. Buffalo	BU
6. Eggs	6. Chicago	CH
7. Entertainment Commodities	7. Cleveland	CL
8. Entertainment Services	8. Cincinnati	CI
9. Fish & Seafood	9. Dallas	DA
10. Food Away from Home	10. Denver	DN
11. Footwear	11. Detroit	DT
12. Fresh Fruits and Vegetables	12. Honolulu	HO
13. Fuels	13. Houston	HS
14. Furniture & Bedding	14. Kansas City	KC
15. Homeowners' Costs	15. Los Angeles	LA
16. Hospital & Related Services	16. Miami	MA
17. Household Maintenance & Repairs	17. Milwaukee	MI
18. Housekeeping Services	18. Minneapolis	MS
19. Housekeeping Supplies	19. New York	NY
20. Infants' and toddlers' Apparel	20. Philadelphia	PH
21. Meats	21. Pittsburgh	PI
22. Medical Care Commodities	22. Portland	PO
23. Men's and Boy's Apparel	23. San Diego	SD
24. Motor Fuel	24. San Francisco	SF
25. New Vehicles	25. Seattle	SE
26. Other Apparel Commodities	26. St. Louis	SL
27. Other Food at Home	27. Tampa Bay	TA
28. Other Furnishings	28. Washington DC	DC
29. Other Renter's Costs		
30. Other Utilities & Public Services	*Benchmark:*	
31. Personal & Educational Expenses	New Orleans	NO
32. Personal Care		
33. Poultry		
34. Processed Fruits and Vegetables		
35. Professional Medical Services		
36. Public Transportation		
37. Residential Rent		
38. Textile Housefurnishings		
39. Tobacco & Smoking Products		
40. Used Cars		
41. Women's and Girl's Apparel		

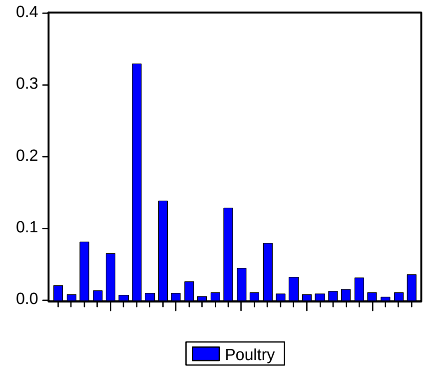
Diagram 2: Threshold Estimates from Symmetric Band-TVECM

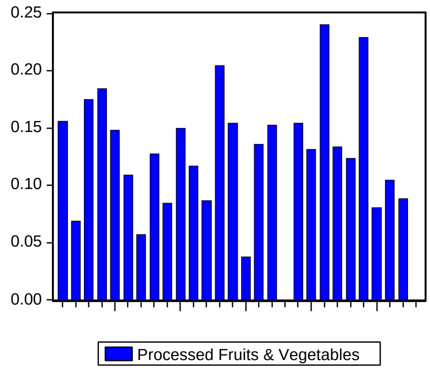


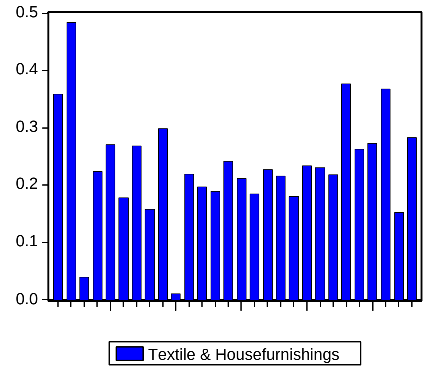
Note: The vertical axis represents the threshold estimates. The order of city (v.s. New Orleans) on the horizontal axis is (AN,AT, BA,BO,BU,CH,CI,CL,DA,DC,DN,CT,HO,HS,KC,LA,MA,MI,MS,NY,PH,PI,PO,SD,SE,SF,SL,TA).


0.20
0.15
0.10
0.05
0.00
Fuels
0.4
0.3
0.2
0.1
0.0
Furniture & Bedding
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Meats
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Men's & Boy's Apparel
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Motor Fuel
0.10
0.08
0.06
0.04
0.02
0.00
Other Food at Home


0.08
0.06
0.04
0.02
0.00
Other Utility & Public Services


0.4
0.3
0.2
0.1
0.0
Poultry


0.25
0.20
0.15
0.10
0.05
0.00
Processed Fruits & Vegetables


0.5
0.4
0.3
0.2
0.1
0.0
Textile & Housefurnishings

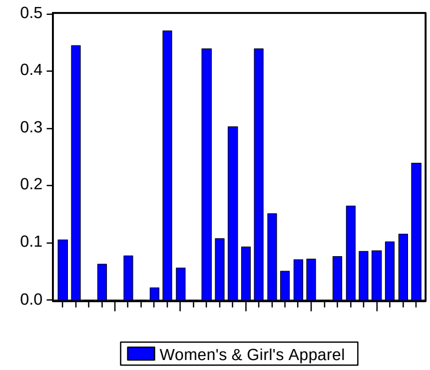

0.5
0.4
0.3
0.2
0.1
0.0
Women's & Girl's Apparel

Table 1: Tests for No-Cointegration
Empirical Rejection Frequency of 5% Tests

		EQ-TAR				BAND-TAR			
c	T	ADF	HW	EG	BVD	ADF	HW	EG	BVD
3	100	0.416	0.944	0.961	1	0.139	0.206	0.221	0.625
3	250	0.989	1	1	1	0.616	0.946	0.983	1
3	500	1	1	1	1	0.997	1	1	1
5	100	0.135	0.272	0.325	0.916	0.092	0.127	0.132	0.383
5	250	0.636	0.996	0.998	1	0.193	0.256	0.273	0.869
5	500	0.996	1	1	1	0.612	0.874	0.949	1
10	100	0.085	0.134	0.155	0.609	0.062	0.104	0.091	0.250
10	250	0.144	0.219	0.265	0.898	0.109	0.121	0.131	0.458
10	500	0.335	0.536	0.658	0.998	0.127	0.159	0.159	0.732

Notes: ADF, HW, EG and BVD denote the augmented Dickey-Fuller, Horvath-Watson, Enders-Granger and Berben-van Dijk tests, respectively. The critical values for the HW test are taken from Table 1 from Horvath and Watson (1995); the values for EG test are taken from Table 1 from Enders and Granger (1998); the values for the BVD test are taken from Table 1 from Berben and van Dijk (1999). The lag lengths for all of the tests are fixed at the true value of 1. Number of simulations = 1000.

Table 2: Tests of Linearity
Empirical Rejection Frequency of 5% Tests

		EQ-TAR						BAND-TAR					
c	T	Tsay-U	Tsay-M	F_{12}	LR_{12}	F_{13}	LR_{13}	Tsay-U	Tsay-M	F_{12}	LR_{12}	F_{13}	LR_{13}
3	100	0.807	0.457	0.827	0.665	0.808	0.600	0.263	0.181	0.305	0.226	0.25	0.187
3	250	0.999	0.887	0.998	0.985	0.996	0.985	0.597	0.463	0.690	0.550	0.665	0.525
3	500	1	1	1	1	1	1	0.918	0.81		0.840		0.860
5	100	0.714	0.457	0.76	0.665	0.718	0.585	0.194	0.159	0.235	0.175	0.175	0.195
5	250	0.953	0.887	0.981	0.965	0.976	0.975	0.548	0.385	0.515	0.450	0.455	0.400
5	500	1	1	0.999	1	0.999	1	0.842	0.697	0.870	0.765	0.845	0.735
10	100	0.442	0.325					0.14	0.133				
10	250	0.668	0.541					0.308	0.236				
10	500	0.900	0.825					0.522	0.442				

Notes: Tsay-U and Tsay-M denote Tsay's univariate and multivariate nonparametric test for threshold nonlinearity, respectively. F12 and F13 denote Hansen's sup-F test of a TAR(1) against a TAR(2) and TAR(3), respectively. LR12 and LR13 denote Hansen's sup-LR test of a TVECM(1) against a TVECM(2) and TVECM(3), respectively. The critical values for the Hansen type tests are computed using the homoskedastic bootstrap procedure of Hansen (1999) based on 200 bootstrap values. The lag lengths for the tests are fixed at the true value of 1. Number of simulations = 1000.

Table 3: Unrestricted TAR(3) Estimates
Monte Carlo Means and Standard Deviations

True Model: Continuous, symmetric threshold and symmetric adjustment BAND-TVECM

c	T	$c^{(1)}$	$c^{(2)}$	$\mu^{(1)}$	$\rho^{(1)}$	$\alpha^{(2)}$	$\rho^{(2)}$	$\mu^{(3)}$	$\rho^{(3)}$
3	100	-1.13	1.20	-1.88	0.51	0.30	0.33	2.88	0.45
		(1.77)	(1.71)	(13.58)	(0.50)	(4.87)	(2.19)	(10.95)	(0.54)
3	250	-1.53	1.53	-2.32	0.47	0.07	0.57	1.10	0.47
		(1.52)	(1.49)	(8.19)	(0.34)	(2.10)	(0.85)	(29.64)	(0.35)
3	500	-1.94	2.00	-2.44	0.47	0.01	0.84	2.46	0.46
		(1.13)	(1.10)	(1.61)	(0.24)	(0.69)	(0.37)	(2.89)	(0.24)
5	100	-1.44	1.44	-2.95	0.54	-0.46	0.48	2.50	0.57
		(3.02)	(2.98)	(8.70)	(0.52)	(7.56)	(2.01)	(15.47)	(0.52)
5	250	-2.11	2.10	-3.76	0.58	0.06	0.68	4.48	0.55
		(2.62)	(2.65)	(37.98)	(0.33)	(3.03)	(0.76)	(21.83)	(0.27
5	500	-2.77	2.80	-3.29	0.56	-0.05	0.86	2.91	0.55
		(2.22)	(2.30)	(7.84)	(0.27)	(1.30)	(0.33)	(15.63)	(0.27)
10	100	-1.39	2.01	-0.67	0.66	0.27	0.58	9.10	0.62
		(5.97)	(5.97)	(80.87)	(0.44)	(10.93)	(1.49)	(268.5)	(0.49)
10	250	-2.48	2.51	-6.02	0.72	0.06	0.76	4.81	0.70
		(6.00)	(5.89)	(64.06)	(0.31)	(5.56)	(0.72)	(13.68)	(0.32)
10	500	-3.38	3.70	-9.60	0.73	0.00	0.85	5.68	0.71
		(5.87)	(5.78)	(115.8)	(0.27)	(3.07)	(0.37)	(34.74)	(0.28)

Notes: Standard deviations are in parentheses. The estimated unrestricted model is (6). The true value of ρ is 0.4. The regime specific means are computed as $\mu^{(j)} = \alpha^{(j)}/(1 - \rho^{(j)})$ for j=1,3. Number of simulations = 1000.

Table 4: Unrestricted TVECM(3) Estimates
Monte Carlo Means and Standard Deviations

True Model: Continuous, symmetric threshold and symmetric adjustment BAND-TVECM

c	T	$c^{(1)}$	$c^{(2)}$	$\mu^{(1)}$	$\gamma_1^{(1)}$	$\gamma_2^{(1)}$	$\gamma_1^{(2)}$	$\gamma_2^{(2)}$	$\mu^{(3)}$	$\gamma_1^{(3)}$	$\gamma_2^{(3)}$
3	100	-1.07	1.07	-2.24	-1.54	-0.53	-2.23	-0.88	5.54	-1.60	-0.55
		(1.83)	(1.80)	(38.2)	(1.87)	(0.73)	(7.98)	(3.14)	(99.2)	(2.10)	(0.83)
3	250	-1.37	1.45	14.06	-1.51	-0.52	-1.66	-0.64	3.92	-1.58	-0.55
		(1.60)	(1.56)	(523)	(1.17)	(0.44)	(3.22)	(1.25)	(18.9)	(1.27)	(0.48)
3	500	-1.78	1.81	-3.78	-1.60	-0.55	-0.90	-0.35	3.34	-1.63	-0.56
		(1.32)	(1.27)	(34.9)	(0.90)	(0.33)	(1.71)	(0.66)	(27.9)	(0.90)	(0.33)
5	100	-1.34	1.42	-4.48	-1.29	-0.45	-1.36	-0.52	5.09	1.47	-0.52
		(3.01)	(3.00)	(159)	(1.75)	(0.67)	(6.84)	(2.69)	(84.5)	(2.01)	(0.79)
5	250	-1.91	1.92	-4.03	-1.27	-0.44	-1.31	-0.50	23.82	-1.27	-0.44
		(2.67)	(2.70)	(18.8)	(1.17)	(0.43)	(2.71)	(1.04)	(628)	(1.13)	(0.42)
5	500	-2.38	2.50	-4.80	-1.25	-0.44	-0.70	-0.27	5.52	-1.28	-0.44
		(2.38)	(2.36)	(14.4)	(0.90)	(0.32)	(1.44)	(0.56)	(12.4)	(0.91)	(0.33)
10	100	-1.27	1.84	6.55	-1.02	-0.36	-1.82	-0.70	7.59	-1.12	-0.40
		(6.22)	(6.12)	(346)	(1.71)	(0.66)	(6.41)	(2.49)	(187)	(1.84)	(0.72)
10	250	-2.17	2.58	-0.03	-0.85	-0.30	-0.99	-0.37	4.21	-0.93	-0.33
		(5.99)	(5.93)	(485)	(1.04)	(0.38)	(2.74)	(1.07)	(185)	(1.12)	(0.42)
10	500	-3.52	3.20	-7.35	-0.84	-0.29	-0.43	-0.17	6.19	-0.78	-0.27
		(5.67)	(5.67)	(112)	(0.86)	(0.31)	(1.37)	(0.53)	(54.5)	(0.83)	(0.30)

Notes: Standard deviations are in parentheses. The estimated unrestricted model is (10). The regime specific mean is computed as $\mu^{(j)} = \beta'\alpha^{(j)}/\beta'\gamma^{(j)}$ for j=1,3. Number of simulations = 1000.

Table 5: Restricted TAR(3) and TVECM(3) Estimates
Monte Carlo Means, Standard Deviations and Specification Tests

True Model: Continuous, symmetric threshold and symmetric adjustment Band-TVECM

		TAR				TVECM				
c	T	c	ρ	LR	Wald	c	γ_1	γ_2	LR	Wald
3	100	2.82 (0.87)	0.25 (0.43)	0.38	0.65	2.80 (0.90)	-2.23 (1.44)	-0.74 (0.52)	0.33	0.58
3	250	2.97 (0.50)	0.32 (0.27)	0.40	0.56	2.89 (0.61)	-1.96 (0.85)	-0.65 (0.30)	0.20	0.52
3	500	2.98 (0.34)	0.37 (0.17)	0.34	0.42	2.97 (0.36)	-1.87 (0.55)	-0.62 (0.19)	0.15	0.45
5	100	4.36 (1.41)	0.31 (0.42)	0.45	0.75	4.21 (1.51)	-2.00 (1.51)	-0.66 (0.54)	0.34	0.68
5	250	4.73 (0.97)	0.35 (0.28)	0.38	0.66	4.71 (0.97)	-1.93 (0.95)	-0.64 (0.34)	0.22	0.60
5	500	4.89 (0.52)	0.38 (0.19)	0.31	0.56	4.90 (0.55)	-1.91 (0.64)	-0.63 (0.23)	0.15	0.53
10	100	7.54 (3.04)	0.47 (0.45)	0.56	0.84	7.34 (3.15)	-1.61 (1.42)	-0.54 (0.52)	0.38	0.71
10	250	8.36 (2.57)	0.54 (0.51)	0.53	0.82	8.26 (2.67)	-1.38 (0.95)	-0.46 (0.34)	0.27	0.64
10	500	9.05 (1.91)	0.51 (0.23)	0.46	0.76	8.96 (1.99)	-1.41 (0.71)	-0.47 (0.25)	0.22	0.71

Note: For c and ρ, the Monte Carlo standard deviations are in parentheses. The estimated restricted TAR(3) and the estimated restricted TVECM are (14) and (11) respectively with the additional restrictions $c=c^{(2)}=-c^{(1)}$, $\mu^{(3)}=c$, $\mu^{(1)}=-c$ and $g_i = g_i^{(3)} = g_i^{(1)}$. The empirical rejection frequencies of 5% tests based on a asymptotic χ^2 distribution are reported for likelihood ratio and Wald tests. The true value of ρ is 0.4, γ_1 -1.8 and γ_2 -0.6. Number of simulations=1000.

Table 6: Number of Rejections for No Cointegration at 10% Level

Categories	ADF	HW	EG	BVD
Alcoholic Beverages	0	2	1	2
Apparel Services	7	7	8	10
Appliances, incl. electronic equipment	5	10	12	12
Automobile Maintenance & Repairs	2	3	4	1
Cereals and Bakery Products	3	12	15	26
Eggs	7	24	26	28
Entertainment Commodities	5	10	11	16
Entertainment Services	9	16	27	24
Fish & Seafood	15	28	28	28
Food Away from Home	4	6	6	6
Footwear	15	26	28	24
Fresh Fruits and Vegetables	20	27	28	28
Fuels	20	24	24	25
Furniture & Bedding	8	20	24	26
Homeowners' Costs	10	5	5	27
Hospital & Related Services	10	16	12	15
Household Maintenance & Repairs	0	0	0	1
Housekeeping Services	1	4	3	10
Housekeeping Supplies	10	12	16	12
Infants' and toddlers' Apparel	3	2	3	4
Meats	7	26	26	28
Medical Care Commodities	3	11	6	13
Men's and Boy's Apparel	25	21	23	28
Motor Fuel	22	23	21	27
New Vehicles	14	21	25	18
Other Apparel Commodities	0	0	1	5
Other Food at Home	8	23	27	28
Other Furnishings	9	18	21	13
Other Renter's Costs	5	8	8	19
Other Utilities & Public Services	7	21	22	21
Personal & Educational Expenses	1	2	3	7
Personal Care	3	7	9	11
Poultry	6	26	27	28
Processed Fruits and Vegetables	0	3	3	26
Professional Medical Services	4	11	8	7
Public Transportation	2	3	3	9
Residential Rent	13	4	4	25
Textile Housefurnishings	6	21	23	28
Tobacco & Smoking Products	3	7	4	8
Used Cars	1	6	6	26
Women's and Girl's Apparel	14	22	26	23

Note: AIC is used to search for the most appropriate number of lags for the ADF and the BVD tests. Given that the sample size is small, the maximum possible lags (at the level) is set at six. We adopt the recommendation of Enders and Granger (1998) and use a nonlinear version of AIC for the EG test, which utilizes the variance from the TAR model. For Horvath-Watson test, we restrict it to a VAR(0) model so as to maintain a reasonable degree of freedom given the small sample size in a bivariate framework.

Table 7: Summary of a Simple AR(1) Estimation on Price Differences

Categories	Mean	Median	Std Dev	Min	City	Max	City	Halflife
Used Cars	1.0044	1.0049	0.0131	0.9707	AN	1.0279	BO	NA
Household Maintenance & Repairs	0.9682	0.9736	0.0238	0.8902	TA	0.9965	NY	21.4608
Personal & Educational Expenses	0.9584	0.974	0.0557	0.7275	MS	1.0041	HO	16.3033
Other Apparel Commodities	0.9526	0.9579	0.021	0.8694	MS	0.9799	MI	14.2723
Automobile Maintenance & Repairs	0.9512	0.9553	0.0383	0.8518	AN	1.0578	MI	13.8671
Alcoholic Beverages	0.9463	0.9445	0.0169	0.9108	PO	0.9856	AN	12.5467
Processed Fruits and Vegetables	0.9382	0.9547	0.0514	0.7527	DA	0.9782	CL	10.8604
Residential Rent	0.9367	0.9511	0.0483	0.7652	MI	0.9892	KC	10.5942
Tobacco & Smoking Products	0.9339	0.9537	0.0693	0.6584	MA	0.9865	TA	10.1304
Infants' and Toddlers' Apparel	0.9329	0.9405	0.0396	0.7704	CI	0.9666	BO	9.9739
Public Transportation	0.9315	0.9423	0.0379	0.8013	AT	0.9653	SD	9.7699
Housekeeping Services	0.9303	0.9428	0.0586	0.6891	AT	0.9979	HO	9.5894
Apparel Services	0.9245	0.9321	0.0477	0.7705	CH	0.9852	MA	8.8276
Homeowners' Costs	0.9229	0.9429	0.06	0.7065	DA	0.9728	SD	8.6373
Food Away from Home	0.921	0.927	0.0534	0.7836	MI	0.9753	KC	8.4194
Medical Care Commodities	0.9133	0.9256	0.0483	0.7812	PH	0.968	DN	7.6444
Personal Care	0.9113	0.9288	0.0753	0.7108	MS	1.012	PI	7.4627
Professional Medical Services	0.9073	0.9114	0.0422	0.8209	HO	0.972	BO	7.1261
Other Renter's Costs	0.9015	0.932	0.0832	0.5778	CI	0.9701	AT	6.6843
Entertainment Commodities	0.8957	0.9118	0.053	0.7787	PO	0.9598	NO	6.2956
Hospital & Related Services	0.8836	0.8904	0.0595	0.7273	SF	0.9927	SD	5.5996
Appliances, incl. electronic equipment	0.8769	0.8989	0.0712	0.7109	NO	0.9711	SD	5.277
Housekeeping Supplies	0.8701	0.8802	0.0679	0.7313	MS	0.9747	SL	4.9801
Cereals and Bakery Products	0.8629	0.8916	0.0883	0.6571	SL	0.956	NY	4.6997
Other Furnishings	0.8565	0.8661	0.0553	0.7291	MA	0.9451	NO	4.4745
Textile Housefurnishings	0.8352	0.828	0.0632	0.7078	HS	0.9492	DA	3.85
Women's and Girl's Apparel	0.8300	0.8488	0.0501	0.6848	PI	0.8882	BA	3.7193
Men's and Boy's Apparel	0.8209	0.8296	0.0672	0.6527	MI	0.9083	NO	3.5132
Motor Fuel	0.8209	0.8458	0.0924	0.6124	CL	0.9695	TA	3.5117
New Vehicles	0.8109	0.8271	0.0934	0.5604	KC	0.9197	MA	3.3066
Furniture & Bedding	0.8008	0.8331	0.1166	0.526	HO	0.9603	BO	3.1201
Meats	0.7921	0.806	0.082	0.5741	TA	0.9413	CI	2.9738
Footwear	0.7814	0.8005	0.0694	0.6126	DC	0.872	DT	2.8098
Other Utilities & Public Services	0.7672	0.8089	0.1839	0.2377	SL	0.9827	BO	2.6157
Other Food at Home	0.7578	0.75	0.0995	0.5634	KC	0.9274	HS	2.499
Fresh Fruits and Vegetables	0.7406	0.7662	0.1063	0.4091	AT	0.8645	DC	2.3081
Eggs	0.7397	0.7477	0.1377	0.3008	DT	0.957	SF	2.2988
Entertainment Services	0.7249	0.7365	0.1262	0.4609	HS	0.9295	PO	2.1543
Fuels	0.6682	0.6748	0.1692	0.1763	KC	0.8922	MA	1.7193
Fish & Seafood	0.6682	0.6768	0.1084	0.4126	CI	0.8551	HS	1.719
Poultry	0.6261	0.6454	0.1751	0.2519	DA	0.9208	CI	1.4801
Average	0.8590	0.8722	0.0737	0.6558		0.9146		6.7274

Note: The mean, the median and the standard deviation among the 28 data series in each category are reported in columns 1-3. The minimum and the maximum value of the AR(1) coefficients (and the corresponding city pair) are reported in columns 4-7. Half-life is computed by the formula ln(0.5)/ln r where r is the estimated coefficient of the model. Unlike models elsewhere in this paper, the dependent variable of this AR regression is the level, not first difference.

Table 8: Asymmetric Adjustment of Prices between New York and New Orleans

Categories	g_{NY}	g_{NO}	$\lvert g_{NY} \rvert - \lvert g_{NO} \rvert$	HW stat
Meats	0.2951	-0.0165	0.2786	20.4732
Fresh Fruits and Vegetables	0.2006	0.0094	0.1912	12.1468
Fuels	0.2659	-0.0469	0.2190	22.1979
Appliance, incl. Electronic equipment	0.1015	-0.0605	0.0410	9.9600
Women's & Girl's Apparel	0.1624	0.0091	0.1533	10.0989
Medical Care Commodities	0.0898	0.0143	0.0755	8.4475
Poultry	0.2579	-0.0932	0.1647	23.9801
Fish & Seafood	0.4328	-0.0197	0.4131	35.6801
Eggs	0.1228	-0.0278	0.0950	12.4674
Alcoholic Beverages	0.0377	-0.0202	0.0175	10.4786
Homeowners' Costs	0.0230	-0.0243	-0.0013	11.4070
Other Utilities & Public Services	0.1137	-0.0396	0.0741	13.4433
Textile Housefurnishings	0.1252	-0.0270	0.0982	9.2462
Footwear	0.2810	-0.0391	0.2419	23.2207
Apparel Services	0.0333	-0.0524	-0.0191	20.8390
New Vehicles	0.1970	0.0172	0.1798	13.9002
Motor Fuels	0.1870	-0.0294	0.1576	15.6482
Professional Medical Services	0.0373	-0.0325	0.0048	10.4430
Hospital & Related Services	0.0380	-0.0618	-0.0238	9.1319
Entertainment Services	0.3013	-0.0300	0.2713	26.5697

Table 9: Number of Rejections of Linearity at 10% Level

Categories	TU	TM	HU	HM
Alcoholic Beverages	2	10	2	2
Apparel Services	2	15	1	4
Appliances, incl. electronic equipment	6	6	4	4
Automobile Maintenance & Repairs	3	5	1	3
Cereals and Bakery Products	18	6	10	3
Eggs	20	18	14	8
Entertainment Commodities	2	8	2	2
Entertainment Services	16	22	4	11
Fish & Seafood	5	13	3	1
Food Away from Home	2	3	3	3
Footwear	6	13	2	4
Fresh Fruits and Vegetables	2	4	5	5
Fuels	12	5	5	5
Furniture & Bedding	17	18	5	6
Homeowners' Costs	21	14	12	12
Hospital & Related Services	4	5	3	3
Household Maintenance & Repairs	0	5	4	4
Housekeeping Services	8	13	2	8
Housekeeping Supplies	4	3	4	4
Infants' and toddlers' Apparel	3	16	2	1
Meats	15	14	17	12
Medical Care Commodities	5	4	7	7
Men's and Boy's Apparel	5	4	4	6
Motor Fuel	4	11	2	13
New Vehicles	5	11	8	8
Other Apparel Commodities	4	12	3	5
Other Food at Home	16	20	13	9
Other Furnishings	7	13	8	8
Other Renter's Costs	11	25	13	15
Other Utilities & Public Services	15	15	5	16
Personal & Educational Expenses	11	12	6	6
Personal Care	10	13	6	4
Poultry	12	18	7	9
Processed Fruits and Vegetables	12	15	15	7
Professional Medical Services	4	5	1	4
Public Transportation	9	7	7	12
Residential Rent	17	19	7	8
Textile Housefurnishings	13	18	5	17
Tobacco & Smoking Products	5	4	12	9
Used Cars	12	22	14	21
Women's and Girl's Apparel	7	7	7	5

Note: Both univariate and multivariate Hansen's tests are restricted with one lag (at the level).

Table 10: Symmetric Band TAR(3) and TVECM(3) Median Estimates within each Selected Group of Categories

Categories	AR	TAR	AR	TAR	TVECM	
	r		Half-life		γ_1	γ_2
Cereals and Bakery Products	-0.0006	-0.3086	1181.229	1.8781	-0.2228	0.1588
Eggs	-0.0563	-0.6191	11.9521	0.7181	-0.3542	0.2437
Entertainment Services	0.0001	-0.8083	NA	0.4197	-0.0625	0.7534
Fish & Seafood	-0.022	-0.4509	31.1405	1.1563	-0.0477	0.3594
Footwear	-0.1621	-0.3042	3.9183	1.911	-0.0921	0.2093
Fresh Fruits and Vegetables	-0.0112	-0.3829	61.5467	1.4361	-0.1163	0.2222
Fuels	-0.0809	-0.4241	8.2133	1.2562	-0.206	0.2071
Furniture & Bedding	-0.0143	-0.6722	48.1034	0.6214	-0.1133	0.5742
Meats	-0.0061	-0.7974	113.111	0.4341	-0.1174	0.6483
Men’s and Boy’s Apparel	-0.1023	-0.2368	6.4239	2.5656	-0.0806	0.1149
Motor Fuel	-0.0047	-0.3044	145.6761	1.9099	0.0297	0.2891
Other Food at Home	-0.0148	-0.5943	46.4553	0.7683	-0.1228	0.43
Other Utilities & Public Services	-0.0031	-0.4431	224.4629	1.184	-0.1559	0.1876
Poultry	-0.018	-0.5762	38.2002	0.8073	-0.3068	0.1881
Processed Fruits and Vegetables	0.0013	-0.3722	NA	1.4892	-0.2157	0.115
Textile Housefurnishings	-0.0438	-0.5242	15.4859	0.9332	-0.0441	0.5171
Women’s and Girl’s Apparel	-0.1495	-0.241	4.2808	2.5141	-0.0078	0.2235

Note: The TAR model and the TVECM are specified in the same way as in Table 5. Number of lags is set at 1 at the level. All coefficients and half-lifes reported are the median among the data series used in each category. It is not clear if the computation of half-lifes for linear model is applicable for nonlinear model. However, we report them here for a simple comparison.