Journal of Economics and Econometrics Vol. 56, No.2, 2013 pp. 54-77

ISSN 2032-9652 E-ISSN 2032-9660

Threshold Cointegration: Model Selection with an Application

PETER SEPHTON, JANELLE MANN†

ABSTRACT

In this article we examine the performance of an extended approach to testing for threshold cointegration that relies on the threshold specification process suggested by Gonzalo and Pitarakis (2002) and the block-bootstrap threshold unit root test of Seo (2008). A topical application demonstrates its merits.

JEL Classification: C23, F35, O23, O55.
Keywords: Threshold cointegration; block bootstrapping; model selection criteria.

†Corresponding author: Peter Sephton, Queen's University, School of Business, Kingston, ON K7L 3N6, Canada. E-mail: psephton@business.queensu.ca. Janelle Mann, Assistant Professor, University of Manitoba, Department of Economics, Winnipeg, MB R3T 2N2, Canada. E-mail: janelle.mann@ad.umanitoba.ca.

1 INTRODUCTION

There is a large literature on nonlinear and asymmetric adjustment in long-run economic relationships (see Li and Lee (2010) for recent examples). One particularly fruitful area of research involves the concept of threshold cointegration, in which departures from long-run equilibrium can persist as long as certain conditions are satisfied. These types of models can help explain, for example, conflicting evidence on the law of one price in spatial arbitrage relationships (Myers and Jayne (2011), Sephton (2011a) and in fundamental macroeconomic relationships such as purchasing power parity (Nam (2011)), exchange rate pass-through (Larue et al (2010)) and the twin-deficits (Holmes (2011)); the apparent lack of a link between wholesale and retail prices (Peri and Baldi (2010), Sephton (2011b)); and divergent views on whether short-term and long-term interest rates are related through the term structure (Clements and Galvao (2003), Krishnakurmar and Neto (2012)).

The traditional approach (Enders and Siklos (2001)) to testing for threshold cointegration requires a multi-step process that searches for the number of thresholds, their location, and some method for testing the null hypothesis of non-cointegration through the use of Engle and Granger (1987) type tests to determine whether the residuals from a cointegrating regression contain a unit root. Since the thresholds are not identified under the null, the solution to the Davies (1987) problem usually involves a sup-Wald test of the restrictions in the testing equation. Alternatively, one could proceed along the lines of Li and Lee (2010) and use a direct test of non-cointegration in the error correction equations, which themselves, depend on the estimated threshold and cointegrating regression parameters, following the so-called BDM approach (named after Banerjee, Dolado and Mestre (1998)).

The purpose of this paper is to suggest an alternative guide to testing for threshold cointegration that endogenizes the search for the optimal number and value of the thresholds, the timing of the indicator variable, whilst bootstrapping the test statistics to ensure robust inference. This blends the work of Gonzalo and Pitarakis (2002), who argued that threshold selection could be viewed from a model-selection perspective, and Seo (2008), who developed a residual based block-bootstrap test for unit roots in threshold models. We examine the performance of this approach and show that it performs well in practice. An empirical application demonstrates its merits.

In the next section we review the residual-based approach to testing for threshold cointegration and both the main findings of Gonzalo-Pitarakis (2002) and Seo (2008). This is followed by evidence that indicates the procedure works very well on simulated data. The fourth section highlights the practical performance of the procedure. Final remarks follow.

2 THRESHOLD COINTEGRATION

Balke and Fomby (1997) introduced threshold cointegration as an attractor to which several series were drawn in the long-run, but when certain conditions were satisfied, there could be persistent departures from the equilibrium without any tendency for the system to return to a state of balance. These conditions were framed in terms of an equilibrium error that followed a threshold autoregression that was mean-reverting outside a certain range, and non-stationary within a certain range. This kind of model might be capable of explaining why arbitrage between markets only becomes profitable after price differences exceed the costs of transacting; why central banks might intervene in markets when prices exceed some threshold; or perhaps why the Federal Reserve recently attempted to “twist” the term structure of interest rates in attempt to stimulate aggregate demand, whilst maintaining some semblance of credibility vis a vis maintaining a relatively low and stable rate of inflation that would be consistent with robust economic growth.

The Engle and Granger (1987) approach to testing for linear cointegration involves estimating a cointegrating regression on series that are integrated of the same order (usually I(1)), and examining the residuals to see if they are integrated of a lower order (usually, I(0)). If there is no stationary linear combination of the series then they are not cointegrated – at least not linearly cointegrated. (There is a large literature on non-linear cointegration, but that is outside the scope of the present analysis). The Engle-Granger test for non-cointegration examines the residuals from the cointegrating regression in equation (1) to see if $\hat{\varepsilon}_t$ has a unit root. Under the null of a unit root, the series are not cointegrated.

$$Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t \tag{1}$$

$$\Delta\varepsilon_t = \delta_1\varepsilon_{t-1} + \nu_t \tag{2}$$

Enders and Siklos (2001) extended this approach to testing for a unit root, allowing the residuals to follow a threshold autoregression based on an indicator variable I_t that takes on the value of 1 if certain conditions are met, and zero otherwise.

$$\Delta\varepsilon_t = \delta_1 I_t \varepsilon_{t-1} + \delta_2 (1 - I_t)\varepsilon_{t-1} + \nu_t \tag{3}$$

Those "certain conditions" were framed in terms of the TAR or MTAR models in which the lagged residuals, or their first difference, was above or below a threshold, denoted by τ_1 according to (4)

TAR MTAR

$$\begin{array}{cccc} I_t = 1 & I_t = 0 & I_t = 1 & I_t = 0 \\ \hat{\varepsilon}_{t-1} \le \tau_1 & \hat{\varepsilon}_{t-1} > \tau_1 & \Delta\hat{\varepsilon}_{t-1} \le \tau_1 & \Delta\hat{\varepsilon}_{t-1} > \tau_1 \end{array} \tag{4}$$

A simple threshold value of zero might be considered, or one could be estimated by searching over a range of possible values using the methods of Chan (1984), in which the residuals were sorted and the top and bottom 15% are excluded from the search. Then, solving the Davies (1987) problem (in which the threshold is not identified under the null), a series of Wald tests could be undertaken, with the "optimal threshold" chosen as the sup-F test of the null of non-cointegration, $\delta_1 = \delta_2 = 0$.

Other approaches to testing for cointegration rely on direct estimation of the error correction model implied by the cointegrated system. Banerjee et al (1998) provide one example, and Li and Lee (2010) employ this approach to develop innovative tests for threshold cointegration.

Our purpose here is to take a different approach to testing for threshold cointegration, one that will allow us to extend previous work in several directions. Close inspection of (4) indicates that the selection of the threshold indicator is based on the first lagged value of the residual in the cointegrating regression. Indeed, any stationary series could act as the variable driving the indicator, and it need not be limited to the first lagged value of the cointegrating regression residual. One might envisage thresholds based on the difference between two series rather than the lagged cointegrating regression residual (for example, Nam (2011) uses the real exchange rate, the difference between the logarithm of the nominal exchange rate and the logarithm of the relative price level; Mann (2012) uses the difference between the tax-adjusted wholesale and retail prices of gasoline) or a situation in which retail prices are revised, but with a lag, when wholesale prices exceed some threshold (Sephton (2011b)); or when local prices in a small market respond, with delay, to changes in the world price, or the price in a dominant market (Sephton (2011a)).

One might also allow for multiple thresholds – there is no reason, *a priori*, to consider a single threshold. Band-TAR adjustment models, for example, allow two thresholds so that behaviour above the upper threshold, or below the lower threshold, should guarantee a return to the long-run equilibrium. Between the two thresholds there may be no pressure on prices to adjust back to the attractor because arbitrage is not profitable within this range. Price behaviour, for example, might appear totally random within this middle range, only returning to a stationary path once taken above the upper bound or below the lower bound. In some cases[1] the Band-TAR process is assumed to have two thresholds with opposite signs but the same absolute value, but there is no reason, *a priori*, to force this to be the case.

Within the framework here, consider equations 5 and 6. We allow for up to two thresholds τ_1, τ_2 and indicator variables $I\{z_{t-d}\}$ denoted by z_{t-d}, the dth lag of stationary variable z_t

$$Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t \tag{5}$$

$$\Delta\varepsilon_t = \delta_1 I\{z_{t-d} < \tau_1\}\varepsilon_{t-1} + \delta_2 I\{\tau_1 \le z_{t-d} < \tau_2\}\varepsilon_{t-1} + \delta_3 I\{z_{t-d} \ge \tau_2\}\varepsilon_{t-1} + \nu_t \tag{6}$$

Seo (2008) provided a test of the unit root hypothesis $\delta_1 = \delta_2 = \delta_3$ against the alternative of a stationary threshold process and demonstrated how residual based block bootstrapping can be used for inference.[2] This test is very powerful relative to the usual unit root test applied in the Engle-Granger (1987) approach, so it offers a natural solution to testing the null of non-cointegration in (6). In Band-TAR models where the middle regime exhibits non-stationary behaviour, the stationarity of the model itself, depends on the coefficients on the outer

---

[1] See Seo (2008) for an example.

[2] Seo (2008) examined a unit root test in the univariate context and suggested that his test can be applied to the residuals from a cointegration regression. "Threshold cointegration entails SETAR to the error-correction term, and thus the unit root test developed in this paper can be applied" (Seo (2008, p 1700).

regimes. Here we do not assume a symmetric Band-TAR specification, so that in the case of a stationary threshold process, the speeds of adjustment back to the attractor can differ depending on the nature of the disequilibrium.

As with similar tests, there is an augmented version of the test that incorporates lagged changes in the residuals in an attempt to "whiten" the covariance matrix, leading to a testing equation of the form (7)

$$\begin{aligned}\Delta\varepsilon_t &= \delta_1 I\{z_{t-d} < \tau_1\}\varepsilon_{t-1} + \delta_2 I\{\tau_1 \le z_{t-d} < \tau_2\}\varepsilon_{t-1} \\ &\quad + \delta_3 I\{z_{t-d} \ge \tau_2\}\varepsilon_{t-1} + \rho_1\Delta\varepsilon_{t-1} + \ldots + \rho_p\Delta\varepsilon_{t-p} + \nu_t \end{aligned} \tag{7}$$

where model selection criteria can be used to choose the appropriate lag length (the p term) in the augmented version of the test. The Wald test statistic (8) is the sup-Wald test statistic for fixed thresholds, $\tau \in \Gamma$. Seo (2008) examines a single threshold model that searches over the full range of the indicator series $\left[-\gamma, \gamma\right]$ where $\gamma = \max\left|z_{t-d}\right|$, but that approach will not accommodate multiple thresholds with a minimum number of observations within each regime. We restrict each threshold search to exclude the top and bottom 15% of observations within each range, and the number of observations within each regime to 10. Under the null hypothesis of a unit root Seo (2008) demonstrates the asymptotic distribution is well defined and depends on nuisance parameters, requiring bootstrapping for inference. Seo (2008) demonstrates that the residual based block bootstrap achieves consistency of the bootstrap, and his Monte Carlo simulations show this test to be much more powerful and better sized than the usual Augmented Dickey-Fuller test, even in relatively small samples.

$$W_n = \sup_{\tau\in\Gamma} n\left\{\frac{\hat{\sigma}^2}{\hat{\sigma}^2(\tau)} - 1\right\} \tag{8}$$

Our approach involves allowing between zero and two thresholds, and applying the Seo (2008) procedure to test for non-cointegration. We then employ the Gonzalo and Pitarakis (2002) approach to choose among the various threshold models. Gonzalo and Pitarakis (2002) framed the problem in terms of a model selection choice, where they maximized a criterion function $Q_T(m)$ involving m thresholds. They examined the log of the estimated residual variance in the no threshold case (denoted $\hat{\sigma}^2$) relative to the threshold case (denoted $\hat{\sigma}^2(\tau_1,\ldots,\tau_m)$), and subtracted a term involving the sample size (T), the number of parameters (K) and a penalty λ_t, where the usual BIC criterion sets $\lambda_T = \log T$. Other criteria considered include the AIC ($\lambda_T = 2$) and the BIC2 ($\lambda_T = 2\log T$).[3]

$$Q_T(m) = \max_{\tau_1,\dots,\tau_m} \log\left[\frac{\hat{\sigma}^2}{\hat{\sigma}^2(\tau_1,\dots,\tau_m)}\right] - \frac{\lambda_T}{T} Km \qquad (9)$$

Our approach is to use the sup-Wald test proposed by Seo (2008) to estimate models containing up to three thresholds and to test for threshold cointegration, and the model selection procedure (9) due to Gonzalo and Pitarakis (2002) to choose the appropriate threshold specification. It is a relatively trivial matter to endogenize the search for the appropriate lag of the indicator variable to allow for the possibility that adjustment is based on a lagged cointegrating regression residual (or some other stationary series) beyond the first lag, as in Gospodinov (2005). For a given lag of the indicator variable, we perform the Seo (2008) test, given the number of thresholds. This yields a matrix of Seo (2008) test values with a typical element given by the indicator lag and the number of thresholds. Optimization of (9) across the number of thresholds and the indicator lag yields the final specification.

The next section provides strong support for this approach and is followed by several empirical applications that demonstrate the benefits to adopting this methodology.

3 MONTE CARLO EVIDENCE

Gonzalo and Pitarakis (2002) report their model selection approach provides relatively high correct decision frequencies (over 90 percent with 600 observations), with little tendency to under-segment, even in a three regime model.[4] Here it is interesting to examine the correct

---

[3] Gonzalo and Pitarakis (2002) proved the consistency of their approach within the context of stationary threshold models. We are not aware of a formal proof extending their results to the case of non-stationary threshold models under the null, but our simulation experiments should highlight any deficiencies in employing their approach.

[4] While Altissimo and Corradi (2002) suggest the penalty term associated with the AIC is too weak, we include it for comparison with the results reported by Gonzalo and Pitarakis (2002).

decision frequencies (the number of times the procedure chooses zero thresholds when there are zero in the data generating process; the number of times it chooses one threshold when there is one threshold in the data generating process, and so on) for a range of settings of the other parameters of the search process. Given a particular sample size, these other settings involve the size of the threshold; the maximum lag in the search of the indicator variable; the minimum number of observations within each regime; the parameters of the data generating process; whether or not we employ the augmented version of the test to “whiten” the covariance matrix; and how non-spherical errors (in this case, those following a GARCH process) affect the correct decision frequencies.

In the context of threshold cointegration testing using the momentum threshold adjustment model, Cook (2007) demonstrated that the general-to-unity detrending approach of Elliot et al (1996) increases the power of unit root tests on the cointegrating regression residuals. Simulations here will also examine whether GLS detrending affects the correct decision frequencies and the power of the Seo (2008) test.

3.1 Correct Decision Frequencies

There are two interesting measures of the proportion of the time that the procedure correctly chooses the actual number of thresholds. The first is simply a measure of the correct decision frequency when there are zero, one, or two thresholds. These do not examine the correct frequency for our proposed methodology – they are simply a measure of the frequency that a given number of thresholds are found when there are that many in the data generating process (DGP). Table 1 provides this information. In addition to the AIC, BIC and BIC2 penalty terms used to select the number of thresholds by the procedure, we also include results from an approach that chooses the number of thresholds on the basis of maximizing the Seo Wald statistic (labeled MAXF). The correct decision frequencies are reported for two sample sizes, T=100 and T=250, using both the level of the series (denoted ES for Enders-Siklos) and the GLS-detrended series (denoted GLS), allowing both a constant and a constant and a trend in the cointegrating regression used to generate the residuals. The data generating process involved creating cumulative sums of normally distributed random variables (denoted r) or cumulative sums of a normally distributed random variable with strong GARCH effects (a unit intercept, an ARCH coefficient equal to 0.3 and a GARCH coefficient equal to 0.6, denoted g).

When there are no thresholds in the data generating process the BIC2 has the highest correct decision frequency relative to all other metrics. This is true whether the sample size is small (T=100) or large (T=250). GLS detrending has little impact on the correct decision frequency in this case.

In the case of a single threshold (set at zero) in the data generating process, the AIC and MAXF criteria appear to be equally capable of capturing the true number of thresholds, with GLS detrending generally leading to a higher correct decision frequency than when including deterministic components in the cointegrating regression. When two thresholds appear in the DGP (set at zero and five), the MAXF approach outperforms all others, again with what appears to be a minor gain to employing GLS detrending.

Table 1 appears to suggest that the BIC2 criterion is best when there are no thresholds, whereas the AIC or MAXF are best when there is one threshold, and the MAXF is best when there are two thresholds. Unfortunately, in practice, one rarely knows *a priori*, the number of thresholds. To guide empirical practice, we require a more general analysis of the model selection procedure. To that end, Tables 2a and 2b and Tables 3a and 3b provide correct decision frequencies associated with allowing the procedure to search for as many as three thresholds, when in fact the DGP had either one or two thresholds, respectively.

Table 2a reports the small sample results. When there is one threshold, the BIC2, with its strong penalty term, underfits the truth whereas the MAXF overfits, as does the AIC. The BIC appears to outperform the other metrics in this case, even when there are strong GARCH effects in the DGP. These results do not appear to extend to the larger sample, with Table 2b suggesting that the BIC2 metric has the highest correct decision frequency when there is a single threshold. The BIC still performs relatively well, but appears to have a tendency to overfit, as do the AIC and MAXF.

The practical guidance from Table 2 is that when there is a single threshold, on a small sample, one might want to employ the BIC to select the number of thresholds, whereas on a larger sample, the stronger penalty of the BIC2 might be preferred.

When there are two thresholds in the DGP, Table 3a suggests that the BIC and BIC2 underfit on small samples, whereas the MAXF overfits. The AIC appears best able to select the true number of thresholds. Table 3b indicates that these findings extend to a larger sample. The AIC appears to be the best decision criterion in this case.

Tables 2 and 3 suggest that one might want to employ both the AIC and the BIC as threshold selection criteria when one expects one or two thresholds, since the correct decision frequencies of the procedure are relatively high. This is true independently of whether the cointegrating regression includes deterministic components or is based on GLS-detrended data, as well as whether or not there are strong GARCH effects.

3.2 Test Size and Power

Table 4 presents information on the size of the test for sample sizes of 100 and 250, for both the AIC and BIC criteria when there are both a constant and a linear trend in the cointegrating regression. The BIC leads to sizes that are much closer to the nominal five percent level, even in small samples. In the case of a linear trend in the cointegrating regression, the AIC leads to a relatively oversized test, even in large samples.

Tables 5 and 6 examine the power of the Seo (2008) test when there is a single threshold in the DGP, using the AIC and BIC for model selection. Power is examined for a variety of settings of the adjustment coefficients when the series are threshold cointegrated. Generally speaking, the stronger the adjustment in both series, the greater is the power of the test. When one series is weakly exogenous, the power of the test can be quite low (as low as 44%) in small samples, but in large samples, the power is substantially higher (rising to 93% under the same conditions). The stronger the adjustment in the endogenous series, even when the other series is weakly exogenous, the greater is the power of the test, with a seemingly minor dominance of the BIC criterion.

In toto, these results suggest the use of either the AIC or the BIC as the model selection criterion and that GLS detrending does not appear to significantly affect correct decision frequency nor test power (Cook (2007) reported these power improvements for the MTAR specification but did not report results for the TAR specification). In practice, we recommend the use of both the AIC and BIC in order to compare test results across potentially different specifications. Ideally, inference will be unaffected by the choice of model selection criterion. In that case, one might be relatively confident concluding the series are either threshold cointegrated or not.

4 APPLICATION

There have been many recent applications of threshold cointegration tests, but few rely on the Gonzalo and Pitarakis (2002) model selection approach. In agricultural economics, Myers and Jayne (2012) examined maize markets in South Africa and Zambia and demonstrated that government intervention reduced the transmission of prices across borders. Adachi and Liu (2011) examined thresholds in the relationship between fluid milk demand and advertising; Boetel et al (2007) examined thresholds in US hog production; and Sephton (2011) examined thresholds between black and white pepper prices in Sarawak. In recent macroeconomic applications, Holmes (2011) investigated the relationship between government budget deficits and trade deficits; Nam (2011) explored the merits to employing threshold cointegration when revisiting the PPP debate. Other recent applications include Esteve and Tamarit (2012) who examined non-linear adjustment between CO2 and income in Spain, and Sephton (2011b, 2012), who reported little evidence of threshold cointegration between the prices of American processed cheddar cheese at the wholesale and retail levels, nor between measures of customer satisfaction and discretionary consumption expenditures in the United States, respectively.

In Sephton and Mann (2013) we employed this procedure to provide support for non-linear cointegration and asymmetric adjustment betweeen per capita GDP and CO2 emissions in Spain. Here, to highlight the merits of employing our methodology, we focus on an empirical application related to Holmes (2011), who examined the twin deficits debate.

4.1 The Twin Deficits Debate

Bartolini and Lahiri (2006) suggest the twin deficits hypothesis appeared to explain the American experience of the 1980s and the first part of the 21st Century, but that it did not seem to capture the behaviour observed in the 1990s. This suggests that there may be an asymmetric link between the current account balance and the fiscal balance. The twin deficits hypothesis suggests that a larger fiscal deficit, either through its affect on national saving (requiring greater borrowing from abroad, higher interest rates, currency appreciation, and capital inflows; the usual Mundell-Flemming effect) or domestic absorption (higher imports resulting from currency appreciation) will exacerbate the current account deficit. The causality is from the fiscal deficit to the current account deficit, that is, the current account deficit is a byproduct of expansionary fiscal policies.

Holmes (2011) examined this relationship using the methods of Hansen and Seo (2002) and reported evidence in favour of a single threshold, but with causality in either direction, depending on the size of the disequilibrium error. The traditional causal link from the fiscal deficit to the current account deficit becomes operative only when the fiscal deficit becomes "too large", corresponding roughly to the 1980-1984 and 1990-1994 periods.

Does our approach provide different results? Using data from the Bureau of Economic Analysis spanning 1947Q1 until 2011Q2 on the federal budget deficit and net foreign investment (both as a proportion of GDP), Table 7 presents the test results and the estimated error correction models. Both the AIC and BIC criteria choose identical specifications. We estimate the cointegrating regression including a constant and trend, and exclude deterministic components in the error correction equations.

There are two thresholds identified in the relationship, as described in Figure 1, which also highlights dates of US recessions. In the lowest regime there are 70 observations, above that 131, above that 47 months (the minimum number of observations was set at 15). The test of non-cointegration has a value of 24.69, and a bootstrap 5% critical value of 18.29 (the associated p-value is very small). This leads us to reject the null of non-cointegration in favour of the twin deficits being threshold cointegrated.

Inspection of the coefficients in the testing equation provides support for a band-TAR specification. In the lowest regime there is tendency to return to equilibrium but it is relatively weak (at -0.066) when compared to the estimated speed of adjustment in the upper region (at -0.117). In the middle regime there is no pressure for the series to restore balance. From the estimated error correction equations, it is clear that the fiscal deficit is the series that is moving to restore balance while the current account deficit appears to be weakly exogenous because it does not react to the disequilibrium in the relationship. This result is at odds with that reported by Holmes (2011), and provides support in favour of the traditonal mechanism describing the twin deficits hypothesis. Fiscal policies react to move the system back to the attractor.

Figure 1 also helps explain conflicting reports of whether the series were cointegrated over the last sixty years. The duration of periods when the disequilibrium was above the highest threshold was relatively short (51 quarters in total), consistent with the relatively fast adjustment back to equilibrium, given the estimated error correction equations. When adjustment was relatively slow, the disequilibria appear to be longer-lived (72 quarters in total). Periods during which there appeared to be no tendency to return to equilibrium span from 1961 until 1975, in the late 1970s, from 1983-1985, between 1989-1992, 1993-2003 (consistent with Bartolini and Lahiri (2006)), and 2009 until 2012. Most likely as a result of unprecedented fiscal expansion related to the latest period of creative destruction, since 2009, the link between the fiscal and current account deficits has diverged from the long-run equilibrium, without tendancy to return to the long-run attractor.

These findings indicate there are benefits to adopting the joint methodology of Gonzalo and Pitarakis (2002) and Seo (2008), as they provide a much richer understanding of the twin deficits debate.

5 CONCLUSIONS

This article provides a novel approach to testing for threshold cointegration by selecting the number of thresholds following the model selection approach advocated by Gonzalo and Pitarakis (2002) and employing the threshold unit root test of Seo (2008). Monte Carlo evidence suggests the approach fares very well vis a vis selecting the number of thresholds and testing the null of non-cointegration, even when the data contains strong GARCH effects. While previous work (Cook (2007)) suggested GLS detrending has the potential to increase test power, it did not appear to provide significant gains in power within the framework of the usual TAR specification. An empirical application demonstrated the merits of our approach and led to somewhat different findings than those reported in previous work.

Extensions of the procedures to the momentum threshold model (MTAR) where cointegration thresholds are determined by the change in the deviation from the long-run attracter, and a wider variety of data generating processes under both the null and the alternative hypotheses, is the subject of ongoing work.

REFERENCES

Adachi, K. and Liu, D. (2011) Estimating Threshold Effects of U.S. Generic Fluid Milk Advertising. American Journal of Agricultural Economics 92, 727-739.

Altissimo, F., and V Corradi. (2002) Bounds for Inference With Nuisance Parameters Present Only Under the Alternative. Econometrics Journal (5), 494-519.

Balke, N. and Fomby, T. (1997) Threshold Cointegration. International Economic Review 38, 627-643.

Banerjee, A., Dolado, J. and Mestre, R. (1998) Error Correction Mechanism Tests for Cointegration in a Single-Equation Framework. Journal of Time Series Analysis 19, 267-283.

Bartolini, L. and Lahiri, A. (2006) Twin Deficits: Twenty Years Later. Current Issues in Economics and Finance 12, 1-7.

Boetel, B., Hoffmann, R. and Liu, D. (2007) Estimating Investment Rigidity with a Threshold Regression Framework: The Case of U.S. Hog Production Sector. American Journal of Agricultural Economics 89, 36-51.

Chan, K. (1993) Consistency and Limiting Distributions of the Least Squares Estimator of a Threshold Autoregressive Model. The Annals of Statistics 21, 520-533.

Clements, M. and Galvao, A. (2003) Testing the Expectations Theory of the Term Structure of Interest Rates in Threshold Models. Macroeconomic Dynamics 7, 567-585.

Cook, S. (2007) A Threshold Cointegration Test with Increased Power. Mathematics and Computers in Simulation 73, 386-392.

Davies, J. (1987) Hypothesis Testing when a Nuisance Parameter is Present Only Under the Alternative. Biometrika 74, 33-43.

Elliott, G., Rothenberg, R. and Stock, J. (1996) Efficient Tests for an Autoregressive Unit Root. Econometrica 64, 813-836.

Enders, W. and Siklos, P. (2001) Cointegration and Threshold Adjustment. Journal of Business & Economic Statistics 19, 156-276.

Engle, R. and Granger, C. (1987) Cointegration and Error Correction: Representation, Estimation and Testing. Econometrica 55, 251-276.

Esteve, V. and Tamarit, C. (2012) Threshold cointegration and nonlinear adjustment between CO2 and income: The environmental Kuznets curve in Spain, 1857–2007. Energy Economics, forthcoming

Gonzalo, J. and Pitarakis, J. (2002) Estimation and Model Selection Based Inference in Single and Multiple Threshold Models. Journal of Econometrics 110, 319-352.

Gospodinov, N. (2005) Testing for Threshold Nonlinearity in Short-Term Interest Rates. Journal of Financial Econometrics 3, 344-371.

Hansen, B. and Seo, B. (2002) Testing for Two-Regime Threshold Cointegration in Vector Error-Correction Models. Journal of Econometrics 110, 293-318.

Holmes, M. (2011) Threshold Cointegration and the Short-Run Dynamics of Twin Deficit Behaviour. Research in Economics 65, 271-277.

Krishnakumar, J. and Neto, D. (2012) Testing Uncovered Interest Rate Parity and Term Structure Using a Three-Regime Threshold Unit Root VECM: An Application to the Swiss 'Isle' Of Interest Rates'. Oxford Bulletin of Economics and Statistics 72, 180-201.

Larue, B., Gervais, J. and,Rancourt, Y. (2010) Exchange rate pass-through, Menu Costs and Threshold Cointegration. Empirical Economics 38, 171-192.

Li., J. and Lee, J. (2010) ADL Tests for Threshold Cointegration. Journal of Time Series Analysis 31, 241-254.

Mann, J. (2012) Threshold Cointegration with Applications to the Oil and Gasoline Industry. Unpublished Ph.D Dissertation, Queen's University

Myers, R. and Jayne, T. (2011) Multiple-Regime Spatial Price Transmission with an Application to Maize Markets in Southern Africa. American Journal of Agricultural Economics 94, 174-188.

Nam, D. (2011) The Roles of Nominal Exchange Rate and Relative Price Adjustments in PPP Reversion. Journal of Money, Credit and Banking 43, 775-785.

Peri, M. and Baldi, L. (2010) Vegetable Oil Market and Biofuel Policy: An Asymmetric Cointegration Approach. Energy Economics 32, 687-693.

Seo, M. (2006) Bootstrap Testing for the Null of No Cointegration in a Threshold Vector Error Correction Model. Journal of Econometrics 134, 129-150.

Seo, M. (2008) Unit Root Test in a Threshold Autoregression: Asymptotic Theory and Residual-Based Block Bootstrap. Econometric Theory 24, 1699-1716.

Sephton, P. (2011a) Threshold Cointegration in Sarawak Pepper Prices. Canadian Journal of Agricultural Economics 59, 405-416.

Sephton, P. (2011b) Who Cut the Cheese? Applied Economics Letters 18, 7-10.

Sephton, P. (2012) Consumer Spending and Customer Satisfaction: Untying the Knot. Economics Research International, 1-7.

Sephton, P., and J. Mann (2013) Further Evidence of the Environmental Kuznets Curve in Spain. Energy Economics 36, 177-181.

Table 1: Correct Frequency Percentages

		AIC			BIC			BIC2			MAX F		
T=100													
		0	1	2	0	1	2	0	1	2	0	1	2
ES Constant	r	0.11	0.96	0.87	0.48	0.82	0.36	0.89	0.50	0.02	0.06	0.97	1.00
	g	0.06	0.97	0.84	0.36	0.85	0.37	0.78	0.52	0.04	0.07	0.98	1.00
ES Constant Trend	r	0.12	0.92	0.86	0.48	0.69	0.38	0.87	0.30	0.02	0.09	0.96	1.00
	g	0.08	0.93	0.81	0.42	0.76	0.29	0.82	0.37	0.02	0.09	0.95	0.98
GLS Constant	r	0.05	1.00	0.87	0.41	0.91	0.45	0.82	0.74	0.04	0.01	0.99	1.00
	g	0.03	0.99	0.81	0.26	0.94	0.41	0.70	0.76	0.02	0.01	0.99	1.00
GLS Constant Trend	r	0.12	0.94	0.86	0.48	0.79	0.33	0.87	0.36	0.03	0.03	0.96	1.00
	g	0.08	0.95	0.81	0.43	0.83	0.31	0.82	0.45	0.02	0.03	0.98	0.99
T=250													
		0	1	2	0	1	2	0	1	2	0	1	2
ES Constant	r	0.16	1.00	0.97	0.59	0.98	0.33	0.95	0.91	0.01	0.04	1.00	1.00
	g	0.07	1.00	0.87	0.41	0.97	0.29	0.78	0.82	0.00	0.03	1.00	1.00
ES Constant Trend	r	0.17	1.00	0.98	0.65	0.98	0.32	0.94	0.79	0.01	0.06	1.00	1.00
	g	0.07	1.00	0.89	0.43	0.96	0.30	0.76	0.68	0.01	0.06	0.97	1.00
GLS Constant	r	0.07	1.00	0.98	0.50	1.00	0.46	0.89	0.96	0.03	0.00	1.00	1.00
	g	0.06	1.00	0.85	0.35	0.99	0.27	0.70	0.95	0.02	0.00	1.00	1.00
GLS Constant Trend	r	0.15	1.00	0.95	0.62	0.96	0.34	0.94	0.84	0.01	0.01	1.00	1.00
	g	0.07	1.00	0.89	0.39	0.97	0.28	0.76	0.79	0.03	0.01	1.00	1.00

Notes: Table entries are the proportion of times the model selection criterion chooses the correct number of thresholds for the zero, one, and two threshold cases. Es denotes the usual Enders Siklos test whereas GLS denotes the tests applied to GLS detrended data. The DGP were based on normally distributed random numbers (r) or those containing GARCH (with ARCH coefficient 0.3 and GARCH coefficient 0.6). Simulations were based on ngrid=50, nboot=200, and 200 replications. Speed of adjustment coefficients in the single threshold case (set at 0) were Y [-.05 -.8] and X [.05 .6] and in the two threshold case where thresholds were set at 0 and +5 were Y [-.05 -.25 -.8] and X [.1 .3 .8].

Table 2a: Threshold Selection Frequencies, Single Threshold

	Zero		One		Two		Three	
T=100	Random	Garch	Random	Garch	Random	Garch	Random	Garch
AIC								
ES Constant	0.010	0.010	0.185	0.265	0.715	0.670	0.090	0.055
ES Constant Trend	0.020	0.040	0.160	0.225	0.800	0.695	0.020	0.040
GLS Constant	0.000	0.005	0.260	0.240	0.725	0.720	0.015	0.035
GLS Constant Trend	0.015	0.020	0.190	0.250	0.775	0.710	0.020	0.020
BIC								
ES Constant	0.115	0.135	0.540	0.610	0.340	0.240	0.005	0.015
ES Constant Trend	0.230	0.210	0.420	0.525	0.350	0.265	0.000	0.000
GLS Constant	0.045	0.065	0.635	0.640	0.315	0.295	0.005	0.000
GLS Constant Trend	0.175	0.190	0.440	0.525	0.385	0.285	0.000	0.000
BIC2								
ES Constant	0.490	0.480	0.480	0.490	0.030	0.030	0.000	0.000
ES Constant Trend	0.670	0.620	0.290	0.355	0.040	0.025	0.000	0.000
GLS Constant	0.255	0.240	0.690	0.720	0.055	0.040	0.000	0.000
GLS Constant Trend	0.590	0.535	0.360	0.415	0.050	0.050	0.000	0.000
MAX F								
ES Constant	0.000	0.000	0.000	0.000	0.000	0.010	1.000	0.985
ES Constant Trend	0.000	0.000	0.000	0.000	0.010	0.005	0.990	0.995
GLS Constant	0.000	0.000	0.000	0.000	0.010	0.020	0.990	0.980
GLS Constant Trend	0.000	0.000	0.000	0.000	0.015	0.005	0.985	0.995

Notes: Table entries are correct decision frequencies for the procedure based on selecting the number of thresholds using each metric. The single threshold value is set at zero and the adjustment coefficients are Y [-.05 -.8] and X [.05 .6]. The double threshold values are at 0 and +5 and the adjustment coefficients are Y [-.05 -.25 -.8] and X [.1 .3 .8].

Table2b: Threshold Selection Frequencies, Single Threshold

	Zero		One		Two		Three	
T=250	Random	Garch	Random	Garch	Random	Garch	Random	Garch
AIC								
ES Constant	0.000	0.000	0.130	0.180	0.655	0.675	0.215	0.145
ES Constant Trend	0.000	0.000	0.155	0.165	0.680	0.720	0.165	0.115
GLS Constant	0.000	0.000	0.145	0.210	0.800	0.720	0.055	0.070
GLS Constant Trend	0.000	0.000	0.125	0.145	0.775	0.760	0.100	0.095
BIC								
ES Constant	0.015	0.040	0.610	0.705	0.345	0.240	0.030	0.015
ES Constant Trend	0.030	0.040	0.710	0.690	0.245	0.270	0.015	0.000
GLS Constant	0.005	0.010	0.740	0.755	0.250	0.235	0.005	0.000
GLS Constant Trend	0.035	0.030	0.705	0.745	0.250	0.225	0.010	0.000
BIC2								
ES Constant	0.100	0.185	0.855	0.785	0.045	0.045	0.000	0.000
ES Constant Trend	0.200	0.345	0.780	0.655	0.020	0.020	0.000	0.000
GLS Constant	0.045	0.060	0.935	0.915	0.020	0.025	0.000	0.000
GLS Constant Trend	0.160	0.210	0.815	0.775	0.025	0.025	0.000	0.000
MAX F								
ES Constant	0.000	0.000	0.000	0.000	0.000	0.005	1.000	0.995
ES Constant Trend	0.000	0.000	0.000	0.000	0.010	0.010	0.990	0.990
GLS Constant	0.000	0.000	0.000	0.000	0.010	0.015	0.990	0.985
GLS Constant Trend	0.000	0.000	0.000	0.000	0.005	0.005	0.995	0.995

Notes: Table entries are correct decision frequencies for the procedure based on selecting the number of thresholds using each metric. The single threshold value is set at zero and the adjustment coefficients are Y [-.05 -.8] and X [.05 .6]. The double threshold values are at 0 and +5 and the adjustment coefficients are Y [-.05 -.25 -.8] and X [.1 .3 .8].

Table 3a: Threshold Selection Frequencies, Two Thresholds

	Zero		One		Two		Three	
T=100	Random	Garch	Random	Garch	Random	Garch	Random	Garch
AIC								
ES Constant	0.040	0.045	0.080	0.125	0.875	0.820	0.005	0.010
ES Constant Trend	0.070	0.030	0.090	0.125	0.835	0.835	0.005	0.010
GLS Constant	0.030	0.020	0.085	0.170	0.880	0.800	0.005	0.010
GLS Constant Trend	0.035	0.015	0.105	0.175	0.855	0.805	0.005	0.005
BIC								
ES Constant	0.405	0.255	0.240	0.370	0.355	0.375	0.000	0.000
ES Constant Trend	0.365	0.265	0.280	0.420	0.355	0.315	0.000	0.000
GLS Constant	0.310	0.195	0.215	0.400	0.475	0.405	0.000	0.000
GLS Constant Trend	0.365	0.280	0.320	0.410	0.315	0.310	0.000	0.000
BIC2								
ES Constant	0.870	0.750	0.105	0.220	0.025	0.030	0.000	0.000
ES Constant Trend	0.850	0.765	0.125	0.215	0.025	0.020	0.000	0.000
GLS Constant	0.830	0.640	0.135	0.300	0.035	0.040	0.000	0.000
GLS Constant Trend	0.845	0.695	0.105	0.290	0.050	0.015	0.000	0.000
MAX F								
ES Constant	0.000	0.000	0.000	0.000	0.000	0.010	1.000	0.985
ES Constant Trend	0.000	0.000	0.000	0.000	0.010	0.005	0.990	0.995
GLS Constant	0.000	0.000	0.000	0.000	0.010	0.015	0.990	0.980
GLS Constant Trend	0.000	0.000	0.000	0.000	0.020	0.005	0.985	0.995

Notes: See Table 2.

Table 3b: Threshold Selection Frequencies, Two Thresholds

	Zero		One		Two		Three	
T=250	Random	Garch	Random	Garch	Random	Garch	Random	Garch
AIC								
ES Constant	0.010	0.005	0.050	0.125	0.930	0.835	0.010	0.035
ES Constant Trend	0.005	0.010	0.030	0.100	0.955	0.890	0.010	0.000
GLS Constant	0.015	0.000	0.025	0.160	0.925	0.805	0.035	0.035
GLS Constant Trend	0.000	0.005	0.035	0.110	0.945	0.865	0.020	0.020
BIC								
ES Constant	0.385	0.145	0.310	0.555	0.305	0.300	0.000	0.000
ES Constant Trend	0.400	0.230	0.260	0.505	0.340	0.265	0.000	0.000
GLS Constant	0.285	0.095	0.285	0.635	0.430	0.270	0.000	0.000
GLS Constant Trend	0.350	0.155	0.305	0.555	0.345	0.290	0.000	0.000
BIC2								
ES Constant	0.855	0.575	0.135	0.420	0.010	0.005	0.000	0.000
ES Constant Trend	0.910	0.705	0.090	0.290	0.000	0.005	0.000	0.000
GLS Constant	0.800	0.500	0.175	0.490	0.025	0.010	0.000	0.000
GLS Constant Trend	0.855	0.560	0.130	0.425	0.015	0.015	0.000	0.000
MAX F								
ES Constant	0.000	0.000	0.000	0.000	0.000	0.005	1.000	0.995
ES Constant Trend	0.000	0.000	0.000	0.000	0.010	0.010	0.990	0.990
GLS Constant	0.000	0.000	0.000	0.000	0.010	0.015	0.990	0.985
GLS Constant Trend	0.000	0.000	0.000	0.000	0.005	0.005	0.995	0.995

Notes: See Table 2.

Table 4: Test Size T = 100 and T = 250

	AIC	BIC
T =100		
Constant	0.045	0.055
Trend	0.185	0.065
T =250		
Constant	0.070	0.065
Trend	0.170	0.085

Size at 5 percent level of significance for cumulative sums of normally distributed random variables under the null of non cointegration for sample sizes T = 100 and T = 250 using a block length of b = 6.

Table 5: Power at 5%, Single Threshold, Test with Constant

						AIC			BIC		
			ES		GLS		ES		GLS		
Y Adjustment		X Adjustment		r	g	r	g	r	g	r	g
T=100											
-0.050	-0.030	0.000	0.000	0.440	0.490	0.450	0.470	0.480	0.500	0.450	0.510
-0.050	-0.030	0.000	0.500	0.670	0.610	0.510	0.510	0.710	0.660	0.610	0.560
-0.050	-0.030	0.000	0.900	0.710	0.680	0.540	0.560	0.800	0.790	0.690	0.670
-0.050	-0.030	0.200	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.050	-0.030	0.200	0.500	0.990	0.950	0.990	0.960	0.990	0.980	0.990	0.990
-0.050	-0.030	0.200	0.900	0.980	0.940	0.950	0.950	0.990	0.940	0.970	0.960
-0.050	-0.800	0.000	0.000	0.590	0.700	0.550	0.660	0.730	0.590	0.670	0.620
-0.050	-0.800	0.000	0.500	0.720	0.720	0.520	0.670	0.790	0.760	0.640	0.700
-0.050	-0.800	0.000	0.900	0.740	0.690	0.530	0.520	0.820	0.760	0.710	0.640
-0.050	-0.800	0.200	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.050	-0.800	0.200	0.500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995
-0.050	-0.800	0.200	0.900	1.000	1.000	0.990	1.000	1.000	1.000	1.000	1.000
T=250											
-0.050	-0.030	0.000	0.000	0.930	0.900	0.840	0.850	0.860	0.940	0.880	0.900
-0.050	-0.030	0.000	0.500	0.990	0.940	0.930	0.920	0.990	0.980	0.910	0.940
-0.050	-0.030	0.000	0.900	0.990	1.000	0.920	0.950	1.000	0.980	0.980	0.960
-0.050	-0.030	0.200	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.050	-0.030	0.200	0.500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.050	-0.030	0.200	0.900	1.000	1.000	1.000	0.995	1.000	1.000	1.000	1.000
-0.050	-0.800	0.000	0.000	0.980	0.950	0.890	0.900	0.990	0.950	0.960	0.910
-0.050	-0.800	0.000	0.500	1.000	0.970	0.920	0.870	0.990	0.990	0.940	0.920
-0.050	-0.800	0.000	0.900	0.990	1.000	0.930	0.940	1.000	0.990	0.940	0.960
-0.050	-0.800	0.200	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.050	-0.800	0.200	0.500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.050	-0.800	0.200	0.900	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Notes: Entries are power at five percent for the test allowing for a constant in the cointegrating regression (Table 5) or constant and trend (Table 6) assuming the threshold is set at zero with speed of adjustment coefficients in Y and X as labeled. Simulations based on ngrid=50, nboot=200 and 200 replications. ES denotes the usual Enders Siklos test whereas GLS denotes the tests applied to GLS detrended data. The DGP were based on integrated sums of normally distributed random numbers (r) or those containing GARCH (with arch coefficient 0.3 and garch coefficient 0.6).

Table 6: Power at 5%, Single Threshold, Test with Constant and Trend

				AIC				BIC			
				ES		GLS		ES		GLS	
Y Adjustment		X Adjustment		r	g	r	g	r	g	r	g
T=100											
-0.050	-0.030	0.000	0.000	0.460	0.490	0.440	0.450	0.440	0.560	0.380	0.530
-0.050	-0.030	0.000	0.500	0.520	0.540	0.420	0.480	0.550	0.540	0.550	0.460
-0.050	-0.030	0.000	0.900	0.570	0.550	0.420	0.420	0.640	0.650	0.510	0.600
-0.050	-0.030	0.200	0.000	0.990	0.940	0.990	0.970	1.000	0.960	0.990	0.970
-0.050	-0.030	0.200	0.500	0.990	0.970	0.960	0.950	0.990	0.960	0.970	0.940
-0.050	-0.030	0.200	0.900	0.980	0.950	0.930	0.920	0.980	0.950	0.940	0.940
-0.050	-0.800	0.000	0.000	0.630	0.690	0.470	0.610	0.750	0.670	1.000	0.600
-0.050	-0.800	0.000	0.500	0.660	0.640	0.520	0.560	0.710	0.630	0.580	0.540
-0.050	-0.800	0.000	0.900	0.640	0.620	0.460	0.490	0.730	0.650	0.600	0.540
-0.050	-0.800	0.200	0.000	1.000	1.000	0.990	0.990	1.000	1.000	1.000	1.000
-0.050	-0.800	0.200	0.500	0.990	0.980	1.000	0.990	1.000	0.980	1.000	0.970
-0.050	-0.800	0.200	0.900	0.990	0.960	0.960	0.940	0.990	0.940	0.950	0.960
T=250											
-0.050	-0.030	0.000	0.000	0.890	0.900	0.860	0.870	0.870	0.930	0.840	0.890
-0.050	-0.030	0.000	0.500	0.920	0.890	0.800	0.820	0.920	0.890	0.830	0.780
-0.050	-0.030	0.000	0.900	0.910	0.920	0.780	0.830	0.950	0.880	0.780	0.790
-0.050	-0.030	0.200	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.050	-0.030	0.200	0.500	1.000	0.990	1.000	0.990	1.000	1.000	0.995	1.000
-0.050	-0.030	0.200	0.900	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.990
-0.050	-0.800	0.000	0.000	0.970	0.930	0.930	0.890	0.990	0.950	0.980	0.890
-0.050	-0.800	0.000	0.500	0.960	0.880	0.870	0.810	0.990	0.920	0.950	0.830
-0.050	-0.800	0.000	0.900	0.960	0.940	0.850	0.850	0.970	0.920	0.930	0.850
-0.050	-0.800	0.200	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.050	-0.800	0.200	0.500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.050	-0.800	0.200	0.900	1.000	1.000	1.000	0.990	1.000	1.000	1.000	1.000

Notes: Entries are power at five percent for the test allowing for a constant in the cointegrating regression (Table 5) or constant and trend (Table 6) assuming the threshold is set at zero with speed of adjustment coefficients in Y and X as labeled. Simulations based on ngrid=50, nboot=200 and 200 replications. ES denotes the usual Enders Siklos test whereas GLS denotes the tests applied to GLS detrended data. The DGP were based on integrated sums of normally distributed random numbers (r) or those containing GARCH (with arch coefficient 0.3 and garch coefficient 0.6).

Table 7: Twin Deficits Results

	Test Equation	Fiscal Deficit	Current Account Deficit
Threshold Indicator Lag	2		
Number of Thresholds	2		
Test Statistic	24.69		
P-Value	0		
Adjustment Coefficients			
Above Highest	-0.1169	-0.1332	0.0015
t statistic	-3.5206	-6.3974	0.0177
Middle	0.1039	0.0920	0.0674
Below Lowest	-0.0660	-0.0526	0.0195
t statistic	-2.4892	-1.6554	0.3299

Notes: Data on the U.S. federal fiscal and current account deficits, as a fraction of GDP was obtain from the Bureau of Economic Analysis database, spanning 1947Q1 to 2011Q2. A constant and trend were included in the cointegrating regression. The test statistic involved 300 bootstrap replications with a blocksize of 6 and a maximum lag length in the test equation and in the error correction models set at 4. The top and bottom 15% of the sample was excluded from the threshold search using a grid span of 100 with a minimum number of observations within a regime set at 15. Adjustment coefficients are the estimated coefficients in the testing equation or the error correction models, as noted, with t-statistics denoted by t. The maximum lag for the threshold indicator was set at 2.