Testing for Threshold Cointegration in Vector Error Correction Models[*]

Bruce E. Hansen

University of Wisconsin[†]

and

Byeongseon Seo

Soongsil University[‡]

July 2000

Revised: June 2001

[†]Department of Economics, Madison, WI 53706. bhansen@ssc.wisc.edu. www.ssc.wisc.edu/~bhansen

[‡]Department of Economics, Seoul, 156-743, Korea. seo@saint.soongsil.ac.kr.

Abstract

This paper proposes a formal test for threshold cointegration and an algorithm to estimate the model parameters. Our model is a vector error correction model (VECM) with a single cointegrating vector, and a threshold effect in the error-correction term.

We propose a relatively simple algorithm to obtain maximum likelihood estimation (MLE) of the complete multivariate threshold cointegration model for the bivariate case. However, we do not provide a proof of consistency, nor a distribution theory for the MLE.

We propose testing for a threshold in this model with a Lagrange Multiplier (LM) test. This is particularly convenient since the LM test can be computed by an ordinary least squares regression involving the conventional maximum likelihood estimate of the cointegrating vector. Since the threshold is not identified under the null hypothesis, our tests take the Sup-LM form. We derive the null asymptotic distribution of the test, and find that it takes the form given in Hansen (1996). We show how to calculate asymptotic critical values by simulation. We also present a bootstrap approximation to the sampling distribution.

We investigate the performance of the test using Monte Carlo simulation, and find that the test works quite well. The Type I error of the test is quite close to the nominal level, even in a small sample with conditional heteroskedasticity.

We apply our methods to the term structure model of interest rates. In several bi-variate cointegrating VECMs, we find strong evidence for a threshold error-correction model.

1 Introduction

Threshold cointegration was introduced by Balke and Fomby (1997) as a feasible means to combine non-linearity and cointegration. In particular, the model allows for nonlinear adjustment to long-run equilibrium. The model has generated significant applied interest, including the following applications: Balke and Wohar (1998), Baum, Barkoulas, and Caglayan (2001), Baum and Karasulu (2001), Enders and Falk (1998), Lo and Zivot (2001), Marterns, Kofman and Vorst (1998), Michael, Nobay, and Peel (1997), O'Connel (1998), O'Connel and Wei (1997), Obstfeld and Taylor (1997), and Taylor (2001). Lo and Zivot (2001) provide an extensive review of this growing literature.

One of the most important statistical issues for this class of models is testing for the presence of a threshold effect (the null of linearity). Balke and Fomby (1997) proposed using the application of the univariate tests of Hansen (1996) and Tsay (1989) to the error-correction term (the cointegrating residual). This is known to be valid when the cointegrating vector is known, but Balke-Fomby did not provide a theory for the case of estimated cointegrating vector. Lo and Zivot (2001) extended the Balke-Fomby approach to a multivariate threshold cointegration model with a known cointegrating vector, using the tests of Tsay (1998) and multivariate extensions of Hansen (1996).

In this paper, we extend this literature by examining the case of unknown cointegrating vector. As in Balke-Fomby, our model is a vector error correction model (VECM) with one cointegrating vector and a threshold effect based on the error-correction term. However, unlike Balke-Fomby who focus on univariate estimation and testing methods, our estimates and tests are for the complete multivariate threshold model. The fact that we use the error-correction term as the threshold variable is not essential to our analysis, and the methods we discuss here could easily be adapted to incorporate other models where the threshold variable is a stationary transformation of the predetermined variables.

This paper makes two contributions. First, we propose a method to implement maximum likelihood estimation (MLE) of the threshold model. This algorithm involves a joint grid search over the threshold and the cointegrating vector. The algorithm is simple to implement in the bivariate case, but would be difficult to implement in higher-dimensional cases.

Furthermore, at this point we do not provide a proof of consistency, nor a distribution theory for the MLE.

Second, we develop a test for the presence of a threshold effect. Under the null hypothesis, there is no threshold, so the model reduces to a conventional linear VECM. Thus estimation under the null hypothesis is particularly easy, reducing to conventional reduced rank regression. This suggests that a test can be based on the Lagrange Multiplier (LM) principle, which only require estimation under the null. Since the threshold parameter is not identified under the null hypothesis we base inference on a SupLM test. (See Davies (1987), Andrews (1993), and Andrews and Ploberger (1994) for motivation and justification for this testing strategy.) Our test takes a similar algebraic form to those derived by Seo (1998) for structural change in error correction models.

We derive the asymptotic null distribution of the Sup-LM test, and find that it is identical to the form found in Hansen (1996) for threshold tests applied to stationary data. In general the asymptotic distribution depends on the covariance structure of the data, precluding tabulation. We suggest using either the fixed regressor bootstrap of Hansen (1996, 2000b), or alternatively a parametric residual bootstrap algorithm, to approximate the sampling distribution.

Section 2 introduces the threshold models and derives the Gaussian quasi-MLE for the models. Section 3 presents our LM test for threshold cointegration, its asymptotic distribution, and two methods to calculate p-values. Section 4 presents simulation evidence concerning the size and power of the tests. Section 5 presents an application to the term structure of interest rates. Proofs of the asymptotic distribution theory are presented in the appendix. Gauss programs which compute the estimates and test, and replicate the empirical work reported in this paper, are available at www.ssc.wisc.edu/~bhansen.

2   Estimation

2.1   Linear Cointegration

Let x_t be a p-dimensional $I(1)$ time series which is cointegrated with one $p \times 1$ cointegrating vector β. Let $w_t(\beta) = \beta' x_t$ denote the $I(0)$ error-correction term. A linear VECM of order $l+1$ can be compactly written as

$$\Delta x_t = A' X_{t-1}(\beta) + u_t \tag{1}$$

where

$$X_{t-1}(\beta) = \begin{pmatrix} 1 \\ w_{t-1}(\beta) \\ \Delta x_{t-1} \\ \Delta x_{t-2} \\ \vdots \\ \Delta x_{t-l} \end{pmatrix}.$$

The regressor $X_{t-1}(\beta)$ is $k \times 1$ and A is $k \times p$ where $k = pl+2$. The error u_t is assumed to be a vector martingale difference sequence (MDS) with finite covariance matrix $\Sigma = E\left(u_t u_t'\right)$.

The notation $w_{t-1}(\beta)$ and $X_{t-1}(\beta)$ indicates that the variables are evaluated at generic values of β. When evaluated at the true value of the cointegrating vector, we will denote these variables as w_{t-1} and X_{t-1}, respectively.

We need to impose some normalization on β to achieve identification. Since there is just one cointegrating vector, a convenient choice is to set one element of β equal to unity, which has no cost when the system is bi-variate ($p = 2$) and for $p > 2$ only imposes the restriction that the corresponding element of x_t enters the cointegrating relationship.

The parameters (β, A, Σ) are estimated by maximum likelihood under the assumption that the errors u_t are iid Gaussian (using the above normalization on β). Let these estimates be denoted $(\tilde{\beta}, \tilde{A}, \tilde{\Sigma})$. Let $\tilde{u}_t = \Delta x_t - \tilde{A}' X_{t-1}(\tilde{\beta})$ be the residual vectors.

2.2 Threshold Cointegration

As an extension of model (1), a two-regime threshold cointegration model takes the form

$$\Delta x_t = \begin{cases} A_1' X_{t-1}(\beta) + u_t & \text{if} \quad w_{t-1}(\beta) \leq \gamma \\ \\ A_2' X_{t-1}(\beta) + u_t & \text{if} \quad w_{t-1}(\beta) > \gamma \end{cases}$$

where γ is the threshold parameter. This may alternatively be written as

$$\Delta x_t = A_1' X_{t-1}(\beta) d_{1t}(\beta, \gamma) + A_2' X_{t-1}(\beta) d_{2t}(\beta, \gamma) + u_t, \tag{2}$$

where

$$\begin{aligned} d_{1t}(\beta, \gamma) &= 1\left(w_{t-1}(\beta) \leq \gamma\right) \\ d_{2t}(\beta, \gamma) &= 1\left(w_{t-1}(\beta) > \gamma\right) \end{aligned},$$

and $1(\cdot)$ denotes the indicator function.

The threshold model (2) has two regimes, defined by the value of the error correction term. The coefficient matrices A_1 and A_2 govern the dynamics in these regimes. Model (2) allows all coefficients (except the cointegrating vector β) to switch between these two regimes. In many cases, it may make sense to impose greater parsimony on the model, by only allowing some coefficients to switch between regimes. This is a special case of (2) where constraints are placed on (A_1, A_2). For example, a model of particular interest only lets the coefficients on the constant and the error-correction w_{t-1} to switch, constraining the coefficients on the lagged Δx_{t-j} to be constant across regimes.

The threshold effect only has content if $0 < P(w_{t-1} \leq \gamma) < 1$, otherwise the model simplifies to linear cointegration. We impose this constraint by assuming that

$$\pi_0 \leq P(w_{t-1} \leq \gamma) \leq 1 - \pi_0, \tag{3}$$

where $\pi_0 > 0$ is a selected trimming parameter. For the empirical application of section 5, we set $\pi_0 = .05$.

We propose estimation of the model (2) by maximum likelihood, under the assumption that the errors u_t are iid Gaussian. The Gaussian likelihood is

$$\mathcal{L}_n(A_1, A_2, \Sigma, \beta, \gamma) = -\frac{n}{2} \log |\Sigma| - \frac{1}{2} \sum_{t=1}^{n} u_t(A_1, A_2, \beta, \gamma)' \Sigma^{-1} u_t(A_1, A_2, \beta, \gamma),$$

where

$$u_t(A_1, A_2, \beta, \gamma) = \Delta x_t - A_1' X_{t-1}(\beta) d_{1t}(\beta, \gamma) - A_2' X_{t-1}(\beta) d_{2t}(\beta, \gamma).$$

The MLE $(\hat{A}_1, \hat{A}_2, \hat{\Sigma}, \hat{\beta}, \hat{\gamma})$ are the values which maximize $\mathcal{L}_n(A_1, A_2, \Sigma, \beta, \gamma)$.

It is computationally convenient to first concentrate out (A_1, A_2, Σ). That is, hold (β, γ) fixed and compute the constrained MLE for (A_1, A_2, Σ). This is just OLS regression, specifically[1],

$$\hat{A}_1(\beta, \gamma) = \left(\sum_{t=1}^{n} X_{t-1}(\beta) X_{t-1}(\beta)' d_{1t}(\beta, \gamma) \right)^{-1} \left(\sum_{t=1}^{n} X_{t-1}(\beta) \Delta x_t' d_{1t}(\beta, \gamma) \right), \qquad (4)$$

$$\hat{A}_2(\beta, \gamma) = \left(\sum_{t=1}^{n} X_{t-1}(\beta) X_{t-1}(\beta)' d_{2t}(\beta, \gamma) \right)^{-1} \left(\sum_{t=1}^{n} X_{t-1}(\beta) \Delta x_t' d_{2t}(\beta, \gamma) \right), \qquad (5)$$

$$\hat{u}_t(\beta, \gamma) = u_t(\hat{A}_1(\beta, \gamma), \hat{A}_2(\beta, \gamma), \beta, \gamma)$$

and

$$\hat{\Sigma}(\beta, \gamma) = \frac{1}{n} \sum_{t=1}^{n} \hat{u}_t(\beta, \gamma) \hat{u}_t(\beta, \gamma)'. \qquad (6)$$

It may be helpful to note that (4) and (5) are the OLS regressions of Δx_t on $X_{t-1}(\beta)$ for the subsamples for which $w_{t-1}(\beta) \leq \gamma$, and $w_{t-1}(\beta) > \gamma$, respectively.

This yields the concentrated likelihood function

$$\begin{aligned} \mathcal{L}_n(\beta, \gamma) &= \mathcal{L}_n(\hat{A}_1(\beta, \gamma), \hat{A}_2(\beta, \gamma), \hat{\Sigma}(\beta, \gamma), \beta, \gamma) \\ &= -\frac{n}{2} \log \left| \hat{\Sigma}(\beta, \gamma) \right| - \frac{np}{2}. \end{aligned} \qquad (7)$$

The MLE $(\hat{\beta}, \hat{\gamma})$ are thus found as the minimizers of $\log \left| \hat{\Sigma}(\beta, \gamma) \right|$ subject to the normalization imposed on β as discussed in the previous section and the constraint

$$\pi_0 \leq n^{-1} \sum_{t=1}^{n} 1\left(x_t' \beta \leq \gamma \right) \leq 1 - \pi_0$$

(which imposes (3)). The MLE for A_1 and A_2 are $\hat{A}_1 = \hat{A}_1(\hat{\beta}, \hat{\gamma})$ and $\hat{A}_2 = \hat{A}_2(\hat{\beta}, \hat{\gamma})$.

This criterion function (7) is not smooth, so conventional gradient hill-climbing algorithms are not suitable for its maximization. In the leading case $p = 2$, we suggest using a grid search over the two-dimensional space (β, γ). In higher dimensional cases, grid search

---

[1]These formula are for the unconstrained model (2). If a constrained threshold model is used, then the appropriate constrained OLS estimates should be used.

becomes less attractive, and alternative search methods (such as a genetic algorithm, see Dorsey and Mayer (1995)) might be more appropriate. Note that in the event that β is known a priori, this grid search is greatly simplified.

To execute a grid search, one needs to pick a region over which to search. We suggest calibrating this region based on the consistent estimate $\tilde{\beta}$ obtained from the linear model (the MLE[2] discussed in section 2.1). Set $\tilde{w}_{t-1} = w_{t-1}(\tilde{\beta})$, let $[\gamma_L, \gamma_U]$ denote the empirical support of $\tilde{w}_{t-1}$, and construct an evenly spaced grid on $[\gamma_L, \gamma_U]$. Let $[\beta_L, \beta_U]$ denote a (large) confidence interval for β constructed from the linear estimate $\tilde{\beta}$ (based, for example, on the asymptotic normal approximation) and construct an evenly spaced grid on $[\beta_L, \beta_U]$. The grid search over (β, γ) then examines all pairs (γ, β) on the grids on $[\gamma_L, \gamma_U]$ and $[\beta_L, \beta_U]$, conditional on $\pi_0 \leq n^{-1} \sum_{t=1}^{n} 1\left(x_t' \beta \leq \gamma\right) \leq 1 - \pi_0$ (the latter to impose the constraint (3)).

In Figures 1 and 2, we illustrate the non-differentiability of the criterion function for an empirical example from section 5. (The application is to the 12-month and 120-month T-Bill rates, setting $l = 1$.) Figure 1 plots the criterion (7) as a function of γ with β concentrated out, and Figure 2 plots the criterion as a function of β with γ concentrated out.

In summary, our algorithm for the $p = 2$ case is:

1. Form a grid on $[\gamma_L, \gamma_U]$ and $[\beta_L, \beta_U]$ based on the linear estimate $\tilde{\beta}$ as described above.

2. For each value of (β, γ) on this grid, calculate $\hat{A}_1(\beta, \gamma)$, $\hat{A}_2(\beta, \gamma)$, and $\hat{\Sigma}(\beta, \gamma)$ as defined in (4), (5), and (6).

3. Find $(\hat{\beta}, \hat{\gamma})$ as the values of (β, γ) on this grid which yields the lowest value of $\log \left| \hat{\Sigma}(\beta, \gamma) \right|$.

4. Set $\hat{\Sigma} = \hat{\Sigma}(\hat{\beta}, \hat{\gamma})$, $\hat{A}_1 = \hat{A}_1(\hat{\beta}, \hat{\gamma})$, $\hat{A}_2 = \hat{A}_2(\hat{\beta}, \hat{\gamma})$, and $\hat{u}_t = \hat{u}_t(\hat{\beta}, \hat{\gamma})$.

It is useful to note that in step 3, there is no guarantee that the minimizers $(\hat{\beta}, \hat{\gamma})$ will be unique, as the log-likelihood function is not concave.

We have described an algorithm to implement the MLE, but it should be emphasized that this is not a theory of inference. We have not provided a proof of consistency of the estimator, nor a distribution theory. In linear models, $\hat{\beta}$ converges to β at rate n, and in

---

[2]Any consistent estimator could in principle be used here. In all our simulations and applications, we use the Johansen MLE.

stationary models, $\hat{\gamma}$ converges to γ at rate n. It therefore seems reasonable to guess that in the threshold cointegration model, $(\hat{\beta}, \hat{\gamma})$ will converge to (β, γ) at rate n. In this case, the slope estimates $\hat{A}_1$ and $\hat{A}_2$ should have conventional normal asymptotic distributions as if β and γ were known. Hence conventional standard errors can be reported for these parameter estimates.

3 Testing for a threshold

3.1 Test Statistics

Let $\mathcal{H}_0$ denote the class of linear VECM models (1) and $\mathcal{H}_1$ denote the class of two-regime threshold models (2). These models are nested, and the restriction $\mathcal{H}_0$ are the models in $\mathcal{H}_1$ which satisfy $A_1 = A_2$.

We want to test $\mathcal{H}_0$ (linear cointegration) versus $\mathcal{H}_1$ (threshold cointegration). We focus on formal model-based statistical tests, as these allow for direct model comparisons, and yield the greatest power for discrimination between models. Alternatively, one might consider the nonparametric nonlinearity tests of Tsay (1989) and Tsay (1998) for the univariate and multivariate case, respectively. As shown in the simulation studies by Balke and Fomby (1997) and Lo and Zivot (2001), these nonparametric tests generally have lower power than comparable model-based tests.

In this paper we consider Lagrange Multiplier (LM) statistics. We do this for two reasons. First, the LM statistic is computationally quick, enabling feasible implementation of the bootstrap. Second, a likelihood ratio or Wald-type test would require a distribution theory for the parameter estimates for the unrestricted model, which we do not yet have. We conjecture, but have no proof, that these tests are asymptotically equivalent to the LM test. We now derive the LM test statistic.

Assume for the moment that (β, γ) are known and fixed. The model under $\mathcal{H}_0$ is

$$\Delta x_t = A' X_{t-1}(\beta) + u_t, \tag{8}$$

and $\mathcal{H}_1$ is

$$\Delta x_t = A_1' X_{t-1}(\beta) d_{1t}(\beta, \gamma) + A_2' X_{t-1}(\beta) d_{2t}(\beta, \gamma) + u_t. \tag{9}$$

Given (β, γ), the models are linear so the MLE is least-squares. As (8) is nested in (9) and the models are linear, an LM-like statistic which is robust to heteroskedasticity can be calculated from a linear regression on the model (9). Specifically, let $X_1(\beta, \gamma)$ and $X_2(\beta, \gamma)$ be the matrices of the stacked rows $X_{t-1}(\beta)d_{1t}(\beta, \gamma)$ and $X_{t-1}(\beta)d_{2t}(\beta, \gamma)$, respectively, let $\xi_1(\beta, \gamma)$ and $\xi_2(\beta, \gamma)$ be the matrices of the stacked rows $\tilde{u}_t \otimes X_{t-1}(\beta)d_{1t}(\beta, \gamma)$ and $\tilde{u}_t \otimes X_{t-1}(\beta)d_{2t}(\beta, \gamma)$, respectively, with $\tilde{u}_t$ the residual vector from the linear model as defined in section 2.1, and define the outer product matrices

$$
\begin{aligned}
M_1(\beta, \gamma) &= I_p \otimes X_1(\beta, \gamma)'X_1(\beta, \gamma) \\
M_2(\beta, \gamma) &= I_p \otimes X_2(\beta, \gamma)'X_2(\beta, \gamma),
\end{aligned}
$$

and

$$
\begin{aligned}
\Omega_1(\beta, \gamma) &= \xi_1(\beta, \gamma)'\xi_1(\beta, \gamma) \\
\Omega_2(\beta, \gamma) &= \xi_2(\beta, \gamma)'\xi_2(\beta, \gamma).
\end{aligned}
$$

Then we can define $\hat{V}_1(\beta, \gamma)$ and $\hat{V}_2(\beta, \gamma)$, the Eicker-White covariance matrix estimators for $\mathrm{vec}\hat{A}_1(\beta, \gamma)$ and $\mathrm{vec}\hat{A}_2(\beta, \gamma)$ as

$$
\hat{V}_1(\beta, \gamma) = M_1(\beta, \gamma)^{-1}\Omega_1(\beta, \gamma)M_1(\beta, \gamma)^{-1} \tag{10}
$$

$$
\hat{V}_2(\beta, \gamma) = M_2(\beta, \gamma)^{-1}\Omega_2(\beta, \gamma)M_2(\beta, \gamma)^{-1}, \tag{11}
$$

yielding the standard expression for the heteroskedasticity-robust LM-like statistic

$$
LM(\beta, \gamma) = \mathrm{vec}\left(\hat{A}_1(\beta, \gamma) - \hat{A}_2(\beta, \gamma)\right)'\left(\hat{V}_1(\beta, \gamma) + \hat{V}_2(\beta, \gamma)\right)^{-1}\mathrm{vec}\left(\hat{A}_1(\beta, \gamma) - \hat{A}_2(\beta, \gamma)\right). \tag{12}
$$

If β and γ were known, (12) would be the test statistic. When they are unknown, the LM statistic is (12) evaluated at point estimates obtained under $\mathcal{H}_0$. The null estimate of β is $\tilde{\beta}$ (section 2.1), but there is no estimate of γ under $\mathcal{H}_0$, so there is no conventionally defined LM statistic. Arguing from the union-intersection principle, Davies (1987) proposed the statistic

$$
\mathrm{SupLM} = \sup_{\gamma_L \leq \gamma \leq \gamma_U} LM(\tilde{\beta}, \gamma). \tag{13}
$$

For this test, the search region $[\gamma_L, \gamma_U]$ is set so that γ_L is the π_0 percentile of $\tilde{w}_{t-1}$, and γ_U is the $(1 - \pi_0)$ percentile. This imposes the constraint (3). For testing, the parameter

π_0 should not be too close to zero, as Andrews (1993) shows that doing so reduces power. Andrews (1993) argues that setting π_0 between .05 and .15 are typically good choices.

Further justification for the statistic (13) is given in Andrews (1993) and Andrews and Ploberger (1994). Andrews and Ploberger (1994) argue that better power may be achieved by using exponentially-weighted averages of $LM(\tilde{\beta}, \gamma)$, rather than the supremum. There is an inherent arbitrariness in this choice of statistic, however, due to the choice of weighting function, so our analysis will remain confined to (13).

As the function $LM(\tilde{\beta}, \gamma)$ is non-differentiable in γ, to implement the maximization defined in (13) it is necessary to perform a grid evaluation over $[\gamma_L, \gamma_U]$.

In the event that the true cointegrating vector β_0 is known a priori, then the test takes the form (13), except that β is fixed at the known value β_0. We denote this test statistic as

$$\text{SupLM}^0 = \sup_{\gamma_L \leq \gamma \leq \gamma_U} LM(\beta_0, \gamma). \tag{14}$$

It is important to know that the values of γ which maximize the expressions in (13) and (14) will be different from the MLE $\hat{\gamma}$ presented in section 2. This is true for two separate reasons. First, (13)-(14) are LM tests, and are based on parameter estimates obtained under the null rather than the alternative. Second, these LM statistics are computed with heteroskedasticity-consistent covariance matrix estimates, and in this case even the maximizers of SupWald statistics are different from the MLE (the latter equal only when homoskedastic covariance matrix estimates are used). This difference is generic in threshold testing and estimation for regression models, and not special to threshold cointegration.

3.2   Asymptotic Distribution

First consider the case that the true cointegrating vector β_0 is known. The regressors are stationary, and the testing problem is a multivariate generalization of Hansen (1996). It follows that the asymptotic distribution of the tests will take the form given in that paper. We require the following standard weak dependence conditions.

Assumption. $\{\beta' x_t, \Delta x_t\}$ is L^{4r}-bounded, strictly stationary and absolutely regular, with mixing coefficients $\eta_m = O(m^{-A})$ where $A > \nu/(\nu - 1)$ and $r > \nu > 1$. Furthermore, the error u_t is an MDS, and the error correction $\beta' x_t$ has a bounded density function.

Under the null hypothesis (8), these conditions are known to hold when u_t is iid with a bounded density and L^{4r}-bounded. Under the alternative hypothesis (9) these conditions are known to hold under further restrictions on the parameters.

Let $F(\cdot)$ denote the marginal distribution of w_{t-1}, and set $\phi_{t-1} = F(w_{t-1})$. We let "$\Rightarrow$" denote weak convergence with respect to the uniform metric on $[\pi_0, 1-\pi_0]$. Define the matrix functionals

$$M(r) = I_p \otimes E\left(X_{t-1}X'_{t-1}1\left(\phi_{t-1} \leq r\right)\right),$$

and

$$\Omega(r) = E\left[1\left(\phi_{t-1} \leq r\right)\left(u_t u'_t \otimes X_{t-1}X'_{t-1}\right)\right].$$

Theorem 1 *Under $\mathcal{H}_0$,*

$$\mathrm{SupLM}^0 \Rightarrow T = \sup_{\pi_0 \leq r \leq 1-\pi_0} T(r)$$

where

$$T(r) = S^*(r)'\Omega^*(r)^{-1}S^*(r),$$

$$\Omega^*(r) = \Omega(r) - M(r)M(1)^{-1}\Omega(r) - \Omega(r)M(1)^{-1}M(r) + M(r)M(1)^{-1}\Omega(1)M(1)^{-1}M(r),$$

and

$$S^*(r) = S(r) - M(r)M(1)^{-1}S(1),$$

where $S(r)$ is a mean-zero matrix Gaussian process with covariance kernel $E\left(S(r_1)S(r_2)'\right) = \Omega\left(r_1 \wedge r_2\right)$.

The asymptotic distribution in Theorem 1 is the same as that presented in Hansen (1996). In general, the asymptotic distribution does not simplify further. However, we discuss one special simplification at the end of this subsection.

Now we consider the case of estimated β. Since $n\left(\tilde{\beta} - \beta_0\right) = O_p(1)$, it is sufficient to examine the behavior of $LM(\beta, \gamma)$ in an n^{-1} neighborhood of β_0.

Theorem 2 *Under $\mathcal{H}_0$, $LM^{\alpha}(\alpha, \gamma) = LM(\beta_0 + \alpha/n, \gamma)$ has the same asymptotic finite dimensional distributions (fidi's) as $LM(\beta_0, \gamma)$.*

If, in addition, we could show that the process $LM^\alpha(\alpha, \gamma)$ is tight on compact sets, it would follow that SupLM and SupLM^0 have the same asymptotic distribution, namely T. This would imply that the use of the estimate $\tilde{\beta}$, rather than the true value β_0, does not alter the asymptotic null distribution of the LM test. Unfortunately, we have been unable to establish a proof of this proposition. The difficulty is two-fold. The process $LM^\alpha(\alpha, \gamma)$ is discontinuous in γ (due to the indicator functions) and is a function of the non-stationary variable x_{t-1}. There is only a small literature on empirical process results for time-series processes, and virtually none for non-stationary data. Furthermore, the non-stationary variable x_{t-1} appears in the indicator function, so Taylor series methods cannot be used to simplify the problem.

It is our view that despite the lack of a complete proof, the fidi result of Theorem 2 is sufficient to justify using the asymptotic distribution T for the statistic SupLM.

Theorem 1 gives an expression for the asymptotic distribution T. It has the expression as the supremum of a stochastic process $T(r)$, the latter sometimes called a "chi-square process" since for each r the marginal distribution of $T(r)$ is chi-square. As T is the supremum of this stochastic process, its distribution is determined by the joint distribution of this chi-square process, and hence depends on the unknown functions $M(r)$ and $\Omega(r)$. As these functionals may take a broad range of shapes, critical values for T cannot in general be tabulated.

In one special case, we can achieve an important simplification. Take model (2) under $E\left(u_t u_t' \mid \mathcal{F}_{t-1}\right) = \Sigma$ with no intercept and no lags of Δx_t, so that the only regressor is the error-correction term w_{t-1}. Then since $M(r)$ is scalar and monotonically increasing, there exists a function $\psi(s)$ such that $M(\psi(s)) = sM(1)$. We can without loss of generality normalize $M(1) = 1$ and $\Sigma = I$. Then $S(\psi(s)) = W(s)$ is a standard Brownian motion, $\overline{S}(\psi(s)) = W(s) - sW(1)$ is a Brownian bridge, and

$$
\begin{aligned}
T &= \sup_{\pi_0 \leq r \leq 1-\pi_0} T(r) \\
&= \sup_{\pi_0 \leq \psi(s) \leq 1-\pi_0} T(\psi(s)) \\
&= \sup_{s_1 \leq s \leq s_2} \frac{(W(s) - sW(1))^2}{s(1-s)},
\end{aligned}
$$

where $s_1 = \psi^{-1}(\pi_0)$ and $s_2 = \psi^{-1}(1-\pi_0)$. This is the distribution given in Andrews (1993)

for tests for structural change of unknown timing, and is a function of only

$$s_0 = \frac{s_2(1 - s_1)}{s_1(1 - s_2)}.$$

3.3   Asymptotic P-Values: The Fixed Regressor Bootstrap

With the exception discussed at the end of section 3.2, the asymptotic distribution of the SupLM test appears to depend upon the moment functionals $M(r)$ and $\Omega(r)$, so tabulated critical values are unavailable. We discuss in this section how the fixed regressor bootstrap of Hansen (1996, 2000b) can be used to calculate asymptotic critical values and p-values, and hence achieve first-order asymptotically correct inference. .

An important implication of Theorem 2 is that the first-step estimation of the cointegrating vector β does not affect the asymptotic distribution of the LM test. We therefore do not need to take the estimation of β into account.

We now describe the fixed regressor bootstrap. Let $\tilde{w}_{t-1} = w_{t-1}(\tilde{\beta})$, $\tilde{X}_{t-1} = X_{t-1}(\tilde{\beta})$, and let $\tilde{u}_t$ be the residuals from the reduced rank regression as described in section 2. For the remainder of our discussion, $\tilde{u}_t$, $\tilde{w}_{t-1}$, $\tilde{X}_{t-1}$ and $\tilde{\beta}$ are held fixed at their sample values.

Let e_{bt} be iid $N(0,1)$ and set $y_{bt} = \tilde{u}_t e_{bt}$. Regress y_{bt} on $\tilde{X}_{t-1}$ yielding residuals $\tilde{u}_{bt}$. Regress y_{bt} on $\tilde{X}_{t-1}d_{1t}(\tilde{\beta}, \gamma)$ and $\tilde{X}_{t-1}d_{2t}(\tilde{\beta}, \gamma)$, yielding estimates $\hat{A}_1(\gamma)_b$ and $\hat{A}_2(\gamma)_b$, and residuals $\hat{u}_{bt}(\gamma)$. Define $\hat{V}_1(\gamma)_b$ and $\hat{V}_2(\gamma)_b$ as in (10) and (11) setting $\beta = \tilde{\beta}$ and replacing $\tilde{u}_t$ with $\tilde{u}_{bt}$ in the definition of $\xi_1(\tilde{\beta}, \gamma)$ and $\xi_2(\tilde{\beta}, \gamma)$. Then set

$$\text{SupLM}^* = \sup_{\gamma_L \leq \gamma \leq \gamma_U} \text{vec}\left(\hat{A}_1(\gamma)_b - \hat{A}_2(\gamma)_b\right)' \left(\hat{V}_1(\gamma)_b + \hat{V}_2(\gamma)_b\right)^{-1} \text{vec}\left(\hat{A}_1(\gamma)_b - \hat{A}_2(\gamma)_b\right)$$

The analysis in Hansen (1996) shows that under local alternatives to $\mathcal{H}_0$, $\text{SupLM}^* \Rightarrow_p T$, so the distribution of SupLM^* yields a valid first-order approximation to the asymptotic null distribution of SupLM. The symbol "$\Rightarrow_p$" denotes weak convergence in probability as defined in Gine and Zinn (1990).

The distribution SupLM^* is unknown, but can be calculated using simulation methods. The description given above shows how to create one draw from the distribution. With independent draws of the errors e_{bt}, a new draw can be made. If this is repeated a large number of times (e.g. 1000), a p-value can be calculated by counting the percentage of simulated SupLM^* which exceed the actual SupLM.

14

The label "fixed regressor bootstrap" is intended to convey the feature that the regressors $\tilde{X}_{t-1}d_{1t}(\tilde{\beta}, \gamma)$ and $\tilde{X}_{t-1}d_{2t}(\tilde{\beta}, \gamma)$ are held fixed at their sample values. As such, this is not really a bootstrap technique, and is not expected to provide a better approximation to the finite sample distribution than conventional asymptotic approximations. The advantage of the method is that it allows for heteroskedasticity of unknown form, while conventional model-based bootstrap methods effectively impose independence on the errors u_t and therefore do not achieve correct first-order asymptotic inference. It allows for general heteroskedasticity in much the same way as White's (1980) heteroskedasticity-consistent standard errors.

3.4   Residual Bootstrap

The fixed regressor bootstrap of the previous section has much of the computational burden of a bootstrap, but only approximates the asymptotic distribution. While we have no formal theory, it stands to reason that a bootstrap method might achieve better finite sample performance than asymptotic methods. This conjecture is not obvious, as the asymptotic distribution of Section 3.2 is non-pivotal, and it is known that the bootstrap in general does not achieve an asymptotic refinement (an improved rate of convergence relative to asymptotic inference) when asymptotic distributions are non-pivotal.

One cost of using the bootstrap is the need to be fully parametric concerning the data generating mechanism. In particular, it is difficult to incorporate conditional heteroskedasticity, and in its presence a conventional bootstrap (using iid innovations) will fail to achieve the first-order asymptotic distribution (unlike the fixed regressors bootstrap, which does).

The parametric residual bootstrap method requires a complete specification of the model under the null. This is equation (1) plus auxiliary assumptions on the errors u_t and the initial conditions. In our applications, we assume u_t is iid from an unknown distribution G, and the initial conditions are fixed (other choices are possible). The bootstrap calculates the sampling distribution of the test SupLM using this model and the parameter estimates obtained under the null. The latter are $\tilde{\beta}$, $\tilde{A}$, and the empirical distribution of the bi-variate residuals $\tilde{u}_t$.

The bootstrap distribution may be calculated by simulation. Given the fixed initial conditions, random draws are made from the residual vectors $\tilde{u}_t$, and then the vector series

x_{bt} are created by recursion given model (1). The statistic SupLM* is calculated on each simulated sample and stored. The bootstrap p-value is the percentage of simulated statistics which exceed the actual statistic.

4 Simulation Evidence

4.1 Threshold Test

Monte Carlo experiments are made to find out the small sample performance of the test. The experiments are based on a bivariate error correction model with two lags. Letting $x_t = (x_{1t}\ x_{2t})'$, the single-regime model $\mathcal{H}_0$ is

$$\Delta x_t = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} (x_{1t-1} - \beta x_{2t-1}) + \Gamma \Delta x_{t-1} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}. \tag{15}$$

The two-regime model $\mathcal{H}_1$ is the generalization of (15) as in (2), allowing all coefficients to differ depending if $x_{1t-1} - x_{2t-1} \le \gamma$ or $x_{1t-1} - x_{2t-1} > \gamma$.

Our tests are based on model (15), allowing all coefficients to switch between regimes under the alternative. The tests are calculated setting $\pi_0 = .10$, using 50 gridpoints on $[\gamma_L, \gamma_U]$ for calculation of (13), and using 200 bootstrap replications for each replication. Our results are calculated from 1000 simulation replications.

We fix $\mu_1 = \mu_2 = 0$, $\beta = 1$, and $\alpha_1 = -1$. We vary α_2 among $(0, -.5, .5)$, and Γ among

$$\Gamma_0 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \Gamma_1 = \begin{pmatrix} -.2 & 0 \\ -.1 & -.2 \end{pmatrix}, \qquad \Gamma_2 = \begin{pmatrix} -.2 & -.1 \\ -.1 & -.2 \end{pmatrix},$$

and consider two sample sizes, $n = 100$ and $n = 250$. We generated the errors u_{1t} and u_{2t} under homoskedastic and heteroskedastic specifications. For a homoskedastic error, we generate u_{1t} and u_{2t} as independent $N(0,1)$ variates. For a heteroskedastic error, we generate u_{1t} and u_{2t} as independent GARCH(1,1) processes, with $u_{1t} \sim N(0, \sigma_{1t}^2)$ and $\sigma_{1t}^2 = 1 + .2u_{1t-1}^2 + \phi\sigma_{1t-1}^2$, and similarly u_{2t}.

We first explored the size of the SupLM and SupLM[0] statistics under the null hypothesis $\mathcal{H}_0$ of a single regime. This involved generating data from the linear model (15). For

each simulated sample, the statistics and p-values were calculated using both the fixed-regressor bootstrap and the residual bootstrap. In Table 1, we report rejection frequencies from nominal 5% and 10% tests for the SupLM statistic (β unknown). The results for the SupLM^0 statistic (β known) were very similar and so are omitted.

For the first five parameterizations, we generate u_{1t} and u_{2t} as independent $N(0,1)$ variates, and vary the parameters α_2 and Γ. The rejection frequencies of the tests are reported in the first five columns of Table 1. (These are the percentage of the simulated p-values which are smaller than the nominal size.) The rejection frequencies are similar across the various parameterizations. Using the fixed-regressor bootstrap, the test somewhat over-rejects, with the rejection rate at the nominal 5% level and $n = 100$ ranging from .071 to .108. If the residual bootstrap is used, the test has much better size, with rejection rates ranging from .048 to .084. If the sample size is increased to $n = 250$, then the size improves considerably, with the 5% rejection rates for the fixed-regressor bootstrap ranging from .070 to .080 and those for the residual bootstrap ranging from .052 to .058.

Table 1
Size of SupLM Tests

Parameters	Homoskedastic Errors					Heteroskedastic		
α_2	0	$-.5$	$.5$	0	0	0	0	0
Γ	Γ_0	Γ_0	Γ_0	Γ_1	Γ_2	Γ_0	Γ_0	Γ_0
ϕ	0	0	0	0	0	.25	.50	.75
	5% nominal size, $n = 100$							
Fixed-regressor bootstrap	0.083	0.072	0.108	0.071	0.074	0.075	0.080	0.085
Residual bootstrap	0.058	0.052	0.084	0.049	0.048	0.054	0.065	0.065
	5% nominal size, $n = 250$							
Fixed-regressor bootstrap	0.075	0.080	0.070	0.067	0.071	0.064	0.067	0.076
Residual bootstrap	0.052	0.058	0.055	0.052	0.053	0.053	0.051	0.059
	10% nominal size, $n = 100$							
Fixed-regressor bootstrap	0.156	0.133	0.186	0.134	0.144	0.162	0.146	0.156
Residual bootstrap	0.125	0.111	0.139	0.100	0.111	0.125	0.117	0.126
	10% nominal size, $n = 250$							
Fixed-regressor bootstrap	0.138	0.135	0.147	0.122	0.127	0.117	0.133	0.134
Residual bootstrap	0.106	0.113	0.109	0.093	0.095	0.089	0.099	0.103

For the remaining three parameterizations, we generate u_{1t} and u_{2t} as independent GARCH(1,1) processes. The other parameters are set as in the first column of Table 1, and the results for the SupLM tests are reported in the final three columns of Table 1. The rejection rates do not appear to be greatly affected by the heteroskedasticity. The rejection rates for both SupLM tests increase modestly, and the best results are obtained again by the residual bootstrap. (This might appear surprising, as the residual bootstrap does not replicate the GARCH dependence structure, but the LM statistics are constructed with heteroskedasticity-robust covariance matrices, so are first-order robust to GARCH.)

We next explored the power of the tests against the two-regime alternative $\mathcal{H}_1$. To keep the calculations manageable, we generate the data from the simple process

$$\Delta x_t = \begin{pmatrix} -1 \\ 0 \end{pmatrix} (x_{1t-1} - x_{2t-1}) + \begin{pmatrix} \delta \\ 0 \end{pmatrix} (x_{1t-1} - x_{2t-1}) \, 1 \, (x_{1t-1} - x_{2t-1} \leq \gamma) + u_t,$$

with u_t iid $N(0, I_2)$. Setting $\delta = 0$ the null hypothesis holds and this process corresponds to the data generated from the first column of Table 1. For $\delta \neq 0$ the alternative of a two-regime model holds. The threshold parameter γ is set so that $P(w_{t-1} \leq \gamma)$ equals either .5 or .25. While the data is generated from the first-order VAR, we used the same tests as described above, which are based on an estimated second-order VAR.

Table 2. Power of SupLM and SupLM^0 Tests, 5% nominal size
Against Two-Regime Alternative

π	0.5				0.25			
δ	.2	.4	.6	.8	.2	.4	.6	.8
$n = 100$								
SupLM, Fixed-regressor bootstrap	0.107	0.207	0.395	0.624	0.119	0.231	0.364	0.468
SupLM, Residual bootstrap	0.081	0.175	0.324	0.587	0.091	0.183	0.322	0.425
SupLM^0, Fixed-regressor bootstrap	0.104	0.212	0.411	0.713	0.117	0.224	0.392	0.517
SupLM^0, Residual bootstrap	0.087	0.166	0.357	0.647	0.088	0.174	0.332	0.453
$n = 250$								
SupLM, Fixed-regressor bootstrap	0.156	0.450	0.878	0.997	0.187	0.527	0.844	0.933
SupLM, Residual bootstrap	0.117	0.399	0.845	0.995	0.152	0.481	0.816	0.914
SupLM^0, Fixed-regressor bootstrap	0.154	0.460	0.896	0.999	0.184	0.546	0.856	0.953
SupLM^0, Residual bootstrap	0.121	0.403	0.852	0.998	0.144	0.492	0.823	0.929

Table 2 reports the rejection frequency of the SupLM and SupLM^0 tests at the 5% size for several values of δ. As expected, the power increases in the threshold effect δ and sample size n. The fixed regressor bootstrap has a higher rejection rate than the parametric bootstrap, but this is a likely artifact of the size distortions shown in Table 1. The SupLM^0 test (known β) has slightly higher power than the SupLM test (unknown β) but the difference is surprising small. At least in these settings, there is little power loss due to estimation of β.

4.2 Parameter Estimates

We next explore the finite sample distributions of the estimators of the cointegrating vector β and the threshold parameter γ. The simulation is based on the following process

$$
\Delta x_t = \begin{pmatrix} -1 \\ 0 \end{pmatrix} (x_{1t-1} - \beta_0 x_{2t-1}) + \begin{pmatrix} -2 \\ 0 \end{pmatrix} 1 (x_{1t-1} - \beta_0 x_{2t-1} \leq \gamma_0)
$$
$$
+ \begin{pmatrix} .5 \\ 0 \end{pmatrix} (x_{1t-1} - \beta_0 x_{2t-1}) 1 (x_{1t-1} - \beta_0 x_{2t-1} > \gamma_0) + u_t,
$$

with $u_t \sim$ iid $N(0, I_2)$. We set the cointegrating coefficient β_0 at 1 and the threshold coefficient γ_0 at 0. This model has threshold effects in both the intercept and in the error-correction. We consider two sample sizes, $n = 100$ and $n = 250$. We varied some of the coefficients, but omit the results since the essential features were unchanged. While the data is generated from the a VAR(1), our estimates our based on a VAR(2).

We consider three estimators of β, and two of γ. The pair $(\hat{\beta}, \hat{\gamma})$ are the unrestricted estimators of (β, γ), using the algorithm[3] of section 2. $\hat{\beta}_0$ is the restricted estimator obtained when the true value γ_0 is known. $\hat{\gamma}_0$ is the restricted estimator of γ when the true value β_0 is known. We should expect $\hat{\beta}_0$ and $\hat{\gamma}_0$ to be more accurate than $\hat{\beta}$ and $\hat{\gamma}$, respectively, and this comparison allows us to assess the cost due to estimating the threshold and cointegrating vector, respectively. We also consider the Johansen MLE, $\tilde{\beta}$, which would be efficient if there were no threshold effect.

In Table 3 we report the mean, root mean squared error (RMSE), mean absolute error (MAE), and selected percentiles of each estimator in 1000 simulation replications.

---

[3] The grid sizes for $\hat{\gamma}$ and $\hat{\beta}$ are 500 and 100, respectively.

The results contain no surprises. The three estimators of β all have approximate symmetric, unbiased, distributions. The restricted estimator $\hat{\beta}_0$ (which exploits knowledge about γ) is the most accurate, followed by the unrestricted estimator $\hat{\beta}$. Both are more accurate than the Johansen $\tilde{\beta}$.

Both estimators of γ have asymmetric distributions. For $n = 100$, the distributions are similar, both are meaningfully biased and the estimators are quite inaccurate. For $n = 250$, the performance of both estimators is much improved, and the restricted estimator $\hat{\gamma}_0$ has less dispersion but slightly higher bias than the unrestricted estimator $\hat{\gamma}$.

Table 3. Distribution of Estimators

	Mean	RMSE	MAE	percentiles(%)				
				5	25	50	75	95
				$n = 100$				
$\hat{\beta} - \beta_0$	-0.0002	0.0729	0.0154	-0.0382	-0.0104	-0.0004	0.0102	0.0322
$\hat{\beta}_0 - \beta_0$	-0.0000	0.0524	0.0100	-0.0266	-0.0056	0.0001	0.0056	0.0234
$\tilde{\beta} - \beta_0$	0.0000	0.0982	0.0234	-0.0493	-0.0196	-0.0004	0.0187	0.0475
$\hat{\gamma} - \gamma_0$	-0.0621	0.9778	0.1460	-0.3462	-0.0940	-0.0145	0.0462	0.2207
$\hat{\gamma}_0 - \gamma_0$	-0.0918	0.9967	0.1221	-0.3351	-0.0773	-0.0320	-0.0050	0.0983
				$n = 250$				
$\hat{\beta} - \beta_0$	0.0000	0.0130	0.0048	-0.0107	-0.0035	-0.0000	0.0035	0.0107
$\hat{\beta}_0 - \beta_0$	-0.0001	0.0091	0.0029	-0.0080	-0.0017	-0.0000	0.0017	0.0064
$\tilde{\beta} - \beta_0$	0.0003	0.0236	0.0088	-0.0187	-0.0060	-0.0001	0.0066	0.0194
$\hat{\gamma} - \gamma_0$	-0.0051	0.1419	0.0441	-0.1109	-0.0323	-0.0020	0.0238	0.0919
$\hat{\gamma}_0 - \gamma_0$	-0.0150	0.0815	0.0259	-0.0752	-0.0272	-0.0113	-0.0014	0.0376

5 Term Structure

Let r_t be the interest rate on a one-period bond, and R_t be the interest rate on a multi-period bond. As first suggested by Campbell and Shiller (1987), the theory of the term structure of interest rates suggests that r_t and R_t should be cointegrated with a unit cointegrating vector.

This has led to a large empirical literature estimating linear cointegrating VAR models such as

$$\begin{pmatrix} \Delta R_t \\ \Delta r_t \end{pmatrix} = \mu + \alpha w_{t-1} + \Gamma \begin{pmatrix} \Delta R_{t-1} \\ \Delta r_{t-1} \end{pmatrix} + u_t. \tag{16}$$

with $w_{t-1} = R_{t-1} - \beta r_{t-1}$. Setting $\beta = 1$, the error-correction term is the interest rate spread.

Linearity, however, is not implied by the theory of the term structure. In this section, we explore the possibility that a threshold cointegration model provides a better empirical description.

To address this question, we estimate and test models of threshold cointegration using the monthly interest rate series of McCulloch and Kwon (1993). Following Campbell (1995), we use the period 1952-1991. The interest rates are estimated from the prices of U.S. Treasury securities, and correspond to zero-coupon bonds. We use a selection of bonds rates with maturities ranging from 1 to 120 months. To select the VAR lag length, we found that both the AIC and BIC, applied either to the linear VECM or the threshold VECM, consistently picked $l = 1$ across specifications. We report our results for both $l = 1$ and $l = 2$ for robustness. We considered both fixing the cointegrating vector $\beta = 1$ and letting $\hat{\beta}$ be estimated.

First, we tested for the presence of (bivariate) cointegration, using the ADF test applied to the error-correction term (this is the Engle-Granger test when the cointegrating vector is estimated). For all bivariate pairs and lag lengths considered, the tests[4] easily rejected the null hypothesis of no cointegration, indicating the presence of bivariate cointegration between each pair.

To assess the evidence for threshold cointegration, we applied several sets of tests. For the complete multivariate specification, we use the SupLM test (estimated β) and the SupLM^0 test ($\beta = 1$) with 300 gridpoints, and the p-values calculated by the parametric bootstrap. For comparison, we also applied the univariate Hansen (1996) threshold autoregressive test to the error-correction term as in Balke and Fomby (1997). All p-values were computed with 5000 simulation replications. The results are presented in Table 4.

The multivariate tests point to the presence of threshold cointegration in some of the

---

[4]Not reported here to conserve space.

bivariate relationships. In six of the nine models, the SupLM^0 statistic is significant at the 10% level when $l = 1$ and β is fixed at unity. If we set $l = 2$, the evidence appears to strengthen, with seven of the nine significant at the 5% level. If instead of fixing β we estimate it freely, the evidence for threshold cointegration is diminished, with only four of nine significant at the 5% level (in either lag specification).

The Balke-Fomby univariate tests are somewhat more ambiguous. The threshold effect is significant at the 10% level for two of the nine models when $l = 1$ and β is fixed at unity, and for five when $l = 2$. When β is estimated rather than fixed, then none of the models are significant for $l = 1$, and only three for $l = 2$. The univariate specification is quite restrictive, and this undoubtedly reduces the power of the test in some settings.

Table 4
Treasury Bond Rates
Tests for Threshold Cointegration (P-Values)

| | | Multivariate | | | | Univariate | | | |
| | | $\beta = 1$ | | $\hat{\beta}$ estimated | | $\beta = 1$ | | $\hat{\beta}$ estimated | |
Short Rate	Long Rate	$l = 1$	$l = 2$	$l = 1$	$l = 2$	$l = 1$	$l = 2$	$l = 1$	$l = 2$
1-month	2-month	.083	.003	.014	.007	.453	.002	.370	.188
1-month	3-month	.030	.009	.117	.188	.283	.085	.245	.044
1-month	6-month	.085	.029	.634	.288	.017	.040	.122	.133
3-month	6-month	.036	.021	.038	.031	.658	.311	.322	.133
3-month	12-month	.047	.032	.161	.198	.121	.091	.122	.083
3-month	120-month	.193	.102	.095	.146	.227	.485	.171	.357
12-month	24-month	.267	.516	.245	.623	.489	.583	.314	.618
12-month	120-month	.018	.022	.023	.016	.109	.119	.251	.228
24-month	120-month	.173	.005	.139	.008	.024	.011	.278	.051

Next, we report the parameter estimates for one of the relatively successful models, the bivariate relationship between the 120-month (10-year) and 12-month (one-year) bond rates (normalized to be percentages). The parameter estimates were calculated by minimization of (7) over a 300×300 grid on the parameters (γ, β). The estimated cointegrating relationship is $w_t = R_t - 0.984 \, r_t$, quite close to a unit coefficient. The results we report are for the case of estimated cointegrating vector, but the results are very similar if the unit coefficient is imposed.

The estimated threshold is $\hat{\gamma} = -0.63$. Thus the first regime occurs when $R_t \leq 0.984 r_t - 0.63$, i.e. when the 10-year rate is more than .6 percentage points below the short rate. This is relatively unusual, with only 8% of the observations in this regime, and we label this as the "extreme" regime. The second regime (with 92% of the observations) is when $R_t > 0.984 r_t - 0.63$, which we label as the "typical" regime.

The estimated threshold VAR is given below

$$
\Delta R_t = \begin{cases}
\begin{array}{cccccc}
.54 & + & .34 \; w_{t-1} + & .35 \; \Delta R_{t-1} - & .17 \; \Delta r_{t-1} + u_{1t}, & w_{t-1} \leq -.63 \\
(.17) & & (.18) & (.26) & (.12) &
\end{array} \\
\\
\begin{array}{cccccc}
.01 & - & .02 \; w_{t-1} - & .08 \; \Delta R_{t-1} + & .09 \; \Delta r_{t-1} + u_{1t}, & w_{t-1} > -.63 \\
(.02) & & (.03) & (.06) & (.05) &
\end{array}
\end{cases} ,
$$

$$
\Delta r_t = \begin{cases}
\begin{array}{cccccc}
1.45 & + & 1.41 \; w_{t-1} + & .92 \; \Delta R_{t-1} - & .04 \; \Delta r_{t-1} + u_{2t}, & w_{t-1} \leq -.63 \\
(.35) & & (.34) & (.62) & (.26) &
\end{array} \\
\\
\begin{array}{cccccc}
- & .04 & + & .04 \; w_{t-1} - & .07 \; \Delta R_{t-1} + & .23 \; \Delta r_{t-1} + u_{2t}, \quad w_{t-1} > -.63 \\
& (.04) & & (.04) & (.13) & (.13)
\end{array}
\end{cases} ,
$$

Eicker-White Standard errors are in parenthesis. However, as we have no formal distribution theory for the parameter estimates and standard errors, these should be interpreted somewhat cautiously.

In the typical regime, ΔR_t and Δr_t have minimal error-correction effects and minimal dynamics. They are close to white noise, indicating that in this regime, R_t and r_t are close to driftless random walks.

Error-correction appears to occur only in the unusual regime (when R_t is much below r_t). There is a strong error-correction effect in the short-rate equation. In the long-rate equation,

the point estimate for the error-correction term is moderately large, and on the borderline of statistical significance. The remaining dynamic coefficients are imprecisely estimated, due to the small sample in this regime.

In Figure 3 we plot the error-correction effect – the estimated regression functions of ΔR_t and Δr_t as a function of w_{t-1}, holding the other variables constant. In the figure, you can see the flat near-zero error-correction effect on the right size of the threshold, and on the left of the threshold, the sharp positive relationships, especially for the short rate equation.

One finding of great interest is that the estimated error-correction effects are positive[5]. As articulated by Campbell and Shiller (1991) and Campbell (1995), the regression lines in Figure 3 should be positive – equivalently, the coefficients on w_{t-1} in the threshold VECM should be positive. This is because a large positive spread $R_t - r_t$ means that the long bond is earning a higher interest rate, so long bonds must be expected to depreciate in value. This implies that the long interest rate is expected to rise. (The short rate is also expected to rise as R_t is a smoothed forecast of future short rates.)

Using linear correlation methods, Campbell and Shiller (1991) and Campbell (1995) found considerable evidence contradicting this prediction of the term structure theory. The found that the changes in the short rate are positively correlated with the spread, but changes in the long rate are negatively correlated with the spread, especially at longer horizons. These authors viewed this finding as a puzzle.

In contrast, our results are roughly consistent with this term structure prediction. In all nine estimated[6] bi-variate relationships, the four error-correction coefficients (for the long and short rate in the two regimes) are either positive or insignificantly different from zero if negative. As expected, the short rate coefficients are typically positive (in six of the nine models the coefficients are positive in both regimes), and the long rate coefficients are much smaller in magnitude and often negative in sign. There appears to be no puzzle.

---

[5]In the typical regime, the long rate has a negative point estimate (-.02), but it is statistically insignificant and numerically very close to zero.

[6]β fixed at unity, $l = 1$.

6 Conclusion

We have presented a quasi-MLE algorithm for constructing estimates of a two-regime threshold cointegration model, presented a SupLM statistic for the null hypothesis of no threshold, an asymptotic distribution theory for this statistic, presented methods to calculate the asymptotic distribution by simulation, and how to calculate a bootstrap approximation. These methods may find constructive use in applications.

Still, there are many unanswered questions for future research. We have not presented a test for the null of no cointegration in the context of the threshold cointegration model[7]. This testing problem is quite complicated, as the null hypothesis implies that the threshold variable (the cointegrating error) is non-stationary, rendering current distribution theory inapplicable. We also have not presented a distribution theory for the parameter estimates for the threshold cointegration model. As shown in Chan (1993) and Hansen (2000a), threshold estimates have non-standard distributions, and working out such distribution theory is challenging. Furthermore, it would be useful to consider three-regime threshold cointegration models. To assess the statistical relevance of such models, we would need a test of the null of a two-regime model against the alternative of a three-regime model. These topics would all be relevant for future research.

7 Appendix

Proof of Theorem 1. An alternative algebraic representation of the pointwise LM statistic is

$$LM(\beta, \gamma) = N_n^*(\beta, \gamma)'\Omega_n^*(\beta, \gamma)^{-1}N_n^*(\beta, \gamma) \tag{17}$$

where

$$
\begin{aligned}
\Omega_n^*(\beta, \gamma) &= \Omega_n(\beta, \gamma) - M_n(\beta, \gamma)M_n(\beta)^{-1}\Omega_n(\beta, \gamma) \\
&\quad -\Omega_n(\beta, \gamma)M_n(\beta)^{-1}M_n(\beta, \gamma) \\
&\quad +M_n(\beta, \gamma)M_n(\beta)^{-1}\Omega_n(\beta)M_n(\beta)^{-1}M_n(\beta, \gamma),
\end{aligned}
$$

---

[7]Pippenger and Goering (2000) present simulation evidence that linear cointegration tests can have low power to detect threshold cointegration.

$$N_n^*(\beta, \gamma) = N_n(\beta, \gamma) - M_n(\beta, \gamma)M_n(\beta)^{-1}N_n(\beta),$$

$$
\begin{aligned}
M_n(\beta, \gamma) &= I_p \otimes \frac{1}{n} \sum_{t=1}^n d_{1t}(\beta, \gamma)X_{t-1}(\beta)X_{t-1}(\beta)', \\
M_n(\beta) &= I_p \otimes \frac{1}{n} \sum_{t=1}^n X_{t-1}(\beta)X_{t-1}(\beta)', \\
\Omega_n(\beta, \gamma) &= \frac{1}{n} \sum_{t=1}^n d_{1t}(\beta, \gamma)\left(\tilde{u}_t \tilde{u}_t' \otimes X_{t-1}(\beta)X_{t-1}(\beta)'\right) \\
\Omega_n(\beta) &= \frac{1}{n} \sum_{t=1}^n \left(\tilde{u}_t \tilde{u}_t' \otimes X_{t-1}(\beta)X_{t-1}(\beta)'\right) \\
N_n(\beta, \gamma) &= \frac{1}{\sqrt{n}} \sum_{t=1}^n d_{1t}(\beta, \gamma)\left(\Delta x_t \otimes X_{t-1}(\beta)\right), \\
N_n(\beta) &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \left(\Delta x_t \otimes X_{t-1}(\beta)\right).
\end{aligned}
$$

Then observe that

$$\mathrm{SupLM}^0 = \sup_{\gamma_L \leq \gamma \leq \gamma_U} LM(\beta_0, \gamma) = \sup_{\pi_0 \leq r \leq 1-\pi_0} LM(\beta_0, F^{-1}(r))$$

where $r = F(\gamma)$. Since $LM(\beta_0, \gamma)$ is a function of γ only through the indicator function

$$1\left(w_{t-1} \leq \gamma\right) = 1\left(\phi_{t-1} \leq r\right),$$

(as $\phi_{t-1} = F(w_{t-1})$) it follows that

$$\mathrm{SupLM}^0 = \sup_{\pi_0 \leq r \leq 1-\pi_0} LM^0(\beta_0, r),$$

where $LM^0(\beta_0, r) = LM(\beta_0, F^{-1}(r))$ is defined as in (17) except that all instances of $1\left(w_{t-1} \leq \gamma\right)$ are replaced by $1\left(\phi_{t-1} \leq r\right)$.

Under $\mathcal{H}_0$ and making these changes, this simplifies to

$$LM^0(r) = S_n^*(r)'\Omega_n^*(r)^{-1}S_n^*(r)$$

where

$$\Omega_n^*(r) = \Omega_n(r) - M_n(r)M_n^{-1}\Omega_n(r) - \Omega_n(r)M_n^{-1}M_n(r) + M_n(r)M_n^{-1}\Omega_n M_n(r)$$

$$S_n^*(r) = S_n(r) - M_n(r)M_n^{-1}S_n,$$

26

$$M_n(r) = I_p \otimes \frac{1}{n} \sum_{t=1}^{n} 1\left(\phi_{t-1} \le r\right) X_{t-1} X'_{t-1},$$

$$M_n = I_p \otimes \frac{1}{n} \sum_{t=1}^{n} X_{t-1} X'_{t-1},$$

$$\Omega_n(r) = \frac{1}{n} \sum_{t=1}^{n} 1\left(\phi_{t-1} \le r\right) \left(\tilde{u}_t \tilde{u}'_t \otimes X_{t-1} X'_{t-1}\right)$$

$$\Omega_n = \frac{1}{n} \sum_{t=1}^{n} \left(\tilde{u}_t \tilde{u}'_t \otimes X_{t-1} X'_{t-1}\right)$$

$$S_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} 1\left(\phi_{t-1} \le r\right) \left(u_t \otimes X_{t-1}\right),$$

$$S_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(u_t \otimes X_{t-1}\right).$$

The stated result then follows from the joint convergence

$$M_n(r) \Rightarrow M(r)$$

$$\Omega_n(r) \Rightarrow \Omega(r)$$

$$S_n(r) \Rightarrow S(r),.$$

which follows from Theorem 3 of Hansen (1996), which holds under our stated Assumptions.
∎

Proof of Theorem 2. First, let $w_{t-1}(\alpha) = \tilde{w}_{t-1}(\beta_0 + \alpha/n) = w_{t-1} + n^{-1}\alpha' x_{t-1}$, and $X_{t-1}(\alpha) = X_{t-1}(\beta_0 + \alpha/n)$. Hence

$$\Delta x_t = A' X_{t-1} + u_t$$

$$= A' X_{t-1}(\alpha) - \delta\left(n^{-1}\alpha' x_{t-1}\right) + u_t,$$

where δ' is the second row of A (the coefficient vector on w_{t-1}). Hence

$$LM^\alpha\left(\alpha, \gamma\right) = N_n^*(\beta_0 + \alpha/n, \gamma)' \Omega_n^*(\beta_0 + \alpha/n, \gamma)^{-1} N_n^*(\beta_0 + \alpha/n, \gamma),$$

and

$$N_n^*(\beta_0 + \alpha/n, \gamma) = S_n^*(\alpha, \gamma) - C_n^*(\alpha, \gamma),$$

where

$$S_n^*(\alpha, \gamma) = S_n(\alpha, \gamma) - M_n(\beta_0 + \alpha/n, \gamma)M_n(\beta_0 + \alpha/n)^{-1}S_n(\alpha),$$

$$S_n(\alpha, \gamma) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} 1\left(w_{t-1}(\alpha) \leq \gamma\right)\left(u_t \otimes X_{t-1}(\alpha)\right),$$

$$S_n(\alpha) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(u_t \otimes X_{t-1}(\alpha)\right),$$

and

$$C_n^*(\alpha, \gamma) = C_n(\alpha, \gamma) - M_n(\beta_0 + \alpha/n, \gamma)M_n(\beta_0 + \alpha/n)^{-1}C_n(\alpha),$$

$$C_n(\alpha, \gamma) = \frac{1}{n^{3/2}}\sum_{t=1}^{n} 1\left(w_{t-1}(\alpha) \leq \gamma\right)\left(\delta\alpha' x_{t-1} \otimes X_{t-1}(\alpha)\right),$$

$$C_n(\alpha) = \frac{1}{n^{3/2}}\sum_{t=1}^{n}\left(\delta\alpha' x_{t-1} \otimes X_{t-1}(\alpha)\right).$$

To complete the proof, we need to show that $|\Omega_n(\beta_0 + \alpha/n, \gamma) - \Omega_n(\beta_0, \gamma)| = o_p(1)$, $|M_n(\beta_0 + \alpha/n, \gamma) - M_n(\beta_0, \gamma)| = o_p(1)$, $|S_n(\beta_0 + \alpha/n, \gamma) - S_n(\beta_0, \gamma)| = o_p(1)$, and $C_n^*(\alpha, \gamma) = o_p(1)$. First, observe that since $|X_{t-1} - X_{t-1}(\alpha)| = |n^{-1}\alpha' x_{t-1}| = O_p(n^{-1/2})$, it is fairly straightforward to see that we can replace the $X_{t-1}(\alpha)$ by X_{t-1} with only $o_p(1)$ error in the above expressions, and we make this substitution for the remainder of the proof.

Let E_Q be the event $\{n^{-1/2}\sup_{t \leq n}|\alpha' x_{t-1}| > Q\}$. For any $\varepsilon > 0$, there is some $Q < \infty$ such that $P(E_Q) \leq \varepsilon$. The remainder of the analysis conditions on the set $\{n^{-1/2}\sup_{t \leq n}|\alpha' x_{t-1}| > Q\}$

We next show that $|M_n(\beta_0 + \alpha/n, \gamma) - M_n(\beta_0, \gamma)| = o_p(1)$. Indeed, on the set E_Q

$$|M_n(\beta_0 + \alpha/n, \gamma) - M_n(\beta_0, \gamma)|^2 = \left|\frac{1}{n}\sum_{t=1}^{n}\left(d_{1t}(\beta_0 + \alpha/n, \gamma) - d_{1t}(\beta_0, \gamma)\right)X_{t-1}X_{t-1}'\right|^2$$

$$\leq \left(\frac{1}{n}\sum_{t=1}^{n}|X_{t-1}|^4\right)\left(\frac{1}{n}\sum_{t=1}^{n}\left|1\left(w_{t-1} + n^{-1}\alpha' x_{t-1} \leq \gamma\right) - 1\left(w_{t-1} \leq \gamma\right)\right|\right)$$

$$\leq \left(\frac{1}{n}\sum_{t=1}^{n}|X_{t-1}|^4\right)\left(\frac{1}{n}\sum_{t=1}^{n} 1\left(w_{t-1} - n^{-1/2}Q \leq \gamma \leq w_{t-1} + n^{-1/2}Q\right)\right)$$

$$\leq o_p(1).$$

The proof that $|\Omega_n(\beta_0 + \alpha/n, \gamma) - \Omega_n(\beta_0, \gamma)| = o_p(1)$ follows similarly.

28

Next, since u_t is an MDS,

$$
\begin{aligned}
E \left(S_n(\beta_0 + \alpha/n, \gamma) - S_n(\beta_0, \gamma) \right)^2 &= E \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(1 \left(w_{t-1}(\alpha) \le \gamma \right) - 1 \left(w_{t-1} \le \gamma \right) \right) X_{t-1} u_t' \right)^2 \\
&= E \left| \left(1 \left(w_{t-1}(\alpha) \le \gamma \right) - 1 \left(w_{t-1} \le \gamma \right) \right) X_{t-1} u_t' \right|^2 \\
&\le E \left| 1 \left(w_{t-1} - n^{-1/2} Q \le \gamma \le w_{t-1} + n^{-1/2} Q \right) X_{t-1} u_t' \right|^2 + \varepsilon \\
&= o(1) + \varepsilon
\end{aligned}
$$

and ε can be made arbitrarily small.

Finally, using similar analysis, $C_n^*(\alpha, \gamma) = C_n^*(0, \gamma) + o_p(1)$. Since x_t is $I(1)$, $n^{-1/2} x_{[nr]} \Rightarrow B(r)$, a vector Brownian motion. We can appear to Theorem 3 of Caner and Hansen (2000) as our assumptions imply theirs (absolute regularity is stronger than strong mixing). Hence

$$
\begin{aligned}
C_n^*(0, \gamma) &= \frac{1}{n^{3/2}} \sum_{t=1}^{n} 1 \left(w_{t-1} \le \gamma \right) X_{t-1} x_{t-1}' \alpha \delta' \\
&\Rightarrow E \left(1 \left(w_{t-1} \le \gamma \right) X_{t-1} \right) \int_0^1 B' \alpha \delta' \\
&= M(\gamma) e_1 \int_0^1 B' \alpha \delta',
\end{aligned}
$$

where e_1 is a p-vector with the first element 1 and the remainder 0.

Similarly,

$$
C_n^*(0) \Rightarrow M e_1 \int_0^1 B' \alpha \delta',
$$

and hence

$$
\begin{aligned}
C_n^*(\alpha, \gamma) &= C_n^*(0, \gamma) + o_p(1) \\
&= C_n(\alpha, \gamma) - M_n(\beta_0 + \alpha/n, \gamma) M_n(\beta_0 + \alpha/n)^{-1} C_n(\alpha) \\
&\Rightarrow M(\gamma) e_1 \int_0^1 B' \alpha \delta' - M(\gamma) M^{-1} M e_1 \int_0^1 B' \alpha \delta' = 0.
\end{aligned}
$$

This completes the proof. ∎

References

[1] Andrews, D.W.K. (1993), "Tests for parameter instability and structural change with unknown change point," *Econometrica*, 61, 821-856.

[2] Andrews, D.W.K. and W. Ploberger (1994): "Optimal tests when a nuisance parameter is present only under the alternative," *Econometrica*, 62, 1383-1414.

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

[4] Balke, N. S. and M.E. Wohar (1998): "Nonlinear dynamics and covered interest rate parity," *Empirical Economics,* 23, 535-559.

[5] Baum, C.F., J.T. Barkoulas, and M. Caglayan (2001): "Nonlinear adjustment to purchasing power parity in the post-Bretton Woods era," Journal of International Money and Finance, 20, 379-399.

[6] Baum, C. F. and M. Karasulu (1998): "Modelling federal reserve discount policy," *Computational Economics,* 11, 53-70.

[7] Campbell, J.Y. (1995): "Some lessons from the yield curve," *Journal of Economic Perspectives*, 129-152.

[8] Campbell, J.Y. and R.J. Shiller (1987): "Cointegration and tests of present value models," *Journal of Political Economy*, 95, 1062-1088.

[9] Campbell, J.Y. and R.J. Shiller (1991): "Yield spreads and interest rate movements: A bird's eye view," *Review of Economic Studies*, 58, 495-514.

[10] Caner, M. and B. E. Hansen (2001): "Threshold autoregression with a unit root," *Econometrica*, forthcoming.

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

[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 B. Falk (1998), "Threshold-autoregressive, median-unbiased, and cointegration tests of purchasing power parity," *International Journal of Forecasting,* 14, 171-86

[14] Dorsey, R.E. and W.J. Mayer (1995): "Genetic algorithms for estimation problems with multiple optima, nodifferentiability, and other irregular features," *Journal of Business and Economic Statistics,* 13, 53-66

[15] Gine, E. and J. Zinn (1990): "Bootstrapping general empirical measures," *The Annals of Probability*, 18, 851-869.

[16] Hansen, B. E. (1996): "Inference when a nuisance parameter is not identified under the null hypothesis," *Econometrica*, 64, 413-430.

[17] Hansen, B.E. (2000a): "Sample splitting and threshold estimation," *Econometrica*, 68, 575-603.

[18] Hansen, B. E. (2000b): "Testing for structural change in conditional models," *Journal of Econometrics*, 97, 93-115.

[19] Lo, M. and E. Zivot (2001): "Threshold cointegration and nonlinear adjustment to the law of one price," *Macroeconomic Dynamics*, forthcoming.

[20] McCulloch, J.H. and H.-C. Kwon (1993): "US term structure data, 1947-1991," Ohio State University Working Paper No. 93-6.

[21] Martens, M., P. Kofman, and T. C. F. Vorst (1998): "A threshold error-correction model for intraday futures and index returns," *Journal of Applied Econometrics*, 13, 245-263.

[22] Michael, M., R. 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.

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

[24] 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," NBER Worker Paper No. 6089.

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

[26] Pippenger, M. K., G.E. Goering (2000): "Additional results on the power of unit root and cointegration tests under threshold processes," *Applied Economics Letters*, 7, 641-44.

[27] Taylor, A. M. (2001): "Potential pitfalls for the purchasing-power-parity puzzle? Sampling and specification biases in mean-reversion tests of the law of one price," *Econometrica*, 69, 473-498.

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

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

[30] Seo, B. (1998): "Tests for structural change in cointegrated systems," *Econometric Theory*, 14, 221-258.

[31] White, H. (1980): "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica*, 48, 817-838.