Testing for Smooth Transition Nonlinearity in Adjustments of Cointegrating Systems

Milan Nedeljkovic

No 876

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

THE UNIVERSITY OF WARWICK

Testing for Smooth Transition Nonlinearity in Adjustments of Cointegrating Systems

Milan Nedeljkovic*

Department of Economics

The University of Warwick†

October 2008

Abstract

This paper studies testing for the presence of smooth transition nonlinearity in adjustment parameters of the vector error correction model. We specify the generalized model with multiple cointegrating vectors and different transition functions across equations. Given that the nonlinear model is highly complex, this paper proposes an optimal LM test based only on estimation of the linear model. The null asymptotic distribution is derived using empirical process theory and since the transition parameters of the model cannot be identified under the null hypothesis bootstrap procedures are used to approximate the limit. Monte Carlo simulations indicate a good performance of the test.

JEL classification: C12, C32.

Keywords: Nonlinearity, Cointegration, Empirical process theory, Bootstrap.

---

*I would like to thank Valentina Corradi, Mike Clements, Rodrigo Dupleich-Ulpoa, Paolo Parente, Mark Taylor and seminar participants at 2008 ESRC Econometric Study Group Conference in Bristol and at the University of Warwick for useful comments. All errors remain my sole responsibility.

†Department of Economics, The University of Warwick, Coventry, CV4 7AL, UK. Phone: +44 24 7652 8418. E-mail: milan.nedeljkovic@warwick.ac.uk.

1 Introduction

Over the past decade two strands of time series literature related to measuring comovements among nonstationary variables and incorporating nonlinear behaviour into traditional linear models have been put together, originally in Granger and Terasvirta (1993) and Balke and Fomby (1997). The combination has a sound economic grounds. Although the concept of cointegration is related to the long-run statistical equilibrium relationship between nonstationary variables and assumes linear adjustment towards the attractor set in the presence of short-run deviations, economic behaviour of agents may lead to a possibility that the adjustment process behaves differently depending on the size of deviation of the system from the equilibrium. The presence of (asymmetric) transaction costs, heterogeneity of agents, institutional rigidities may all lead to such situations. For example, in the framework of De Long, Shleifer, Summers and Waldmann (1990) with heterogeneity of agents' expectations the evidence that noise traders generally engage in momentum trading (Hong and Stain, 1999), may lead to market overreactions following the arrival of news. This in turn pushes the asset values further away from the fundamental equilibrium. In their model, Campbell and Kyle (1993) show that the larger the expected degree of arbitrage is, the faster the price response to the disequilibrium will be, implying different rates of the error-correction. On the other hand, a vast literature in macroeconomics also documents the presence of asymmetries in the money-output and unemployment-output relationships when they are modelled in univariate framework. Since the causality is expected to run in both directions the next step to investigate is whether the nonlinearity is present in the system framework.

Estimation and inference in the class of stationary smooth transition autoregression models has been thoroughly investigated and is extensively reviewed in Granger and Terasvirta (1993) and Van Dijk, Terasvirta and Franses (2001). However, little was done in the nonstationary vector framework. Popularity of smooth transition vector error correction models (STVECM) in applied studies has grown over the past decade even though the underlying statistical theory was undeveloped. Swanson (1998), Rothman, Franses and Van Dijk (2001), Chen and Wu (2005) among others, estimated the models assuming the existence of the VECM representation and employed the standard Luukonen, Saikkonen and Terasvirta (LST) (1988) type of test for nonlinearity, which was developed in stationary univariate setting. The justification in terms of stability of the Markovian representation of the model and existence of the VECM representation came recently in studies by Saikkonen (2005, 2008) and Kristensen and Rahbek (2008), but testing procedure for detecting the presence of smooth-type nonlinearity is still an open area of research. Hansen and Seo (2002) is an early contribution, although they consider threshold-type nonlinear VECM. Kapetanios, Snell and Shin (2006) proposed Lagrange multiplier (LM)-type of test which is based on the LST-type of Taylor series approximation of the nonlinear function. They derive a non-standard limiting distribution of the test, but this is achieved under very strong assumptions on the long-run and short-run exogeneity of all but one variable, hence their test is essentially single-equation oriented. Moreover, their test is based on a polynomial approximation of the transition function and statistical inference may be affected by the approximation errors depending on the true values of the nonlinear parameters (i.e., how distant the true data generating process is from the local alternative). The proposed test also does not solve the non-indentifiability of the delay parameter related to the lag value of the variable within the nonlinear function under the null hypothesis, where the standard suggestion in univariate stationary situation is to choose the value of the parameter which minimizes the overall p-value from the linearity test. However, the limiting distribution of the test statistic is obtained such that it does not take into account the nuisance character of the delay parameter and Kapetanios, Snell and Shin (2006) set the value of the delay parameter to be equal to one.

In the present paper we extend this literature by examining the LM-type test against exact specification of nonlinearity under the alternative in a more general setup. We consider the generalized model with multiple cointegrating vectors and allow for different transition functions across equations under the alternative, providing a lot of flexibility in the model design. Since the transition parameters of the model cannot be identified under the null hypothesis, this paper proposes an optimal LM test in the spirit of Hansen (1996), also allowing to select the delay parameter in a consistent way. The null asymptotic distribution is derived using empirical process theory of Doukhan, Massart and Rio (1995) and bootstrap procedures were used to approximate the limit. Monte Carlo simulations show a relatively good performance of the test even for sample sizes as small as 100 observations. In situations when the power of the test is small, the procedure can also be seen as more conservative in a model building framework since the non-rejections happen when the values of the parameters in the nonlinear function are such that the large portion of observations is within one extreme linear regime. In such instances the parameters of the nonlinear model will be estimated with a large error (for STVECM simulations see Kristensen and Rahbek, 2008), which in turn results in an inferior forecast relative to linear models. The benefits of using nonlinear models from the forecast perspective in such situations are rather small (see for example Chapters 7 and 8 in Terasvirta, 2006) and the present procedure can be seen as a useful first step in the model selection.

Another study that uses direct LM-type test for nonlinearity is Seo (2004), who considers testing for nonlinearity in systems with a single cointegrating vector and unique transition function across equations. Relative to their approach, we consider a more general model under the alternative, which results in a different formulation of the test statistic. But more importantly, we derive the statistic under a different set of assumptions (with weaker conditions on the moments of the error process, but assuming its independence), which are common in the nonlinear literature. This allows us to obtain a test with respect to a well-defined nonlinear alternative, whereas the test proposed in Seo (2004) does not have the same property since under the formulated assumptions the stability and hence the existence of the nonlinear VECM is not ensured. Simulation evidence also shows that the transformation of the nuisance space proposed in their paper does not affect the performance of the test, but that interpretation of the true value of the nuisance parameter on the basis of transformation may be .

The structure of the paper is the following. Section 2 introduces the STVEC model and discusses existing results on the stability of the model and asymptotic properties of the estimators. In section 3 the LM test is formulated, asymptotic distribution of the test is derived and methods for its approximation are discussed. Section 4 presents finite sample properties of the test. Section 5 contains an empirical application and section 6 concludes. All proofs are presented in the appendix.

A few words on notation. In the following $|\bullet|$ denotes the scalar Euclidean norm, $||\bullet||$ denotes matrix/vector Euclidean norm and $||\bullet||_p$ denotes Lp-norm. For some matrix β, $\overline{\beta} = \beta\left(\beta'\beta\right)^{-1}$ and orthogonal complement of β is defined as $\beta_\perp$. Also, to save space in all proofs unless stated specifically $Z_{t-1} = Z_{t-1}(\beta_0)$ is $(r \times 1)$ vector, $\Psi_t(\gamma) = \Psi(Z_{t-1}(\beta_0), \gamma)$ is $(k \times k)$ matrix and z_{t-1} and $\psi_t(\gamma)$ are respective scalar elements.

2 The Model

Let the $k-$dimensional discrete time vector process $\{X_t\}$ be generated on the probability space $(\Omega, \mathcal{F}, P)$ according to the following data generating process (DGP):

$$\Delta x_t = G(Z_{t-1}(\beta); \eta) + \sum_{j=1}^{p} \Phi_j \Delta x_{t-j} + \epsilon_t \tag{1}$$

or in an extended version:

$$\Delta x_t = \left[\Psi\left(\begin{bmatrix} I_r \\ \beta' \end{bmatrix} x_{t-l}, \gamma \right) d + \alpha \right] \begin{bmatrix} I_r \\ \beta' \end{bmatrix} x_{t-1} + \sum_{j=1}^{p} \Phi_j \Delta x_{t-j} + \epsilon_t \tag{2}$$

where $\Delta x_t, \epsilon_t \in R^k$ and the $(k \times r)$ cointegrating vector β^* is assumed to be uniquely normalized such that $\beta^* = \begin{bmatrix} I_r & \beta \end{bmatrix}'$ and hence $Z_{t-1}(\beta) = \begin{bmatrix} I_r & \beta \end{bmatrix}' x_{t-1}$, and $Z_{t-1}(\beta) \in R^r$. l is the delay parameter which is assumed to belong to a compact set L. We set $l = 1$ in the subsequent analysis since the simplification does not affect any of the results, but when performing the test, l should be included in the nuisance parameter space.

Further, $\Psi(Z_{t-1}(\beta), \gamma)$ is a $(k \times k)$ diagonal matrix of transition functions:

$\Psi(Z_{t-1}(\beta), \gamma) = diag\{\ \psi_1(Z_{t-1}(\beta), \gamma_1), ..\psi_k(Z_{t-1}(\beta), \gamma_k)\}$, where each ψ_j is a (Borel) measurable function: $\psi_j : R^r \to [0, 1]$, and is assumed to be a three times continuously differentiable function bounded between zero and one. For example, in the case of most commonly used exponential transition function with two cointegrating vectors we have:

$$\psi_j(Z_{t-1}(\beta),\gamma_j) = 1 - \exp\{-Z_{t-1}(\beta)' \begin{bmatrix} \gamma_{j1} & 0 \\ 0 & \gamma_{j2} \end{bmatrix} Z_{t-1}(\beta)\}, \qquad \gamma_j \in R^r. \tag{3}$$

Similarly, for the logistic function with two cointegrating vectors:

$$\psi_j(Z_{t-1}(\beta),\gamma_j) = \left(1 + \exp\{- \begin{bmatrix} \gamma_{j1} & \gamma_{j2} \end{bmatrix} Z_{t-1}(\beta)\}\right)^{-1}, \qquad \gamma_j \in R^r \tag{4}$$

The model parameters are then:

$\beta \in R^{(k-r)\times r}$; d, $\alpha \in R^{k\times r}$;$\Phi_j \in R^{p\times p}$; $\gamma \in R^{kr\times 1}$, such that parameter η defined in (1) is:

$\eta = (vec(d)', vec(\alpha)', vec(\gamma))'$, $\eta \in R^{3kr}$.

Finally, errors are assumed to be an i.i.d. process with $E[\epsilon_t] = 0$ and $E[\epsilon_t\epsilon_t'] = \Omega$, where Ω is finite, $k-$dimensional positive definite covariance matrix.

The specification in (2) and (3)/(4) shows that the model allows for a great source of flexibility since each cointegrating vector may induce a different nonlinear response in variables and the response can vary over the variables. For example, in the case of the monetary model of the exchange rate determination, which is analyzed in section 5, it can be expected that the exchange rate adjusts to disequilibrium in a symmetric manner since the agents should react in the same way to overvaluation and undervaluation situations, implying exponential type of adjustment to disequilibrium. On the other hand, changes in money supply in response to disequilibrium can be asymmetric since the overvaluation and undervaluation of the exchange rate may have different effects given the state of economy, which gives a possibility for the logistic type of adjustment. Another difference in comparison to conventional smooth transition type of models is that threshold parameter is set to zero, which reduces the dimensionality of the nuisance parameter space. This can be done since we analyze cointegrating setup where the transition variable is the lagged value of the cointegrating relationship so that transition function should (in most of the applications) be centered to the equilibrium in order to capture the effect of deviations from the equilibrium (see more in Van Dijk, Franses and Terasvirta, 2002). The cointegration space is assumed to be known and equal to r.

Stability properties of the model in (1) have been recently investigated by Saikkonen (2005, 2008) and Kristensen and Rahbek (2008). They show that under mild assumptions on boundness and asymptotic domination of the nonlinear functions $\psi(\bullet)$ and existence of moments of the random process $\{\epsilon_t\}$, the stochastic process $Y_t = (X_t'\beta^*, \Delta X_t'\beta_\perp^*, ...X_{t-p+1}'\beta^*, \Delta X_{t-p+1}'\beta_\perp^*)'$is geometrically ergodic and consequently there exist a set of initial values of Y_t such that the process Y_t is strictly stationary and absolutely regular with geometrically decaying mixing numbers. The sufficient condition for stability is expressed in terms of the joint spectral radius $\rho(M_1, M_2)$, where $(kp \times kp)$ matrices M_1, M_2 are formed from transformation of the characteristic polynomials from both extreme (linear) regimes, see Saikkonen (2008):

$$M_1(L) = \left[(1-L)\left(I_k - \textstyle\sum_{j=1}^p \Phi_j L^j\right)\overline{\beta}^* - \alpha L : (1-L)\left(I_k - \textstyle\sum_{j=1}^p \Phi_j L^j\right)\beta_\perp^*\right]\begin{bmatrix} \beta^{*\prime} \\ \beta_\perp^{*\prime} \end{bmatrix}$$

$$M_2(L) = \left[(1-L)\left(I_k - \textstyle\sum_{j=1}^p \Phi_j L^j\right)\overline{\beta}^* - (\alpha+d)L : (1-L)\left(I_k - \textstyle\sum_{j=1}^p \Phi_j L^j\right)\right]\begin{bmatrix} \beta^{*\prime} \\ \beta_\perp^{*\prime} \end{bmatrix}.$$

The stability condition is that $\rho(M_1, M_2) < 1$.

Kristensen and Rahbek (2008) have derived asymptotic theory for the likelihood estimators of STVEC models. They prove under mild restrictions, which are satisfied by the transition functions of the type under analysis, that the estimated parameter $\widehat{\beta}$ is superconsistent in all directions apart from one where it is $T^{\frac{3}{2}}$, whereas estimates of all other parameters are $\sqrt{T}$-consistent. In addition, they showed that the short-run and long-run parameters are not asymptotically orthogonal unless the sufficient (orthogonality) condition holds. The condition will be satisfied provided that the series are demeaned prior to estimation, as in De Jong (2002).

The maximum likelihood (ML) estimates can be obtained as follows. Let the set of all parameters in (1) be defined as $\vartheta = (\theta, d, \beta, \gamma, \Omega)$, where $\theta = (\alpha, \Phi)$.The log-likelihood (abstracting from constant term) is given as:

$$l(\vartheta) = -\frac{T}{2}\log|\Omega| - \frac{1}{2}\sum_{t=1}^{T}\epsilon_t(\vartheta)'\,\Omega^{-1}\epsilon_t(\vartheta) \tag{5}$$

ML estimation of STVECMs is non-standard due to the presence of cointegrating vector within the transition function and due to a well-documented problem with estimation of the transition parameters γ, see Van Dijk, Franses and Terasvirta (2002). The problem arises since the estimate of γ tends to be inflated and convergence of the likelihood can be problematic if it is maximized with respect to all parameters. To reduce dimensionality, the MLE estimates can be obtained using either the grid search over the set of possible values of (β, γ) as in Hansen and Seo (2002) or using the direct search methods such as the genetic algorithm or the simulated annealing, and then other parameter will be obtained as a simple OLS estimates for the fixed (β, γ). The procedure is repeated until convergence is achieved. Finally, note that even though we consider the LM statistic which is based on the restricted linear VECM, the limiting distribution of the test statistic is derived assuming the existence of the stable STVECM under the alternative and the stability conditions should be checked prior to any further inference on STVECM following the rejection of linearity.

3 Testing Procedure

The question of interest is whether the nonlinear part of the model is significant, which can be formulated as H_0: d=0 in equation (2). However, under the null hypothesis the transition parameter γ will be unidentified and likelihood function will be flat. As shown in Davies (1987), Andrews and Ploberger (1994) and Hansen (1996) the classical likelihood ratio (LR), LM and Wald statistics will not have the usual chi-square asymptotic null distribution in such a case. In order to find the limiting distribution the general idea is to treat some statistic $Q_n(\gamma)$ as a random function on the space $\gamma \in \Gamma$ and then use a functional $h(Q_n)$ and the continuous mapping theorem (CMT) to obtain the limit under the null, where $h(\bullet) : \Gamma \to R$ is a random function on Γ, continuous with respect to the uniform metric and monotonic in the sense that if some $Y_1(\gamma) < Y_2(\gamma)$, then $h(Y_1) < h(Y_2)$ and if $Y(\gamma) \to \infty$ for some subset of Γ with a positive measure μ, then $h(Y) \to \infty$. Davies proposed using $\sup_{\gamma\in\Gamma} Q_n(\gamma)$ as $h(\bullet)$, whereas Andrews and Ploberger suggested using average and exponential versions $\int_\Gamma Q_n(\gamma)dJ(\gamma)$ and $\ln(\int_\Gamma \exp(\frac{1}{2}Q_n(\gamma))dJ(\gamma))$ for some chosen integrable weight function of the values of γ (see also Altissimo and Corradi, 2002 for further extensions). In the present paper we apply their approach and provide conditions under which the limiting distribution is derived for the STVEC models.

The standard LM statistics is defined as:

$$LM_T(\gamma) = vec(\widetilde{s_T}(\gamma))' \left(\widetilde{V}_T(\gamma)\right)^{-1} vec(\widetilde{s_T}(\gamma)) \tag{6}$$

where $\widetilde{s_T}(\gamma)$ is the (concentrated with respect to γ) score function evaluated at the null hypothesis with the covariance matrix $\widetilde{V}_T(\gamma)$. The score[1] vector $s_T(\gamma)_{(1x(k-r)r+(pk+2r)k\ )}$ for model in (1) is defined as:

$$\begin{aligned}
s_{T,\beta}(\gamma)_{(1x(k-r)r\ )} &= \frac{1}{T}\sum_{t=1}^{T}\epsilon_t'\Omega^{-1}\left[\begin{array}{c}\left(\Psi(Z_{t-1}(\beta),\gamma)d+\alpha\right)\left(x_{2,t-1}'\otimes I_r\right)+\\ \left(Z_{t-1}'(\beta)d'\otimes I_k\right)\left(x_{2,t-1}'\otimes\frac{\partial vec(\Psi(Z_{t-1}(\beta),\gamma)}{\partial Z}\right)\end{array}\right]\\
s_{T,\Phi_j}(\gamma)_{(1xk^2)} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_t'\Omega^{-1}(\Delta x_{t-j}'\otimes I_k)\\
s_{T,\alpha}(\gamma)_{(1xkr)} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_t'\Omega^{-1}\left(Z_{t-1}'(\beta)\otimes I_k\right)\\
s_{T,d}(\gamma)_{(1xkr)} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_t'\Omega^{-1}\left(Z_{t-1}'(\beta)\otimes\Psi(Z_{t-1}(\beta),\gamma)\right)
\end{aligned}$$

where the chain rule for matrix differentiation is used to obtain $s_{T,\beta}(\gamma)$ and vector of variables x_t is partitioned as $x_t = [x_{1,t}' x_{2,t}']'$ according to the normalization of the cointegrating vector.

Evaluated at the null hypothesis the only non-zero part of the score $\widetilde{s_T}(\gamma)$ will be:

$$\widetilde{s_T}(\gamma) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{\epsilon}_t'\widetilde{\Omega}^{-1}\left(Z_{t-1}'(\widetilde{\beta})\otimes\Psi(Z_{t-1}(\widetilde{\beta}),\gamma)\right) \tag{7}$$

We impose the following assumptions:

Assumption 1:

1. The k-dimensional discrete time vector process $\{\epsilon_t\}$ is i.i.d. with $E[\epsilon_t] = 0$ and $E[\epsilon_t\epsilon_t'] = \Omega$, where Ω is a finite, k-dimensional positive definite covariance matrix. The marginal distribution has a (Lebesgue) density which is bounded away from zero on compact subsets

[1] Note that the asymptotics of Ω are standard and does not influence the asymptotics of the mean equation parameters so that for the sake of clarity we will not consider part of the score with respect to this parameter.

of R^k and for all moments up to $4+\delta$, $\sup_t E\left[||\epsilon_t||^{4+\delta}\right] < \infty$ for some $\delta > 0$;

2. $\{Z_t, \Delta x_t\}$ is a vector sequence of strictly stationary absolutely regular random variables with geometrically decaying mixing numbers: $\beta_n = O(b^n)$, for some $0 < b < 1$;

3. Let (θ_0, β_0) be the true parameter value under the null for which the likelihood is uniquely maximized. Then, under the null hypothesis $\left\|(\widetilde{\theta}-\theta_0)\right\|$ and $\left\|\sqrt{T}(\widetilde{\beta}-\beta_0)\right\|$ are $op(1)$;

4. $\psi_j(Z_t, \gamma_j)$ is three times continuously differentiable in γ, Z and has bounded partial derivatives such that: $\sup_{\gamma\in\Gamma^*} |\psi_j(Z_t, \gamma_j)| \in [0,1]$, $\sup_{\gamma\in\Gamma^*} ||\nabla_\gamma \psi_j(Z_t, \gamma_j)|| \leq c||Z_t||$, $\sup_{\gamma\in\Gamma^*} ||\nabla_{\gamma\gamma'} \psi_j(Z_t, \gamma_j)|| \leq c||Z_t||$, $\sup_{\gamma\in\Gamma^*} || \nabla_Z \psi_j(Z_t, \gamma_j)|| \leq c||Z_t||$ for different constants c;

5. $V(\theta, d, \beta, \gamma) = U_T \sum_{t=1}^{T} E\left[s_T(\gamma)s_T(\gamma)'\right] U_T'$ has the smallest eigenvalue bounded away from zero in probability on a compact sets $\Theta_0 \times 0 \times B_0$ and $\Theta_0 \times d_0 \times B_0$ for all $\gamma \in \Gamma^*$, where $\beta_0 \in B_0$ and U_T is diagonal normalization matrix containing convergence rates.

6. $\frac{1}{\sqrt{T}} x_{[Ts]}$ converges weakly to a vector Brownian motion W^* with a covariance matrix Σ^*, which has rank equal to $p-r$. Further, $\frac{1}{\sqrt{T}}\ x_{2,[Ts]}$ converges weakly to a vector Brownian motion W with a positive definite covariance matrix Σ.

All the assumptions will be satisfied under both hypothesis. Assumption 1.1. is common in the literature on stability of the nonlinear VECM. It allows x_t to be embedded in a Markov chain which can be shown to be geometrically ergodic under the joint spectral radius condition. The condition translates to the standard eigenvalue condition on the characteristic polynomial for the linear VECM. Assumption 1.2 follows from the stability of the Markov chain, as shown in Doukhan (1994) and Saikkonen (2005, 2008). Moreover, stability of the Markov chain together with Assumption 1.1. implies the existence of moments of the processes $\{Z_t, \Delta x_t\}$ (Theorem 5 in Feigin and Tweedie, 1985, see also Kristensen and Rahbek, 2008) - $\sup_t E\left[||Z_t||^{4+\delta}\right] < \infty$ and $\sup_t E\left[||\Delta x_t||^{4+\delta}\right] < \infty$. Assumption 1.3. follows from the standard linear cointegration literature since the VECM estimates under the null do not depend on γ. Assumption 1.4. will always be satisfied by logistic and exponential functions given in equations (3) and (4). The definition of the Eucledian (Frobenius) norm implies that $\sup_{\gamma\in\Gamma^*} \|\Psi(Z_t,\gamma)\| \in [0, r]$ and $\sup_{\gamma\in\Gamma^*} \|\nabla_\gamma \Psi(Z_t,\gamma)\|$,

$\sup_{\gamma\in\Gamma^*} \|\nabla_z \Psi(Z_t,\gamma)\|$ are $O(\|Z\|)$. Assumption 1.5. implies that the limit of the covariance matrix of the score under the null for normalized cointegrating vector is positive definite, which under suitable conditions specified bellow ensures that $\widetilde{V}_T(\gamma)$ is positive definite with probability approaching 1. Under the null, the assumption follows from the results in linear cointegration literature and imposes only an additional requirement that the diagonal elements of $\Psi(Z_t,\gamma)$ are not equal to zero in at least one point in time. Under the alternative, the results hold on the basis of Theorem 5 in Kristensen and Rahbek (2008) under appropriate scaling. Assumption 1.6. under the alternative follows from Theorem 3 in Saikkonen (2005) given that nonlinearity is present only in the adjustment parameters and here we assume that the cointegrating vector is uniquely normalized. Note that Assumptions 1.4. and 1.5. are imposed on Γ^*, which is a compact subspace of Γ. This is required in order to prove the tightness of the finite dimensional distribution of the score since contrary to threshold models the nuisance parameter space is infinite. As discussed in Hansen (1996), the validity of the assumption in applications depends on whether Γ^* is sufficiently dense in Γ, otherwise the power of the test may be significantly affected. Seo (2004) proposed a one-to-one transformation of γ, which yields a compact subspace of the nuisance parameter by definition (although the transformation implicitly restricts to subspace $\gamma \in [0.5,19]$). The simulation evidence, however, does not show any gain from using the aforementioned transformation. On the other hand, if the transformation is applied following the rejection of linearity, then the estimated values γ are interpreted in the wrong way, since, as it will be shown in the simulation section, the exponential-type nonlinearity is closely approximated by linear model for values of γ larger than 2 and even with the transformation the power of the test remains weak.

In order to obtain the asymptotic distribution of the test statistic, weak convergence of the (normalized) score vector under the null hypothesis and uniform convergence of the covariance matrix $\widetilde{V}_T(\gamma)$ over $\gamma \in\Gamma^*$ are first established.

To establish weak convergence of $s_T(\gamma)$, the score is seen as an empirical process indexed by $\gamma \in\Gamma^*$. Sufficient conditions for convergence of the process (Andrews, 1994) are:

i. Finite dimensional convergence (fidi) holds - for each finite subset $(\gamma_1, ...\gamma_j)$ of Γ^*, $\widetilde{s_T}(\gamma_1)....\widetilde{s_T}(\gamma_j)$

converge in distribution;

ii. For totally bounded (pseudo) metric space (Γ^*, ρ), stochastic equicontinuity holds: For all $\varphi > 0$,

$$\lim_{\tau \to 0} \lim \sup_{T \to \infty} P\left(\sup_{\rho(f_1, f_2) < \tau} \|s_T(f_1) - s_T(f_2)\| > \varphi \right) = 0 \tag{8}$$

for some $f \in \mathcal{F}$, where $\mathcal{F}$ is a class of functions to be restricted later.

The proof of the weak convergence is based on Doukhan, Massart and Rio (DMR) (1995)'s invariance principle for $\beta-$ mixing processes. The proof is essentially related to verifying stochastic equicontinuity since finite dimensional convergence readily holds from CLT for $\beta-$ mixing processes. First, we provide some preliminary results which will be employed to verify conditions of DMR's Theorem 1.

Proposition 1 *Under Assumptions 1.1, 1.2 and 1.4 the following results hold:*

1. $E\left[sup_{\gamma \in \Gamma^*} ||\epsilon_t' \Omega^{-1} \left(Z_{t-1}' \otimes \Psi(Z_{t-1}, \gamma)\right)||^{2+\delta}\right] < \infty$

2. $\sup_t \left(E \sup_{\gamma_1 \in \Gamma^* : ||\gamma_1 - \gamma|| < \varpi} ||\epsilon_t' \Omega^{-1} \left(Z_{t-1}' \otimes (\Psi(Z_{t-1}, \gamma_1) - \Psi(Z_{t-1}, \gamma))\right)||^{2+\delta}\right)^{\frac{1}{2+\delta}} < C\varpi^{\upsilon}$, *for each* $\gamma \in \Gamma^*$, *some small positive constants* C *and* υ *and for any* $\varpi > 0$

Given the results of the proposition above, the following proposition characterizes the limiting behaviour of the score under the null.

Proposition 2 *Under Assumptions 1.1, 1.2 and 1.4 and under the null hypothesis:*

$vec(\widetilde{s_T}(\gamma)) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(Z_{t-1}(\widetilde{\beta}) \otimes \Psi(Z_{t-1}(\widetilde{\beta}), \gamma)\right) \Omega^{-1} \epsilon_t \Rightarrow S(\gamma)$, *where* $S(\gamma)$ *is the Gaussian process indexed by* γ *with covariance matrix* $V(\gamma)$ *and a.s. uniformly continuous sample paths.*

Uniform convergence of the covariance matrix $\widetilde{V}_T(\gamma)$ over $\gamma \in \Gamma^*$follows by showing the point-wise weak law of large numbers and stochastic equicontinuity of $V_T(\gamma)$ over the set $\gamma \in \Gamma^*$and verifying the uniform convergence of the sample analog $\widetilde{V}_T(\gamma)$.

Proposition 3 *Under Assumptions 1.1 - 1.6 and under the null:*

$\sup_{\gamma \in \Gamma^*} \left|\left|\widetilde{V}_T(\gamma) - V(\gamma)\right|\right| = op(1)$

Combining the results from Proposition 2 and 3 it follows that under the null hypothesis:

$$LM_T(\gamma) \Longrightarrow (V(\gamma)^{-\frac{1}{2}}S(\gamma))' \left(V(\gamma)^{-\frac{1}{2}}S(\gamma)\right) = G(\gamma)'G(\gamma)$$

This implies that by the continuous mapping theorem any functional $h(LM_T(\gamma))$:

$$h(LM_T(\gamma)) \Longrightarrow h(G(\gamma)'G(\gamma)),$$

where $G(\gamma)'G(\gamma)$ is a chi-squared process for each γ.

Since the limiting process depends on the underlying distribution of the data $\{X_t\}$ and its moments through the covariance $V(\gamma)$, the asymptotic distribution cannot be tabulated and critical values need to be obtained through simulations. Hansen (1996) provided a proof of asymptotic validity of conditional transformation of the test statistic $h(LM_T(\gamma))$, which yields an approximation of the asymptotic p-values of the statistic via bootstrap. We employ residual bootstrap to approximate the sampling distribution of the statistic. This is in line with the assumption that the errors are i.i.d. and the procedure should perform well if the actual data is in fact homoscedastic, as is commonly assumed in the nonlinear VECM literature (Kristensen and Rahbek, 2008).

However, if the true DGP is subject to conditional heteroscedasticity, the approximation may be adversely affected and in such cases the use of a heteroscedasticity robust statistic is recommended. The standard residual bootstrap is replaced by the fixed design wild bootstrap presented in Hansen (1996), Kreiss (1997) and Goncalves and Kilian (2004). The first order asymptotic validity of the bootstrap is proven under assumptions of the present model in Theorem 4.2. in Kreiss (1997, pp. 8). Note that the presence of ARCH type of conditional heteroscedasticity is admissable for the presented test statistic provided that additional assumptions are imposed.

Assumption 2:

1. The k-dimensional discrete time vector process $\{\epsilon_t\}$ is i.i.d. with $E[\epsilon_t] = 0$ and $E[\epsilon_t \epsilon_t'] = \Omega$, where Ω is finite, k-dimensional positive definite covariance matrix. The marginal distribution has a (Lebesgue) density which is bounded away from zero on compact subsets of R^k and for all moments up to $8 + \delta$, $\sup_t E\left[||\epsilon_t||^{8+\delta}\right] < \infty$ for some $\delta > 0$;

2. The conditonal variance Ξ_t satisfies the condition $\|\Xi_t\| \leq c\,\|\epsilon_t\|^2$, for a generic constant c.

Assumption 2.1 strengthens Assumption 1.1. requiring boundness of the unconditional moments of the error process of order slightly higher than eight, which is a common assumption in the literature on bootstrapping the conditionally heteroscedastic models. Assumption 2.2. restricts the class of permitted conditionally heteroscedastic models such that the model under alternative exists (Saikkonen, 2005, 2008), which will be satisfied for example by any BEKK (Engle and Kroner, 1995) ARCH formulation (an example is given in Saikkonen, 2007). The heteroscedasticity robust statistic is obtained by correcting the score vector in (7):

$$\widetilde{s_T}(\gamma) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \widetilde{\epsilon}_t' \left(Z_{t-1}'(\widetilde{\beta}) \otimes \Psi(Z_{t-1}(\widetilde{\beta}), \gamma) \right) - \widetilde{\epsilon}_t' \widetilde{H}_t \left(\widetilde{H}_t' \widetilde{H}_t \right)^{-1} \widetilde{H}_t' \left(Z_{t-1}'(\widetilde{\beta}) \otimes \Psi(Z_{t-1}(\widetilde{\beta}), \gamma) \right),$$

where: $\widetilde{H}_t = \begin{bmatrix} Z_{t-1}'(\widetilde{\beta}) \otimes I_k & \Delta x_{t-1}' \otimes I_k \end{bmatrix}$.

4 Finite sample properties

In this section the finite sample performance of the statistic is examined via a Monte Carlo experiment. The performance of three statistics - $sup_{\gamma \in \Gamma^*} LM_T(\gamma)$, $ave_{\gamma \in \Gamma^*} LM_T(\gamma)$ and $\exp_{\gamma \in \Gamma^*} LM_T(\gamma)$ is investigated using 1000 repetitions and 200 bootstrap replications within each experiment. The number of bootstrap replications should be higher in practice, but the choice of 200 was made in order to save on computation time. We restrict the nuisance parameter space to $[0, 50]$,[2] and obtain the statistic via grid search with 60 grid points. The experiment is performed for sample sizes T=100 and T =250. Sup statistic can also be obtained using simulated annealing procedure instead of the grid search, but assessing the merits of such approach is beyond the current scope of the paper.

First, the nominal size of the test is analyzed. We generate two-variable VECM(1) for simplicity:

$$\Delta x_t = \alpha \begin{bmatrix} 1 \\ \beta \end{bmatrix} x_{t-1} + \Phi \Delta x_{t-1} + \epsilon_t, \text{ with } \Phi = \begin{bmatrix} 0.5 & -0.2 \\ 0 & 0.5 \end{bmatrix}.$$

We set $\beta = -1, \alpha_2 = 0.2$ and vary α_1 between (-0.1,-0.4,-0.7).

The errors are assumed to be homoscedastic first and drawn from the independent $N(0, 1)$ distribution. Since the model is rich in possibilities the nominal size is examined against all possible specifications - ESTR (exponential smooth transition) and LSTR (logistic smooth transition) nonlinearity in both variables as well as against the specification that contains ESTR adjustment

[2] We have also experimented as in Seo (2004) with restricting the grid search over the set $[0.05, 1]$ with 60 grid points and the obtained results were qualitatively similar to the ones presented here.

in one and LSTR-type adjustment in the second variable.

P-values are obtained by a simple algorithm where the linear VECM is first estimated and new samples of data x_t^B are constructed using the parameter estimates, resampled residuals and initial values. The test statistic is calculated for each resampled data and p-values are obtained as the percentage of times the bootstrap statistics exceeds the sample statistic.

The results in Table 1 show satisfying properties of the test since the simulated p-values are close to its nominal size for both samples. There are no clear differences between the specifications, although the test tends to overreject slightly against both the models for high values of the linear adjustment parameter. *Sup* statistic also slightly overrejects relatively to other two statistics in the case of exponential model.

Next, the power of the test is examined by specifying the following simple two-variable DGP:

$$\Delta x_t = \left[\Psi \left(\begin{bmatrix} 1 \\ \beta \end{bmatrix} x_{t-l}, \gamma \right) d + \alpha \right] \begin{bmatrix} 1 \\ \beta \end{bmatrix} x_{t-1} + \Phi \Delta x_{t-1} + \epsilon_t, \;\; \Phi = \begin{bmatrix} 0.54 & -0.45 \\ -0.4 & 0.45 \end{bmatrix}.$$

where diagonal elements of $\Psi(\bullet)$ are given in equations (3) and (4). The errors are again first assumed to be homoscedastic and drawn from the independent $N(0,1)$ distribution. We fix $\alpha = \begin{bmatrix} -0.2 & 0 \end{bmatrix}'$, $\beta = -1$, and reported elements of Φ were chosen randomly such that the stability condition is satisfied for the choice of d.[3] The nonlinear adjustment parameter in second equation is set to ($d_2 = 0.3$) and d_1 varies between (-0.1,-0.4,-0.6). The experiment is designed to capture the power of the test in a less than ideal situation when the adjustment parameters in one (extreme) regime are small and the adjustment of one of the variables in the second extreme regime is moderate. Transition parameter γ_1 vary between $(0.1, 0.4, 0.7, 1, 2)$ and we set $\gamma_2 = 0.2$. We do not consider higher values of γ since they imply that a large portion of data is (or very close) to one of the extreme regimes, which affects negatively the power of the test and the general intuition can be inferred from the present choice. The size is set at 5%.

The results from Table 2a and Table 2b show distinct behaviour between the logistic and exponential specification. The test against ESTR-type nonlinearity has non-negligible power, which increases when the nonlinearity is stronger (d_1 is higher), but it falls down with the increase

[3] Results in Seo (2004) were obtained without taking into consideration the stability of the underlying system.

in the speed of transition γ_1. This is a consequence of the model specification since the higher value of γ_1 implies faster transition between the regimes and more data points at the extremes. Unless the sample size is large, this will mask the presence of the two regimes' nonlinearity. For example, when γ is set to $\gamma = 0.7$, on average roughly 70% of observations of $\psi(Z_{t-1}, \gamma)$ are very close to 1, implying that the system is close to the extreme linear regime. On the other hand, the logistic specification by definition implies a slower rate of transition, although it may suffer from identification problems on both ends, which explains the loss of power when $\gamma \rightarrow 0$ and/or when γ increases. As γ increases, the LSTR model becomes closer to threshold specification, where the data is concentrated at two extremes, although this happens at a much slower rate relative to ESTR (for $\gamma = 0.7$, 20 % of observations are close to the extremes). However, contrary to ESTR, the LSTR model is subject to identification problems when γ is close to zero since the logistic function becomes flat and concentrated around 1/2, implying the existence of one regime. This in turn implies that identification of LSTR models and hence ability to direct power against them also requires a larger number of data points (although the problem is less severe relative to ESTR), which was also documented in Kristensen and Rahbek (2008).

Four more points are noticeable from the simulation study. First, the power of the test increases with the sample size and degree of nonlinearity (d_1 higher), as expected. Second, the $aveLM_T(\gamma)$ and $\exp LM_T(\gamma)$ formulations perform better than the *sup* test for ESTR specification as γ_1 increases. This is expected since as discussed in Andrews and Ploberger (1994) the former tests direct power towards alternatives closer to the null, whereas the power of the *sup* test is with respect to the distant alternatives. Since the ESTR model for larger values of γ becomes closer to the linear specification, the former tests are expected to have a higher power. Third, the test performs relatively well in choosing the correct nonlinear specification. The additional experiment is performed to illustrate this property of the test. The parameters of the model remain the same, but the simulated model is now tested against all four possible specifications (ESTR-ESTR, LSTR-LSTR,LSTR-ESTR,ESTR-LSTR) and the number of times the correct specification is selected given the rejection of linearity is reported in Table 2c. As we can see, the test performs well when the generated model is logistic adjustment in both variables as well as when different adjustments across the variables are present, selecting the correct type of nonlinearity more than 90% of times. The exponential adjustment is correctly selected in at least 80% of cases, where the selection frequency decreases with the weaker evidence of nonlinearity and/or with the increase in the speed of transition. This is expected since an ESTR models can be well approximated by an LSTR when a large portion of observations is close to the upper (extreme) regime since then only the increasing part of the transition function becomes relevant and hence closer to the monotonically increasing logistic function. However, as discussed in previous section, the power of the test in such situations is lower, decreasing the probability of selecting the LSTR model. The experiment thus implies that in practice one should be relatively confident to run the test under different alternatives and to choose the one which minimizes the p-value of the test. Fourth, the presence of conditional heteroscedasticity in one of the series does not significantly deteriorate the size of the test. The simulation results available from the author show that the size of the test remains close to the reported one in Table 1 for the range of ARCH(1) specifications in one of the residual series.

Next, we investigate the test properties under conditional heteroscedasticity in both residual series. The mean parameters remain unchanged and we specify the conditional variance as a simple diagonal BEKK ARCH (1) process:

$\Omega_{t|t-1} = A_0' A_0 + A_1' \epsilon_{t-1} \epsilon_{t-1}' A_1,$

where the parameter matrices are set as $A_0 = \begin{bmatrix} 1 & 0 \\ 0.5 & 1 \end{bmatrix}$ and $A_1 = \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$ and we vary $a \in [0.2, 0.5, 0.9]$. We perform the experiment over the slightly restricted set of parameter values (d_1, γ_1) compared to the previous experiment due to high dimensionality, but the results are general enough to capture the main intuition. We consider the sample size T =250, since it can be expected that for smaller samples in practice ARCH will not be present (corresponding to quarterly frequency). P-values are obtained by resampling the estimated residuals from the linear VECM and creating a new set of residuals $\epsilon_t^B = \widetilde{\epsilon}_t \xi_t$, where $\xi_t \backsim N\ (0,1)$. The bootstraped residuals are used to construct a new series of data Δx_t^B given the estimated parameters from the linear model, but keeping the observations $\Delta x_{t-1}, \widetilde{Z}_{t-1}$ fixed. The bootstrapped estimates are obtained by regressing Δx_t^B on $\Delta x_{t-1}, \widetilde{Z}_{t-1}$ and the test statistic is calculated using the boostrapped estimates for each resampled data where p-values are again obtained as the percentage of times the bootstrap statistics exceeds the sample statistic.

The results in Table 3 and 4 show that the robust LM statistic possess good properties in the presence of multivariate ARCH volatility. When the level of volatility persistence is moderate, the empirical size is close to the significance level for exponential model, but it tends to slightly underreject the null hypothesis in the case of the LSTR model. On the other hand, for large values of $a = 0.9$, size distortions are larger in the case of the ESTR model, where the *sup* statistic tends to overreject more relative to other two statistics.

The power of the test remains slightly affected relative to the white noise case. The power to reject linearity in favour of LSTR nonlinearity remains high and decreases slowly only when the nonlinearity parameters are small ($d_1 = 0.1$). The power against exponential nonlinearity decreases more with the level of persistence in ARCH specification. The *sup* statistic dominates other two statistics in terms of power against the ESTR, although the difference decreases as the nonlinearity becomes stronger and transition between the regimes faster, similar to the i.i.d. case.

5 An Application

The presence of smooth type nonlinearity in behaviour of the real exchange rate was heavily investigated over the past decade, as documented by Taylor and Taylor (2004). However, little was progressed in examining whether the nonlinearity is present in the relationship between the nominal exchange rate and macro fundamentals and whether this can help in solving a long-standing puzzle in international economics related to the lack of strong empirical evidence on the relationship between floating exchange rates and a set of underlying macroeconomic variables. Even though theories of the exchange rate determination imply that the exchange rate is determined by such fundamental variables, the empirical evidence on the existence of the relationship and/or predictability of the future changes in exchange rate on the basis of fundamentals is still scarce (Cheung, Chinn and Garzia-Pasqual, 2005).

One possibility is that the deviations of the exchange rate from the fundamentally determined level may be governed by a nonlinear adjustment process which is difficult to capture using linear methods. Thus, even though a stable long-run relationship exists, the short-run dynamics will not be correctly modeled, which will affect the obtained results. It is well-known that different types of nonlinearity may arise in the foreign exchange market due to heterogeneity of agents. Overall, when the exchange rate is close to its fundamental value the influence of technical analysis and other types of trading strategies dominates the market. Conversely, when the exchange rate becomes increasingly misaligned with the fundamentals, the pressure from both policy makers and agents for returning the exchange rate to the neighborhood of the long-run level becomes stronger and eventually dominates the market. Taylor and Peel (2000) found significant evidence of nonlinearity in the deviations from the monetary fundamental equilibrium, showing that the exchange rate follows near-unit process for small deviations (implying lack of cointegration found in previous studies), but fast mean-reversion for large departures from the long-run equilibrium. They used a LST type of the test applied on the implied cointegration residual and their results are univariate. Given the level of interrelations in macro series, imposing a priori exogenenity may seem as a strong assumption. We, therefore, consider testing for the presence of nonlinearity in the system framework, providing at the same time the robustness check of the above results.

We consider a classical monetary model of the exchange rate determination, which implies the existence of the relationship between the exchange rate, money differential and output differential between two countries. Monthly data on the nominal exchange rate, monetary aggregate and industrial production for the United States, the United Kingdom and Japan is collected from the IMF's International Financial Statistics for the period 1980-2006. The sample includes 324 observations.

All series are expressed in logarithmic terms and they are first checked for a unit root, which, as expected, shows evidence of the nonstationarity. The VAR is then fit to the data, where the number of lags is computed using AIC criterion, which implied choosing 4 lags for the United Kingdom and 5 lags for Japan.[4] Johansen's cointegration test finds (weak) evidence of the unique cointegrating vector[5] for both countries.

[4] Multivariate LM tests were also performed to ensure that VAR residuals are indeed white noise.

[5] The results for both unit root tests and Johansen's cointegration test are not reported in order to save space, but are available from the author.

We then apply the test for nonlinearity. We allow for the presence of both types of "usual suspects" nonlinearity (exponential and logistic) in all variables. Although the logistic specification is less likely to be present in the exchange rate adjustment due to a lack of theoretical underpinnings as discussed in section 2, we still consider this as an additional check of both model and test properties. We consider all possible combinations within the system yielding in total 8 different alternatives and results are obtained through the grid search with 60 grid points over the nuisance transition parameter space to $[0, 50]$, and over the delay parameter set between $[1, 12]$. The p-values are obtained using 1000 i.i.d. bootstrap replications. On the basis of simulation results, the non-robust statistic was employed given that multivariate ARCH LM test did not find an evidence of ARCH effects and univariate ARCH LM tests showed some evidence of conditional heteroscedasticity only in the money differential series.

Table 5 shows the p-values from performing all three tests. We only report the minimum p-value per each specification. The number in brackets corresponds to the lagged value of the transition variable (cointegrating relationship) for which the minimum is attained. As we can see, linearity is rejected for the monetary model of the British pound using three different specifications, but the strongest rejection is obtained for the ESTR type nonlinearity in all variables. Following the evidence from the simulation study, we can be relatively confident to conclude that the previously obtained result by Taylor and Peel (2000) are confirmed in a more direct way, but, in addition, the results imply that macro fundamentals also respond in a nonlinear manner to the disequilibrium in the long-run relationship from 10 months before, where the response in monetary and real variables depends on the magnitude of over/under pricing of the exchange rate. These links can be exploited further to investigate whether the evidence of nonlinearity can be used to improve the forecast of the exchange rate and, at the same time, whether this also helps in predicting future levels of fundamentals, adding to recently proposed theoretical approach of Engel and West (2005). On the other hand, no evidence of nonlinearity is found for the Japanese yen, implying that the nonlinear response in variables is not general and hence the benefits of employing the nonlinear analysis for forecast purposes may differ on a country basis. The raised issues warrants further analysis which is beyond the scope of the paper and is left for future research.

6 Conclusions

In this paper we have proposed a direct test for the presence of nonlinearity in the adjustment terms of the cointegrating systems that is consistent under the nonlinear alternative. The critical values are obtained via bootstrap simulations. Monte Carlo simulations show encouraging results in the presence of homoscedasticity and certain types of heteroscedasticity, but future work is required to incorporate the broader volatility models, both in terms of the statistical representation of the nonlinear system and for testing against nonlinearity in such framework.

Another possible extension in both directions is to allow for the richer dependence structure in the errors of the process. This does not represent a serious deficiency of the current testing procedure since choosing the appropriate VAR order will "whiten" the errors. However, once the progress is made in terms of the Markovian representation of the system under such an assumption, the changes in the testing procedure would be straightforward using the battery of asymptotic results for dependent (but not necessarily absolutely regular with geometrically decaying mixing numbers) processes. The robust type of tests following White (1994) could also be considered in such framework.

Relaxing the assumption of the known number of cointegrating vectors and testing directly for it is another fruitful area of research. Some progress has been made for threshold vector error correction models by Gonzalo and Pitarakis (2006) who proposed the model selection based approach for determining the rank of the system. In the present context this also requires the "full representation" proof on the number of nonstationary elements, which is available from Saikkonen (2005) under the assumption that the nonlinearity is present only in the adjustment term.

Appendix I: Results:

Table 1: The size of the test:

Model\ α_1		-0.1			-0.4			-0.7		
	Size	sup	ave	exp	sup	ave	exp	sup	ave	exp
T=100:										
ESTR 2	**0.05**	0.063	0.052	0.049	0.060	0.054	0.057	0.069	0.073	0.072
ESTR 2	**0.10**	0.118	0.103	0.101	0.117	0.111	0.111	0.124	0.132	0.126
LSTR 2	**0.05**	0.060	0.057	0.059	0.042	0.042	0.042	0.040	0.041	0.039
LSTR 2	**0.10**	0.104	0.100	0.101	0.095	0.088	0.090	0.082	0.080	0.080
Estr-Lstr	**0.05**	0.047	0.048	0.045	0.045	0.047	0.044	0.044	0.040	0.041
Estr-Lstr	**0.10**	0.111	0.108	0.106	0.122	0.115	0.112	0.105	0.102	0.106
T=250:										
ESTR2	**0.05**	0.055	0.047	0.054	0.064	0.055	0.055	0.072	0.054	0.053
ESTR2	**0.10**	0.108	0.102	0.102	0.122	0.120	0.117	0.128	0.121	0.121
LSTR2	**0.05**	0.058	0.057	0.057	0.064	0.067	0.065	0.067	0.068	0.067
LSTR2	**0.10**	0.119	0.121	0.122	0.125	0.124	0.122	0.116	0.116	0.115
Estr-Lstr	**0.05**	0.059	0.061	0.062	0.058	0.051	0.048	0.060	0.065	0.064
Estr-Lstr	**0.10**	0.104	0.102	0.101	0.125	0.124	0.123	0.102	0.114	0.116

Notes: ESTR2: Estr adjustment in both equations; LSTR2: Lstr adjustment in both; Estr-Lstr: Estr adjustment in one and Lstr adjustment in the other variable.

Table 2a: Power of the test against a specific nonlinear alternative:

Model\\d		-0.1			-0.4			-0.6		
T=100:	γ	sup	ave	exp	sup	ave	exp	sup	ave	exp
	0.1	0.113	0.097	0.096	0.616	0.522	0.527	0.873	0.808	0.817
	0.4	0.113	0.103	0.103	0.175	0.166	0.165	0.238	0.275	0.277
ESTR2	0.7	0.101	0.090	0.089	0.128	0.117	0.119	0.166	0.188	0.181
	1	0.112	0.109	0.107	0.116	0.115	0.113	0.146	0.146	0.144
	2	0.118	0.096	0.098	0.128	0.107	0.111	0.123	0.115	0.114
	0.1	0.748	0.730	0.730	0.791	0.772	0.770	0.815	0.799	0.797
	0.4	0.784	0.753	0.752	0.911	0.878	0.875	0.939	0.910	0.909
LSTR2	0.7	0.776	0.743	0.743	0.877	0.831	0.831	0.865	0.822	0.082
	1	0.756	0.723	0.723	0.853	0.812	0.813	0.844	0.814	0.813
	2	0.773	0.749	0.748	0.847	0.793	0.793	0.883	0.843	0.843
	0.1	0.229	0.233	0.233	0.742	0.733	0.733	0.868	0.862	0.862
	0.4	0.618	0.616	0.614	0.996	0.996	0.996	1	1	1
EstrLstr	0.7	0.717	0.707	0.707	1	1	1	1	1	1
	1	0.710	0.705	0.704	1	1	1	1	1	1
	2	0.757	0.758	0.758	0.999	0.999	0.999	1	1	1

Notes: ESTR2: Estr adjustment in both equations; LSTR2: Lstr adjustment in both; Estr-Lstr: Estr adjustment in one and Lstr adjustment in the other variable.

Table 2b: Power of the test (continuation):

Model\\d		-0.1			-0.4			-0.6		
T=250:	γ	sup	ave	exp	sup	ave	exp	sup	ave	exp
	0.1	0.438	0.446	0.447	0.995	0.996	0.995	1	1	1
	0.4	0.359	0.385	0.381	0.579	0.628	0.627	0.747	0.819	0.811
ESTR2	0.7	0.355	0.373	0.378	0.458	0.519	0.52	0.530	0.614	0.611
	1	0.376	0.396	0.398	0.439	0.475	0.475	0.446	0.508	0.507
	2	0.352	0.362	0.364	0.434	0.455	0.457	0.442	0.455	0.456
	0.1	0.988	0.987	0.987	0.996	0.995	0.995	0.996	0.996	0.995
	0.4	0.999	0.999	0.999	1	0.999	0.999	1	1	1
LSTR2	0.7	0.996	0.994	0.994	1	1	1	1	1	1
	1	0.996	0.993	0.993	1	1	1	1	0.999	0.999
	2	0.996	0.996	0.996	1	1	1	1	0.999	0.999
	0.1	0.467	0.458	0.455	0.983	0.979	0.979	0.998	0.998	0.998
	0.4	0.949	0.942	0.941	1	1	1	1	1	1
EstrLstr	0.7	0.984	0.981	0.981	1	1	1	1	1	1
	1	0.978	0.978	0.978	1	1	1	1	1	1
	2	0.986	0.986	0.986	1	1	1	1	1	1

Notes: ESTR2: Estr adjustment in both equations; LSTR2: Lstr adjustment in both; Estr-Lstr: Estr adjustment in one and Lstr adjustment in the other variable.

Table 2c: Empirical frequency of choosing the nonlinear model:

Model\d		-0.1			-0.4			-0.6		
T=250:	γ	sup	ave	exp	sup	ave	exp	sup	ave	exp
	0.1	0.833	0.833	0.830	0.992	0.992	0.992	1	1	1
	0.4	0.785	0.799	0.797	0.864	0.894	0.891	0.923	0.952	0.951
ESTR2	0.7	0.833	0.838	0.837	0.841	0.848	0.847	0.870	0.898	0.897
	1	0.793	0.812	0.812	0.835	0.861	0.861	0.828	0.851	0.854
	2	0.807	0.810	0.810	0.839	0.858	0.855	0.832	0.850	0.852
	0.1	0.914	0.909	0.907	0.938	0.922	0.924	0.958	0.930	0.929
	0.4	0.953	0.942	0.942	0.989	0.965	0.965	0.998	0.975	0.975
LSTR2	0.7	0.947	0.930	0.930	0.982	0.954	0.953	0.992	0.964	0.965
	1	0.924	0.918	0.918	0.961	0.936	0.936	0.968	0.942	0.940
	2	0.920	0.909	0.910	0.950	0.933	0.933	0.960	0.936	0.936
	0.1	0.594	0.551	0.549	0.935	0.933	0.933	0.977	0.976	0.977
	0.4	0.852	0.849	0.845	1	1	1	1	1	1
EstrLstr	0.7	0.906	0.902	0.903	1	1	1	1	1	1
	1	0.917	0.913	0.911	1	1	1	1	1	1
	2	0.920	0.914	0.913	0.995	0.994	0.992	1	1	1

Notes: ESTR2: Estr adjustment in both equations; LSTR2: Lstr adjustment in both; Estr-Lstr: Estr adjustment in one and Lstr adjustment in the other variable.

Table 3: Size of the test in the presence of ARCH effects:

Model\α				-0.1			-0.4	
T=250:	a	**size**	sup	ave	exp	sup	ave	exp
	0.2	**0.05**	0.060	0.055	0.059	0.073	0.071	0.074
	0.2	**0.10**	0.094	0.112	0.101	0.141	0.123	0.128
ESTR2	0.5	**0.05**	0.058	0.055	0.053	0.064	0.055	0.060
	0.5	**0.10**	0.116	0.105	0.108	0.115	0.114	0.112
	0.9	**0.05**	0.082	0.071	0.070	0.070	0.051	0.066
	0.9	**0.10**	0.138	0.110	0.114	0.148	0.114	0.130
	0.2	**0.05**	0.040	0.033	0.033	0.033	0.029	0.030
	0.2	**0.10**	0.082	0.073	0.073	0.067	0.063	0.063
LSTR2	0.5	**0.05**	0.052	0.045	0.045	0.025	0.028	0.028
	0.5	**0.10**	0.079	0.076	0.077	0.067	0.059	0.060
	0.9	**0.05**	0.042	0.041	0.040	0.055	0.046	0.047
	0.9	**0.10**	0.088	0.072	0.071	0.101	0.086	0.085

Notes: ESTR2: Estr adjustment in both equations; LSTR2: Lstr adjustment in both.

Table 4: Power of the test in the presence of ARCH effects:

Model\\d				-0.1			-0.4			-0.6	
T=250:	a	γ	sup	ave	exp	sup	ave	exp	sup	ave	exp
		0.1	0.680	0.350	0.518	1	0.919	0.997	1	0.999	1
	0.2	0.4	0.514	0.304	0.428	0.898	0.704	0.833	0.982	0.914	0.970
		0.7	0.518	0.303	0.431	0.781	0.535	0.677	0.879	0.726	0.822
		0.1	0.556	0.289	0.399	1	0.850	0.984	1	0.994	1
ESTR2	0.5	0.4	0.417	0.247	0.343	0.801	0.566	0.711	0.931	0.807	0.893
		0.7	0.372	0.242	0.317	0.609	0.405	0.542	0.796	0.617	0.747
		0.1	0.259	0.145	0.190	0.779	0.420	0.643	0.948	0.691	0.889
	0.9	0.4	0.199	0.130	0.156	0.375	0.216	0.310	0.591	0.387	0.509
		0.7	0.228	0.137	0.168	0.292	0.162	0.227	0.387	0.233	0.311
		0.1	0.942	0.898	0.902	0.977	0.946	0.949	0.988	0.977	0.978
	0.2	0.4	0.993	0.976	0.978	1	1	1	1	1	1
		0.7	0.987	0.972	0.972	1	1	1	1	1	1
		0.1	0.964	0.933	0.935	0.993	0.980	0.981	0.995	0.984	0.986
LSTR2	0.5	0.4	0.982	0.969	0.972	1	1	1	1	1	1
		0.7	0.991	0.978	0.979	1	1	1	1	1	1
		0.1	0.993	0.986	0.987	0.997	0.995	0.995	0.998	0.998	0.998
	0.9	0.4	0.989	0.980	0.982	0.991	0.988	0.988	0.985	0.981	0.981
		0.7	0.987	0.977	0.977	0.988	0.986	0.986	0.975	0.967	0.967

Notes: ESTR2: Estr adjustment in both equations; LSTR2: Lstr adjustment in both.

Table 5: Testing the monetary model of exchange rates:

	UK			Japan		
Model	sup	ave	exp	sup	ave	exp
Estr Estr Estr	0.023(10)	0.026(10)	0.025(10)	0.595(12)	0.679(12)	0.690(12)
Estr Estr Lstr	0.055(7)	0.052(7)	0.052(7)	0.287(3)	0.322(3)	0.305(3)
Estr Lstr Estr	0.066(3)	0.042(3)	0.043(3)	0.645(1)	0.521(12)	0.521(12)
Estr Lstr Lstr	0.067(4)	0.044(4)	0.045(4)	0.310(3)	0.307(1)	0.294(1)
Lstr Estr Estr	0.384(3)	0.263(3)	0.284(3)	0.213(1)	0.231(1)	0.231(1)
Lstr Lstr Estr	0.301(3)	0.239(3)	0.239(3)	0.263(1)	0.178(1)	0.181(1)
Lstr Estr Lstr	0.385(2)	0.305(3)	0.311(3)	0.262(4)	0.247(4)	0.247(1)
Lstr Lstr Lstr	0.248(3)	0.170(3)	0.170(3)	0.255(3)	0.250(3)	0.264(12)

Notes: The variables are in the following order: exchange rate, money differential, output. The number in brackets signals the delay parameter at which the p-value is minimized.

Appendix II: Proofs

To keep exposition simpler, all proofs are stated for the case of conditional homoscedasticity, but can be easily amended using the triangular inequality to include permitted classes of conditionally heteroscedastic models under the existence of higher moments of the error process and using the facts that the process H_t is stationary, ergodic with finite moments and that $\left\| H_t \left(H_t' H_t\right)^{-1} H_t' \epsilon_t \right\| \leq \|\epsilon_t\|$ by the standard result for projection matrices. In the following to save space $\sum_t = \sum_{t=1}^T$ and K is a generic constant.

Proof of Proposition 1.1. Consider first expression within the norm. Since $Z_{t-1}' Z_{t-1}$ is a scalar term and $\Psi(Z_{t-1}, \gamma) = \Psi(Z_{t-1}, \gamma)'$, this can be rearranged as follows:

$$||\epsilon_t' \Omega^{-1} \left(Z_{t-1}' \otimes \Psi(Z_{t-1}, \gamma)\right)|| = \left(tr\left\{\epsilon_t' \Omega^{-1} \left(Z_{t-1}' \otimes \Psi(Z_{t-1}, \gamma)\right) \left(Z_{t-1} \otimes \Psi(Z_{t-1}, \gamma)\right) \Omega^{-1'} \epsilon_t\right\}\right)^{\frac{1}{2}}$$

$$= \left(tr\left\{\epsilon_t' \Omega^{-1} \left(Z_{t-1}' Z_{t-1} \otimes \Psi(Z_{t-1}, \gamma) \Psi(Z_{t-1}, \gamma)\right) \Omega^{-1'} \epsilon_t\right\}\right)^{\frac{1}{2}}$$

$$= \left(tr\left\{Z_{t-1}' Z_{t-1} \epsilon_t' \Omega^{-1} \Psi(Z_{t-1}, \gamma) \Psi(Z_{t-1}, \gamma) \Omega^{-1'} \epsilon_t\right\}\right)^{\frac{1}{2}}$$

$$= \left(Z_{t-1}' Z_{t-1}\right)^{\frac{1}{2}} \left(tr\left\{\epsilon_t' \Omega^{-1} \Psi(Z_{t-1}, \gamma) \Psi(Z_{t-1}, \gamma) \Omega^{-1'} \epsilon_t\right\}\right)^{\frac{1}{2}}$$

$$||\epsilon_t' \Omega^{-1} \left(Z_{t-1}' \otimes \Psi(Z_{t-1}, \gamma)\right)||^{2+\delta} = ||Z_{t-1}||^{2+\delta} ||\Psi(Z_{t-1}, \gamma) \Omega^{-1'} \epsilon_t||^{2+\delta}$$

Substituting back into expression and using the Cauchy-Schwarz (C-S) inequality:

$$E\left[sup_{\gamma \in \Gamma^*} ||\epsilon_t' \Omega^{-1} \left(Z_{t-1}' \otimes \Psi(Z_{t-1}, \gamma)\right)||^{2+\delta}\right] = E\left[||Z_{t-1}||^{2+\delta} sup_{\gamma \in \Gamma^*} ||\Psi(Z_{t-1}, \gamma) \Omega^{-1'} \epsilon_t||^{2+\delta}\right]$$

$$\leq \left(E||Z_{t-1}||^{2(2+\delta)}\right)^{\frac{1}{2}} \left(E sup_{\gamma \in \Gamma^*} ||\Psi(Z_{t-1}, \gamma) \Omega^{-1'} \epsilon_t||^{2(2+\delta)}\right)^{\frac{1}{2}} \tag{9}$$

Boundness of the first term in (10) now follows from Assumptions 1.1 and 1.2 given the existence of moments of the Markov chain and choosing δ arbitrarily. Consider the second term:

$$||\Psi(Z_{t-1}, \gamma) \Omega^{-1'} \epsilon_t||^{2+\delta} = \left(tr\left\{\Omega^{-1} \Psi(Z_{t-1}, \gamma) \Psi(Z_{t-1}, \gamma) \Omega^{-1'} \epsilon_t \epsilon_t'\right\}\right)^{\frac{1}{2}(2+\delta)}$$

By Hölder's inequality for matrices (Magnus and Neudecker, 2007, pp. 250), where s is slightly higher than 1 and q is large:

$$\leq \left(tr\left\{\left(\Omega^{-1} \Psi(Z_{t-1}, \gamma) \Psi(Z_{t-1}, \gamma) \Omega^{-1'}\right)^s\right\}\right)^{\frac{1}{2} \frac{(2+\delta)}{s}} \left(tr\left\{(\epsilon_t' \epsilon_t)^q\right\}\right)^{\frac{1}{2} \frac{(2+\delta)}{q}}$$

$$= \left(tr\left\{\left(\Omega^{-1} \Psi(Z_{t-1}, \gamma) \Psi(Z_{t-1}, \gamma) \Omega^{-1'}\right)^s\right\}\right)^{\frac{1}{2} \frac{(2+\delta)}{s}} \left((\epsilon_t' \epsilon_t)^q\right)^{\frac{1}{2} \frac{(2+\delta)}{q}}$$

$$= \left(tr\left\{\left(\Omega^{-1} \Psi(Z_{t-1}, \gamma) \Psi(Z_{t-1}, \gamma) \Omega^{-1'}\right)^s\right\}\right)^{\frac{1}{2} \frac{(2+\delta)}{s}} (\epsilon_t' \epsilon_t)^{\frac{1}{2}(2+\delta)}$$

Substituting back in (10):

$$E\left[sup_{\gamma\in\Gamma^*}||\epsilon_t'\Omega^{-1}\left(Z_{t-1}'\otimes\Psi(Z_{t-1},\gamma)\right)||^{2+\delta}\right]\leq$$

$$K_1\left(E\left[sup_{\gamma\in\Gamma^*}\left(tr\left\{\left(\Omega^{-1}\Psi(Z_{t-1},\gamma)\Psi(Z_{t-1},\gamma)\Omega^{-1'}\right)^s\right\}\right)^{\frac{1}{2}\frac{(2+\delta)}{s}}||\epsilon_t||^{2(2+\delta)}\right]\right)^{\frac{1}{2}}$$

Given the independence of errors, constancy of the error's covariance matrix and boundness of $\Psi(Z_{t-1},\gamma)$ between $[0,r]$, Proposition 1.1. then follows provided that $sup_t\ E||\epsilon_t||^{2(2+\delta)}$ is bounded, which follows from Assumption 1.1. choosing δ arbitrarily. ■

Proof of Proposition 1.2. First we will show sufficient condition to satisfy the Proposition and then prove that the condition is satisfied in current model. Set $\upsilon=\frac{1}{2+\delta}$ and similarly to the proof of Proposition 2.1. rearrange the norm using properties of trace of the Kronecker product:

$$||\epsilon_t'\Omega^{-1}\left(Z_{t-1}'\otimes(\Psi(Z_{t-1},\gamma_1)-\Psi(Z_{t-1},\gamma))\right)||^{2+\delta}=$$

$$=||Z_{t-1}||^{2+\delta}||\epsilon_t'\Omega^{-1}\left(\Psi(Z_{t-1},\gamma_1)-\Psi(Z_{t-1},\gamma)\right)||^{2+\delta}\tag{10}$$

Consider only the second norm. By the mean value theorem for $\overline{\gamma}\in[\gamma,\gamma_1]$, Hölder's inequality[6] for matrices with s slightly higher than 1 and q large:

$$||\left(\epsilon_t'\Omega^{-1}\otimes I_k\right)vec\left(\Psi(Z_{t-1},\gamma_1)-\Psi(Z_{t-1},\gamma)\right)||^{2+\delta}=||\left(\epsilon_t'\Omega^{-1}\otimes I_k\right)\triangledown\Psi\left(Z_{t-1},\overline{\gamma}\right)(\gamma_1-\gamma)||^{2+\delta}$$

$$=\left(tr\left\{\triangledown\Psi\left(Z_{t-1},\overline{\gamma}\right)'\left(\Omega^{-1\prime}\epsilon_t\epsilon_t'\Omega^{-1}\otimes I_k\right)\triangledown\Psi\left(Z_{t-1},\overline{\gamma}\right)(\gamma_1-\gamma)(\gamma_1-\gamma)'\right\}\right)^{\frac{1}{2}(2+\delta)}$$

$$\leq\left(tr\left\{\left(\triangledown\Psi\left(Z_{t-1},\overline{\gamma}\right)'\left(\Omega^{-1\prime}\epsilon_t\epsilon_t'\Omega^{-1}\otimes I_k\right)\triangledown\Psi\left(Z_{t-1},\overline{\gamma}\right)\right)^s\right\}\right)^{\frac{1}{2s}(2+\delta)}\left(tr\left\{\left((\gamma_1-\gamma)'(\gamma_1-\gamma)\right)^q\right\}\right)^{\frac{1}{2q}(2+\delta)}$$

$$=\left(tr\left\{\left(\triangledown\Psi\left(Z_{t-1},\overline{\gamma}\right)'\left(\Omega^{-1\prime}\epsilon_t\epsilon_t'\Omega^{-1}\otimes I_k\right)\triangledown\Psi\left(Z_{t-1},\overline{\gamma}\right)\right)^s\right\}\right)^{\frac{1}{2s}(2+\delta)}\left((\gamma_1-\gamma)'(\gamma_1-\gamma)\right)^{\frac{1}{2}(2+\delta)}$$

Define the first term above as $\Upsilon\left(Z_{t-1},\overline{\gamma}\right)$, such that we can write:

$$||\epsilon_t'\Omega^{-1}\left(\Psi(Z_{t-1},\gamma_1)-\Psi(Z_{t-1},\gamma)\right))||^{2+\delta}=\Upsilon\left(Z_{t-1},\overline{\gamma}\right)||(\gamma_1-\gamma)||^{2+\delta}\tag{11}$$

Substituting (11) in (10) and using (10) the condition can be expressed as:

$$\sup_t\ E\left[||Z_{t-1}||^{2+\delta}\sup_{\gamma_1\in\Gamma^*:||\gamma_1-\gamma||<\varpi}\Upsilon\left(Z_{t-1},\overline{\gamma}\right)||(\gamma_1-\gamma)||^{2+\delta}\right]\leq$$

[6] Note that Hölder's inequality holds only for positive semi-definite matrices of the same order. $\triangledown\Psi\left(Z_{t-1},\overline{\gamma}\right)'\left(\Omega^{-1\prime}\epsilon_t\epsilon_t'\Omega^{-1}\otimes I_k\right)\triangledown\Psi\left(Z_{t-1},\overline{\gamma}\right)$ is positive definite since by construction ($k^2\ x\ kr$) matrix $\triangledown\Psi\left(Z_{t-1},\overline{\gamma}\right)$ is of full column rank and positive definiteness of the middle term follows from Assumption 1.1. and the fact that if matrices A and B are positive definite $A\otimes B$ is positive definite by Theorem 1 in Magnus and Neudecker(2007, pp.33).

$$\sup_t E\left[||Z_{t-1}||^{2+\delta} \sup_{\gamma\in\Gamma^*} \Upsilon\left(Z_{t-1},\gamma\right)\right] \varpi^{2+\delta}$$

Hence, Proposition 1.2. follows by setting $\upsilon = \frac{1}{2+\delta}(2+\delta) = 1$ and choosing C arbitrarily provided that $\sup_t E\left[||Z_{t-1}||^{2+\delta} \sup_{\gamma\in\Gamma^*} \Upsilon\left(Z_{t-1},\gamma\right)\right] < \infty$

By C-S inequality boundness of the first term again follows from assumptions 1.1 and 1.2:

$$\sup_t E\left[||Z_{t-1}||^{2+\delta} \sup_{\gamma\in\Gamma^*} \Upsilon\left(Z_{t-1},\gamma\right)\right] \leq \sup_t \left(E||Z_{t-1}||^{2(2+\delta)}\right)^{\frac{1}{2}} \left(E|\sup_{\gamma\in\Gamma^*} \Upsilon\left(Z_{t-1},\gamma\right)|^2\right)^{\frac{1}{2}}$$

$$\leq K_2 \sup_t \left(E\left|\sup_{\gamma\in\Gamma^*}\left(tr\left\{\left(\bigtriangledown\Psi\left(Z_{t-1},\gamma\right)'\left(\Omega^{-1\prime}\epsilon_t\epsilon_t'\Omega^{-1}\otimes I_k\right)\bigtriangledown\Psi\left(Z_{t-1},\gamma\right)\right)^s\right\}\right)^{\frac{1}{2s}(2+\delta)}\right|^2\right)^{\frac{1}{2}}$$

Using the fact that the Frobenius norm is always positive and rearranging:

$$= K_2 \sup_t \left(E \sup_{\gamma\in\Gamma^*}\left(tr\left\{\left(\left(\Omega^{-1\prime}\epsilon_t\epsilon_t'\Omega^{-1}\otimes I_k\right)\bigtriangledown\Psi\left(Z_{t-1},\gamma\right)\bigtriangledown\Psi\left(Z_{t-1},\gamma\right)'\right)^s\right\}\right)^{\frac{1}{2s}(2+\delta)2}\right)^{\frac{1}{2}}$$

Applying the Hölder's inequality with w slightly higher than 1 and m large:

$$\leq K_2 \sup_t \begin{pmatrix} E\left(tr\left\{\left(\Omega^{-1\prime}\epsilon_t\epsilon_t'\Omega^{-1}\otimes I_k\right)^{sw}\right\}\right)^{\frac{1}{2sw}(2+\delta)2} \\ \sup_{\gamma\in\Gamma^*}\left(tr\left\{\left(\bigtriangledown\Psi\left(Z_{t-1},\gamma\right)\bigtriangledown\Psi\left(Z_{t-1},\gamma\right)'\right)^{sm}\right\}\right)^{\frac{1}{2sm}(2+\delta)2} \end{pmatrix}^{\frac{1}{2}}$$

Using Assumption 1.4. and after that Assumption 1.1 and properties of the Kronecker product, it follows that:

$$\leq K_2 \sup_t \left(E\left[\left(tr\left\{\left(\Omega^{-1\prime}\epsilon_t\epsilon_t'\Omega^{-1}\otimes I_k\right)^{sw}\right\}\right)^{\frac{1}{2sw}(2+\delta)2} K_3\left(\left(Z_{t-1}'Z_{t-1}\right)^{sm}\right)^{\frac{1}{2sm}(2+\delta)2}\right]\right)^{\frac{1}{2}}$$

$$= K_4 \sup_t \left(E\left[\left(tr\left\{\left(\Omega^{-1\prime}\epsilon_t\epsilon_t'\Omega^{-1}\right)^{sw}\right\}\right)^{\frac{1}{2sw}(2+\delta)2} ||Z_{t-1}||^{(2+\delta)2}\right]\right)^{\frac{1}{2}}$$

$$= K_4 \sup_t \left(E\left(tr\left\{\left(\epsilon_t\epsilon_t'\Omega^{-1}\Omega^{-1\prime}\right)^{sw}\right\}\right)^{\frac{1}{2sw}(2+\delta)2}\right)^{\frac{1}{2}} \left(E\,||Z_{t-1}||^{(2+\delta)2}\right)^{\frac{1}{2}}$$

where the last line follows from the error independence assumption. The boundness of the first term above now follows using Assumption 1.1, while the boundness of the second term follows from assumptions 1.1. and 1.2. ■

Proof of Proposition 2. First, note that due to superconsistency of $\widetilde{\beta}$:

$\frac{1}{\sqrt{T}}\sum_t \epsilon_t'\Omega^{-1}\left(Z_{t-1}'(\widetilde{\beta})\otimes\Psi(Z_{t-1}(\widetilde{\beta}),\gamma)\right) = \frac{1}{\sqrt{T}}\sum_t \epsilon_t'\Omega^{-1}\left(Z_{t-1}'\otimes\Psi(Z_{t-1},\gamma)\right) + o_p(1)$ and in the remaining part of the proof only the score evaluated at Z_{t-1} is considered.

The proof follows by verifying the finite dimensional (fidi) convergence of $\widetilde{s_T}(\gamma)$ and stochastic equicontinuity of the empirical process. Both follow from the Theorem 1 Application 1 of DMR (1995). First note that choosing $\phi(x) = x^{1+\frac{\delta}{2}}$, $x > 0$, in DMR it can be shown (Cho and White, 2007, pp.1705) that condition (2.9a) in Lemma 2 of DMR will be satisfied for β-mixing processes with geometrically decaying numbers. Lemma 2 of DMR then implies that the conditions on the class of functions of the empirical process and entropy with bracketing can be established with respect to the $\|\cdot\|_{2\left(1+\frac{\delta}{2}\right)} = \|\cdot\|_{2+\delta}$ norm instead of the $\|\cdot\|_{2,\beta}$ norm proposed in their paper. Fidi convergence then follows from Theorem 1 of DMR given the Assumption 1.2, if the class of functions $\mathcal{F}$ of the empirical process belongs to the class whose envelope has bounded $2+\delta$ moments. Proposition 1.1. ensures that the condition is satisfied in our case.

Furthermore, condition (2.4) in DMR on summability of β-mixing coefficients $\sum_t \beta_n$ is trivially satisfied by Assumption 1.2. on the rate of mixing numbers: $\sum_t \beta_n = \sum_t Cb^n$ and since $0 < b < 1$ $= C\frac{b}{1-b} < \infty$. It remains to show that the entropy with bracketing satisfies the integrability condition (2.11) in DMR. This follows from Theorem 5 in Andrews (1994, see also Andrews, 1993, pp.201) if the L^p-continuity condition holds which is established in Proposition 1.2. for $p = 2+\delta$.

■

Proof of Proposition 3. By triangular inequality:

$\sup_{\gamma\in\Gamma^*}\left|\left|\widetilde{V}_T(\gamma) - V(\gamma)\right|\right| \leq \sup_{\gamma\in\Gamma^*}\left|\left|\widetilde{V}_T(\gamma) - V_T(\gamma)\right|\right| + \sup_{\gamma\in\Gamma^*}||V_T(\gamma) - V(\gamma)||$

Start with the second term on the right hand side, where the uniform convergence of $V_T(\gamma)$ over $\gamma \in\Gamma^*$ is established. Note that under the null hypothesis and Assumptions 1.1 and 1.2. pointwise convergence for each γ follows from the ergodic theorem since all elements of $V_T(\gamma) = \frac{1}{T}\sum_t Z_{t-1}Z'_{t-1}\otimes\left(\Psi_{t-1}(\gamma)\Omega^{-1}\Psi_{t-1}(\gamma)\right)$ are strictly stationary and mixing sequences and hence ergodic sequences (Proposition 3.44 in White, 2001). The convergence follows given the existence of suitable moments, which follows from Assumptions 1.1 and 1.2. We then need to establish conditions under which $V_T(\gamma)$ is stochastically equicontinious over $\gamma \in\Gamma^*$ and uniform convergence follows from Theorem 2.1 in Newey (1991). $V_T(\gamma) - V(\gamma)$ is stochastic equicontinious over Γ^* in the sense that:

$\sup_t P\left[\sup_{\gamma\in\Gamma^*}\sup_{\gamma_1\in B(\gamma,\varpi)}||V_T(\gamma) - V(\gamma_1)|| > \upsilon\right] \xrightarrow{p} 0$, as $\varpi\to 0$,

where $B(\gamma,\varpi)$ is a ball of radius ϖ around $\gamma : \|\gamma - \gamma_1\| \leq\varpi$

By Markov's inequality:

$$\leq \frac{1}{\upsilon}\sup_t E\left[\sup_{\gamma\in\Gamma^*}\sup_{\gamma_1\in B(\gamma,\varpi)}||V_T(\gamma) - V(\gamma_1)||\right] \tag{12}$$

Consider now $||V_T(\gamma) - V(\gamma_1)||$:

$=\left|\left|\frac{1}{T}\sum_t Z_{t-1}Z_{t-1}'\otimes\left(\Psi_{t-1}(\gamma)\Omega^{-1}\Psi_{t-1}(\gamma)-\Psi_{t-1}(\gamma_1)\Omega^{-1}\Psi_{t-1}(\gamma_1)\right)\right|\right|$

$\leq\frac{1}{T}\sum_t\left|\left|Z_{t-1}Z_{t-1}'\otimes\left(\Psi_{t-1}(\gamma)\Omega^{-1}\Psi_{t-1}(\gamma)-\Psi_{t-1}(\gamma_1)\Omega^{-1}\Psi_{t-1}(\gamma_1)\right)\right|\right|$

$$\leq\sup_t\left|\left|Z_{t-1}Z_{t-1}'\otimes\left(\Psi_{t-1}(\gamma)\Omega^{-1}\Psi_{t-1}(\gamma)-\Psi_{t-1}(\gamma_1)\Omega^{-1}\Psi_{t-1}(\gamma_1)\right)\right|\right| \tag{13}$$

Using the fact that $tr\left\{A\otimes C\right\}=tr\left\{A\right\}tr\left\{C\right\}$ and that $Z_{t-1}'Z_{t-1}$ is a scalar term, this can be rearranged as:

$=\sup_t\left\|Z_{t-1}\right\|^2\left|\left|\Psi_{t-1}(\gamma)\Omega^{-1}\Psi_{t-1}(\gamma)-\Psi_{t-1}(\gamma_1)\Omega^{-1}\Psi_{t-1}(\gamma_1)\right|\right|$

Since $\left(tr\left\{A'A\right\}\right)^{1/2}=\left(vecA'vecA\right)^{1/2}$:

$=\sup_t\left\|Z_{t-1}\right\|^2\left|\left|vec\left(\Psi_{t-1}(\gamma)\Omega^{-1}\Psi_{t-1}(\gamma)\right)-vec\left(\Psi_{t-1}(\gamma_1)\Omega^{-1}\Psi_{t-1}(\gamma_1)\right)\right|\right|$

By the mean value theorem for vector functions (Magnus and Neudecker, 2007, pp.110) and for $\overline{\gamma}\in[\gamma,\gamma_1]$

$=\sup_t\left\|Z_{t-1}\right\|^2\left|\left|\left(I_k\otimes\Psi_{t-1}(\gamma)\Omega^{-1}\right)\nabla\Psi_{t-1}(\overline{\gamma})(\gamma-\gamma_1)+\left(\Psi_{t-1}(\gamma)\Omega'^{-1}\otimes I_k\right)\nabla\Psi_{t-1}(\overline{\gamma})(\gamma-\gamma_1)\right|\right|$

$$=\sup_t\sup_{\gamma\in\Gamma^*}\left\|Z_{t-1}\right\|^2\left|\left|\left(I_k\otimes\Psi_{t-1}(\gamma)\Omega^{-1}+\Psi_{t-1}(\gamma)\Omega'^{-1}\otimes I_k\right)\nabla\Psi_{t-1}(\gamma)(\gamma-\gamma_1)\right|\right| \tag{14}$$

where $\nabla\Psi_{t-1}(\overline{\gamma})$ is (k^2xkr) matrix of partial derivatives for some $\overline{\gamma}\in[\gamma,\gamma_1]$.

Define $C(\gamma)=I_k\otimes\Psi_{t-1}(\gamma)\Omega^{-1}+\Psi_{t-1}(\gamma)\Omega'^{-1}\otimes I_k$ and note that (12) is majorized by

$\leq\frac{1}{\upsilon}\sup_t E\left[\left\|Z_{t-1}\right\|^2\sup_{\gamma\in\Gamma^*}\sup_{\gamma_1\in B(\gamma,\varpi)}\left|\left|C(\gamma)\nabla\Psi_{t-1}(\gamma)(\gamma-\gamma_1)\right|\right|\right]$

By the C-S inequality:

$\leq\frac{1}{\upsilon}\sup_t\left(E\left\|Z_{t-1}\right\|^4\right)^{\frac{1}{2}}\left(E\left|\sup_{\gamma\in\Gamma^*}\sup_{\gamma_1\in B(\gamma,\varpi)}\left|\left|C(\gamma)\nabla\Psi_{t-1}(\gamma)(\gamma-\gamma_1)\right|\right|\right|^2\right)^{\frac{1}{2}}$

Given the Assumptions 1.1. and 1.2. $\sup_t\left(E\left\|Z_{t-1}\right\|^4\right)<\infty$, such that by the positiveness of the Frobenius norm:

$\leq\frac{1}{\upsilon}K_5\sup_t\left(E\left|\sup_{\gamma\in\Gamma^*}\sup_{\gamma_1\in B(\gamma,\varpi)}\left|\left|C(\gamma)\nabla\Psi_{t-1}(\gamma)(\gamma-\gamma_1)\right|\right|\right|^2\right)^{\frac{1}{2}}$

$=\frac{1}{\upsilon}K_5\sup_t\left(E\sup_{\gamma\in\Gamma^*}\sup_{\gamma_1\in B(\gamma,\varpi)}\left|\left|C(\gamma)\nabla\Psi_{t-1}(\gamma)(\gamma-\gamma_1)\right|\right|^2\right)^{\frac{1}{2}}$

By repeated application of the Hölder's inequality[7] with s,w slightly higher than 1 and q and m large:

---

[7] Note that $C(\gamma)$ has full rank given the Assumption 1.1. using the fact that rank$(A\otimes B)=rank(A)rank(B)$. Full rank of $C(\gamma)$ implies that $C(\gamma)C(\gamma)'$ is positive definite.

$\leq \frac{1}{\upsilon} K_5 \sup_t \left(E[\sup_{\gamma \in \Gamma^*} \sup_{\gamma_1 \in B(\gamma, \varpi)} \left(tr\left\{\left(C(\gamma)' C(\gamma)\right)^q\right\}\right)^{\frac{2}{2q}}\right.$

$\left(tr\left\{\left((\gamma - \gamma_1)(\gamma - \gamma_1)' \nabla \Psi_{t-1}(\gamma)' \nabla \Psi_{t-1}(\gamma)\right)^s\right\}\right)^{\frac{2}{2s}}])^{\frac{1}{2}}$

$\leq \frac{1}{\upsilon} K_5 \sup_t \left(E[\sup_{\gamma \in \Gamma^*} \sup_{\gamma_1 \in B(\gamma, \varpi)} \left(tr\left\{\left(C(\gamma)' C(\gamma)\right)^q\right\}\right)^{\frac{2}{2q}} \left(tr\left\{\left((\gamma - \gamma_1)(\gamma - \gamma_1)'\right)^{sm}\right\}\right)^{\frac{2}{2sm}}\right.$

$\left(tr\left\{\left(\nabla \Psi_{t-1}(\gamma)' \nabla \Psi_{t-1}(\gamma)\right)^{sw}\right\}\right)^{\frac{2}{2sw}}])^{\frac{1}{2}}$

$= \frac{1}{\upsilon} K_5 \sup_t \left(E[\sup_{\gamma \in \Gamma^*} \sup_{\gamma_1 \in B(\gamma, \varpi)} \left(tr\left\{\left(C(\gamma)' C(\gamma)\right)^q\right\}\right)^{\frac{1}{q}} \|\gamma - \gamma_1\|^2\right.$

$\left(tr\left\{\left(\nabla \Psi_{t-1}(\gamma)' \nabla \Psi_{t-1}(\gamma)\right)^{sw}\right\}\right)^{\frac{2}{2sw}}])^{\frac{1}{2}}$

Using Assumption 1.4. and rearranging:

$\leq \frac{\varpi}{\upsilon} K_5 \sup_t \left(E\left[\sup_{\gamma \in \Gamma^*} \left(tr\left\{\left(C(\gamma)' C(\gamma)\right)^q\right\}\right)^{\frac{1}{q}} \left(\left(Z'_{t-1} Z_{t-1}\right)^{sw}\right)^{\frac{2}{2sw}}\right]\right)^{\frac{1}{2}}$

Since $C(\gamma)$ consists of elements which are constant and bounded between zero and one, stochastic equicontinuity follows by letting $\varpi \rightarrow 0$ under assumptions 1.1 and 1.2. which establish boundness of the last term.

It remains to show that the first term is $op(1)$ uniformly in γ. This follows easily under Assumption 1.3. after taking into account different rates of convergence in the long-run and short-run parameters as in Saikkonen (1995).

$\sup_{\gamma \in \Gamma^*} \left\| \widetilde{V}_T(\gamma) - V_T(\gamma) \right\| =$

$$\sup_{\gamma \in \Gamma^*} \left\| \begin{array}{l} \frac{1}{T} \sum_t Z_{t-1}(\widetilde{\beta}) Z'_{t-1}(\widetilde{\beta}) \otimes \Psi_{t-1}(\widetilde{\beta}, \gamma) \widetilde{\Omega}^{-1} \Psi_{t-1}(\widetilde{\beta}, \gamma) - \\ \frac{1}{T} \sum_t Z_{t-1}(\beta_0) Z'_{t-1}(\beta_0) \otimes \Psi_{t-1}(\beta_0, \gamma) \Omega^{-1} \Psi_{t-1}(\beta_0 \gamma) \end{array} \right\| \qquad (15)$$

By triangular inequality (15) can be rearranged as:

$\leq \sup_{\gamma \in \Gamma^*} \left\| \frac{1}{T} \sum_t \sqrt{T}(\widetilde{\beta} - \beta)' \frac{1}{\sqrt{T}} X_{2,t-1} \frac{1}{\sqrt{T}} X'_{2,t-1} \sqrt{T}(\widetilde{\beta} - \beta) \otimes \Psi_{t-1}(\widetilde{\beta}, \gamma) \widetilde{\Omega}^{-1} \Psi_{t-1}(\widetilde{\beta}, \gamma) \right\|$

$+ \sup_{\gamma \in \Gamma^*} \left\| \frac{1}{T} \sum_t \sqrt{T}(\widetilde{\beta} - \beta)' \frac{1}{\sqrt{T}} X_{2,t-1} Z'_{t-1}(\beta_0) \otimes \Psi_{t-1}(\widetilde{\beta}, \gamma) \widetilde{\Omega}^{-1} \Psi_{t-1}(\widetilde{\beta}, \gamma) \right\|$

$+ \sup_{\gamma \in \Gamma^*} \left\| \frac{1}{T} \sum_t Z_{t-1}(\beta_0) \frac{1}{\sqrt{T}} X'_{2,t-1} \sqrt{T}(\widetilde{\beta} - \beta) \otimes \Psi_{t-1}(\widetilde{\beta}, \gamma) \widetilde{\Omega}^{-1} \Psi_{t-1}(\widetilde{\beta}, \gamma) \right\|$

$+ \sup_{\gamma \in \Gamma^*} \left\| \frac{1}{T} \sum_t Z_{t-1}(\beta_0) Z'_{t-1}(\beta_0) \otimes \left(\Psi_{t-1}(\widetilde{\beta}, \gamma) - \Psi_{t-1}(\beta_0, \gamma)\right) \widetilde{\Omega}^{-1} \Psi_{t-1}(\widetilde{\beta}, \gamma) \right\|$

$+ \sup_{\gamma \in \Gamma^*} \left\| \frac{1}{T} \sum_t Z_{t-1}(\beta_0) Z'_{t-1}(\beta_0) \otimes \Psi_{t-1}(\beta_0, \gamma) \widetilde{\Omega}^{-1} \left(\Psi_{t-1}(\widetilde{\beta}, \gamma) - \Psi_{t-1}(\beta_0, \gamma)\right) \right\|$

$+ \sup_{\gamma \in \Gamma^*} \left\| \frac{1}{T} \sum_t Z_{t-1}(\beta_0) Z'_{t-1}(\beta_0) \otimes \Psi_{t-1}(\beta_0, \gamma) \left(\widetilde{\Omega}^{-1} - {\Omega_0}^{-1}\right) \Psi_{t-1}(\beta_0, \gamma) \right\|$

Now we need to show that each term is $op(1)$.

Using the properties of the trace of Kronecker product $tr\{A \otimes C\} = tr\{A\} tr\{C\}$ and of

Frobenius norm for matrices $\|AC\| \leq \|A\| \|C\|$, the first term can be rearranged as:

$$\left\| \frac{1}{T} \sum_t \sqrt{T}(\widetilde{\beta} - \beta)' \sqrt{T}(\widetilde{\beta} - \beta) \right\| \left\| \frac{1}{T} \sum_t \frac{1}{\sqrt{T}} X_{2,t-1} \frac{1}{\sqrt{T}} X_{2,t-1}' \right\| \sup_{\gamma \in \Gamma^*} \left\| \frac{1}{T} \sum_t \Psi_{t-1}(\widetilde{\beta}, \gamma) \widetilde{\Omega}^{-1} \Psi_{t-1}(\widetilde{\beta}, \gamma) \right\| \tag{16}$$

which is $op(1)$ since the first term in (16) is $op(1)$ under Assumption 1.3, the second is $Op(1)$ by Assumption 1.6. and CMT applied to continuous functional $x \to \int_0^1 \|x(u)x(u)'\| \, du$, and the last term is bounded using Assumption 1.4. $\sup_{\gamma \in \Gamma^*} \left\| \Psi_{t-1}(\widetilde{\beta}, \gamma) \right\| \in [0, r]$ and the fact that under Assumption 1.3. $\left\| \widetilde{\Omega} - \Omega \right\| \xrightarrow{p} 0$, which implies that under Assumption 1.1. maximum eigenvalue $\lambda_{\max}\left(\widetilde{\Omega}^{-1}\right) \leq K_6$ with probability approaching one.

By the same norm properties, applying C-S and noting that $X_{2,t-1}' X_{2,t-1}$and $Z_{t-1}' Z_{t-1}$are scalars, the second and third term can be rearranged to obtain:

$$\left\| \frac{1}{T} \sum_t \sqrt{T}(\widetilde{\beta} - \beta)' \right\| \left\| \frac{1}{T} \sum_t \frac{1}{\sqrt{T}} X_{2,t-1} \right\| \left\| \frac{1}{T} \sum_t Z_{t-1}'(\beta_0) \right\| \sup_{\gamma \in \Gamma^*} \left\| \frac{1}{T} \sum_t \Psi_{t-1}(\widetilde{\beta}, \gamma) \widetilde{\Omega}^{-1} \Psi_{t-1}(\widetilde{\beta}, \gamma) \right\| \tag{17}$$

which is $op(1)$ since the first term in (17) is $op(1)$ under Assumption 1.3, the second is $Op(1)$ by Assumption 1.6. and CMT applied to mapping $x \to \int_0^1 \|x(u)\| \, du$, the third is $Op(1)$ by the law of large numbers for ergodic processes under Assumptions 1.1. and 1.2. and the boundeness of the last is established above.

Similarly, the fourth and fifth term can be rearranged as:

$$\left\| \frac{1}{T} \sum_t Z_{t-1}(\beta_0) Z_{t-1}'(\beta_0) \right\| \sup_{\gamma \in \Gamma^*} \left\| \frac{1}{T} \sum_t \left(\Psi_{t-1}(\widetilde{\beta}, \gamma) - \Psi_{t-1}(\beta_0, \gamma) \right) \right\| \left\| \frac{1}{T} \sum_t \widetilde{\Omega}^{-1} \Psi_{t-1}(\widetilde{\beta}, \gamma) \right\| \tag{18}$$

The first and the last term in (18) are $Op(1)$ by the law of large numbers for ergodic processes under Assumptions 1.1. and 1.2. for the former and previous argument for the latter. Consider only the remaining part. By the mean value theorem, C-S and using Assumption 1.4:

$$\begin{aligned}
&\sup_{\gamma \in \Gamma^*} \left\| \tfrac{1}{T} \textstyle\sum_t \left(\Psi_{t-1}(\widetilde{\beta}, \gamma) - \Psi_{t-1}(\beta_0, \gamma) \right) \right\| = \sup_{\gamma \in \Gamma^*} \left\| \tfrac{1}{T} \textstyle\sum_t \nabla_z \Psi(\overline{Z}_{t-1}, \gamma) \left(\widetilde{Z}_{t-1} - Z_{t-1} \right) \right\| \\
&\leq \sup_{\gamma \in \Gamma^*} \left\| \tfrac{1}{T} \textstyle\sum_t \nabla_z \Psi(\overline{Z}_{t-1}, \gamma) \right\| \left\| \tfrac{1}{T} \textstyle\sum_t \left(\widetilde{Z}_{t-1} - Z_{t-1} \right) \right\| \\
&\leq K_7 \left\| \tfrac{1}{T} \textstyle\sum_t Z_{t-1} \right\| \left\| \tfrac{1}{T} \textstyle\sum_t \sqrt{T}(\widetilde{\beta} - \beta)' \tfrac{1}{\sqrt{T}} X_{2,t-1} \right\| \\
&\leq K_7 \left\| \tfrac{1}{T} \textstyle\sum_t Z_{t-1} \right\| \left\| \tfrac{1}{T} \textstyle\sum_t \sqrt{T}(\widetilde{\beta} - \beta)' \right\| \left\| \tfrac{1}{T} \textstyle\sum_t \tfrac{1}{\sqrt{T}} X_{2,t-1} \right\|
\end{aligned}$$

which is $op(1)$ by the previous arguments. Finally, the last term above is $op(1)$ by the same arguments. ■

References

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

Andrews, D.W.K. (1993): "An Introduction to Econometric Applications of Empirical Process Theory for Dependent Random Variables", *Econometric Reviews,12(2),* 183-216.

Andrews, D.W.K. (1994): "Empirical Process Methods" in *Handbook of Econometrics 4,* Chapter 37, North-Holland, New York.

Andrews, D.W.K. and W.Ploberger (1994): "Optimal Test when a Nuisance Parameter is Present Only Under Alternative", *Econometrica 62,* 1383-1414.

Balke, N.S. and T.Fomby (1997): "Threshold Cointegration", *International Economic Review, 38,* 627-645.

Campbell, J. and A. Kyle (1993): "Smart Money, Noise Trading and Stock Price Behaviour", *Review of Economic Studies 60,* 1-34.

Cheung, Y, M. Chinn, and A.Garcia-Pasqual (2005), "Empirical Exchange Rate Models of the Nineties: Are any Fit to Survive", *Journal of International Money and Finance 24*, 1150-1175;

Cho, J-S. and H. White (2007): "Testing for Regime Switching", *Econometrica 75(6),* 1671-1720.

Chen, S-L. and J-L. Wu (2005): "Long-run Money Demand Revisited: Evidence from a Nonlinear Approach", *Journal of International Money and Finance 24,* 19-37.

De Jong, R.M. (2002): "Nonlinear Minimization Estimators in the Presence of Cointegrating Relations": *Journal of Econometrics 110,* 241-259.

De Long, J. B., A. Shleifer, L.H Summers and R.J. Waldmann (1990), "Noise Trader Risk in Financial Markets", *Journal of Political Economy, 98*, 703–738.

Doukhan, P, P. Massart and E.Rio (1995): "Invariance Principles for Absolutely Regular Empirical Processes", *Annales de l'l. H.Poincare, section B, 31(2),* 393-427.

Doukan, P. (1994): "Mixing", Springer-Verlag.

Engel, C. and K. West (2005), "Exchange Rates and Fundamentals", *Journal of Political Economy 113(3)*, 485-517;

Engle R and K. Kronner (1995): "Multivariate Simultaneous Generalized GARCH", *Econometric Theory 11,* 122-150.

Feigin, P.D. and R.L. Tweedie (1985): "Random Coefficient Autoregressive Process: A Markov Chain Analysis of Stationarity and Finiteness of Moments", *Journal of Time Series Analysis 6,* 1-14.

Granger, C.W.J and T.Terasvirta (1993): *Modelling Nonlinear Dynamic Relationships,* Oxford University Press.

Goncalves, S. and L. Killian (2004): "Bootstrapping Autoregressions with Conditional Heteroscedasticity of Unknown Form": *Journal of Econometrics 123,* 89-120.

Gonzalo, J. and J.Y Pitarakis (2006): "Threshold Effects in Multivariate Error Correction Models" in *Palgrave's Handbook of Econometrics,* Chapter 15.

Hansen, B. (1996): "Inference when the Nuisance Parameter is not Identified Under the Null Hypothesis", *Econometrica 64,* 413-430.

Hansen, B. and B.Seo (2002): "Testing for Two-Regime Threshold Cointegration in Vector Error-Correction Models": *Journal of Econometrics 110, 2*93-318.

Hong, H, and J.C. Stein (1999), "A unified theory of underreaction, momentum trading, and overreaction in asset markets", *Journal of Finance, 54,* 2143–2184.

Kapetanios, G, Y. Shin and A. Snell (2006): "Testing for Cointegration in Nonlinear Smooth Transition Error Correction Models", *Econometric Theory 22,* 279-303.

Kreiss, J.P. (1997): "Asymptotic Properties of Residual Bootstrap for Autoregressions", Institute for Mathematical Stochastics, Technical University of Braunschweig.

Kristensen, D. and A. Rahbek (2008, in press): "Likelihood Based Inference in Nonlinear Error Correction Models", *Journal of Econometrics.*

Luukkonen, R, P. Saikkonen, and T. Terasvirta (1988): "Testing linearity against smooth transition autoregressive models", *Biometrika 75,* 491–499.

Magnus, J. R, and H. Neudecker (2007): *Matrix Differential Calculus with Applications in Statistics and Econometrics,* John Wiley & Sons.

Newey, W.K (1991): "Uniform Convergence in Probability and Stochastic Equicontinuity", *Econometrica 59,* 1161-1167.

Rothman, P, D. van Dijk, and P.H. Franses (2001): "Multivariate STAR Analysis of the Relationship between Money and Output", *Macroeconomic Dynamics 5,* 506-532.

Saikkonen, P. (1995): "Problems with the Asymptotic Theory of Maximum Likelihood Estimation in Integrated and Cointegrated Systems", *Econometric Theory 11,* 888-911.

Saikkonen, P. (2005): "Stability Results for Nonlinear Error Correction Models", *Journal of Econometrics, 127,* 69-81.

Saikkonen, P. (2007): "Stability of Mixtures of Vector Autoregressions with Autoregressive Conditional Heteroscedasticity", *Statistica Sinica 17,* 221-239.

Saikkonen, P. (2008, in press) "Stability of Regime Switching Error Correction Models under Linear Cointegration", *Econometric Theory.*

Seo, B. (2004): "Testing for Nonlinear Adjustment in Smooth Transition Vector Error Correction Models", Texas A&M University, mimeo.

Swanson, N. (1998): "Finite Sample Properties of a simple LM Test for Neglected Nonlinearity in Error Correcting Regression Equations", *Statistica Neerlandica 53,* 76-95.

Taylor, M.P, and A.M. Taylor (2004): "Purchasing Power Parity Debate", NBER Working Paper No. 10607.

Taylor, M.P, and D.A. Peel (2000): "Nonlinear adjustment, long-run equilibrium and exchange rate fundamentals", *Journal of International Money and Finance 19(1),* 33-53.

Terasvirta, T. (2007): "Forecasting Economic Variables with Nonlinear Models" in *Handbook of Forecasting,* Nort-Holland, New York.

Van Dijk, D, T. Terasvirta and P.H. Franses (2002): "Smooth Transition Autoregressive Models – A Survey of Recent Developments", *Econometric Reviews 21(1),* 1-47.

White, H. (1994): *Estimation, Inference and Specification Analysis,* Cambridge University Press.

White, H. (2001): *Asymptotic Theory for Econometricians,* Academic Press.