Tests for Unit-Root Versus Threshold Specification With an Application to the Purchasing Power Parity Relationship

Frédérique BEC
CREST–ENSAE, Malakoff, France, F-92245

Mélika BEN SALEM
OEP, University of Marne-la-Vallée, France, F-77454

Marine CARRASCO
Economics Department, University of Rochester, NY, 14627

Last version : February 2004

Abstract

We consider modeling the real exchange rate by a stationary three-regime self-exciting threshold autoregressive (SETAR) model with possibly a unit root in the middle regime. This representation is consistent with purchasing power parity in the presence of trading costs. Our main contribution is to provide statistical tools for testing unit root versus a SETAR. First, we show that a SETAR with a unit root in the middle regime is stationary and mixing under reasonable assumptions. Second, we derive analytically the asymptotic distribution of our unit-root test under the null. Using monthly real exchange rate data, our test rejects the null of unit-root against a threshold process for five European series.

Keywords: Mixing conditions; Real exchange rate; Threshold autoregressive process.

1. INTRODUCTION

Despite two decades of intensive empirical exploration, the results regarding the real exchange rate stationarity are, at best, mitigated (see the reviews by e.g. Froot and Rogoff (1985) or MacDonald (1995)). Consequently, the possibility for the Purchasing Power Parity (hereafter PPP) deviations to follow a random walk cannot be ruled out. In this case, national price levels have no tendency to be equal, even in the long-run. Many explanations of temporary PPP deviations have been suggested, such as nominal rigidities, tariffs and shipping costs. However, in a world where a large range of individual goods are traded across countries, it is difficult to understand why arbitrage in the goods market would not make domestic and foreign price levels come closer to one another, at least in the long-run.

Three theoretical arguments are often put forward to explain this phenomenon. The first one relies on the so-called Balassa-Samuelson effect, see Balassa (1964) and Samuelson (1964). These authors conjectured that rich countries experience higher price levels than poor ones not only because their productivity is higher, but also because the former are relatively stronger than the latter in the traded goods sector. A related theory is developed by Kravis and Lipsey (1983) and Bhagwati (1984), according to which the discrepancy between rich and poor countries exchange-rate adjusted price levels is due to higher capital-labor ratio in the former. This explanation is not very convincing for countries with similar productivity levels. A related argument is put forward by Engel (2000) who attributes the failure of PPP to the presence of a nonstationary nontradable relative component in the price index. Nevertheless, Ng and Perron (2002) show that this result relies on simulated data whose properties do not mimic the ones of observed real exchange rate data. The second argument is financial markets incompleteness. Bec and Hairault (1997) show that under this assumption, a two-country intertemporal stochastic general equilibrium model predicts the random walk behavior of the real exchange rate. However, this explanation is not relevant if the real exchange rate dynamics is found to be nonlinear. The third argument, first proposed by Krugman (1987), puts forward a "pricing-to-market" behavior of exporting firms in an imperfect competition setup, according to which they may not wish to pass the cost of nominal exchange rate fluctuations on to the foreign customers. However, Chang and Devereux (1998) among others find that without price-stickiness, a general equilibrium model of pricing-to-market cannot produce the observed persistence of real exchange rate movements. So, this argument is confined to the short-term analysis, and may not justify the unit-root process of the real exchange rate.

Recently, the debate has shifted from the issue of (non) stationarity of PPP deviations to the issue of modeling the real exchange rate as a nonlinear stochastic process. This specification is a natural one in the presence of trading costs for goods, as underlined by e.g. Obstfeld and Rogoff (2000). Indeed, one can show that in a two-country stochastic general equilibrium model, such costs create a region of no trade where the PPP relationship does not hold (see Appendix A for a proof). Outside this area, international goods market arbitrage provides an error correction mechanism which brings national price levels back to equality, apart from a constant term reflecting the trading costs. See for instance Sercu, Uppal and Van Hulle (1995). So, if the stationarity of the real exchange rate is interpreted in terms of cointegration between the nominal exchange rate and the national price levels, this mechanism involves some kind of threshold cointegration (as defined in Balke and Fomby (1997)) with the adjustment towards the long-run relationship being active only when the deviation exceeds a threshold value. Following this idea, Michael, Nobay and Peel (1997) and Kilian and Taylor (2003) among others estimate an exponential smooth transition autoregressive (ESTAR) model. In the same spirit, Obstfeld and Taylor (1997) fit a threshold autoregressive model (TAR) to the real exchange rate. According to these studies, the null of a linear autoregressive process is rejected against the nonlinear alternative.

This theoretical approach together with the recent developments in nonlinear econometrics leads us to reconsider the very nature of the debate: Is the real exchange rate process really integrated of order one? In other words, are the data really incompatible with (a soft version of) PPP? The Dickey-Fuller test has been shown to lack power against a nonlinear alternative by e.g. Pippenger and Goering (1993), Pippenger and Goering (2000), and Taylor (2001). This leads us to devise a specific test for unit root versus a stationary Self Exciting TAR (SETAR) model. Recently, a stream of research has focused on models where the threshold variable is stationary and differs from the dependent variable — see e.g. Caner and Hansen (2001), Shin and Lee (2001) or Gonzalez and Gonzalo (1998). This approach is of reduced interest for studying the PPP puzzle. Testing unit-root versus a threshold alternative where the threshold variable is the same as the dependent variable is both better adapted to the issue and more challenging. Enders and Granger (1998) and Lo and Zivot (2001) study by Monte Carlo the empirical properties of such tests but they do not provide a theoretical justification. Our main contribution is to provide such a justification.

First, we show that in a three-regime SETAR model allowing for general autoregressive orders, the process may be stationary and mixing even though there is a unit or explosive root in the middle regime. Although this result was known for some special cases as the SETAR(1) model, it is new for a general SETAR(p). We follow the idea of Enders and Granger (1998) and test the null hypothesis of a unit-root versus a threshold process. Instead of a F-test for an arbitrary threshold value as in Enders and Granger, we propose a supWald test allowing for an unknown threshold value. We derive analytically its asymptotic distribution and show that it is nuisance parameter free. This test is specifically designed for our alternative and, as we demonstrate in a small Monte Carlo experiment, has more power than the usual ADF test.

We apply our test to two sets of monthly real exchange rate data. The first set includes the G7 currencies apart from Japan vis-à-vis the US dollar. For these five real exchange rates, our test does not reject the null of a unit-root process, thus confirming the ADF test conclusions. The second set includes a panel of European currencies, namely the French franc, lira, Belgian franc, Dutch guilder, peseta and escudo vis-à-vis the Deutschmark. Here, our test rejects the null for five series out of six whereas the ADF rejects the unit-root hypothesis for the French franc only.

The rest of the paper is structured as follows. In Section 2, we describe the methodological framework used in our empirical work and study the properties of a three-regime TAR model. Section 3 presents the tests. Section 4 describes the data and presents the results. Section 5 concludes our analysis.

2. THE SETAR(p) MODEL

2.1 Presentation

The model under study is the following SETAR(p):

$$y_t = \begin{cases} a_{10} + a_{11}y_{t-1} + ... + a_{1p}y_{t-p} + \eta_t & \text{if } y_{t-d} \leq \lambda_1 \\ a_{20} + a_{21}y_{t-1} + ... + a_{2p}y_{t-p} + \eta_t & \text{if } \lambda_1 < y_{t-d} \leq \lambda_2 \\ a_{30} + a_{31}y_{t-1} + ... + a_{3p}y_{t-p} + \eta_t & \text{if } y_{t-d} > \lambda_2 \end{cases} \tag{1}$$

where p and d are any positive integers and $-\infty < \lambda_1 < \lambda_2 < +\infty$. This model is quite general. Indeed, it allows for a general autoregressive order p intended to account for possible serial correlation. Moreover, all the slope coefficients are allowed to switch between the regimes. As such, model (1) includes a lot of constrained versions, either linear or not. For example, the stationary AR(p) model corresponds to the case where $a_{ij} = a_j$, for $i = 1, 2, 3$ and the roots of the characteristic polynomial $1 - a_1 z - \cdots - a_p z^p = 0$ lie outside the unit circle; the random walk with or without drift is obtained by setting $a_{ij} = a_j$, for $i = 1, 2, 3$ and one root of the characteristic polynomial lies on the unit circle. It also includes the Equilibrium-TAR, Band-TAR and Returning Drift-TAR models presented in Balke and Fomby (1997). For instance, the Band-TAR(1) model studied by Taylor (2001) among others is obtained by imposing $\lambda_1 = -\lambda_2$, $a_{11} = a_{31}$, $a_{21} = 1$ and $a_{10} = -a_{30} = -\lambda_2(1 - a_{11})$.

We address two major questions: *(i)* Is y_t stationary? *(ii)* Is the model linear? The order in which these questions are addressed is essential. Indeed one should not try to answer the second one before establishing stationarity. The tests of linearity versus a threshold alternative proposed by Andrews and Ploberger (1994) and Hansen (1996) assume that the series is stationary and even absolutely regular (also named $\beta-$mixing). The distribution of the linearity tests given by these authors would be different if the process were a random walk under the null. For this reason, we adopt the following two-step strategy:

1st step: Establish whether the process is H_0: "unit root". To this end, we propose a test that is specifically designed for a threshold alternative.

2nd step: Once stationarity is established, estimate the SETAR(p) and test the linearity hypothesis using Hansen (1996).

2.2 Stationarity and Mixing Properties

Because our aim is to test unit root versus stationary threshold model, it is essential to know under which conditions Model (1) is stationary. Moreover, Bierens and Guo (1993) stationarity tests require that the process be strong mixing (also called $\alpha-$mixing) with a fast enough decay rate. The conditions given in this subsection guarantee the applicability of their tests. To prove our result, we need the following assumption on the distribution of η_t.

Assumption η : η_t is i.i.d.$(0, \sigma^2)$, independent of y_0. η_t has a distribution that is absolutely continuous and its density is positive everywhere.

Note that Assumption η is satisfied if η_t is i.i.d. $\mathcal{N}(0, \sigma^2)$. Theorem 1 below gives a sufficient condition for geometric ergodicity of the Markov process $X_t = (y_t, y_{t-1}, ..., y_{t-m+1})$ where $m = \max(p, d)$. Geometric ergodicity implies the existence of a (unique) stationary distribution of X_t and that X_t becomes stationary exponentially fast when initialized from an arbitrary value. It also implies that once initialized from its stationary distribution, X_t is stationary and $\beta - mixing$ with geometric decay. As the notion of geometric ergodicity applies only to Markov (of order 1) processes, we can not say that y_t itself is geometrically ergodic. However as y_t is a component of X_t, it inherits some of its properties, in particular, y_t is $\beta - mixing$ with geometric decay. Denote A_i, $i = 1, 2, 3$ the $m \times m-$matrices that appear in the recursion of X_t. For $m > p$, we define

$$A_i = \begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{ip-1} & a_{ip} & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \ddots & & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & & \ddots & \ddots & 0 & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 1 & 0 \end{bmatrix}. \tag{2}$$

For $m \leq p$, we define

$$A_i = \begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{ip-1} & a_{ip} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}. \tag{3}$$

Let $\rho(Q)$ be the largest eigenvalue in absolute value of the square matrix Q. As A_i has the specific form (2) and (3), $\rho(A_i)$ is actually the smallest root in absolute value of the characteristic polynomial $1 - a_{i1}\delta - a_{i2}\delta^2 ... - a_{ip}\delta^p = 0$.

The proof of Theorem 1 relies on the Schur decomposition (see Horn and Johnson (1985)) of the matrix A_1. By the Schur triangularization theorem, there exists a unitary matrix U and an upper triangular matrix Δ such that $A_1 = U\Delta U^{-1}$, and the diagonal elements of Δ are $(\delta_{11}, \delta_{21}, ..., \delta_{m1})$, the eigenvalues of A_1. Suppose that the eigenvalues are ordered so that $\rho(A_1) = |\delta_{11}| \geq |\delta_{21}| \geq ...$ Let $D_t \equiv diag(t, t^2, ..., t^m)$ and define the norm

$$\|B\|_{M_1} = \left\| D_t U^{-1} B U D_t^{-1} \right\|_1 \tag{4}$$

where $\|Q\|_1 = \max_{j=1,...,m} \sum_{i=1}^{m} |q_{ij}|$ for any $m \times m-$matrix $Q = [q_{ij}]$. Remark that $\|.\|_{M_1}$ is a matrix norm (in the sense of Horn and Johnson, 1985, page

290); it is a function of t and A_1 (through U). Moreover it is known that for any $\varepsilon > 0$, and for t sufficiently large, the following inequality holds

$$\|A_1\|_{M_1} \leq \rho(A_1) + \varepsilon.$$

Theorem 1 *Suppose that $\{y_t\}$ satisfies (1) and η_t satisfies Assumption η. Let $m = \max(d, p)$. Assume that there exists a value of t such that*

(a) for $d \geq p$,

$$\max_{\pi_{1i},\pi_{2i}} \left\| \prod_{i=1}^{m} A_1^{\pi_{1i}} A_2^{\pi_{2i}} A_3^{1-\pi_{1i}-\pi_{2i}} \right\|_{M_1} < 1 \tag{5}$$

with $\sum_{i=1}^{m} \pi_{2i} < m$, where $\pi_{1i} = 1$ or 0, $\pi_{2i} = 1$ or 0, and $\pi_{1i}\pi_{2i} = 0$.

(b) for $d < p$, besides (5), the following is true

$$\|A_2^p\|_{M_1} \leq 1. \tag{6}$$

If, moreover, $\{y_t\}$ is initialized from its stationary distribution, then $\{y_t\}$ is stationary $\alpha-$ and $\beta-$mixing with geometric decay.

Conditions (5)-(6) are not obvious, however they are easy to verify on a computer. The command "schur" in GAUSS and Matlab permits one to compute U and hence the matrix norm $\|.\|_{M_1}$. As $\rho(Q) \leq \|Q\|_{M_1}$, Condition (5) requires $\rho(A_1) < 1$, $\rho(A_3) < 1$, $\max_{i,j} \rho\left(A_1^i A_2^j A_3^{d-i-j}\right) < 1$ with $i \leq d$ and $j < d$. However, these latter conditions might not be sufficient. As it will be illustrated shortly, (5) does not exclude the possibility of an explosive root in the middle regime. This means that the smallest root in absolute value of the characteristic polynomial in the middle regime, that is $1 - a_{21}\delta - a_{22}\delta^2 ... - a_{2p}\delta^p = 0$, may be equal to 1 (unit root) or less than 1 (explosive root) while the process is globally stationary. For $d < p$, an explosive root is ruled out by assumption (6) but the possibility of a unit root in the middle regime remains. The following corollary illustrates Conditions (5)-(6) on a simple example. Its proof, given in Appendix B, highlights the basic properties of Schur triangularization.

Corollary 1 *Assume $d = p = 2$ and $A_1 = A_3$. Denote δ_{12} and δ_{22} the eigenvalues of A_2 so that $\rho(A_2) = |\delta_{12}| \geq |\delta_{22}|$. Assume moreover $\delta_{22} = \delta_{11}$. Then, Condition (5) is satisfied as soon as*

$$\rho(A_1) < 1 \text{ and } |\delta_{21}| \rho(A_2) < 1.$$

This simple example shows that the root in the middle regime may be explosive as long as the second root of the outside regime compensates for it. Note that since $|\delta_{21}| \leq |\delta_{11}|$, the condition $|\delta_{21}| \rho(A_2) < 1$ is a weaker requirement than $\rho(A_1) \rho(A_2) < 1$. In the case where $\delta_{21} = 0$, $\rho(A_2)$ can be arbitrarily large and the process is still stationary (as long as $\rho(A_1) < 1$).

We now briefly discuss previous results. Chen and Tsay (1991) give necessary and sufficient conditions for the ergodicity of the two-regime TAR(1) and Chan, Petruccelli, Tong and Woolford (1985) give necessary and sufficient conditions for a multiple-regime TAR(1) with $d = 1$. However, necessary and sufficient conditions for TAR(p) are not yet available. Chan and Tong (1985) give a sufficient condition when $d \leq p$, namely $\max_i \sum_{j=1}^{p} |a_{ij}| < 1$, that is much stronger than our Conditions (5)-(6). Note that the model considered by Tjøstheim (1990, Theorem 4.4) is different from Model (1) because he assumes that the change of regime is dictated by all the lag values $(y_{t-1}, y_{t-2}, ..., y_{t-p})$.

2.3 A Model for the Real Exchange Rate Dynamics

The dynamics of the real exchange rate in the presence of trading costs for goods provides a simple illustration of the usefulness of this kind of threshold autoregression. Such costs generate a region within which the PPP deviations are not corrected because international arbitrage does not occur. In Appendix A, we develop a basic two-country one-sector model with shipping costs, like the one of Sercu et al. (1995), and show that such a model implies a threshold process for the real exchange rate. It is characterized by the existence of a region with no trade, $[-\lambda, \lambda]$, where λ corresponds to the proportional transaction costs. Outside this region, the real exchange rate is stationary. Therefore, a three-regime SETAR model seems particularly well-suited to model the real exchange rate $y_t = \ln(e_t) + \ln(p_t^*) - \ln(p_t)$. From now on, the threshold autoregression we will focus on is the following version of Model (1):

$$\Delta y_t = \begin{cases} \alpha_{11}\Delta y_{t-1} + ... + \alpha_{1p-1}\Delta y_{t-p+1} + \mu_1 + \rho_1 y_{t-1} + \varepsilon_t & \text{if } y_{t-1} \leq -\lambda \\ \alpha_{21}\Delta y_{t-1} + ... + \alpha_{2p-1}\Delta y_{t-p+1} + \mu_2 + \rho_2 y_{t-1} + \varepsilon_t & \text{if } |y_{t-1}| < \lambda \\ \alpha_{31}\Delta y_{t-1} + ... + \alpha_{3p-1}\Delta y_{t-p+1} + \mu_3 + \rho_3 y_{t-1} + \varepsilon_t & \text{if } y_{t-1} \geq \lambda \end{cases} \tag{7}$$

with ε_t iid $\mathcal{N}(0,\sigma^2)$. In accordance with the theoretical model, the band of inaction is symmetric around zero: $\lambda_1 = -\lambda_2 = -\lambda$. Within this band, PPP deviations may persist (case where $\rho_2 = 0$) because prices are too small compared to the shipping costs. Due to the lack of a precise theoretical prior about the delay parameter d, we set it to unity following existing studies. Indeed, even though the theoretical model developed in Appendix A predicts an instantaneous adjustment towards the band, a less stylized model allowing for e.g. price stickiness or time to collect information would likely imply a lagged adjustment. It would be desirable to estimate d. But, unfortunately, while d can be accurately estimated in theory, this is not the case in practice. By performing simulations (not reported here), we have found that the estimator of d is very unreliable when there is a lot of persistence in the data. This is due to the fact that y_{t-1}, y_{t-2}, y_{t-3} etc... behave very similarly. Nevertheless, Model (7) is more general than the SETAR analyzed in existing empirical studies since it allows for autoregressive lags and for the possibility that all the slope coefficients switch between regimes, and it does not impose any symmetry in the outer regimes dynamics.

3. UNIT-ROOT TESTS AGAINST A THRESHOLD ALTERNATIVE

3.1 Asymptotic Properties of the Unit-Root Tests

Denoting $\alpha_j = (\alpha_{j1}, ..., \alpha_{jp-1})'$, $j = 1, 2, 3$ and $I_{t<} = I\{y_{t-1} \leq -\lambda\}$, $\bar{I}_t = I\{|y_{t-1}| < \lambda\}$, $I_{t>} = I\{y_{t-1} \geq \lambda\}$, $u_t = \Delta y_t$ and $u_{t-1}^p = (\Delta y_{t-1}, ..., \Delta y_{t-p+1})$, Model (7) can be rewritten as:

$$u_t = x_t'\beta + \varepsilon_t \tag{8}$$

with:

$$\begin{aligned} \beta &= (\alpha_1', \alpha_2', \alpha_3', \mu_1, \rho_1, \mu_2, \rho_2, \mu_3, \rho_3)' \\ x_t &= \left(I_{t<} u_{t-1}^p, \bar{I}_t u_{t-1}^p, I_{t>} u_{t-1}^p, I_{t<}, I_{t<} y_{t-1}, \bar{I}_t, \bar{I}_t y_{t-1}, I_{t>}, I_{t>} y_{t-1}\right)'. \end{aligned}$$

In this model, one tests the null hypothesis

$$H_0 : \rho_1 = \rho_2 = \rho_3 = 0.$$

But we derive the distribution of our tests under the stronger assumption

$$H_0' : \rho_1 = \rho_2 = \rho_3 = 0 = \mu_1 = \mu_2 = \mu_3 \text{ and } \alpha_1 = \alpha_2 = \alpha_3.$$

Under H_0', the model is a random walk without drift, moreover, it is assumed that the roots of $1 - \alpha_{11} z ... - \alpha_{1p-1} z^{p-1} = 0$ lie outside the unit circle. The reason why we test H_0 and not H_0' is that testing H_0' is testing linearity in addition to unit root. We might reject H_0' because the model is nonlinear but still nonstationary. It is not uncommon in the unit-root literature to test a weaker restriction than the assumption under which the distribution of the test is derived. This is the case for example with the Dickey-Fuller t-tests described in Cases 2 and 4 of Hamilton (1994, Section 19.4).

To study the asymptotic properties of our tests, we proceed in two steps. First, we fix the threshold value λ, second we take the supremum over an interval of values for λ. First assume that λ is not estimated but is chosen a priori by the econometrician. The test statistics we consider are the usual Wald ($W_T(\lambda)$), Lagrange Multiplier ($LM_T(\lambda)$), and Likelihood Ratio ($LR_T(\lambda)$) tests. Let $\hat{\beta}$ be the OLS estimator of β in the unrestricted regression (8), $\hat{\varepsilon}_t = u_t - x_t'\hat{\beta}$, and $\hat{\sigma}^2 = \sum_{t=1}^T \hat{\varepsilon}_t^2 / T$. Let $\tilde{\beta}$ be the restricted OLS estimator of β in (8) with the constraint $\rho_1 = \rho_2 = \rho_3 = 0$, denote

$\tilde{\varepsilon}_t = u_t - x_t'\tilde{\beta}$, and $\tilde{\sigma}^2 = \sum_{t=1}^{T} \tilde{\varepsilon}_t^2/T$. We have

$$\begin{aligned}
W_T(\lambda) &= \frac{1}{\hat{\sigma}^2}\hat{\rho}'\left[R\left(\sum_{t=1}^{T} x_t x_t'\right)^{-} R'\right]^{-}\hat{\rho}, \\
LM_T(\lambda) &= \frac{1}{\tilde{\sigma}^2}\left[\sum_{t=1}^{T} x_t\tilde{\varepsilon}_t\right]'\left[\sum_{t=1}^{T} x_t x_t'\right]^{-}\left[\sum_{t=1}^{T} x_t\tilde{\varepsilon}_t\right], \\
LR_T(\lambda) &= T\ln\left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2}\right)
\end{aligned}$$

where $\hat{\rho} = (\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3)'$ and R is the $3 \times (3p+6)$-selection matrix so that $R\hat{\beta} = \hat{\rho}$. The notation Q^- denotes the Moore-Penrose generalized inverse of the matrix Q. It is possible that no data lie in one of the regimes and hence the matrix $\sum_{t=1}^{T} x_t x_t'$ is singular.

Below we consider the limiting behavior of the test statistics computed for a specific λ.

Theorem 2 *Assume that λ is chosen a priori so that $\lambda = \sqrt{T}\pi > 0$. Under H_0', we have*

$$W_T(\lambda), LM_T(\lambda), LR_T(\lambda) \xrightarrow{L} T(k) \sim \sum_{j=1}^{3} \frac{(N_j)^2}{D_j \int_0^1 I_j(r)\,dr}$$

with $k = \pi/\delta$, $\delta = \sigma/(1 - \alpha_{11} - \alpha_{12} - ... - \alpha_{1p-1})$ and

$$\begin{aligned}
&I_1(r) = I(B(r) \leq -k),\; I_2(r) = I(|B(r)| < k),\; I_3(r) = I(B(r) \geq k), \\
&N_j = \textstyle\int_0^1 I_j(r)\,dr \int_0^1 I_j(r)\,B(r)dB(r) - \int_0^1 B(r)\,I_j(r)\,dr \int_0^1 I_j(r)\,dB(r), \\
&D_j = \textstyle\int_0^1 I_j(r)\,dr \int_0^1 B^2(r)I_j(r)\,dr - \left[\int_0^1 B(r)I_j(r)\,dr\right]^2,\; j = 1, 2, 3,
\end{aligned}$$

where $B(.)$ denotes a standard Brownian motion on $[0,1]$. The convention 0/0=0 is used.

Assume now that the DGP is a stationary SETAR model with unknown (fixed) threshold λ_0 and that the Wald and Lagrange Multiplier and Likelihood Ratio statistics are computed for an arbitrary fixed value of λ (not necessarily equal to λ_0) then we have

$$W_T(\lambda), LM_T(\lambda), LR_T(\lambda) \xrightarrow{P} \infty.$$

The proof of Theorem 2 is given in Appendix B. Remark that the distribution $T(k)$ depends on the nuisance parameters λ, σ, α_{11}, ..., α_{1p} only through the ratio $k = \pi/\delta$. Note that $P[I_1(r) = 0, \forall r \in [0,1]] > 0$ and $P[I_3(r) = 0, \forall r \in [0,1]] > 0$. If $I_j(r) = 0, \forall r \in [0,1]$ then $D_j = 0 = N_j$, hence we obtain 0/0 in $T(k)$. The generalized inverse imposes the convention 0/0=0.

Now we will justify choosing $\lambda = \sqrt{T}\pi$. Note that this assumption is relevant only under the null of a unit root since under the alternative we take λ fixed. Under this assumption and when the data follows a non-stationary process, the probability that a given observation lies in Regime 2 is

$$P(|y_t| \leq \lambda) = P\left(\frac{|y_t|}{\sqrt{T}} \leq \frac{\lambda}{\sqrt{T}}\right) \to P[|B(r)| \leq \pi/\delta]$$

with $r = t/T$. On the other hand, if λ is fixed, this probability becomes $P[|B(r)| \leq 0] = 0$ for $r > 0$. Hence, the assumption $\lambda/\sqrt{T} = \pi$ guarantees that the probability of being in Regime 2 is positive at all times. The same type of assumption is made in the Structural Change literature where it is assumed that a jump occurs at a time $T\pi$ with $\pi \in (0,1)$. This guarantees that the proportions of observations before and after the jump are constant.

In practice, λ is unknown. Moreover, under the null hypothesis, it is not identified and hence can not be estimated consistently. We suggest to use sup tests of the type

$$\text{supLR} \equiv \sup_{\lambda \in [\underline{\lambda}_T, \bar{\lambda}_T]} LR_T(\lambda).$$

Similarly we use the notation supLM and supW for the Lagrange Multiplier and Wald tests. These tests are directly inspired by Davies (1987) sup test for testing the presence of a structural change. To guarantee that the sup tests have a distribution that is nuisance parameter free, we need to choose the interval $[\underline{\lambda}_T, \bar{\lambda}_T]$ in an appropriate manner. First we order the absolute values of the data by increasing value: $|y|_{(1)} < |y|_{(2)} <...< |y|_{(T)}$ and select $\underline{\lambda}_T = |y|_{([10T/100])}$ and $\bar{\lambda}_T = |y|_{([90T/100])}$. This choice guarantees that at least 20% of the observations lie outside and inside the band, so that the estimated

SETAR is not driven by, say, a few important outliers. The limits of the normalized percentiles, denoted as

$$\begin{aligned} \underline{k} &\stackrel{d}{=} \lim_{T\to\infty} \frac{|y|_{[10T/100]}}{\delta\sqrt{T}} &= k_{0.1}, \\ \bar{k} &\stackrel{d}{=} \lim_{T\to\infty} \frac{|y|_{[90T/100]}}{\delta\sqrt{T}} &= k_{0.9}, \end{aligned}$$

satisfy

$$\int_0^1 I\left[|B(r)| \leq k_p\right] dr = p, \;\; p \in [0,1].$$

As a result the asymptotic distributions of the sup tests are pivotal, that is, nuisance parameter free.

Theorem 3 *Under H_0', we have*

$$\sup_{\lambda\in\left[\underline{\lambda}_T,\bar{\lambda}_T\right]} W_T(\lambda), \;\; \sup_{\lambda\in\left[\underline{\lambda}_T,\bar{\lambda}_T\right]} LM_T(\lambda), \;\; \sup_{\lambda\in\left[\underline{\lambda}_T,\bar{\lambda}_T\right]} LR_T(\lambda) \stackrel{L}{\to} \sup_{k\in\left[\underline{k},\bar{k}\right]} T(k),$$

moreover the distribution of $\sup_{k\in\left[\underline{k},\bar{k}\right]} T(k)$ *is pivotal.*

Since the sup tests have asymptotically pivotal distributions under the null hypothesis, we can use empirical critical values. A similar approach has been taken independently by Berben and van Dijk (1999) when testing a unit root versus a Continuous two-regime TAR(1) model. Instead of taking the maximum over λ, they take the Wald test $W_T(\lambda)$ at the value of λ corresponding to $(1-\tau)\, y_{(0)} + \tau y_{(T-1)}$ for some $\tau \in (0,1)$. This choice guarantees that λ increases with the sample size. Hence, our approach is close to theirs in spirit but is adapted to a three-regime model. More recently, Bec, Guay, and Guerre [2002] proposed to take the sup of $W_T(\lambda), LM_T(\lambda), LR_T(\lambda)$ over an interval $\left[\underline{\lambda}_T, \bar{\lambda}_T\right]$ that remains bounded under H_0. Their asymptotic theory strongly relies on techniques developed by Park and Phillips (2001) on local time.

3.2 Empirical Critical Values

In the empirical study below, we found that $p = 2$, so that our model becomes:

$$\Delta y_t = \begin{cases} \alpha_1 \Delta y_{t-1} + \mu_1 + \rho_1 y_{t-1} + \sigma\varepsilon_t, \text{ if } y_{t-1} \leq -\lambda \\ \alpha_2 \Delta y_{t-1} + \mu_2 + \rho_2 y_{t-1} + \sigma\varepsilon_t, \text{ if } |y_{t-1}| < \lambda \\ \alpha_3 \Delta y_{t-1} + \mu_3 + \rho_3 y_{t-1} + \sigma\varepsilon_t, \text{ if } y_{t-1} \geq \lambda \end{cases} \tag{9}$$

where ε_t are iid $\mathcal{N}(0,1)$. This is the model we choose to retain under H_1 in order to compute the empirical critical values. Under H_0, we generate the model

$$\Delta y_t = \alpha \Delta y_{t-1} + \mu + \sigma\varepsilon_t \tag{10}$$

where ε_t are drawn from iid $\mathcal{N}(0,1)$, $\alpha = 0.3$, $\mu = 0$ and $\sigma = 0.02$. This choice of the parameters is dictated by the data. When fitting (10) on the real exchange rates, we obtain for most of the series a value of μ very close to zero, suggesting there is no trend in the data. We also obtain estimates for α around 0.3, with a range of $0.13 \leq \alpha < 0.4$, and of σ around 0.02. In Table 1, we report the empirical critical values from 10,000 replications of samples of size 325 which corresponds to our sample length.

Table 1: Empirical critical values of the unit-root test (α=0.3, $\sigma = 0.02$)

	1%	5%	10%	15%	85%	90%	95%	99%
SupW	3.905	5.151	6.053	13.730	14.751	16.161	18.400	23.010
SupLM	3.907	5.133	6.016	13.337	14.285	15.587	17.630	21.756
SupLR	3.882	5.110	5.998	13.448	14.426	15.772	17.898	22.232

3.3 Size-Corrected Power

In the sequel, we focus on the supLR test because the LR test is known to be more reliable than the LM and Wald tests (see Dufour (1997)). To assess the performance of this test in practice, we generate the model under the null (Model (10)) for $\mu \in \{-0.1, 0, 0.1\}$ using a different seed for the random number generator from that used to compute the empirical critical values. In Table 2, we report the rejection rate from 10,000 replications with $n = 325$. For comparison purpose, the empirical size of the Augmented Dickey-Fuller statistics is also reported. The latter is also calculated using empirical critical values obtained by generating 10,000 replications of Model (10) — for the same parameter values as given above — which yields -2.628, -2.895 and -3.533 at the 10, 5 and 1% level respectively.

Table 2: Empirical size of the unit-root test (α=0.3, $\sigma = 0.02$)

Theoretical size	ADF	supLR
	$\mu = -0.01$	
1%	0.000	0.004
5%	0.004	0.021
10%	0.008	0.047
	$\mu = 0$	
1%	0.008	0.009
5%	0.048	0.047
10%	0.089	0.098
	$\mu = 0.01$	
1%	0.000	0.004
5%	0.003	0.021
10%	0.008	0.046

We see that the sup test exhibits accurate empirical sizes under the null without drift. In presence of a drift, this test is excessively conservative, but less so than the ADF test. This is a desirable property — by contrast with an excessively liberal test — since we do not want to conclude wrongly in favor of a TAR when the data actually exhibits a trend.

To assess the power of these statistics, we generate 3000 series under the alternative (9). This model is invariant to a rescaling of the variables, indeed we may replace y by y/σ, and λ by λ/σ so that σ can be normalized to unity. Due to the number of parameters involved in Model (9), we have chosen to restrict the simulation experiments to the symmetric version of the model considered by Taylor (2001), i.e. to cases where $\alpha_1 = \alpha_2 = \alpha_3 = \alpha$,

$\mu_1 = -\mu_3$ and $\rho_1 = \rho_3$. When estimated using our dataset, this constrained version of Model (9) suggests that in the outer regimes, μ_1 is roughly equal to $1.3\lambda\rho_1$. Hence, we retain this value for μ_1, set μ_2 to zero, and σ to unity, and we report below the rejection rate for a theoretical size of 5% and for various values of $(\alpha, \rho_1, \lambda)$. The choice of λ is again based on the data, since our estimates suggest that the ratio λ/σ lies between 5.6 and 9.6.

Table 3: Size-corrected power of the unit-root test ($\mu_2 = \rho_2 = 0$ and $\sigma = 1$)

$(\alpha, \rho_1, \lambda)$	ADF	supLR
(0,-0.30,10)	26.5	82.8
(0,-0.10,10)	12.6	17.1
(0,-0.05,10)	09.2	08.3
(0.3,-0.3,10)	31.7	94.1
(0.3,-0.1,10)	16.6	25.5
(0,-0.30,5)	20.2	70.4
(0,-0.10,5)	14.6	15.7
(0,-0.05,5)	11.4	09.2
(0.3,-0.3,5)	27.7	87.8
(0.3,-0.1,5)	19.2	25.8
(0,-0.30,2)	97.1	80.0
(0,-0.10,2)	41.9	21.4
(0,-0.05,2)	17.9	10.9

The power of the supLR test is increasing in α, λ and $|\rho_1|$. For the cases close to our point estimates — lines 4 and 9 of the table — the sup test clearly outperforms the ADF test. It is worth noting that the power of both the supLR and ADF tests falls dramatically as ρ_1 approaches zero. For a small threshold ($\lambda = 2$), the Dickey-Fuller test outperforms the supLR test.

4. AN APPLICATION TO THE PPP RELATIONSHIP

For this application, we use two sets of monthly real exchange rates using data from Datastream. The nominal exchange rate data are monthly averages, and the nominal price data are consumption price indices. The first set includes the Canadian dollar (CAD), sterling (GBP), Deutschmark (DEM), French franc (FRF) and lira (ITL) vis-à-vis the US dollar from 1973:09 to 2000:09. The second one includes a panel of European currencies, namely the French franc, lira, Belgian franc (BEF), Dutch guilder (NLG), peseta (ESP) and escudo (PTE) vis-à-vis the Deutschmark. For these data we consider observations up to 1998:12 only, since the Euro was introduced in January 1999. The study of these European data is motivated by the close monetary and trade links existing between European countries, which are likely to favor the PPP hypothesis.

This empirical study will be conducted as follows. First, we will perform the supLR unit-root test presented above. Since we assume that the inaction band is symmetric around zero, the real exchange rate data used for the TAR estimation are demeaned. Then, for the series which reject the null, we will test the null of linearity using the Sup Wald statistic corresponding to the hypothesis $\alpha_i = \alpha$, $\mu_i = \mu$ and $\rho_i = \rho$, $\forall i = 1, 2, 3$ in Model (9). The distribution of this test under H_0 depends on nuisance parameters, therefore the p-values are computed by simulations along the lines proposed by Hansen (1996). Finally, we will report the SETAR estimates for the non-linear stationary pairs only.

Table 4: SupLR unit-root tests

	CAD	GBP	DEM	FRF	ITL	
/USD	9.69	9.45	8.73	9.51	13.36	
	FRF	ITL	BEF	NLG	ESP	PTE
/DEM	20.35*	38.63**	26.70**	22.24**	18.01*	10.50

NOTE: ** and * denote the 1% and 5% levels resp. Data are centered.

The order of the lag polynomial in (7) is chosen according to the BIC criterion, which leads to $p = 2$ for each pair considered. Model (9) is estimated by the nonlinear least squares method. Our supLR unit-root test rejects the null for five pairs out of eleven, as can be seen from Table 4. Using the same sample, the ADF test (Dickey and Fuller (1981)) fails to reject the null for every pair but the FRF/DEM one. The PP (Phillips and Perron (1988)) and KPSS (Kwiatkowski, Phillips, Schmidt and Shin (1992)) tests both lead to the same conclusion. This finding confirms the lack of power of the ADF unit-root test for the kind of process we consider. A striking feature of these results is that the five stationary real exchange rates are European ones, namely FRF/DEM, ITL/DEM, BEF/DEM, NLG/DEM and ESP/DEM. Moreover, the Sup Wald linearity test rejects the null for these series at the 14%, 1%, 15%, 5% and 3% level respectively.

The results obtained from the estimation of Model (9) are reported in Table 5. The threshold parameter λ is estimated by minimizing the sum of squared residuals over $[\underline{\lambda}_T, \bar{\lambda}_T]$.

Basically, these results are compatible with the theoretical model involving transaction costs. Nevertheless, a careful inspection of Table 5 reveals that the results obtained for ESP/DEM must be cautiously interpreted. Indeed, due to the lack of accuracy in the estimation for this series, one cannot rule out the possibility that $\widehat{\rho}_1$ is zero.

The general form of Model (9) allows us to test restrictions often imposed *a priori* in existing empirical studies devoted to the SETAR modelling of the real exchange rates. For instance, Obstfeld and Taylor (1997) and Kapetanios and Shin (2002) impose $\rho_2 = 0$. As can be seen from Table 5, this restriction does not hold for the BEF/DEM real exchange rate, with a t-statistics rejecting the null at the 1% level. Kapetanios and Shin (2002) further impose that H_0^1: $\mu_i = 0$ for $i = 1, 2, 3$, an assumption which is clearly at odds with our data, as can be seen from the bottom part of Table 5. Model (9) also embeds the widely studied BAND-TAR model — see e.g. Balke and Fomby (1997), Obstfeld and Taylor (1997), Taylor (2001), Lo and Zivot (2001) — which imposes symmetric dynamics in the outer regimes, i.e. $\mu_1 = -\mu_3$ and $\rho_1 = \rho_3$, as well as $\alpha_i = \alpha$ for $i = 1, 2, 3$ if a lag polynomial is introduced (H_0^2), otherwise $\alpha_i = 0$ for $i = 1, 2, 3$ (H_0^3). The Wald test of H_0^2 does not reject the null for the ITL/DEM and FRF/DEM pairs only, and the Wald test

Table 5: TAR estimates

	FRF/ DEM	ITL/ DEM	BEF/ DEM	NLG/ DEM	ESP/ DEM
$\widehat{\alpha}_1$	0.249	0.430	-0.043	0.421	0.179
	(0.154)	(0.360)	(0.122)	(0.124)	(0.132)
$\widehat{\alpha}_2$	0.302	0.376	0.202	0.141	0.438
	(0.077)	(0.065)	(0.103)	(0.072)	(0.139)
$\widehat{\alpha}_3$	0.391	0.424	0.220	0.053	0.079
	(0.086)	(0.093)	(0.076)	(0.096)	(0.067)
$\widehat{\mu}_1$	-0.012	-0.023	-0.005	-0.024	0.001
	(0.007)	(0.027)	(0.003)	(0.006)	(0.004)
$\widehat{\mu}_2$	0.000	-0.000	0.001	0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.000)	(0.002)
$\widehat{\mu}_3$	0.011	0.095	0.005	0.005	0.012
	(0.005)	(0.016)	(0.002)	(0.003)	(0.004)
$\widehat{\rho}_1$	-0.191	-0.134	-0.075	-0.577	-0.001
	(0.097)	(0.049)	(0.049)	(0.132)	(0.029)
$\widehat{\rho}_2$	0.001	-0.020	-0.240	-0.005	-0.127
	(0.026)	(0.014)	(0.058)	(0.019)	(0.089)
$\widehat{\rho}_3$	-0.228	-0.560	-0.126	-0.141	-0.131
	(0.066)	(0.090)	(0.046)	(0.071)	(0.034)
$\widehat{\lambda}$	0.045	0.143	0.026	0.035	0.040
$\widehat{\sigma}$	0.010	0.015	0.007	0.005	0.019
p_1	15.56	14.24	27.48	15.23	30.46
p_2	70.20	69.54	36.75	62.91	38.08
	$H_0^1 : \mu_1 = \mu_2 = \mu_3 = 0$				
Wald	8.84	37.18	10.75	19.72	7.79
p-val.	0.03	0.00	0.01	0.00	0.05
	$H_0^2 : \mu_1 = -\mu_3,\ \rho_1 = \rho_3$ and $\alpha_1 = \alpha_2 = \alpha_3$				
Wald	3.56	6.91	9.37	11.86	14.65
p-val.	0.47	0.14	0.05	0.02	0.00
	$H_0^3 : \mu_1 = -\mu_3,\ \rho_1 = \rho_3$ and $\alpha_1 = \alpha_2 = \alpha_3 = 0$				
Wald	41.00	58.20	17.84	20.39	21.22
p-val.	0.00	0.00	0.00	0.00	0.00

NOTE: p_1 and p_2 denote the % of observations in the lower and middle regimes. Standard errors are in parentheses.

of H_0^3 rejects the null in every case. These results thus suggest that testing for a unit root against an excessively constrained nonlinear alternative might lead to fallacious conclusions. More precisely, it might lead to underreject the null even though the process is stationary.

So, apart from the escudo, all the European real exchange rates considered here appear to be nonlinear and stationary. This suggests that other forces — e.g. exchange rates or international trade agreements which have been part of the European unification project — could be important in explaining the PPP relationship, in addition to international arbitrage on goods markets.

From Figures C.1 through C.5, reported in Appendix C, it can be seen that the observations lying outside the band mainly correspond to two periods characterized by a high volatility in the exchange rates market and/or in nominal prices. The first one is the 1973-1979 period corresponding to the end of the Bretton Woods agreements and the two oil price shocks. The second one is the 1992-1996 period characterized by the 1992 and 1993 European Monetary System crisis. The sterling and lira left the EMS in September 1992 — the lira returned in 1996 after a depreciation of more than 30% vis-a-vis the Deutschemark. The Belgian franc, peseta and escudo experienced a lot of significant devaluations in early 1993. When finally the French franc came under attack during that summer, the Exchange Rate Mechanism's bands were widened to 15%. By contrast to these agitated periods, the eighties are more quiet. They are characterized by two major European decisions. The first one is the setting of the EMS in March 1979, which succeeded in stabilizing the European exchange rates for one decade, while the second one is the 1985 Delors report aiming at reinforcing the European Common Market in the perspective of the Maastricht treaty to be signed in 1992. As a result, the price differentials began to decrease between the twelve European Community members.

5. CONCLUSION

Although widely studied over the last decade, the SETAR(1) model is of little empirical interest due to the serial correlation exhibited by most economic time series. This paper provides two advances in the analysis of the more general SETAR(p) models whose economic relevance has been well-documented by e.g. Balke and Fomby (1997). First, we give sufficient conditions for a three-regime SETAR(p) to be stationary. Theorem 1 states that the stationarity of such a process is compatible with a unit root in the middle regime, as soon as the roots of the outer regimes lie outside the unit circle and some extra condition is satisfied. Second, we derive analytically the asymptotic distribution of a unit-root test against the SETAR(p) alternative. The power of this test is found to be greater than the one of the ADF test against a wide range of stationary threshold alternatives. Establishing stationarity is important as it warrants the usual properties of the SETAR(p) NLS estimates and the corresponding linearity test.

We then apply these tools to real exchange rates data in order to investigate the PPP empirical puzzle. While the standard ADF test fails to reject the null of a unit-root for all the real exchange rates but the FRF/DEM one, our supLR test rejects the null of a unit-root against the stationary three-regime TAR alternative for five out of six European pairs. Basically, our results suggest that the European Monetary System and then Monetary Unification have provided a rather strong correction mechanism. On the other hand, one cannot reject the null for the real exchange rates vis-à-vis the US dollar. A recent empirical study by Henry and Shields (2003) sheds some light on the time series properties of the US Consumer Price Index. By contrast with e.g. the Japanese or UK price indices, the US one seems to follow a linear process. This feature could explain the failure to find stationary real exchange rate vis-à-vis the US dollar.

Mainly two generalizations of our work should be the matter of future research. First, although the stationarity conditions are derived without imposing the symmetry of the thresholds, i.e. $\lambda_1 = -\lambda_2$, this assumption has been maintained when deriving the asymptotic distribution of our sup test statistics. Relaxing it constitutes a rather straightforward extension, but greatly complicates the writing of the proof. Second, we derived the result of

Theorem 2 assuming that d was known, equal to one. Extending this result to an arbitrary given value of $d > 1$ may require different tools than the ones used here, but we expect that the distribution of our unit-root test will still only depend on the nuisance parameters through k.

Acknowledgments

We are particularly grateful to E. Guerre for his valuable help and to A. Rahbek for pointing out mistakes in a previous version. We thank Eric Ghysels and two anonymous referees for their useful suggestions. We also gratefully acknowledge helpful discussions with F. Collard, A. Guay, S. Johansen, R. de Jong, L. Kilian, G. Laroque, C. Perraudin, and seminar and conference participants at Queen's University (Canada), the University of Toronto, the University of Copenhagen, the University of Paris 1 and the CREST-INSEE (France) and also at CEA (Montreal, 2001), ESEM (Lausanne, 2001), EEA (Lausanne, 2001), and the Econometric Society Meeting (Los Angeles, 2002). Of course, all errors are ours. Carrasco thanks the National Science Foundation for partial financial support under grant No. SES 0211418.

Appendix A

In this appendix, we demonstrate that a basic two-country one-sector model with shipping costs — very close to the one developed by Sercu et al. (1995) — predicts the existence of a region of no trade, in which it is possible for the real exchange rate to follow a unit-root process, whereas it is stationary elsewhere. Let us consider a pure exchange two-country model. Each country has a single representative household. The country-i household is endowed each period with q_{it} units of the same non-storable consumption good. These endowments are stochastic and their process will be described later on. Due to the single-good nature of this model, international trade takes place only to smooth consumption. There are costs to shipping the good internationally, modeled as a waste of resources. Following Obstfeld and Rogoff (1996) for instance, we assume that a given fraction of any good shipped between countries evaporates in transit. So, if one unit is shipped, only $\frac{1}{(1+\lambda)}$ units actually arrive. In what follows, indices 1 and 2 refer to domestic and foreign variables respectively, and all variables are in per capita terms unless otherwise stated. The country-i household wishes to maximize

$$E\left\{\Sigma_{t=0}^{\infty}\beta^t U(c_{it})\right\}, \qquad 0<\beta<1,$$

where c_{it} is consumption in country i in period t. The utility function and the discount factor are assumed to be identical across countries. For simplicity, we retain the logarithmic utility function $U(c_{it}) = \log c_{it}$.

Let $nx_{it} = x_{it} - z_{it}$ denote the country-i net exports, where x_{it} and z_{it} represent country-i exports and imports respectively. Since both countries are endowed with the same good, international trade is always one-sided. Exports and imports cannot be simultaneously positive in a country. In this setup, if $x_{1t} > 0$, then $z_{1t} = x_{2t} = 0$ and $z_{2t} = \frac{x_{1t}}{1+\lambda}$. The national resource constraints are given by

$$q_{it} = c_{it} + nx_{it}, \qquad \text{for } i = 1, 2. \tag{A.1}$$

Under complete and frictionless financial markets, it can easily be shown that the competitive equilibrium and the Pareto optimum coincide, so that the solution may be found by solving the following central planner problem :

$$\max_{c_{1t},c_{2t}} E\left\{\Sigma_{t=0}^{\infty}\beta^t(U(c_{1t})+U(c_{2t}))\right\}$$

subject to the national resource constraints. The absence of weights preceding the national utility functions is allowed by the assumption that the national initial wealths are identical. This intertemporal problem may be rewritten as a static one. Substituting (A.1) in the logarithmic utility functions, the social planner problem is

$$\max_{nx_{1t},nx_{2t}} \left(\log(q_{1t}-nx_{1t})+\log(q_{2t}-nx_{2t})\right).$$

If Country 1 exports then the objective function becomes

$$\max_{x_{1t}} \left(\log(q_{1t}-x_{1t})+\log(q_{2t}+\frac{x_{1t}}{1+\lambda})\right).$$

The first order condition is

$$\frac{U'(c_{1t})}{U'(c_{2t})}=\frac{1}{1+\lambda}\Leftrightarrow x_{1t}=\frac{q_{1t}-(1+\lambda)q_{2t}}{2},$$

and x_{1t} is positive if and only if:

$$\frac{q_{1t}}{q_{2t}}>1+\lambda.$$

Symmetrically, if Country 2 exports, one finds that

$$\frac{U'(c_{1t})}{U'(c_{2t})}=1+\lambda\Leftrightarrow x_{2t}=\frac{q_{2t}-(1+\lambda)q_{1t}}{2}.$$

So, x_{2t} is positive if and only if:

$$\frac{q_{1t}}{q_{2t}}<\frac{1}{1+\lambda}.$$

Therefore, in the region of no trade, the real exchange rate defined by the relative marginal utility of consumption is bounded above and below by the shipping cost :

$$\frac{1}{1+\lambda}<\frac{U'(c_{2t})}{U'(c_{1t})}=\frac{q_{1t}}{q_{2t}}<1+\lambda.$$

In this region, the stochastic behavior of the real exchange rate is entirely determined by the stochastic process underlying the national endowments. If they are integrated of order one but not cointegrated, then the real exchange rate follows a unit-root process. However, this unit root behavior is local only, since as soon as the national outputs ratio exceeds the bounds, international trade is activated and brings the real exchange rate back to the nearest bound.

Appendix B

Proof of Theorem 1. Let φ be Lebesgue measure. The following lemma gives a sufficient condition for the geometric ergodicity of a Markov chain.

Lemma 1 *Assume* $\{X_t\}$ *is an aperiodic* $\varphi-$*irreducible Markov chain,* $X_t \in R^m$*.* $\{X_t\}$ *is geometrically ergodic if there exists a norm-like function* $V : R^m \to R^+$*, a positive integer* h*, a small set* K *with complement* K^c*,* $0 < C < 1$*, and positive constants* M*,* L *such that*

$$E\left[V\left(X_{t+h}\right)|X_t = x\right] \leq M,\ x \in K, \tag{B.1}$$

$$E\left[V\left(X_{t+h}\right)|X_t = x\right] \leq CV\left(x\right) + L,\ \ x \in K^c. \tag{B.2}$$

This lemma follows from Tjøstheim (1990, p. 591) and condition DD4 in Meyn-Tweedie (1992, p.564).

To verify the inequalities (B.1) and (B.2), we use the following lemma from Horn and Johnson (1985, p. 306):

Lemma 2 *For any matrix norm* $\|.\|_M$ *, there exist (several) vector norms* $\|.\|_V$ *such that* $\|Qx\|_V \leq \|Q\|_M \|x\|_V$ *for all square matrices* Q *and vectors* x*.*

Let $m = \max(d, p)$ and $X_t = (y_t, y_{t-1}, ..., y_{t-m+1})'$. Using Lemma 1, we are going to show that X_t is geometrically ergodic. Then we use the fact that geometric ergodicity implies strong mixing with geometric decay (see Meyn and Tweedie (1992)) and $\beta-$mixing with geometric decay (Pham (1986)). Finally, the $\alpha-$ and $\beta-$mixing property of X_t implies that of y_t.

Chan and Tong (1985) show that under Assumption η (a) $\{X_t\}$ is an aperiodic $\varphi-$irreducible Markov chain and (b) all compact sets are small sets. Therefore, we just need to verify (B.1), (B.2) for an appropriately chosen $h = m$ and a compact set $K = [\lambda_1, \lambda_2]^m$.

$$
\begin{aligned}
X_{t+1} = {} & d_1 I\left(y_{t+1-d} \leq \lambda_1\right) + d_2 I\left(\lambda_1 < y_{t+1-d} \leq \lambda_2\right) + d_3 I\left(y_{t+1-d} > \lambda_2\right) + \\
& A_1 X_t I\left(y_{t+1-d} \leq \lambda_1\right) + A_2 X_t I\left(\lambda_1 < y_{t+1-d} \leq \lambda_2\right) + A_3 X_t I\left(y_{t+1-d} > \lambda_2\right) \\
& + \varepsilon_{t+1}
\end{aligned}
$$

where $\varepsilon_t = (\eta_t, 0, ...0)$, I denotes the indicator function, $d_i = (a_{i0}, 0, ..., 0)'$, $i = 1, 2, 3$, and A_i are $m \times m$ matrices defined in (2) and (3). In the following, we use the notation

$$
I_t^- = I\left(y_t \leq \lambda_1\right), \; I_t^+ = I\left(y_t > \lambda_2\right), \; \bar{I}_t = I\left(\lambda_1 < y_t \leq \lambda_2\right)
$$

and

$$
\begin{aligned}
D_t &= d_1 I_t^- + d_2 \bar{I}_t + d_3 I_t^+, \\
G_t &= A_1 I_t^- + A_2 \bar{I}_t + A_3 I_t^+.
\end{aligned}
$$

Computing recursively, we obtain

$$
\begin{aligned}
& X_{t+m} = \left(\Pi_{i=1}^m G_{t+i-d}\right) X_t \\
& + \left(D_{t+m-d} + \varepsilon_{t+m}\right) + G_{t+m-d}\left(D_{t+m-1-d} + \varepsilon_{t+m-1}\right) \\
& + \left(\Pi_{i=m-1}^m G_{t+i-d}\right)\left(D_{t+m-2-d} + \varepsilon_{t+m-2}\right) + ... \\
& + \left(\Pi_{i=2}^m G_{t+i-d}\right)\left(D_{t+1-d} + \varepsilon_{t+1}\right).
\end{aligned}
$$

The coefficient of X_t can be rewritten as

$$
\begin{aligned}
\Pi_{i=1}^m G_{t+i-d} = {} & A_1^m I_{t+m-d}^- ... I_{t+1-d}^- \\
& + A_2^m \overline{I}_{t+m-d} ... \overline{I}_{t+1-d} \\
& + A_3^m I_{t+m-d}^+ ... I_{t+1-d}^+ \\
& + A_1 A_2 A_1^{m-2} I_{t+m-d}^- \overline{I}_{t+m-1-d} I_{t+m-2-d}^- ... I_{t+1-d}^- \\
& + ...
\end{aligned}
\tag{B.3}
$$

We need to choose an appropriate drift function V so that Equations (B.1), (B.2) are satisfied. Let $\|.\|_{M_1}$ be the matrix norm defined in (4). We choose as drift function V the vector norm $\|.\|_{V_1}$ (possibly dependent on A_1), $V(x) = \|x\|_{V_1}$, such that

$$\|Qx\|_{V_1} \leq \|Q\|_{M_1} \|x\|_{V_1}$$

for all square matrices Q and vectors x. The existence of such a vector norm is guaranteed by Lemma 2.

Case where $d \geq p$. Here $m = d$. Note that when $X_t \in K^c$ then the product

$$\bar{I}_t \bar{I}_{t-1} \cdots \bar{I}_{t-d+1} = 0. \tag{B.4}$$

Hence the coefficient of A_2^m in Equation (B.3) is zero when $X_t \in K^c$. Consequently, (B.1), (B.2) are satisfied for

$$\max_{\pi_{1i},\pi_{2i}} \left\| \prod_{i=1}^{m} A_1^{\pi_{1i}} A_2^{\pi_{2i}} A_3^{1-\pi_{1i}-\pi_{2i}} \right\|_{M_1} < 1 \tag{B.5}$$

with $\sum_{j=1}^{m} \pi_{2i} < m$, where $\pi_{1i} = 1$ or 0, $\pi_{2i} = 1$ or 0, and $\pi_{1i}\pi_{2i} = 0$. Finally, the mixing property of y_t follows from that of X_t.

Case where $d < p$. Here $m = p$. Conditionally to X_t, the terms $\bar{I}_t$, $\bar{I}_{t-1}$, $\bar{I}_{t+1-d}$ are deterministic while the terms $\bar{I}_{t+p-d}, ..., \bar{I}_{t+1}$ are random. By the irreducibility of the Markov chain, $P\left[\bar{I}_{t+m-d} \cdots \bar{I}_{t+1-d} = 1 | X_t = x\right] < 1$. The conditional expectation of the term involving A_2^h is

$$\begin{aligned}
&E\left[\left\|\bar{I}_{t+m-d} \cdots \bar{I}_{t+1-d} A_2^m X_t\right\|_{V_1} | X_t = x\right] \\
&= \|A_2^m x\|_{V_1} E\left[\bar{I}_{t+m-d} \cdots \bar{I}_{t+1-d} | X_t = x\right] \\
&= \|A_2^m x\|_{V_1} P\left[\bar{I}_{t+m-d} \cdots \bar{I}_{t+1-d} = 1 | X_t = x\right] \\
&< \|A_2^m x\|_{V_1} \\
&\leq \|A_2^m\|_{M_1} \|x\|_{V_1}
\end{aligned}$$

for any $x \in K^c$. The conditions for (B.1), (B.2) are $\|A_2^m\|_{M_1} \leq 1$ and (B.5). The end of the proof is the same as that for $d \geq p$.

■

Proof of Corollary 1. Here $p = d = 2$, hence $m = 2$. Condition (5) is equivalent to

$$\begin{aligned} \left\|A_1^2\right\|_{M_1} &< 1, \\ \left\|A_1 A_2\right\|_{M_1} &< 1, \\ \left\|A_2 A_1\right\|_{M_1} &< 1. \end{aligned}$$

We proceed to construct the matrix norm $\|.\|_{M_1}$. Let U be defined as

$$U = \frac{1}{\sqrt{\delta_{11}^2 + 1}} \begin{pmatrix} \delta_{11} & -1 \\ 1 & \delta_{11} \end{pmatrix}.$$

Note that U is unitary as $U'U = I_2$. It is easy to check that

$$U^{-1} A_1 U = \begin{pmatrix} \delta_{11} & \left(\delta_{11}^2 a_{12} - \delta_{11} a_{11} - 1\right) / \left(\delta_{11}^2 + 1\right) \\ 0 & \delta_{21} \end{pmatrix} \equiv \Delta.$$

Let $D_t = diag\left(t, t^2\right)$. We have

$$U^{-1} A_1 A_2 U \equiv \frac{1}{\delta_{11}^2 + 1} \begin{pmatrix} g_1 & g_2 \\ g_3 & g_4 \end{pmatrix}$$

and

$$\begin{aligned} D_t U^{-1} A_1 A_2 U D_t^{-1} &= \frac{1}{\delta_{11}^2 + 1} \begin{pmatrix} g_1 & \frac{1}{t} g_2 \\ t g_3 & g_4 \end{pmatrix} \\ &\equiv \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix}. \end{aligned}$$

Note that as δ_{11} is an eigenvalue of A_1 and A_2, we have $\delta_{11}^2 - \delta_{11} a_{11} - a_{12} = \delta_{11}^2 - \delta_{11} a_{21} - a_{22} = 0$. Using $trace\left(A_2\right) = a_{21} = \delta_{12} + \delta_{22} = \delta_{12} + \delta_{11}$ and

$\det(A_2) = -a_{22} = \delta_{12}\delta_{22} = \delta_{12}\delta_{11}$, we have

$$\begin{aligned}
g_1 &= \delta_{11}a_{11}(\delta_{11}a_{21} + a_{22}) + \delta_{11}^2 a_{12} + \delta_{11}a_{21} + a_{22} \\
&= \delta_{11}^2 \left(\delta_{11}^2 + 1\right); \\
g_3 &= -\delta_{11}(a_{11}a_{21} + a_{12}) + \delta_{11}^2 a_{21} + \delta_{11}a_{22} - a_{11}a_{22} \\
&= -a_{11}(\delta_{11}a_{21} + a_{22}) + \delta_{11}^2 a_{21} + \delta_{11}a_{22} - \delta_{11}a_{12} \\
&= -a_{11}\delta_{11}^2 - \delta_{11}a_{22} + \delta_{11}^3 \\
&= 0; \\
g_4 &= a_{11}a_{21} + a_{12} - \delta_{11}a_{21} - \delta_{11}a_{11}a_{22} + \delta_{11}^2 a_{22} \\
&= (\delta_{11} + \delta_{21})a_{21} - \delta_{11}\delta_{21} - \delta_{11}a_{21} - \delta_{11}(\delta_{11} + \delta_{21})a_{22} + \delta_{11}^2 a_{22} \\
&= \delta_{21}(\delta_{12} + \delta_{11}) - \delta_{11}\delta_{21} + \delta_{11}\delta_{21}\delta_{12}\delta_{12} \\
&= \delta_{21}\delta_{12}\left(\delta_{11}^2 + 1\right).
\end{aligned}$$

Therefore, $q_{11} = \delta_{11}^2, q_{21} = 0$, and $q_{22} = \delta_{21}\delta_{12}$. Hence

$$\|A_1A_2\|_{M_1} = \left\|D_tU^{-1}A_1A_2UD_t^{-1}\right\|_1 = \max\left\{\left|\delta_1^2\right| + |q_{12}|, |\delta_{21}\delta_{12}|\right\}$$

As t can be taken arbitrary large, the term $|q_{12}|$ can be made as small as we like. Finally it can be checked that $U^{-1}A_1A_2U = U^{-1}A_2A_1U$. Hence, for any $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$, there exists a t (large), such that $\|A_1\|_{M_1} \leq \rho(A_1) + \varepsilon_1$ and $\|A_1A_2\|_{M_1} = \|A_2A_1\|_{M_1} \leq \max\{|\delta_{11}^2| + \varepsilon_2, |\delta_{21}\delta_{12}|\}$. To complete the proof, note that $\|A_1^2\|_{M_1} \leq \|A_1\|_{M_1}^2$.

■

Proof of Theorem 2. We are in a case close to the case 2 of Hamilton (1994). Using the notation of Model (8), we have

$$\hat{\beta} - \beta_0 = \left[\sum_{t=1}^{T} x_t x_t'\right]^{-} \sum_{t=1}^{T} x_t \varepsilon_t. \tag{B.6}$$

$$\sum_t x_t x_t' = \begin{bmatrix} M_{11} & M_{21}' \\ M_{21} & M_{22} \end{bmatrix},$$

$$M_{11} = \begin{bmatrix} \sum_t \left(u_{t-1}^p\right)\left(u_{t-1}^p\right)' I_{t<} & 0 & 0 \\ 0 & \sum_t \left(u_{t-1}^p\right)\left(u_{t-1}^p\right)' \bar{I}_t & 0 \\ 0 & 0 & \sum_t \left(u_{t-1}^p\right)\left(u_{t-1}^p\right)' I_{t>} \end{bmatrix},$$

$$M_{21} = \sum_t \begin{bmatrix} u_{t-1}^p I_{t<} & 0 & 0 \\ u_{t-1}^p y_{t-1} I_{t<} & 0 & 0 \\ 0 & u_{t-1}^p \overline{I}_t & 0 \\ 0 & u_{t-1}^p y_{t-1} \overline{I}_t & 0 \\ 0 & 0 & u_{t-1}^p I_{t>} \\ 0 & 0 & u_{t-1}^p y_{t-1} I_{t>} \end{bmatrix},$$

$$M_{22} = \begin{bmatrix} M_{22}^{<} & 0 & 0 \\ 0 & \bar{M}_{22} & 0 \\ 0 & 0 & M_{22}^{>} \end{bmatrix},$$

$$M_{22}^{<} = \begin{bmatrix} \sum_t I_{t<} & \sum_t y_{t-1} I_{t<} \\ \sum_t y_{t-1} I_{t<} & \sum_t y_{t-1}^2 I_{t<} \end{bmatrix}.$$

$\bar{M}_{22}$ and $M_{22}^{>}$ are the same as $M_{22}^{<}$ where $I_{t<}$ is replaced by $\bar{I}_t$ and $I_{t>}$ respectively.

$$\sum x_t \varepsilon_t = \begin{bmatrix} \sum \left(u_{t-1}^p\right) I_{t<} \varepsilon_t \\ \sum \left(u_{t-1}^p\right) \bar{I}_t \varepsilon_t \\ \sum \left(u_{t-1}^p\right) I_{t>} \varepsilon_t \\ \sum I_{t<} \varepsilon_t \\ \sum y_{t-1} I_{t<} \varepsilon_t \\ \sum \bar{I}_t \varepsilon_t \\ \sum y_{t-1} \bar{I}_t \varepsilon_t \\ \sum I_{t>} \varepsilon_t \\ \sum y_{t-1} I_{t>} \varepsilon_t \end{bmatrix}.$$

We use as the scaling matrix the following $(3p+3)\times(3p+3)$ diagonal matrix, Γ_T, with diagonal $\left(\sqrt{T},..., \sqrt{T}, \sqrt{T}, T, \sqrt{T}, T, \sqrt{T}, T\right)$.

Premultiplying (B.6) by Γ_T, we obtain:

$$\Gamma_T\left(\hat{\beta} - \beta_0\right) = \left[\Gamma_T^{-1}\left[\sum x_t x_t'\right]\Gamma_T^{-1}\right]^{-}\left\{\Gamma_T^{-1}\left[\sum x_t \varepsilon_t\right]\right\}. \tag{B.7}$$

Under H_0', $y_t = \sum u_{t-j}$ is a random walk and $y_t/\sqrt{T}$ converges to a $\delta B(r)$, $r = t/T$, where $B(.)$ is a standard Brownian motion, on the other hand $\frac{1}{\sqrt{T}}\sum_{t=1}^{Tr} \varepsilon_t$ converges to $\sigma B(r)$. The functions $T_1(x) = I(|x| < k)$, $T_2(x) = xI(|x| < k)$, and $T_3(x) = x^2 I(|x| < k)$ are piecewise continuous in x and therefore regular, we can apply Lemma A2 of Park and Phillips (2001):

$$\frac{1}{T}\sum I\left(\frac{|y_{t-1}|}{\sqrt{T}}<\pi\right)\xrightarrow{L}\int_0^1 I\left(|B(r)|<\pi/\delta\right)dr$$
$$\frac{1}{T}\sum \frac{y_{t-1}}{\sqrt{T}}I\left(\frac{|y_{t-1}|}{\sqrt{T}}<\pi\right)\xrightarrow{L}\delta\int_0^1 B(r)I\left(|B(r)|<\pi/\delta\right)dr$$
$$\frac{1}{T}\sum \frac{y_{t-1}^2}{T}I\left(\frac{|y_{t-1}|}{\sqrt{T}}<\pi\right)\xrightarrow{L}\delta^2\int_0^1 B^2(r)I\left(|B(r)|<\pi/\delta\right)dr$$
$$\frac{1}{\sqrt{T}}\sum I\left(\frac{|y_{t-1}|}{\sqrt{T}}<\pi\right)\varepsilon_t\xrightarrow{L}\sigma\int_0^1 I\left(|B(r)|<\pi/\delta\right)dB(r)$$
$$\frac{1}{\sqrt{T}}\sum \frac{y_{t-1}}{\sqrt{T}}I\left(\frac{|y_{t-1}|}{\sqrt{T}}<\pi\right)\varepsilon_t\xrightarrow{L}\sigma\delta\int_0^1 B(r)I\left(|B(r)|<\pi/\delta\right)dB(r).$$

Using again Park and Phillips, it can be shown that

$$T^{-1}\sum I\left(|y_{t-1}|<\lambda\right)u_{t-j}\xrightarrow{P}0,$$
$$T^{-3/2}\sum y_{t-1}I\left(|y_{t-1}|<\lambda\right)u_{t-j}\xrightarrow{P}0.$$

The same types of results apply to terms involving $I\left(y_{t-1}\geq\lambda\right)$ and $I\left(y_{t-1}\leq-\lambda\right)$ but are not detailed here. Therefore, we have

$$\left[\Gamma_T^{-1}\left[\sum x_t x_t'\right]\Gamma_T^{-1}\right]\xrightarrow{L}\begin{bmatrix} V & 0\\ 0 & Q\end{bmatrix}\tag{B.8}$$

with

$$V = \begin{bmatrix} \Omega\int_0^1 I_1(r)\,dr & 0 & 0\\ 0 & \Omega\int_0^1 I_2(r)\,dr & 0\\ 0 & 0 & \Omega\int_0^1 I_3(r)\,dr\end{bmatrix}$$
$$Q = \begin{bmatrix} Q_1 & 0 & 0\\ 0 & Q_2 & 0\\ 0 & 0 & Q_3\end{bmatrix},$$
$$Q_j = \begin{bmatrix} \int_0^1 I_j(r)\,dr & \delta\int_0^1 B(r)I_j(r)\,dr\\ \delta\int_0^1 B(r)I_j(r)\,dr & \delta^2\int_0^1 B^2(r)I_j(r)\,dr\end{bmatrix},\ j=1,2,3,$$

and $\Omega = E\left[\left(u_{t-1}^p\right)\left(u_{t-1}^p\right)'\right]$. From (B.8), we see that the estimators of $(\alpha_1,\alpha_2,\alpha_3)'$ are orthogonal to the other estimators. We focus on the latter ones.

One can decompose $\Gamma_T^{-1}\left[\sum x_t\varepsilon_t\right]$ into two pieces. The first part satisfies

$$\left[\begin{array}{c} \frac{1}{\sqrt{T}}\sum u_{t-1}^{p} I\left(y_{t-1}\leq -\lambda\right)\varepsilon_t \\ \frac{1}{\sqrt{T}}\sum u_{t-1}^{p} I\left(\left|y_{t-1}\right|< \lambda\right)\varepsilon_t \\ \frac{1}{\sqrt{T}}\sum u_{t-1}^{p} I\left(y_{t-1}\geq \lambda\right)\varepsilon_t \end{array}\right] \xrightarrow{L} h_1 \sim \left[\begin{array}{c} \sigma\Omega^{1/2}\int_0^1 I_<\left(r\right)dB(r) \\ \sigma\Omega^{1/2}\int_0^1 \bar{I}\left(r\right)dB(r) \\ \sigma\Omega^{1/2}\int_0^1 I_>\left(r\right)dB(r) \end{array}\right]$$

and the second part follows asymptotically

$$\left[\begin{array}{l} \frac{1}{\sqrt{T}}\sum I\left(y_{t-1}\leq -\lambda\right)\varepsilon_t \\ \frac{1}{T}\sum y_{t-1} I\left(y_{t-1}\leq -\lambda\right)\varepsilon_t \\ \frac{1}{\sqrt{T}}\sum I\left(\left|y_{t-1}\right|< \lambda\right)\varepsilon_t \\ \frac{1}{T}\sum y_{t-1} I\left(\left|y_{t-1}\right|< \lambda\right)\varepsilon_t \\ \frac{1}{\sqrt{T}}\sum I\left(y_{t-1}\geq \lambda\right)\varepsilon_t \\ \frac{1}{T}\sum y_{t-1} I\left(y_{t-1}\geq \lambda\right)\varepsilon_t \end{array}\right] \xrightarrow{L} h_2 \sim \left[\begin{array}{l} \sigma\int_0^1 I_<\left(r\right)dB(r) \\ \sigma\delta\int_0^1 B(r)I_<\left(r\right)dB(r) \\ \sigma\int_0^1 \bar{I}\left(r\right)dB(r) \\ \sigma\delta\int_0^1 B(r)\bar{I}\left(r\right)dB(r) \\ \sigma\int_0^1 I_>\left(r\right)dB(r) \\ \sigma\delta\int_0^1 B(r)I_>\left(r\right)dB(r) \end{array}\right].$$

Substituting into (B.7), we get

$$\Gamma_T\left(\hat{\beta}-\beta_0\right) \xrightarrow{L} \left[\begin{array}{c} V^{-}h_1 \\ Q^{-}h_2 \end{array}\right].$$

We want to test $R\beta = 0$ where R is the appropriate selection matrix. The Wald test is given by

$$\begin{aligned} W_T\left(\lambda\right) &= \left(R\hat{\beta}\right)'\left[\widehat{\sigma}^2 R\left[\sum x_t x_t'\right]^{-} R'\right]^{-} R\hat{\beta} \\ &= \left(\tilde{R}\widehat{\gamma}\right)'\left[\widehat{\sigma}^2 \tilde{R}\left[\sum \tilde{x}_t\tilde{x}_t'\right]^{-} \tilde{R}'\right]^{-} \tilde{R}\widehat{\gamma} \end{aligned}$$

where $\widehat{\gamma}$ is the estimator of $\gamma = \left(\mu_1,\rho_1,\mu_2,\rho_2,\mu_3,\rho_3\right)'$, $\tilde{x}_t$ are the regressors associated with γ, $\widehat{\sigma}^2$ is a consistent estimator of σ^2 and

$$\tilde{R} = \left[\begin{array}{cccccc} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right].$$

Hence the asymptotic distribution of $W_T\left(\lambda\right)$ is given by

$$\left(\tilde{R}Q^{-}h_2\right)'\left[\widehat{\sigma}^2\tilde{R}Q^{-}\tilde{R}'\right]^{-}\tilde{R}Q^{-}h_2,$$

which simplifies to the formula given in the Proposition 2. By equivalence between the test statistics, the limiting distributions of LM_T and LR_T are the same as that of W_T.

Under the alternative of a stationary TAR model, that is when $\rho_1 < 0$ and $\rho_3 < 0$, $\hat{\rho}_1$ and $\hat{\rho}_3$ converge in $\sqrt{T}$ to their pseudo true values that are functions of λ and λ_0 and strictly negative. Hence the Wald, LM, and LR test statistics diverge.

■

Proof of Theorem 3. As before we focus on supW test. The proof for supLR and supLM would be identical. We first introduce new notation

$$\begin{aligned}\tilde{W}_T(k) &\equiv W_T\left(\sqrt{T}k\delta\right), \\ \underline{k}_T &= \frac{\underline{\lambda}_T}{\delta\sqrt{T}}, \\ \bar{k}_T &= \frac{\bar{\lambda}_T}{\delta\sqrt{T}}.\end{aligned}$$

Note that

$$\sup_{k\in\left[\underline{k}_T,\bar{k}_T\right]} \tilde{W}_T(k) = \sup_{\lambda\in\left[\underline{\lambda}_T,\bar{\lambda}_T\right]} W_T(\lambda).$$

To establish the result, we need to prove the following four points:

1) Weak convergence of $\tilde{W}_T(k)$ to $T(k)$ for k in any bounded interval $[a,b]$ with $0 < a < b < \infty$.

2) $\left(\underline{k}_T, \bar{k}_T\right)$ converges in distribution to $\left(\underline{k}, \bar{k}\right)$ jointly with the convergence of $\tilde{W}_T(k)$ to $T(k)$.

3) $\sup_{k\in\left[\underline{k}_T,\bar{k}_T\right]} \tilde{W}_T(k) \xrightarrow{L} \sup_{k\in\left[\underline{k},\bar{k}\right]} T(k)$.

4) The distribution of $\sup_{k\in\left[\underline{k},\bar{k}\right]} T(k)$ is pivotal.

1) By the proposition (page 2251) of Andrews (1994), the weak convergence of $\tilde{W}_T(k)$ on a finite interval $a < k < b$ follows from (i) finite dimensional convergence (fidi) and (ii) stochastic equicontinuity.

(i) is established by Theorem 2,

(ii) follows from Proposition 2 of Bec, Guay and Guerre (2002).

2) From Theorem 3.1 of Park and Phillips (2001), we have

$$\frac{1}{T}\sum I\left(\frac{|y_{t-1}|}{\delta\sqrt{T}} < k\right) \xrightarrow{L} \int_0^1 I\left(|B(r)| < k\right)dr$$

uniformly in $k \in [a, b]$. By continuity and as the limiting function is increasing, it follows that the pair of percentiles $\left(\underline{k}_T, \bar{k}_T\right)$ converges in distribution to $\left(\underline{k}, \bar{k}\right)$. Moreover this convergence jointly holds with that of point 1).

3) This part is delicate because $\left(\underline{k}, \bar{k}\right)$ is random and the support of its distribution is not bounded. Since $\left(\underline{k}_T, \bar{k}_T\right)$ converges in distribution, the sequence $\left(\underline{k}_T, \bar{k}_T\right)$ is uniformly tight (see e.g. Theorem 2.4 of van der Vaart (1998)). Moreover $P\left(\underline{k}_T > 0\right) = 1$. Therefore,

$$\lim_{a \to 0, b \to \infty} \sup_T P\left(a < \underline{k}_T < \bar{k}_T < b\right) = 1$$

where the limit is taken over $0 < a < b < \infty$. Hence, it follows that for any $\epsilon \in (0, 1)$, there are some $a > 0$ small enough and b large enough such that

$$\sup_T P\left(a < \underline{k}_T < \bar{k}_T < b\right) \geq 1 - \epsilon. \tag{B.9}$$

Let $\underline{k}_T^a = \max\left(\underline{k}_T, a\right)$ and $\bar{k}_T^b = \min\left(\bar{k}_T, b\right)$, which are such that $\underline{k}_T \leq \underline{k}_T^a < \bar{k}_T^b \leq \bar{k}_T$. By (B.9), we have

$$\sup_T P\left(\sup_{k \in \left[\underline{k}_T^a, \bar{k}_T^b\right]} \tilde{W}_T(k) \neq \sup_{k \in \left[\underline{k}_T, \bar{k}_T\right]} \tilde{W}_T(k)\right) \leq \epsilon. \tag{B.10}$$

Denote $\underline{k}^a = \max\left(\underline{k}, a\right)$ and $\bar{k}^b = \min\left(\bar{k}, b\right)$. By the continuous mapping theorem (see Theorem 5.5 of Billingsley (1968)), we have

$$\sup_{k \in \left[\underline{k}_T^a, \bar{k}_T^b\right]} \tilde{W}_T(k) \xrightarrow{L} \sup_{k \in \left[\underline{k}^a, \bar{k}^b\right]} T(k) \tag{B.11}$$

as T goes to infinity and for any $0 < a < b < \infty$.

Remark that

$$\sup_{k \in \left[\underline{k}^a, \bar{k}^b\right]} T(k) \xrightarrow{a.s.} \sup_{k \in \left[\underline{k}, \bar{k}\right]} T(k) \tag{B.12}$$

as $a \to 0$, $b \to \infty$ with $0 < a < b < \infty$, by continuity of $T(k)$ over $\left[\underline{k}, \bar{k}\right]$ and because $\underline{k}^a$ decreases to $\underline{k}$ when a decreases to 0 and $\bar{k}^b$ increases to $\bar{k}$ when b goes to infinity.

Consider $\epsilon \in (0,1)$ and any $y \in \mathbf{R}$. Choose a and b as in (B.9), and such that

$$\left| P\left(\sup_{k\in\left[\underline{k}^a,\bar{k}^b\right]} T(k) \leq y\right) - P\left(\sup_{k\in\left[\underline{k},\bar{k}\right]} T(k) \leq y\right)\right| \leq \epsilon,$$

which is possible due to (B.12). By (B.11), there is a T_ϵ such that

$$\sup_{T\geq T_\epsilon}\left| P\left(\sup_{k\in\left[\underline{k}_T^a,\bar{k}_T^b\right]} \tilde{W}_T\left(k\right) \leq y\right) - P\left(\sup_{k\in\left[\underline{k}^a,\bar{k}^b\right]} T(k) \leq y\right)\right| \leq \epsilon$$

for any $y \in \mathbf{R}$. Now by the triangular inequality, (B.10) and (B.12) yield that

$$\sup_{T\geq T_\epsilon}\left| P\left(\sup_{k\in\left[\underline{k}_T^a,\bar{k}_T^b\right]} \tilde{W}_T\left(k\right) \leq y\right) - P\left(\sup_{k\in\left[\underline{k}^a,\bar{k}^b\right]} T(k) \leq y\right)\right| \leq 3\epsilon.$$

Point 3) follows.

4) The distribution of $\sup_{k\in\left[\underline{k},\bar{k}\right]} T(k)$ is pivotal because the distributions of $T\left(k\right)$ and $\left(\underline{k},\bar{k}\right)$ are nuisance parameter free.

■

Appendix C

Figure C.1: FRF/DEM real exchange rate (in log, demeaned)

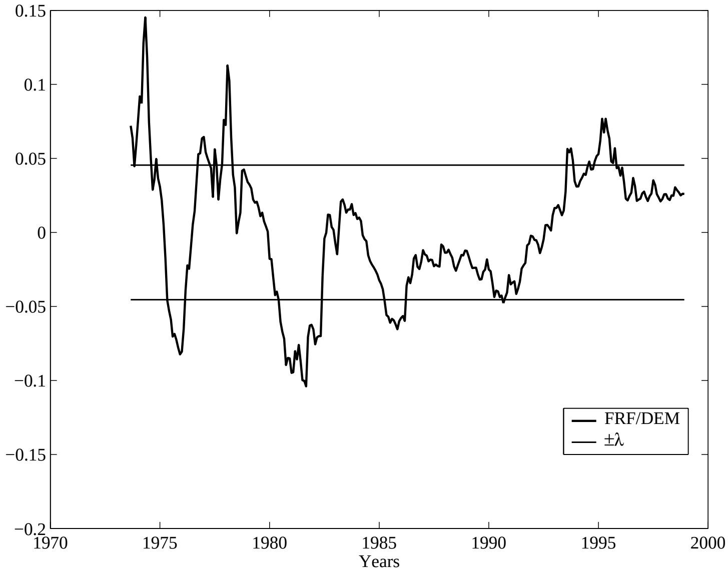


Figure C.2: ITL/DEM real exchange rate (in log, demeaned)

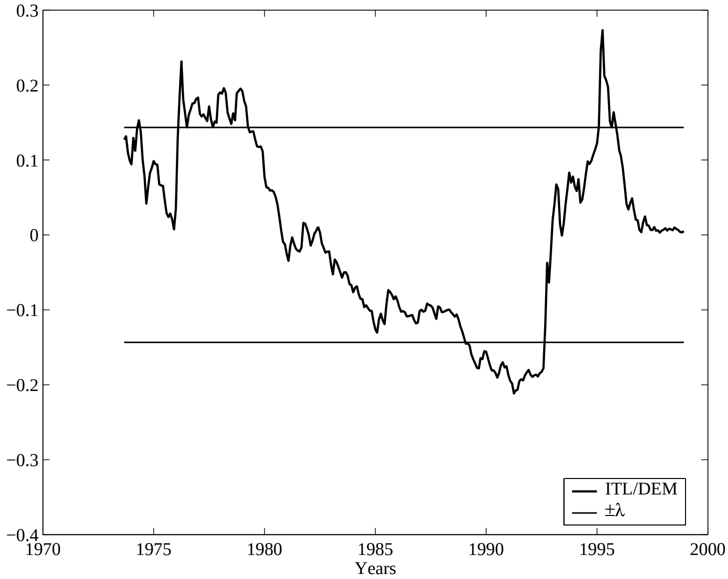


Figure C.3: BEF/DEM real exchange rate (in log, demeaned)

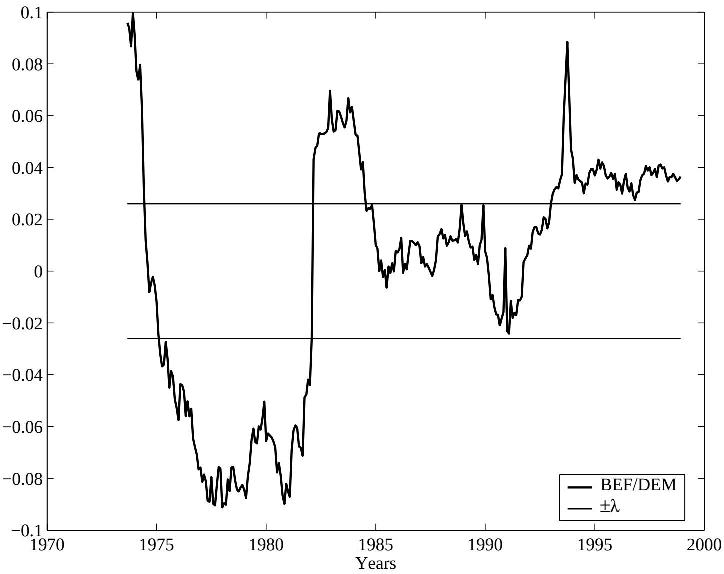


Figure C.4: NLG/DEM real exchange rate (in log, demeaned)

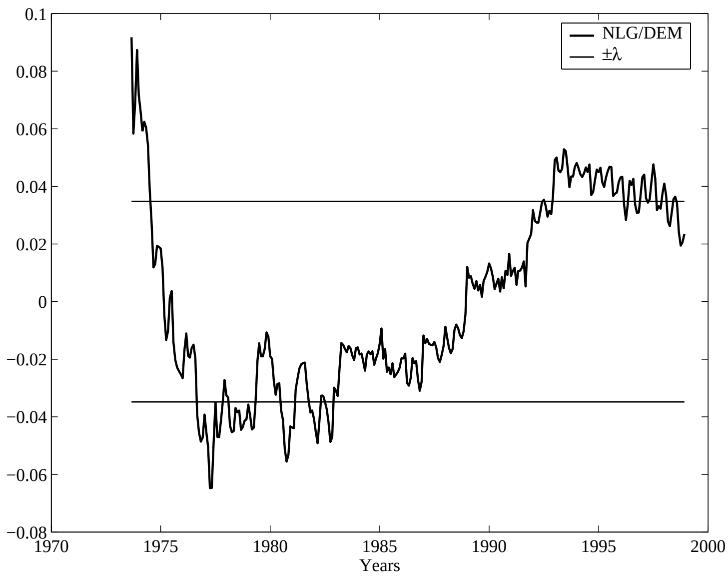
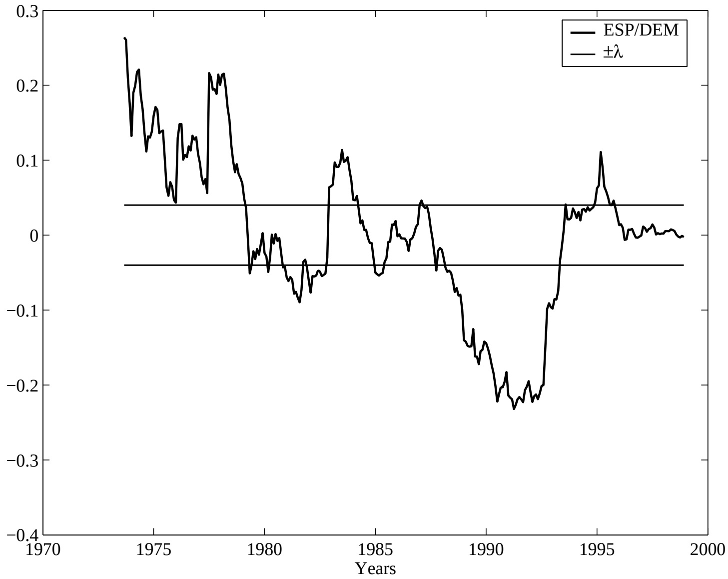


Figure C.5: ESP/DEM real exchange rate (in log, demeaned)



References

Andrews, D. W. K. (1994), "Empirical Process Methods in Econometrics," in *Handbook of Econometrics, IV*, eds R.F. Engle and D.L. McFadden.

—— and Ploberger, W. (1994), "Optimal Tests When A Nuisance Parameter Is Present Only Under The Alternative," *Econometrica*, 62, 1383–1414.

Balassa, B. (1964), "The Purchasing Power Parity Doctrine : A Reappraisal," *Journal of Political Economy*, 72, 584–96.

Balke, N. S., and Fomby, T. B. (1997), "Threshold Cointegration," *International Economic Review*, 38, 627–45.

Bec, F., Guay, A., and Guerre, E. (2002), "An Adaptive Consistent Unit-Root Test Based on Autoregressive Threshold Model," Working Paper, CREST, France.

—— and Hairault, J. O. (1997), "Les Implications de la Structure des Marchés Financiers pour la Dynamique des Modèles d'Équilibre Général à Deux Pays," *Revue d'Economie Politique*, 4, 479–494.

Berben, R. P., and van Dijk, D. (1999), "Unit Root Tests and Asymmetric Adjustment : A Reassessment," Research Report EI-9902/A, Econometeric Institute, Rotterdam.

Bhagwati, J. (1984), "Why Are Services Cheaper in Poor Countries ?," *Economic Journal*, 94, 279–86.

Bierens, H., and Guo, S. (1993), "Testing Stationarity and Trend Stationarity Against the Unit Root Hypothesis," *Econometric Reviews*, 12, 1–32.

Billingsley, A. W. (1968), *Convergence of Probability Measures*, New York: John Wiley.

Caner, M., and Hansen, B. (2001), "Threshold Autoregression With a Unit Root," *Econometrica*, 69, 1555–1596.

Chan, K. S., Petruccelli, J., Tong, H., and Woolford, S. (1985), "A Multiple-Threshold AR(1) Model," *Journal of Applied Probability*, 22, 267–279.

Chan, K. S., and Tong, H. (1985), "On the Use of Lyapunov Functions for the Ergodicity of Stochastic Difference Equations," *Advances in Applied Probability*, 17, 666–678.

Chang, R., and Devereux, M. (1998), "Dynamic Collusion, Pricing to Market, and Real Exchange Rates," Working Paper, University of British Columbia.

Chen, R., and Tsay, R. (1991), "On the Ergodicity of TAR(1) Processes," *The Annals of Applied Probability*, 1, 613–634.

Davies, R. B. (1987), "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative," *Biometrika*, 74, 33–43.

Dickey, D., and Fuller, W. (1981), "Likelihood Ratio Statistics for Autoregressive Time Series With an Unit Root," *Econometrica*, 49, 1057–1072.

Dufour, J.-M. (1997), "Some Impossibility Theoremes in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1387.

Enders, W., and Granger, C. W. J. (1998), "Unit-Root Tests and Asymmetric Adjustment With an Example Using the Term Structure of Interest Rates," *Journal of Business and Economic Statistics*, 16, 304–11.

Engel, C. (2000), "Long-Run PPP May Not Hold After All," *Journal of International Economics*, 57, 243–273.

Froot, K. A., and Rogoff, K. (1985), "Perspectives on PPP and Long-Run Real Exchange Rates," in *Handbook of International Economics, III*, eds. R.W. Jones and P.B. Kenen, pp. 679–747.

Gonzalez, M., and Gonzalo, J. (1998), "Threshold Unit Root Models," Working Paper, University Carlos III.

Hamilton, J. (1994), *Time Series Analysis*, Princeton, New Jersey: Princeton University Press.

Hansen, B. E. (1996), "Inference When a Nuisance Parameter is not Identified Under the Null Hypothesis," *Econometrica*, 64, 413–430.

Henry, Ó. T., and Shields, K. (2003), "Is There a Unit-Root in Inflation," *Journal of Macroeconomics*, forthcoming.

Horn, R. A., and Johnson, C. R. (1985), *Matrix Analysis*, Cambridge, UK: Cambridge University Press.

Kapetanios, G., and Shin, Y. (2002), "Unit Root Tests in Three-Regime SETAR Models," Unpublished Manuscript, Bank of England.

Kilian, L., and Taylor, M. (2003), "Why Is It so Difficult to Beat the Random Walk Forecast of Exchange Rates?," *Journal of International Economics*, 60, 85–107.

Kravis, I., and Lipsey, R. (1983), "Toward an Explanation of National Price Levels," Working Paper, Princeton University.

Krugman, P. R. (1987), "Pricing to Market When the Exchange Rate Changes," in *Real Financial Linkages Among Open Economies*, eds. S.W. Arndt and J.D. Richardson, Cambridge Mass:MIT Press, pp. 389–406.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Shin, Y. (1992), "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root. How Sure Are We That Economic Time Series Have a Unit Root ?," *Journal of Econometrics*, 54, 159–178.

Lo, M. C., and Zivot, E. (2001), "Threshold Cointegration and Nonlinear Adjustment to the Law of One Price," *Macroeconomic Dynamics*, 5, 533–576.

MacDonald, R. (1995), "Long-Run Exchange Rate Modeling: A Survey of the Recent Evidence," *IMF Staff Papers*, 42, 437–489.

Meyn, S., and Tweedie, R. L. (1992), "Stability of Markovian Processes I: Criteria for Discrete-Time Chains," *Advances in Applied Probability*, 24, 542–574.

Michael, P., Nobay, A., and Peel, D. (1997), "Transactions Costs and Nonlinear Adjustment in Real Exchange Rates : An Empirical Investigation," *Journal of Political Economy*, 105, 862–79.

Ng, S., and Perron, P. (2002), "PPP May Not Hold After All: A Further Investigation," *Annals of Economics and Finance*, 3, 43–64.

Obstfeld, M., and Taylor, A. (1997), "Nonlinear Aspects of Goods-Market Arbitrage and Adjustment: Heckscher's Commodity Points Revisited," *Journal of the Japanese and International Economies*, 11, 441–479.

—— and Rogoff, K. (1996), *Foundations of International Macroeconomics*, USA: MIT Press.

—— and —— (2000), "The Six Major Puzzles in International Macroeconomics : Is There a Common Cause ?," *NBER Macroeconomics Annual*, 15, 339–390.

Park, J., and Phillips, P. (2001), "Nonlinear Regressions with Integrated Time Series," *Econometrica*, 69, 117–161.

Pham, D. T. (1986), "The Mixing Property of Bilinear and Generalised Random Coefficient Autoregressive Models," *Stochastic Processes and their Applications*, 23, 291–300.

Phillips, P., and Perron, P. (1988), "Testing for Unit Root in Time Series Regression," *Biometrika*, 75, 335–346.

Pippenger, M., and Goering, G. (1993), "A Note on the Empirical Power of Unit Root Tests Under Threshold Processes," *Oxford Bulletin of Economics and Statistics*, 55, 473–481.

—— and —— (2000), "Additional Results on the Power of Unit root and Cointegration Tests Under Threshold Process," *Applied Economics Letters*, 7, 641–644.

Rogoff, K. (1996), "The Purchasing Power Parity Puzzle," *Journal of Economic Literature*, 34, 647–68.

Samuelson, P. (1964), "Theoretical Notes on Trade Problems," *Review of Economics and Statistics*, 46, 145–64.

Sercu, P., Uppal, R., and Van Hulle, C. (1995), "The Exchange Rate in the Presence of Transaction Costs : Implications for Tests of Purchasing Power Parity," *The Journal of Finance*, 50, 1309–1319.

Shin, D. W., and Lee, O. (2001), "Tests for Asymmetry in Possibly Nonstationary Time Series Data," *Journal of Business and Economic Statistics*, 19, 233–244.

Taylor, A. (2001), "Potential Pitfalls for the PPP Puzzle ? Sampling and Specification Biases in Mean-Reversion Tests of the LOOP," *Econometrica*, 69, 473–498.

Tjøstheim, D. (1990), "Non-Linear Time Series and Markov Chains," *Advances in Applied Probability*, 22, 587–611.

van der Vaart, A. W. (1998), *Asymptotic statistics*, Cambridge, UK: Cambridge University Press.