Detecting Mean Reversion in Real Exchange Rates from a Multiple Regime STAR Model

Frédérique Bec, Mélika Ben Salem, Marine Carrasco

Detecting Mean Reversion in Real Exchange Rates from a Multiple Regime STAR Model*

Frédérique Bec †
Mélika Ben Salem ‡
and Marine Carrasco §

February 27, 2010

Abstract

Recent studies on general equilibrium models with transaction costs show that the dynamics of the real exchange rate are necessarily nonlinear. Our contribution to the literature on nonlinear price adjustment mechanisms is threefold. First, we model the real exchange rate by a Multi-Regime Logistic Smooth Transition AutoRegression (MR-LSTAR), allowing for both ESTAR-type and SETAR-type dynamics. This choice is motivated by the fact that even the theoretical models, which predict a smooth behavior for the real exchange rate, do not rule out the possibility of a discontinuous adjustment as a limit case. Second, we propose two classes of unit-root tests against this MR-LSTAR alternative, based respectively on the likelihood and on an auxiliary model. Their asymptotic distributions are derived analytically. Third, when applied to 28 bilateral real exchange rates, our tests reject the null hypothesis of a unit root for eleven series bringing evidence in favor of the purchasing power parity.

Keywords: Half-life, purchasing power parity, mixing conditions, smooth transition autoregressive model, unit-root test, real exchange rate.

JEL Codes: C12, C22, F31.

*We wish to thank, for helpful comments, the associate editor and an anonymous referee, Mark Bils and Alan Stockman, and also the participants of the Symposium of the Society for Nonlinear Dynamics and Econometrics (Atlanta 2002), ESEM (Venice, 2002), and the Midwest Econometrics Group conference (Chicago, 2004). Carrasco gratefully acknowledges financial support from the National Science Foundation, grant SES 0211418.

†Université de Cergy Pontoise, Thema, Crest. Email: bec@ensae.fr

‡OEP, Université de Marne-la-Vallée. Email: melika.bensalem@univ-mlv.fr

§University of Montreal, Département de Sciences Economiques. Email: marine.carrasco@umontreal.ca.

1 Introduction

The issue of whether the long-run purchasing power parity (PPP) relationship holds is still largely unsettled[1]. The most widespread test for the long-run PPP consists in testing for a unit root in the real exchange rate within a linear framework. So far, the conclusions emerging from this approach are mixed[2]. Over the last decade, empirical and theoretical analysis of the real exchange rate have shifted toward a nonlinear framework.

From a theoretical point of view, introducing shipping costs into two-country general equilibrium models (e.g. Dumas [1992], Uppal [1993], Sercu, Uppal and Van Hulle [1995] or Berka [2004]) generates two regimes for the real exchange rate process. The "outer" regime corresponds to PPP departures, which are greater than the shipping costs in absolute value. In this regime, shipping takes place to exploit the profit opportunities and the PPP deviations are corrected by international trade. The "inner" regime is associated with PPP differentials that are smaller than the transaction costs in absolute value. In this regime, no shipping takes place and PPP deviations are not corrected for, and hence may persist for quite some time. As pointed out by Taylor [2001], this theoretical outcome could explain the failure of standard unit root tests to reject the null hypothesis. He shows that the power of the Augmented Dickey-Fuller (ADF) test falls dramatically when the fraction of observations lying in the "inner" regime increases, even if the process is globally stationary.

From an empirical point of view, the nonlinear dynamics predicted by the theoretical models with transaction costs have been accounted for by threshold models. More precisely, two kinds of threshold models have been, concurrently but independently, explored to this end. The first one, namely the Self-Exciting Threshold Autoregressive (SETAR) model, retains a discontinuous transition function between regimes (see e.g. Obstfeld and Taylor [1997]). The second one is the Smooth Transition Autoregressions (STAR), which, contrary to the SETAR, allow for smooth regime changes. So far, the Exponential Smooth Transition Autoregression (ESTAR) has been retained to capture this kind of smooth adjustment (see e.g. Michael, Nobay and Peel [1997], Baum, Barkoulas and Caglayan [2001], Taylor, Peel and Sarno [2001], O'Connell and Wei [2002] and Kilian and Taylor [2003]). In order to test for long-run PPP, Enders and Granger [1998], Lo and Zivot [2001] and Bec, Ben Salem and Carrasco [2004] among others have proposed unit root tests, which are specifically devised to have power against a stationary SETAR alternative. Similarly,

[1] See Rogoff [1996] for a survey.

[2] See for instance the recent empirical studies by Papell [1997], Lothian and Taylor [2000], Murray and Papell [2002a] or Murray and Papell [2002b].

Kapetanios, Shin and Snell [2003] have developed tests to test unit-root against a stationary ESTAR alternative. All these tests were found to be more powerful than the ADF test. Two recent papers are closely related to the present paper. Bec, Guay and Guerre [2008] propose a new sup Wald test with the specific feature that the set of thresholds is selected differently under the null and the alternative. This adaptive selection aims to improve the power of the test. Park and Shintani [2005] derive the asymptotic distribution of a inf-t test for various alternatives including TAR and LSTAR models, however they limit their attention to testing a simple null hypothesis and their results do not directly apply to our setting.

Our analysis departs from existing work in three dimensions. *First*, it relies on a general model, the Multi-Regime Logistic Smooth Transition AutoRegression (MR-LSTAR), allowing for both ESTAR-type and SETAR-type dynamics. Indeed, even though the ESTAR model is often considered as the smooth transition analogue of the SETAR model, the former does not nest the latter. Yet, as will be stressed in the next section, neither the discontinuous nor the continuous adjustment cases can be ruled out *a priori* on theoretical grounds. *Second*, we develop two classes of unit-root tests against this MR-LSTAR alternative, based respectively on the likelihood and on an auxiliary model. The asymptotic distribution of each test is derived analytically and is shown to be nuisance parameter free. A Monte Carlo experiment reveals that contrary to the ADF test, the power of our tests remains high for large values of the threshold parameter. *Third*, we apply our tests on monthly data from 8 countries leading to 28-real exchange rates for the post-Bretton Woods period. The null of a unit root is rejected for eleven pairs of currencies, while the ADF test rejects only for one series. The half-lives we obtain are much smaller for large shocks than for small shocks, which supports the theory of PPP in the presence of transaction costs. Another interesting result is that the shapes of the estimated transition functions are only slightly smoother than the discontinuous transition function characterizing the SETAR model.

The paper is organized as follows. In Section 2 we discuss the real exchange rate nonlinear dynamics implied by existing theoretical models. This motivates the choice of our MR-LSTAR model, which is then presented and compared to the ESTAR. Section 3 describes the unit root tests and their asymptotic distributions before reporting their small sample properties. The data and the empirical results are presented in Section 4. Section 5 concludes.

2 The Nonlinear Continuous Adjustment Specification

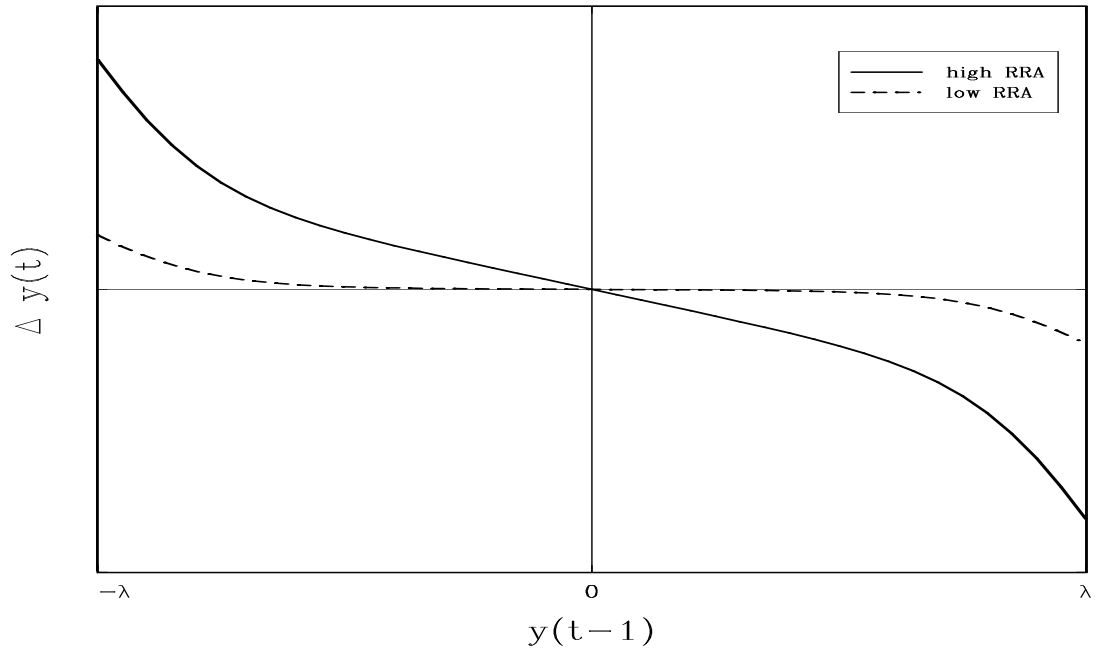
2.1 Theoretical backgrounds

Recent general equilibrium models with proportional transport costs — see e.g. Sercu et al. [1995], Bec et al. [2004], Dumas [1992], Uppal [1993] and Berka [2004] — outline multi-regime dynamics for the real exchange rate process. Assuming a symmetrical two-country setup, these models predict the existence of a no-trade region within which the real exchange rate adjustment toward the PPP equilibrium, if any, is expected to be slow. Let y_t denote the logarithm of the real exchange rate (the price of a unit of domestic goods in units of foreign goods). The region of no trade, called the inner regime, is defined by $y_t \in (-\lambda; \lambda)$ where $\lambda \in (0, 1)$ is the proportional shipping cost. Outside this region, international arbitrage forces the real exchange rate back toward the band. In these models, the real exchange rate is a nonlinear monotone function of the physical imbalance, ω, as measured by the difference in endowments between home and abroad. Therefore, the behavior of ω determines the dynamics of y_t. When ω is exogenous, as in Sercu et al. [1995] and Bec et al. [2004], the implied dynamics of the real exchange rate can be represented by a discontinuous adjustment threshold model, namely the SETAR model. However, these very simple models may be viewed as a crude version of the more sophisticated setups proposed by Dumas [1992] or Berka [2004]. More particularly, Dumas' model features country-specific productivity shocks and the dynamics of ω are endogenously determined through the capital accumulation process.

Replicating Figure 5 of Dumas [1992], Figure 1, below, shows the conditional expected change of the real exchange rate as a function of its lagged value in the inner regime. From Figure 1, we see that the real exchange rate process is mean reverting, i.e. its conditional expected change is negative (positive) for positive (negative) values, and the mean reversion is strongest when the deviation from parity is largest[3]. Hence, from an empirical point of view, these features point to a smooth transition autoregression. The question is which STAR model to choose. As stressed by Dumas [1992], the shape of the conditional expected change function depends crucially on the relative risk aversion (RRA) parameter. The dotted line in Figure 1 represents the conditional expected change associated with a low degree of risk aversion. As one approaches risk neutrality, the function stays longer on the zero axis. Indeed, the lower the risk aversion, the less sensitive

[3]Even though the process for ω is exogenous in Berka [2004], the same kind of dynamics for the real exchange rate are obtained. This results from the multi-sector assumption, each type of good involving a different loss-shipping factor. Aggregation over the different goods then provides a smooth adjustment of the real exchange rate.

Figure 1: Conditional expected change of the real exchange rate in the inner regime



the agents are to the ex ante benefits of diversification achieved by shipping. Consequently, a low degree of risk aversion makes rebalancing of physical capital less desirable, which in turn implies a slower mean reversion of the real exchange rate. In the limit case of risk neutrality, the conditional expected change function lies on the zero axis, which corresponds to a SETAR-type discontinuous adjustment. Note that the possibility of risk neutrality cannot be ruled out according to the findings of, e.g., Hansen and Singleton [1982], Hansen and Singleton [1984], and Epstein and Zin [1991].

2.2 Comparison between ESTAR and MR-LSTAR models

According to the discussion above, it seems highly desirable to empirically analyze the nonlinear dynamics of the real exchange rate in a framework which encompasses both discontinuous and continuous types of adjustments. As will be stressed below, the most popular model, namely the Exponential STAR model, cannot account for the discontinuous case. In this subsection, we compare the ESTAR and MR-LSTAR models in the simplest setup, one that includes only one autoregressive lag.

Consider first the ESTAR model given by:

$$y_t = \phi_1 y_{t-1} G + \phi_2 y_{t-1}(1 - G) + \varepsilon_t, \tag{1}$$

where the transition function G is given by:

$$G = G(y_{t-1}, \gamma, \kappa) = 1 - \exp(-\gamma(y_{t-1} - \kappa)^2). \tag{2}$$

The positive parameter γ governs the speed of transition between regimes and is called the speed parameter. Note that when y_{t-1} goes to $\pm\infty$, G goes to one so that the ESTAR model becomes:

$$y_t = \phi_1 y_{t-1} + \varepsilon_t, \tag{3}$$

the parameter ϕ_1, hence, characterizing the "outer" regime dynamics. When $y_{t-1} = \kappa$, then $G = 0$ and the ESTAR model dynamics are now governed by ϕ_2:

$$y_t = \phi_2 y_{t-1} + \varepsilon_t.$$

Hence, the weight of the "outer" regime parameter, ϕ_1, decreases as y_{t-1} approaches κ. Finally, the dynamics generated by the transition function, (2), are symmetrical around κ.

An undesirable feature of model (1) is that it collapses to the linear process given in (3) when the speed parameter, γ, goes to infinity. This should correspond to a sudden shift between regimes. To overcome this issue, a natural candidate is the MR-LSTAR model mentioned in van Dijk, Terasvirta and Franses [2002]:

$$y_t = \phi_1 y_{t-1} G_1 + \phi_2 y_{t-1} G_2 + \phi_3 y_{t-1} G_3 + \varepsilon_t, \tag{4}$$

where the transition functions are now defined by:

$$\begin{aligned} G_1 &= [1 + \exp(\gamma(y_{t-1} + \lambda)]^{-1}, \\ G_2 &= 1 - G_1 - G_3, \\ G_3 &= [1 + \exp(-\gamma(y_{t-1} - \lambda)]^{-1}, \end{aligned}$$

and λ is the threshold parameter. Note that when y_{t-1} goes to $-\infty$, G_1 goes to unity while G_3 goes to zero, so that the MR-LSTAR dynamics are governed by ϕ_1. On the other hand, when y_{t-1} goes to $+\infty$, G_1 goes to zero whereas G_3 goes to one, so that the MR-LSTAR dynamics are governed by ϕ_3. One reason which motivates the choice of this MR-LSTAR model is that it becomes a SETAR model as the speed parameter, γ, goes to ∞. In this case, note that $G_1 \to I(y_{t-1} < -\lambda)$ and $G_3 \to I(y_{t-1} > \lambda)$. Hence, when γ goes to ∞, Model (4) rewrites as the following 3-regime SETAR:

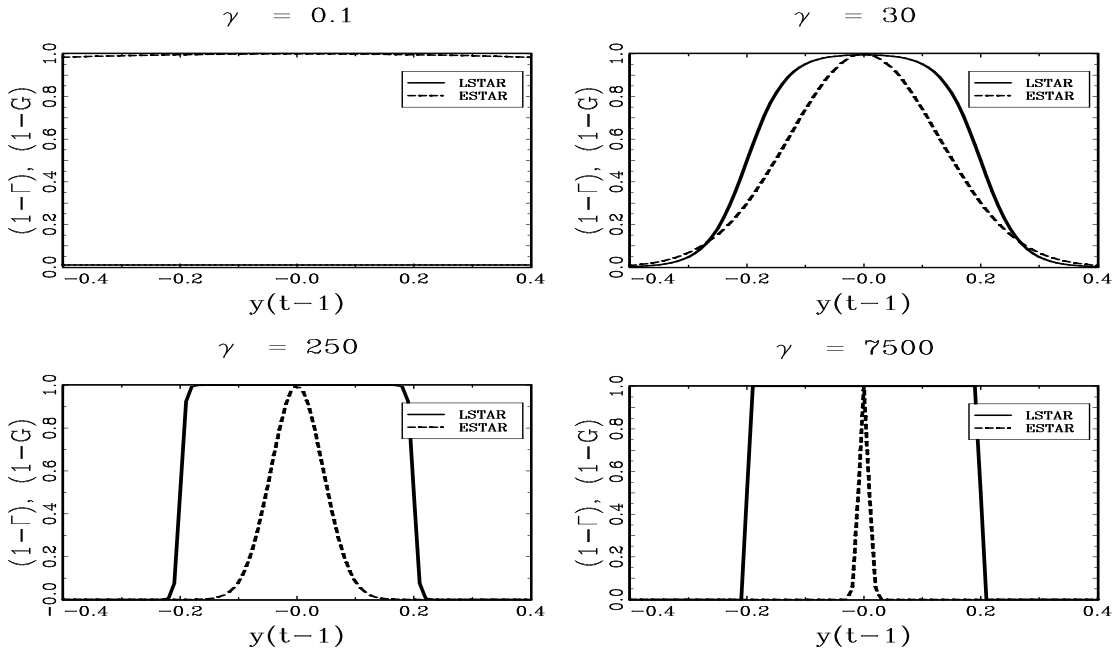
$$y_t = \phi_1 y_{t-1} I(y_{t-1} < -\lambda) + \phi_2 y_{t-1} I(-\lambda \leq y_{t-1} \leq \lambda) + \phi_3 y_{t-1} I(y_{t-1} > \lambda) + \varepsilon_t.$$

In order to compare this MR-LSTAR model with the ESTAR above, suppose that $\phi_1 = \phi_3$, i.e., the dynamics are assumed to be symmetrical in the 'lower' and 'upper' regimes, which is a maintained assumption in the ESTAR model (1). Consequently, the MR-LSTAR dynamics will be symmetrical around zero, which would correspond to $\kappa = 0$ in model (1). Model (4) then becomes:

$$y_t = \phi_1 y_{t-1}\Gamma + \phi_2 y_{t-1}(1-\Gamma) + \varepsilon_t, \tag{5}$$

where $\Gamma = G_1 + G_3$. Hence, Models (5) and (1) are quite similar to each other, except for the definition of the transition function, which is logistic in the former and exponential in the latter. To illustrate the properties of the transition functions Γ and G of models (5) and (1), we consider a sequence of $y_{t-1} \in [-0.4; +0.4]$, a threshold parameter $\lambda = 0.2$ and various values of the speed parameter, γ, ranging from 0.1 to 7500.[4] The ESTAR and MR-LSTAR functions associated with the 'inner' regime, namely $(1-G)$ and $(1-\Gamma)$, respectively, are reported in Figure 2. As the speed parameter approaches zero, both

Figure 2: MR-LSTAR and ESTAR transition functions



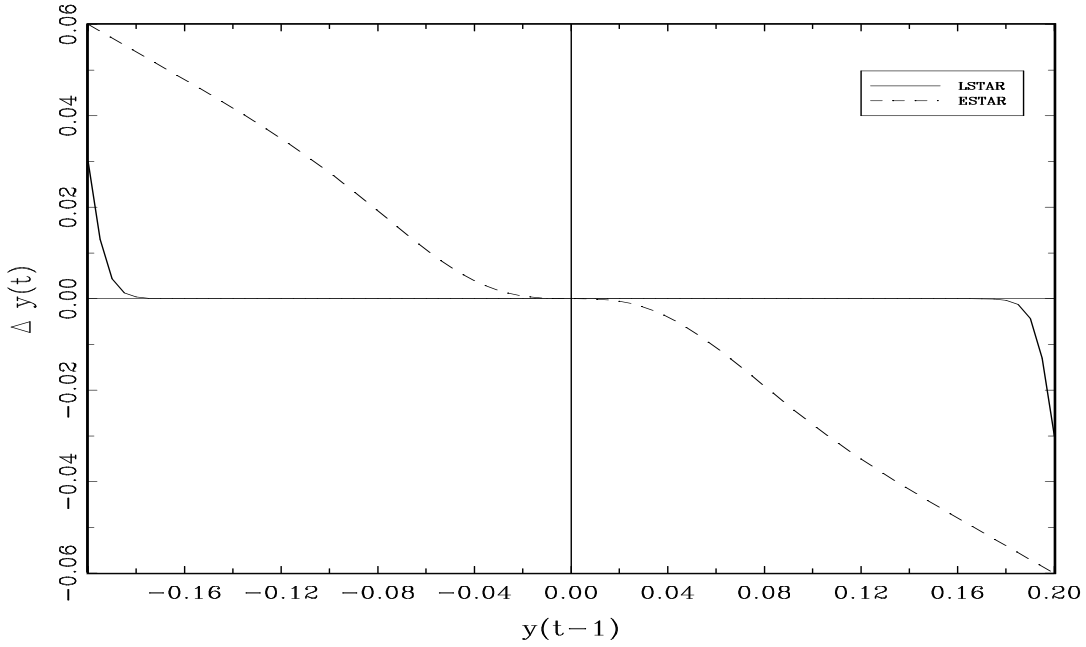
functions tend to become flat, as shown at the top left panel of this figure. For medium values, such as $\gamma = 30$, these functions are quite similar and, hence, similarly mimic the smooth transition adjustment. However, the shapes of $(1-G)$ and $(1-\Gamma)$ become different

[4]The values for y_{t-1} and λ are chosen according to the model estimates for the Italian Lira real exchange rate vis-à-vis the US Dollar.

as γ increases. The LSTAR function tends to become discontinuous, defining a central area for $y_{t-1} \in (-\lambda, +\lambda)$ exactly as a 3-regime SETAR model would do. On the contrary, the inner regime in the ESTAR model tends to shrink to a single point.

We now discuss how these properties translate in terms of conditional expected change functions. As the functions $(1 - G)$ and $(1 - \Gamma)$ behave very similarly to each other for small and medium values of γ, the conditional expected change functions will also be very close to each other for ESTAR and LSTAR models. However their shapes will differ dramatically for large values of γ. Figure 3 reports the conditional expected change function obtained from the symmetrical ESTAR and MR-LSTAR models for the following parameters values: $\phi_1 = 0.7$, $\phi_2 = 1$, $\kappa = 0$, $\lambda = 0.2$ and $\gamma = 250$. Figure 3 shows that the

Figure 3: Simulated conditional expected change functions



MR-LSTAR model has the advantage of mimicking the behavior of the real exchange rate predicted by Dumas' model when the level of relative risk aversion is low. On the other hand, the ESTAR is not able to capture these dynamics under any parametrization.

2.3 The proposed MR-LSTAR model

The general MR-LSTAR model we will study in the remainder of the paper writes:

$$\begin{aligned} \Delta y_t &= (-\mu_1 + \rho_1 y_{t-1})\, G_1 + (\mu_2 + \rho_2 y_{t-1})\, G_2 + (\mu_1 + \rho_1 y_{t-1})\, G_3 \\ &\quad + a_1 \Delta y_{t-1} + ... + a_{p-1} \Delta y_{t-p+1} + \varepsilon_t, \end{aligned} \tag{6}$$

with $\varepsilon_t \sim iid\left(0, \sigma^2\right)$, $G_1 = G\left(y_{t-1}, -\gamma, -\lambda\right)$, $G_2 = 1 - G_1 - G_3$, and $G_3 = G\left(y_{t-1}, \gamma, \lambda\right)$ where

$$G\left(y_{t-1}, \gamma, \lambda\right) = \left[1 + \exp\left(-\gamma\left(y_{t-1} - \lambda\right)\right)\right]^{-1}, \ \gamma > 0, \ \lambda > 0. \tag{7}$$

This symmetrical MR-LSTAR model generalizes model (5) by including lagged values of Δy_t to remove some of the autocorrelations in ε_t, and by allowing for drifts in the outer and inner regimes, denoted μ_1 and μ_2, respectively. Following the theoretical models discussed above, we maintain the assumption that the autoregressive coefficients are the same in the two outer regimes and we assume that $\mu_3 = -\mu_1$, as in e.g. Obstfeld and Taylor [1997].

As the focus of the paper is on testing unit root versus a stationary LSTAR alternative, we need to determine under which conditions the MR-LSTAR process is stationary and "well behaved". We briefly discuss sufficient conditions for y_t, defined in (6), to be $\beta - mixing$ with geometric decay. This property implies that (a) the stationary distribution of y_t exists, (b) starting from an arbitrary value y_0 the process y_t becomes stationary exponentially fast, and (c) y_t is $\alpha - mixing$ with geometric decay which is a desirable property to do inference.

Bec et al. [2004] study the mixing properties of a SETAR(p). We give here an intuitive argument that shows that the mixing properties of SETAR and MR-LSTAR are essentially the same. The mixing property (see e.g. Tjøstheim [1990]) of y_t is dictated by what happens as y_{t-1} goes to infinity. As y_{t-1} goes to plus infinity, G_1 converges to 0 and G_3 to 1, as y_{t-1} goes to minus infinity, G_3 goes to 0 and G_1 to 1, finally as $|y_{t-1}|$ goes to infinity, $1 - G_1 - G_3$ goes to zero. Therefore, for large values of $|y_{t-1}|$, the MR-STAR model behaves like a SETAR model where G_1 is replaced by $I\left(y_{t-1} < -\lambda\right)$ and G_3 by $I\left(y_{t-1} > \lambda\right)$. So the conditions on the parameters that guarantee the mixing property of a MR-LSTAR are the same as those for the mixing property of its SETAR counterpart. We refer the reader to the mixing conditions for a SETAR(p) model given in Theorem 1 of Bec et al. [2004] and do not reproduce them here. The more striking result is that the coefficient in the middle regime, ρ_2, may be equal to 0 (corresponding to a unit root) or positive (explosive root) while the model remains globally stationary.

For the data at hand, the paper tries to answer the following questions: *(i)* Is Model (6) stationary? *(ii)* Is it linear? The order in which these questions are addressed is essential. The next section highlights this point.

2.4 Testing linearity

The order in which the unit root test and linearity test are performed is crucial. The linearity tests proposed by, e.g., Hansen [1996] or Luukkonen, Saikkonen and Terasvirta

[1988] requires that the series be stationary. Consequently, one must establish stationarity before turning to linearity tests. In this section, we illustrate the way a linearity test may lead to fallacious inference in the presence of a unit root. Consider a simple LSTAR model given by:

$$y_t = \phi_1 y_{t-1}\left[1 - G\left(y_{t-1};\gamma,\lambda\right)\right] + \phi_2 y_{t-1} G\left(y_{t-1};\gamma,\lambda\right) + \varepsilon_t, \tag{8}$$

$$\varepsilon_t \sim iid\left(0,\sigma^2\right). \tag{9}$$

The hypothesis of interest is $H_0 : \phi_1 = \phi_2$. Under H_0, the model is linear and γ and λ are not identified, therefore the usual properties of the Wald test no longer hold. As an alternative, Luukkonen et al. [1988] suggest to use an auxiliary model,

$$y_t = \beta_0 y_{t-1} + \beta_1 y_{t-1}^2 + e_t, \tag{10}$$

and to test $H_0' : \beta_1 = 0$. The Wald test of this hypothesis (denoted WL in the sequel) has power against an alternative of type (8). Under the null hypothesis $H_0 : \phi_1 = \phi_2 = \phi$ where $|\phi| < 1$, the Wald test statistic converges asymptotically to a chi-square with one degree of freedom. But what is its limit if the process is a random walk, that is $\phi = 1$, under H_0? This is an important issue as the linearity is often tested on series for which there is no strong evidence of stationarity, see for instance Michael et al. [1997]. In the appendix, we show the following result.

Proposition **1** *If y_t is a random walk with $y_0 = 0$, then*

$$WL \xrightarrow{L} \frac{\left\{\frac{1}{3}B\left(1\right)^3 \int_0^1 B\left(r\right)^2 dr - \frac{1}{2}\left(B\left(1\right)^2 - 1\right)\int_0^1 B\left(r\right)^3 dr\right\}^2}{\int_0^1 B\left(r\right)^2 dr \left\{\int_0^1 B\left(r\right)^2 dr \int_0^1 B\left(r\right)^4 dr - \left(\int_0^1 B\left(r\right)^3 dr\right)^2\right\}},$$

where $B(.)$ is a standard Brownian motion.

From Proposition 1, we see that the asymptotic distribution of WL is very different from a chi-square distribution, when the DGP is non-stationary. Using 10,000 replications from a sample of size 10,000, we computed the fractiles of the distribution of WL and compared them with those of a chi-square with 1 degree of freedom. The results are summarized in Table 1. The line labelled "p-value" gives the probability of rejecting H_0 obtained when using the critical values given by the chi-square when the data follow a random walk. We see that the distribution of WL has a much thicker right tail than the $\chi^2\left(1\right)$. Using the critical values of the chi-square might result in rejecting wrongly the linear model. A 5% level for the chi-square corresponds to a 16.4% level for WL.

Table 1: Fractiles of WL and $\chi^2(1)$

	1%	5%	10%	50 %	90%	95%	99%
WL	.0001	.0054	.0215	.8012	5.1344	6.8564	10.9005
$\chi^2(1)$	-	.004	.016	.455	2.71	3.84	6.63
p-value WL	-	.9573	.9131	.5976	.2416	.1643	.0552

We illustrated our point on a simple model but we expect the same conclusions to hold for more general models. This is the reason why one should test for stationarity prior to testing for linearity and not the other way around.

3 Testing Unit Root versus MR-LSTAR

3.1 Likelihood-based unit-root tests

In model (6), we want to test $H_0 : \mu_1 = \mu_2 = \rho_1 = \rho_2 = 0$ "random walk without drift", against the alternative H_1 : stationary MR-LSTAR model. Under the null hypothesis, it is assumed that the roots of $1 - a_1 z - ... - a_p z^p = 0$ lie outside the unit circle. For convenience, we reparametrize the model in terms of $\beta = \lambda\gamma$ and λ, so that $G(y_{t-1}, \gamma, \lambda) = \left[1 + \exp\left(-\frac{\beta}{\lambda} y_{t-1} + \beta\right)\right]^{-1}$. Note that under H_0, the nuisance parameters β (or γ) and λ are not identified. It is therefore impossible to find consistent estimators of β and λ under the null hypothesis. For β and λ given, we can estimate the unrestricted and restricted regressions by OLS. The vector of unrestricted residuals obtained from (6) is denoted $\hat{\varepsilon}$. The restricted regression is given by

$$\Delta y_t = a(L)\Delta y_t + \varepsilon_t. \tag{11}$$

Denote the vector of restricted residuals by $\tilde{\varepsilon}$. In absence of heteroskedasticity, the trilogy of tests can be written in terms of the residual sum of squares:

$$\begin{aligned} W_T(\beta, \lambda) &= T\left[\frac{\tilde{\varepsilon}'\tilde{\varepsilon} - \hat{\varepsilon}'\hat{\varepsilon}}{\hat{\varepsilon}'\hat{\varepsilon}}\right], \\ LM_T(\beta, \lambda) &= T\left[\frac{\tilde{\varepsilon}'\tilde{\varepsilon} - \hat{\varepsilon}'\hat{\varepsilon}}{\tilde{\varepsilon}'\tilde{\varepsilon}}\right], \\ LR_T(\beta, \lambda) &= T\ln\left[\frac{\tilde{\varepsilon}'\tilde{\varepsilon}}{\hat{\varepsilon}'\hat{\varepsilon}}\right]. \end{aligned}$$

Proposition **2** *Let $\beta > 0$ and $\tilde{\lambda} = \lambda/\sqrt{T} > 0$ be fixed. Suppose $\pi = \left(\beta, \tilde{\lambda}\right)$ belongs to Π where Π is a compact set of R_*^{+2}. Under H_0, the Wald, Lagrange Multiplier and Likelihood*

Ratio tests satisfy

$$W_T(\beta,\lambda), LM_T(\beta,\lambda), LR_T(\beta,\lambda) \xrightarrow[T\to\infty]{L} \mathcal{D}(k) \tag{12}$$

uniformly in π, *where* $k = \left(\beta, \tilde{\lambda}/\delta\right)$, $\delta = \sigma/(1 - a_1 - a_2... - a_p)$ *and* $\mathcal{D}(k)$ *is a complicated function of Brownian motions given in Equation (22) in Appendix A. Under the alternative of a stationary MR-LSTAR model, the test statistics diverge.*

Next, we discuss the assumptions on β (or γ) and λ. The assumption $\lambda = \sqrt{T}\tilde{\lambda}$ is reminiscent of the assumption made in the structural change literature that the breakpoint is supposed to be equal to $T\pi$ with $\pi \in (0,1)$, which guarantees fixed proportions of observations before and after the break. Here, the assumption is that λ grows at the same rate as the standard error of an integrated process, so that under the null the proportions of observations in the lower and upper regimes have a lower bound. Note that because γ and λ are not identified under H_0, we are free to make any assumptions on them. Moreover under H_0, $y_t/\sqrt{T}$ converges to a Brownian motion, $\delta B(r)$, with $r = t/T$. First assume that γ and λ are fixed, then we obtain

$$\begin{aligned} G(y_t,\gamma,\lambda) &= [1+\exp(-\gamma(y_t-\lambda))]^{-1} \\ &= \left[1+\exp\left(-\gamma\sqrt{T}\left(\frac{y_t}{\sqrt{T}}-\frac{\lambda}{\sqrt{T}}\right)\right)\right]^{-1} \\ &\xrightarrow[T\to\infty]{} I\{\delta B(r)>0\}. \end{aligned}$$

This means that testing with γ and λ fixed is equivalent to testing in the context of a two-regime SETAR model with $\lambda = 0$. If moreover the symmetry of the outer regimes is assumed, then the resulting model is actually linear. This will result inevitably in a great loss of power. On the other hand, assuming that $\left(\beta, \tilde{\lambda}\right)$ is fixed, we obtain

$$G(y_t,\gamma,\lambda) = \left[1+\exp\left(-\frac{\sqrt{T}\beta}{\lambda}\frac{y_t}{\sqrt{T}}+\beta\right)\right]^{-1} \xrightarrow[T\to\infty]{L} \left[1+\exp\left(-\frac{\beta}{\tilde{\lambda}}\delta B(r)+\beta\right)\right]^{-1}. \tag{13}$$

Note that λ is assumed to diverge with $\sqrt{T}$ only under the null hypothesis. This assumption is used to derive the distribution of the test statistics under H_0. Under H_1, λ is of course assumed to be fixed.

As β and λ are not identified under the null hypothesis, the choice of β and λ is arbitrary. To select β and λ, we use the same strategy as in testing linearity against a SETAR model (see Tong [1990]), namely we take the supremum of the test statistics with

respect to the nuisance parameters. The tests under consideration are therefore:

$$\begin{aligned} \text{SupW} &\equiv \sup_{(\beta,\lambda)\in\mathcal{B}\times\Lambda} W_T(\beta,\lambda), \\ \text{SupLM} &\equiv \sup_{(\beta,\lambda)\in\mathcal{B}\times\Lambda} LM_T(\beta,\lambda), \\ \text{SupLR} &\equiv \sup_{(\beta,\lambda)\in\mathcal{B}\times\Lambda} LR_T(\beta,\lambda), \end{aligned}$$

where $\mathcal{B} = \left[\underline{b}, \bar{b}\right]$ and $\Lambda = \left[\underline{\lambda}, \bar{\lambda}\right]$. Since λ plays the role of a threshold, we adopt the same approach as in the SETAR literature. We order the absolute value of $y_t : |y|_{(1)} < |y|_{(2)} < ... < |y|_{(T)}$ and we discard 15% of the highest and smallest values. Hence, $\underline{\lambda} = |y|_{([15T/100])}$ and $\bar{\lambda} = |y|_{([85T/100])}$. For $\mathcal{B}$, we choose any arbitrary fixed interval. The test will have power even if a single value for β is used. However, the test will have more power if a range of values is considered. The interval should not be too wide either because for large β, G becomes flat. Note that $(\beta,\lambda) \in \mathcal{B}\times\Lambda$ implies that

$$0 < \underline{b} < \beta < \bar{b},\ 0 < \underline{\ell}_T < \frac{\tilde{\lambda}}{\delta} < \bar{\ell}_T, \tag{14}$$

where $\underline{\ell}_T = \underline{\lambda}/(\sqrt{T}\delta)$ and $\bar{\ell}_T = \bar{\lambda}/(\sqrt{T}\delta)$. Whereas $\underline{b}$ and $\bar{b}$ are fixed numbers, $\underline{\ell}_T$ and $\bar{\ell}_T$ are random variables. Their limiting distributions are given by

$$\begin{aligned} \underline{\ell} &\overset{L}{=} \lim_{T\to\infty} \frac{|y|_{([15T/100])}}{\sqrt{T}\delta} = \ell_{0.15}, \\ \bar{\ell} &\overset{L}{=} \lim_{T\to\infty} \frac{|y|_{([85T/100])}}{\sqrt{T}\delta} = \ell_{0.85}, \end{aligned}$$

where ℓ_p is the random variable that solves

$$\ell_p = \inf\left\{ l : \int_0^1 I\left[|B(r)| \le \ell\right] dr > p \right\}.$$

This choice of Λ and $\mathcal{B}$ insures that the asymptotic distributions of the sup tests are nuisance parameter free. This justifies the use of empirical critical values obtained by simulations.

Proposition **3** *Under* H_0,

$$\textit{SupW, SupLM, SupLR} \xrightarrow[T\to\infty]{L} \sup_{k\in K} D(k)$$

where $K = \left[\underline{b}, \bar{b}\right] \times \left[\underline{\ell}, \bar{\ell}\right]$. *Moreover, the limiting distributions of the sup tests are nuisance parameter free.*

The proofs of Propositions 2 and 3 are given in Appendix A. Note that the limiting parameter space, K, is random which is unusual and complicates the asymptotic theory. Our proof builds on results by Bec, Guay, and Guerre (2008, Theorems 2, 3, Section 2.3).

3.2 Unit-root tests based on an auxiliary model

Luukkonen et al (1988) introduced a linearity test based on an approximation of the function G by linearization. We propose to use the same approach to construct a unit-root test. This idea has been exploited before by Kapetanios, Shin and Snell (2003) for testing unit-root against a ESTAR model.

Using a ith order Taylor expansion of G around $\gamma = 0$, we get the following auxiliary model

$$\Delta y_t = a_1 \Delta y_{t-1} + ... + a_{p-1} \Delta y_{t-p+1} + \sum_{j=1}^{i} \beta_j y_{t-1}^{j+1} + \varepsilon_t. \tag{15}$$

In this case, a Wald test of $H_0 : \beta_1 = ... = \beta_i = 0$ will have in general power against a LSTAR (this test is referred to as F_i).

The question is which order i to choose. As a test F_i uses an auxiliary regression, there may be LSTAR models against which this test has no power. This problem is particularly serious for $i = 1$. Consider a simple illustration. Assume, to simplify, that $a_1 = ... = a_{p-1} = 0$, $\mu_1 = \mu_2 = \mu_3$, $\rho_1 = \rho_3$, and $\rho_2 = 0$. The OLS estimate, $\hat{\beta}_1$, of β_1 in

$$\Delta y_t = \beta_1 y_{t-1}^2 + \varepsilon_t \tag{16}$$

satisfies

$$\hat{\beta}_1 \xrightarrow{P} \frac{E\left(\Delta y_t y_{t-1}^2\right)}{E\left(y_{t-1}^4\right)},$$

and

$$E\left(\Delta y_t y_{t-1}^2\right) = \rho_1 E\left(\left(G_1 + G_3\right) y_{t-1}^3\right).$$

Remark that the function $\left(G_1 + G_3\right) y_{t-1}^3$ is a odd function of y_{t-1}. Therefore, if the density of y_t is symmetric around 0[5], we have $E\left(\left(G_1 + G_3\right) y_{t-1}^3\right) = 0$, and then $\hat{\beta}_1 \xrightarrow{P} 0$. Hence F_1 does not have much power against such a LSTAR. This is however a special case, this lack of power is by no means the rule. F_2 would have power in this example.

The higher the order i of the Taylor expansion, the larger the range of possible alternatives against which the test will have power. For instance, F_2 has, in general, power not only against LSTAR but also against ESTAR models (see Kapetanios et al. 2003). As estimating parameters is costly, we choose to adopt F_2 which is more parsimonious than F_3.

Proposition **4** *Under H_0, F_2 has the nonstandard distribution given in (24) in Appendix A, which is nuisance parameter free.*

[5]This could be expected if the density of ε_t is itself symmetric around 0.

It is expected that F_2 will have less power than our sup test because F_2 is not specifically designed to test LSTAR. Note however that, because of the presence of nuisance parameters that are not identified under the null, there is no uniformly most powerful test.

3.3 Empirical critical values of the unit-root tests

In this section, we compute the empirical critical values of tests described earlier, SupW, SupLM, SupLR, F_2, and another test, SupLRh. SupLRh is defined as

$$\text{SupLRh} \equiv \sup_{(\beta,\lambda)\in\mathcal{B}\times\Lambda} LRh_T\left(\beta,\lambda\right),$$

where LRh_T is the heteroskedasticity-robust version of the likelihood ratio test, which exact expression is given in Appendix B.

In the empirical study below, we have found that $p = 2$, so that model (6) may be rewritten as follows :

$$\begin{aligned}\Delta y_t &= a\Delta y_{t-1} + \mu_1(G_3 - G_1) + \mu_2 G_2 \\ &\quad +\rho_1 y_{t-1}(G_1 + G_3) + \rho_2 y_{t-1} G_2 + \sigma\varepsilon_t, \end{aligned} \tag{17}$$

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

$$\Delta y_t = a\Delta y_{t-1} + \sigma\varepsilon_t, \tag{18}$$

where the ε_t's are drawn from an iid $\mathcal{N}(0,1)$, $a = 0.3$, and $\sigma = 0.02$. This choice of the parameters is dictated by the data. When fitting (18) on the real exchange rates, we obtain, for most of the series, a close to 0.3 with a range $0.13 \leq a < 0.4$ and σ around 0.02. In Table 2, we report the empirical critical values from 10,000 replications of samples of size 200, 300 and 400. The two-dimensional search grid in β and λ was performed for the following sets of values : $\beta \in \{0.2, 0.3, 0.4, 0.5, 0.6\}$ and $\lambda \in [\underline{\lambda}, \overline{\lambda}]$ with $\underline{\lambda}$ and $\overline{\lambda}$ such that 15% of the smallest and highest values of $|y_{t-1}|$ are excluded from the grid.

3.4 Size and power analysis of the unit-root tests

In order to examine the size and power of the proposed tests, we perform a small sample study. For both Tables 3 and 4 and later when we analyze the data, we use the empirical critical values obtained in Table 2. Hence, the power is actually a size-corrected power. First, we generate the model under the null, i.e., model (18) for $a = 0.3$ and $\sigma = 0.02$ and ε_t iid $\mathcal{N}(0,1)$ using a different seed for the random number generator from the one used to compute the empirical critical values. In Table 3, we report the empirical rejection

Table 2: Empirical critical values of the unit-root test (a=0.3, $\sigma = 0.02$)

	20 %	15%	10%	5%	1%
			$n = 200$		
SupW	10.32	11.35	12.64	14.81	19.46
SupLM	10.02	10.96	12.13	14.07	18.09
SupLR	10.07	11.04	12.26	14.29	18.57
SupLRh	9.66	10.52	11.69	13.48	17.58
F_2	7.02	7.75	8.71	10.33	14.33
			$n = 300$		
SupW	10.22	11.18	12.44	14.40	18.61
SupLM	10.02	10.93	12.11	13.92	17.76
SupLR	10.05	10.98	12.19	14.06	18.05
SupLRh	9.71	10.93	11.58	13.37	17.31
F_2	7.01	7.77	8.79	10.35	13.61
			$n = 400$		
SupW	10.08	10.99	12.10	14.02	18.20
SupLM	9.94	10.81	11.86	13.68	17.59
SupLR	9.96	10.84	11.92	13.78	17.80
SupLRh	9.59	10.42	11.49	13.36	17.29
F_2	6.99	7.70	8.74	10.21	13.60

frequencies from 5,000 Monte Carlo replications with $n = 200$, 300 and 400 successively. For comparison purpose, the empirical size of the Augmented Dickey-Fuller statistics is also reported. Here and in the rest of the paper, we do not report the SupW and SupLM tests because it is well-known that the Wald test exhibits important size distortions and the LM test is usually very close to the LR test.

Table 3: Empirical size of the unit-root tests

Theoretical size	ADF	SupLR	SupLRh	F_2
		$n = 200$		
1%	0.009	0.008	0.008	0.007
5%	0.048	0.045	0.045	0.035
10%	0.100	0.094	0.089	0.066
		$n = 300$		
1%	0.009	0.010	0.008	0.007
5%	0.048	0.053	0.047	0.032
10%	0.097	0.097	0.100	0.063
		$n = 400$		
1%	0.009	0.008	0.008	0.008
5%	0.050	0.053	0.048	0.032
10%	0.095	0.104	0.097	0.061

From Table 3, we see that the empirical size of our tests is quite accurate. Nevertheless, the SupLRh and the F_2 tests appear slightly more conservative than the SupLR.

Next, we explore the power of the tests by generating 1,000 series under the alternative, (17), for various parameter values. In all the following experiments, we normalize σ to unity. We also set $a = 0.3$, $\mu_2 = \rho_2 = 0$ and $\mu_1 = \lambda\rho_1$, which is consistent with our MR-LSTAR estimates (see next section). In Table 4, we report the size-corrected power of the sup tests described in Subsection 3.1, the test F_2 described in Subsection 3.2, the Adjusted Dickey Fuller test, and the test proposed by Kapetanios et al. (denoted KSS). The theoretical size of these tests is $\alpha = 5\%$. The power of the sup tests is increasing in λ, $|\rho_1|$ and n. In most of the MR-LSTAR estimates reported in the next section, the λ/σ ratio is greater than eight, which is close to the case with $\lambda = 10$ in Table 4. In those cases, our SupLR test clearly outperforms the ADF one. As expected, SupLR clearly dominates the tests based on auxiliary models: KSS and F_2. Although the KSS test has been developed to test against ESTAR alternative, it seems to have comparable power to

Table 4: Size-corrected power of the unit root tests

$(\rho_1, \lambda/\sigma, \gamma/\sigma)$	n	SupLR	SupLRh	F_2	K.S.S	ADF
(-0.05,2,10)	200	52.7	51.6	54.9	50.4	82.1
	300	88.0	86.4	74.8	72.1	99.2
	400	98.8	98.5	89.3	87.4	100.0
(-0.10,2,10)	200	99.0	98.1	93.5	91.0	99.9
	300	100.0	100.0	99.0	98.0	100.0
	400	100.0	100.0	99.6	99.4	100.0
(-0.30,2,10)	200	100.0	100.0	100.0	100.0	100.0
	300	100.0	100.0	100.0	100.0	100.0
	400	100.0	100.0	100.0	100.0	100.0
(-0.05,10,10)	200	55.4	25.4	37.0	17.1	16.9
	300	83.3	44.8	67.7	69.0	19.3
	400	95.7	64.6	92.9	95.2	28.0
(-0.10,10,10)	200	83.2	24.7	57.3	48.9	16.8
	300	96.8	41.5	90.0	92.1	21.4
	400	99.6	61.9	99.5	99.4	35.6
(-0.30,10,10)	200	97.2	19.6	95.7	95.3	31.8
	300	99.6	34.2	99.7	99.8	66.4
	400	99.9	56.1	100.0	100.0	93.2
(-0.05,2,200)	200	52.0	52.1	54.5	50.5	82.2
	300	88.0	86.4	74.6	71.8	99.2
	400	98.9	98.1	89.4	87.4	100.0
(-0.10,2,200)	200	98.6	98.0	93.4	90.8	99.9
	300	100.0	100.0	99.0	98.0	100.0
	400	100.0	100.0	99.6	99.6	100.0
(-0.30,2,200)	200	100.0	100.0	100.0	100.0	100.0
	300	100.0	100.0	100.0	100.0	100.0
	400	100.0	100.0	100.0	100.0	100.0
(-0.05,10,200)	200	56.5	26.0	37.1	19.1	17.2
	300	82.2	44.3	70.3	71.6	19.1
	400	96.6	65.9	94.2	96.4	28.7
(-0.10,10,200)	200	85.6	22.5	62.0	56.1	16.3
	300	97.8	39.4	92.9	95.0	21.7
	400	99.8	59.5	99.9	99.8	39.3
(-0.30,10,200)	200	98.7	19.9	97.8	97.9	40.7
	300	100.0	35.7	99.9	99.9	82.1
	400	100.0	98.4	100.0	100.0	98.6

F_2. We also performed some power simulations (not reported here) on the tests F_1 and F_3 described in Subsection 3.2. The performance of F_1 is very poor, while that of F_3 is comparable to that of F_2. Finally, the SupLRh test outperforms the KSS test for a small value of λ but its power deteriorates a lot for large threshold values, even when ρ_1 is large in absolute value. For instance, with $n = 100$ and $\rho_1 = -0.30$, its power is 100% for $\lambda = 2$ but drops to around 35% for $\lambda = 10$, no matter the value of γ. Motivated by these results, the unit root tests retained in the following empirical study are the SupLR and SupLRh, the former being privileged in presence of large threshold values.

4 The empirical results

The data set comprises monthly observations[6] spanning 1973:09 to 2009:12 for eight countries : United-States, Germany, United-Kingdom, Italy, Canada, France, Belgium and Finland. The corresponding currencies are denoted USD, DEM, GBP, ITL, CAD, FRF, BEF, FIM. The nominal exchange rate data are monthly averages, and the nominal price data are consumption price indices. Overall, we have twenty-eight real exchange rates. For the pairs of currencies involving a country belonging to the Euro zone, we used data up to December 1998, since the Euro was introduced in January 1999. So, the sample size is only 304 for these pairs, whereas it is 466 for the remainders.

4.1 Standard unit root tests

First, we check the order of integration of the real exchange rates in the linear autoregressive model using three statistics, namely ADF (Dickey and Fuller [1981]), PP (Phillips and Perron [1988]) and KPSS (Kwiatkowski, Phillips, Schmidt and Shin [1992])[7]. The corresponding results are reported in Table 5[8]. These tests fail to reject the unit root for every pair, except for GBP/USD and FRF/DEM. The BEF/DEM pair is the only one for which a deterministic time trend is significant in the ADF regression. When allowing for this trend, we find that ADF and PP statistics are respectively -3.59 and -3.51, which are both significant at the 5% level. Since the following analysis is not suited to handle that case, the further results obtained from this pair should be interpreted with caution. Since ADF and PP tests were shown to have low power against nonlinear alternatives by Pippenger and Goering [2000] among others, these results can not constitute evidence.

[6] The data were obtained from Datastream.

[7] We include at most a constant term in the deterministic component under the null.

[8] The lag length for the ADF (k) is chosen according to the Ljung-Box statistic. It is always equal to 1. The size of the Bartlett windows for PP and KPSS tests (resp. ℓ and m) is obtained following Andrews [1991].

Table 5: Tests of I(1) or I(0) for y_t

currency	ADF(k)	PP(ℓ)	KPSS(m)	ADF(k)	PP(ℓ)	KPSS(m)
	versus USD			versus DEM		
DEM	-1.68	-1.33	0.58			
GBP	-3.24*	-2.95*	0.24*	-2.12	-1.82	1.39
ITL	-1.75	-1.28	0.76	-1.93	-1.66	1.75
CAD	-0.48	-0.36	4.06	-1.99	-1.65	1.58
FRF	-1.51	-1.14	0.48	-2.98*	-2.62	0.67
BEF	-1.40	-1.20	0.70	-2.17	-2.07	2.64
FIM	-1.24	-0.80	0.55	-2.04	-1.84	0.94
	versus GBP			versus ITL		
ITL	-2.19	-1.77	0.73			
CAD	-2.51	-2.48	0.95	-2.27	-1.83	3.22
FRF	-2.06	-1.60	1.92	-1.85	-1.58	2.21
BEF	-1.61	-1.22	2.05	-1.49	-1.24	2.31
FIM	-0.84	-0.52	2.04	-2.44	-2.10	2.13
	versus CAD			versus FRF		
FRF	-2.07	-1.73	1.36			
BEF	-1.95	-1.58	0.90	-2.48	-2.03	1.51
FIM	-2.57	-2.19	1.32	-2.23	-1.77	0.81
	versus BEF					
FIM	-1.69	-1.29	1.12			

The critical values at the 5 % level are -2.88 for ADF and PP, and 0.463 for KPSS.

4.2 Testing for unit-root and for linearity against MR-LSTAR

Before proceeding to the supLR test, we test the homoskedasticity of the residuals in the linear model. According to the tests of homoskedasticity proposed by White, Engle and Pagan[9], the null of homoskedasticity is rejected for the following pairs : GBP/USD, CAD/GBP, ITL/GBP, BEF/GBP, FRF/DEM, ITL/DEM, BEF/DEM, FRF/ITL, BEF/ITL and BEF/FRF. For these series, we will also consider the heteroskedasticity-robust version of the LR tests and the standard-errors of the parameters estimates will be corrected using White [1980]'s consistent estimator of the covariance matrix.

Table 6 reports the results of unit-root and linearity tests calculated from the MR-LSTAR models, (17), for which the unit root hypothesis can be rejected. The real exchange rate data used for the MR-LSTAR estimation are demeaned because model (17) implies symmetrical behavior around zero.

According to the unit-root tests' statistics given in the first two columns of Table 6, eleven real exchange rates reject the null of random walk against our MR-LSTAR alternative, namely GBP/USD, GBP/FRF, GBP/DEM, CAD/GBP, FRF/DEM, ITL/DEM, BEF/DEM, FRF/ITL, FIM/ITL, BEF/FRF and FIM/DEM. In this table, exponents $^{-}$, * and **, respectively, denote the 15, 10 and 5 percent significance levels. The conclusions arising from

Table 6: SupLR Unit-root, and LM linearity tests

	SupLR	SupLRh	LML
GBP/USD	12.99^{*}	7.89	3.87^{-}
GBP/DEM	13.55^{*}	7.86	6.25^{**}
GBP/FRF	12.69^{*}	10.86	6.05^{**}
CAD/GBP	10.63^{-}	7.08	1.54
FRF/DEM	25.45^{**}	4.83	74.87^{**}
ITL/DEM	27.26^{**}	4.53	73.16^{**}
BEF/DEM	12.32^{*}	9.44	8.84^{**}
FRF/ITL	19.07^{**}	3.47	91.62^{**}
FIM/ITL	16.56^{**}	4.93	12.94^{**}
BEF/FRF	16.02^{**}	2.99	95.10^{**}
FIM/DEM	11.75^{*}	5.70	5.47^{*}

Note : Data are centered.

the SupLRh statistics are more conservative of the unit root null than the ones of the SupLR. Nevertheless, as will be seen in the next table, our MR-LSTAR estimates of λ

[9]In order to save space, these results are not reported but are available upon request.

and γ parameters are typically large and hence correspond to the cases where the power of the SupLRh statistics may fall under 40%. For this reason, we decide to keep the series rejecting the null according to the SupLR in the subsequent analysis.

Before performing the corresponding MR-LSTAR maximum likelihood estimations, it is necessary to test linearity against the MR-LSTAR representation. To this end, we proceed along the lines suggested by Luukkonen et al. [1988], which consist in approximating the transition function $G(.)$ by a suitable Taylor series expansion. Accordingly, Model (6) is rewritten by replacing the transition functions by their second-order approximations, which are preferred to the first-order ones in order to capture the possible nonlinearity arising from the intercepts :

$$\Delta y_t = a\Delta y_{t-1} + \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-1}^2 + \beta_3 y_{t-1}^3 + e_t,$$

where the β_i's, $i = 0, 1, 2, 3$ are functions of the parameters ϕ_1, ϕ_2, γ, and λ. We use a Lagrange Multiplier test (denoted LML) to test $H_0 : \beta_2 = \beta_3 = 0$ in the regression above. This test follows a $\chi^2(2)$ distribution under H_0 and diverges under the MR-LSTAR alternative. We report the values of this statistic in the third column of Table 6. For the pairs displaying heteroskedasticity, the LM statistic is corrected along the lines described in Appendix B. The null of linearity is rather strongly rejected for all these pairs, except for the CAD/GBP one.

4.3 The constrained maximum likelihood estimation

The estimation of the MR-LSTAR representation given in (17) was performed using the constrained maximum likelihood method[10]. We imposed the constraints $\gamma > 0$ and $\lambda \in [\underline{\lambda}, \overline{\lambda}]$, with $\underline{\lambda}$ and $\overline{\lambda}$ such that 10% of the observations, in absolute value, are below $\underline{\lambda}$ and 10% are above $\overline{\lambda}$. Following van Dijk et al. [2002], we simplify the estimation problem by concentrating the sum of squares function to be minimized. Indeed, for known and fixed γ and λ, our MR-LSTAR model is linear in the other parameters. So, conditional upon γ and λ, the estimates of the other parameters can be obtained by ordinary least squares. This property allows us to reduce the nonlinear estimation problem, since the sum of squared residuals has to be minimized with respect to γ and λ only. The starting values are then obtained from a two-dimensional grid over these two parameters. The set of grid values for the threshold λ is $[\underline{\lambda}, \overline{\lambda}]$ as defined above. The choice of a grid for γ is less obvious. Once again, we follow the advice given by van Dijk et al. [2002] which consists

[10] We used the CML library of Gauss, with the Newton-Raphson optimization algorithm.

in reparameterizing the transition function as follows:

$$G(s_t, \gamma, \lambda) = \left[1 + \exp\left(-\frac{\gamma}{\hat{\sigma}_{y_{t-1}}}(s_t - \lambda)\right)\right]^{-1},$$

where $\hat{\sigma}_{y_{t-1}}$ is the sample standard deviation of the switching variable, so as to make γ approximately scale-free. Then, the grid for γ was arbitrarily set to $\{2.5, 5, 7.5, 10, ..., 25\}$.

The MR-LSTAR estimates for these eleven pairs are reported in Table 7. The standard error estimates of $\widehat{\gamma}$ are not reported. Indeed, due to an identification problem, this parameter estimate does not have a standard asymptotic distribution. Moreover, this parameter cannot be accurately estimated when it is large, since in that case, the transition function is close to a step function, and one would need many observations in the neighborhood of λ to estimate it accurately. In fact, outside of this area, large changes in γ have only small effects on the shape of the transition function.

Table 7: MR-STAR estimates

	$\widehat{a}$	$\widehat{\mu}_1$	$\widehat{\mu}_2$	$\widehat{\rho}_1$	$\widehat{\rho}_2$	$\frac{\widehat{\gamma}}{\hat{\sigma}_{y_{t-1}}}$	$\widehat{\lambda}$	$\widehat{\sigma}_\varepsilon$
GBP/USD	**0.351**	0.006	-0.002	**-0.062**	-0.071	157.1	**0.073**	0.024
($n = 436$)	(0.045)	(0.005)	(0.002)	(0.028)	(0.070)		(0.021)	[0.25,0.47]
GBP/FRF	**0.358**	**0.046**	0.001	**-0.299**	-0.023	4358.0	**0.131**	0.020
($n = 304$)	(0.053)	(0.016)	(0.001)	(0.098)	(0.018)		(0.006)	[0.16,0.69]
GBP/DEM	**0.354**	**0.048**	-0.001	**-0.255**	-0.014	1891.2	**0.176**	0.021
($n = 304$)	(0.060)	(0.016)	(0.001)	(0.077)	(0.012)		(0.014)	[0.08,0.77]
CAD/GBP	**0.280**	**0.046**	0.001	**-0.168**	**-0.028**	987.8	**0.229**	0.025
($n = 436$)	(0.052)	(0.022)	(0.001)	(0.077)	(0.011)		(0.005)	[0.10,0.84]
FRF/DEM	**0.345**	**0.045**	0.000	**-0.549**	-0.025	4884.2	**0.071**	0.010
($n = 304$)	(0.075)	(0.015)	(0.001)	(0.163)	(0.017)		(0.001)	[0.06,0.89]
ITL/DEM	**0.374**	**0.076**	0.000	**-0.438**	-0.019	4243.1	**0.142**	0.015
($n = 304$)	(0.104)	(0.023)	(0.001)	(0.127)	(0.014)		(0.11)	[0.14,0.69]
BEF/DEM	0.152	0.003	0.001	**-0.060**	**-0.209**	1035.0	**0.028**	0.008
($n = 304$)	(0.102)	(0.002)	(0.001)	(0.029)	(0.062)		(0.005)	[0.27,0.37]
FRF/ITL	**0.294**	**0.076**	0.000	**-0.466**	-0.004	517.1	**0.152**	0.015
($n = 304$)	(0.105)	(0.044)	(0.001)	(0.257)	(0.011)		(0.010)	[0.03,0.90]
FIM/ITL	**0.307**	**0.075**	-0.001	**-0.447**	-0.027	2056.4	**0.138**	0.017
($n = 304$)	(0.079)	(0.038)	(0.001)	(0.206)	(0.015)		(0.004)	[0.06,0.87]
BEF/FRF	**0.341**	0.020	0.000	-0.187	-0.02	698.6	**0.085**	0.010
($n = 304$)	(0.081)	(0.020)	(0.001)	(0.159)	(0.025)		(0.009)	[0.11,0.87]
FIM/DEM	**0.351**	**0.165**	0.000	**-0.773**	-0.012	125.5	**0.201**	0.016
($n = 304$)	(0.054)	(0.087)	(0.001)	(0.387)	(0.011)		(0.019)	[0.07,0.90]

Notes: Data are centered. The numbers in parentheses are the standard errors. The figures in brackets in the last column are the percentages of observations below $-\hat{\lambda}$ and between $-\hat{\lambda}$ and $\hat{\lambda}$, respectively.

The estimated parameters for these eleven pairs share a lot of common features.

Firstly, the estimates of $\frac{\widehat{\gamma}}{\hat{\sigma}_{y_{t-1}}}$ are very large, ranging from 125.5 for FIM/DEM to 4884.2 for FRF/DEM, thus providing support for a SETAR specification. This is confirmed by the plots of the transition functions: in Appendix C, we report the G_2 function of model (17) corresponding to the smallest ratio $\frac{\widehat{\gamma}}{\hat{\sigma}_{y_{t-1}}}$ obtained for the FIM/DEM pair, and to the largest one which is obtained for the FRF/DEM pair. Actually, most of the series exhibit an almost discontinuous adjustment. Secondly, the parameter estimates always have the expected sign, and relative size. Indeed, in every case μ_1 is positive while μ_2 is negative or null, and ρ_1 is rather strongly negative — ranging from -0.062 for the GBP/USD pair to -0.773 for the FIM/DEM pair — while ρ_2 is never significantly different from zero, except for the CAD/GBP and BEF/DEM pairs.

The results obtained for the CAD/GBP pair are amazingly similar to the other ones, although we were not able to reject the linearity hypothesis in that case. Its parameter estimates appear to be significantly different across regimes, the threshold estimate is significantly different from zero and the smoothness parameter is not close to zero. Finally, the only striking result is the one obtained for BEF/FRF, for which no other parameter than $\hat{a}$ and $\hat{\lambda}$ seems to be significantly different from zero, even at the 10% level.

4.4 Half-lives

One of the PPP puzzle is the high degree of persistence in the real exchange rate (Rogoff [1996]). In this subsection, we compute the half-lives resulting from the MR-LSTAR model in order to compare them with existing results.

First we compute the impulse response function following the Monte Carlo method described in Gallant, Rossi and Tauchen [1993] and Taylor et al. [2001]. We set the initial values of y_{t-1} and y_{t-2} to the value given in the column labelled "starting value". Then, we generate two series of length $T = 100$ with identical errors except that the first series has an extra additive shock at time t of the form $\ln(1 + k/100)$ with $k = 40, 20, 5$, and 1. This corresponds to shocks of k percent. Then, we compute the difference between the two series and repeat this procedure 5000 times. We average out the differences to obtain an estimate of the impulse response function. This function is used to compute the half-lives reported below.

As the model is nonlinear, the impulse response function and hence the half-lives depend crucially on the amplitude of the shock and on the starting values. We consider a set of four starting values. One corresponds to the sample mean of the data, 0, as suggested by Gallant et al. [1993]. The other ones have been chosen to get some insights on the impact of a shock if the starting values are on the edge of the band. As we consider

only positive shocks, we also investigate a starting value far below the left boundary, namely -2λ. Table 8 reports half-lives (in months) for three representative real exchange rates: GBP/USD, FIM/DEM, and FRF/DEM[11].

Table 8: Half-lives (in months)

	shock size (%)			
starting value	40	20	5	1
	GBP/USD			
-2λ	24	27	22	19
$-\lambda$	23	29	25	24
0	21	28	30	26
λ	18	24	33	35
	FIM/DEM			
-2λ	35	50	12	2
$-\lambda$	44	55	46	24
0	4	41	59	56
λ	2	2	2	4
	FRF/DEM			
-2λ	5	27	5	1
$-\lambda$	4	27	34	33
0	2	4	33	33
λ	1	1	2	14

It is also informative to compute the half-lives associated with the outside regime. They are obtained from the formula given by Hamilton (1994, page 10) by doing as if the model were an AR(2) with autoregressive coefficients $\phi_1 = 1 + a + \rho_1$ and $\phi_2 = -a$. We find the following half lives: 12 months for GBP/USD and only 1 month for FIM/DEM and FRF/DEM. These values provide a lower bound for the half-lives. Such values will result either from a very large shock that brings the real exchange rate into the outside regime (see FRF/DEM, cell $(\lambda, 40)$), or from a small shock in the outside regime so that the real exchange rate remains in this regime (see FRF/DEM, cell $(-2\lambda, 1)$).

Note that the dissymmetry of the results between the case where the starting value is λ and the case where it is -λ is due to the fact that the shock is positive. Starting from λ, the shock pushes toward the outside regime and hence dies out faster than starting from

[11] GBP/USD has been selected because it also appears in Taylor et al. [2001]. FIM/DEM, and FRF/DEM correspond to the series exhibiting the smoothest and most discountinuous adjustments, respectively.

-λ. For the same reason, a large shock may be very persistent if it occurs below the left boundary and it pushes the real exchange rate into the middle regime (see cell $(-2\lambda, 40)$ of FIM/DEM).

When the starting value is set at the mean (0), we clearly see that the mean reversion for large shocks is stronger than for small shocks. This is consistent with the presence of transaction costs. The half-lives we obtain for GBP/USD range from 18 to 35 months, which are comparable to those obtained by Taylor et al. [2001] using a ESTAR model. While most empirical studies based on a linear model agree on a half life for GBP/USD of about 4.6 years (see Rogoff [1996]), our results suggest a much shorter half-life. In a recent paper, Taylor [2001] finds half-lives equal to 1.7 years for GBP/USD using an AR(2) model. These different results can be reconciled in the following manner. In a linear model, the half life is computed for a shock of size 1 and is independent of the starting values. This half-life can be thought of as a weighted average of the half-lives obtained for various starting values.

5 Conclusion

This paper explores the possibility that, in presence of transaction costs, a nonlinear MR-LSTAR representation is more relevant for the real exchange rate than a linear specification. While most linear models predict a half-life for PPP deviations ranging from 3 to 5 years, our model shows that large shocks adjust much faster than small shocks, resulting in a half-life as short as two months for FRF/DEM.

Whereas the classical tests (PP, ADF) fail to reject the null of a unit-root for all of the real exchange rates but GBP/USD and FRF/DEM, our SupLR unit-root test rejects the null of a unit-root in favor of a three-regime MR-LSTAR model for eleven pairs. For these series, the estimation results support the PPP hypothesis by exhibiting strong mean reversion for large PPP departures. Seven out of these eleven pairs involve currencies of the European Monetary System member states. This finding may indeed be related to the close trade links developed by the European Economic Community. The European build up has probably made international arbitrage in goods markets easier inside the EEC.

Another contribution of the paper lies in the modeling of the real exchange rate. While former empirical studies focused on either the SETAR or ESTAR models, our paper uses a MR-LSTAR specification, which is able to mimic both the abrupt adjustments of the SETAR and the smooth adjustments of the ESTAR. It is worth noting that in the case of FRF/DEM, the shape of the estimated transition functions is very close to the shape of the indicator function characterizing the SETAR model. Consequently, this more

parsimonious model may be considered for this particular series as a good approximation of the more general MR-LSTAR model we have studied in this paper.

References

Andrews, D.W.K., Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica*, 1991, *59* (3), 817–858.

Baum, C., J. Barkoulas, and M. Caglayan, Nonlinear adjustment to purchasing power parity in the post-Bretton Woods era, *Journal of International Money and Finance*, 2001, *20*, 379–399.

Bec, F., A. Guay, and E. Guerre, Adaptive consistent unit root tests based on autoregressive threshold model, *Journal of Econometrics*, 2008, *142*, 94–133.

___ , M. Ben Salem, and M. Carrasco, Test for Unit-root Versus Threshold Specification with an Application to the PPP, *Journal of Business and Economic Statistics*, 2004, *22*, 382–395.

Berka, M., *General Equilibrium Model of Arbitrage Trade and Real Exchange Rate Persistence*, Manuscript, University of British Columbia 2004.

Dickey, D. and W. Fuller, Likelihood ratio statistics for autoregressive time series with an unit root, *Econometrica*, 1981, *49* (4), 1057–1072.

Dumas, B., Dynamic equilibrium and the real exchange rate in a spatially separated world, *Review of Financial Studies*, 1992, *5* (2), 153–80.

Enders, W. and C.W.J. Granger, Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates, *Journal of Business and Economic Statistics*, 1998, *16* (3), 304–11.

Epstein, L. and S. Zin, Substitution, risk aversion and the temporal behavior of consumption and asset returns: an empirical analysis, *Journal of Political Economy*, 1991, *99*, 263–268.

Gallant, A., P. Rossi, and G. Tauchen, Nonlinear Dynamic Structures, *Econometrica*, 1993, *61* (4), 871–907.

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

Hansen, B.E., Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica*, 1996, *64* (2), 413–430.

Hansen, L.P. and K. J. Singleton, Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica*, 1982, *50*, 1269–1286.

___ and ___ , Errata: Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica*, 1984, *52*, 267–268.

Kapetanios, G., Y. Shin, and A. Snell, Testing for a unit root in the nonlinear STAR framework, *Journal of Econometrics*, 2003, *112* (2), 359–379.

Kilian, L. and M. Taylor, Why Is It So Difficult to Beat the random Walk Forecast of Exchange Rates ?, *Journal of International Economics*, 2003, *60* (1), 85–107.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin, 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*, 1992, *54*, 159–178.

Lo, M. C. and E. Zivot, Threshold Cointegration and Nonlinear Adjustment to the Law of One Price, *Macroeconomic Dynamics*, September 2001, *5* (4).

Lothian, J. and M. Taylor, Purchasing Power Parity Over Two Centuries: Strengthening the Case for Real Exchange Stability, *Journal of International Money and Finance*, 2000, *19*, 759–764.

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

Michael, P., A. Nobay, and D. Peel, Transactions costs and nonlinear adjustment in real exchange rates : an empirical investigation, *Journal of Political Economy*, 1997, *105* (4), 862–79.

Murray, C. and D. Papell, The Purchasing Power Parity Persistence Paradigm, *Journal of International Economics*, 2002, *56*, 1–19.

___ and ___ , *The Purchasing Power Parity Puzzle is Worse Than You Think*, Manuscript, University of Huston 2002.

Newey, W. and K. West, Hypothesis Testing with Efficient Method of Moments Estimation, *International Economic Review*, 1987, *55* (3), 703–708.

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

O'Connell, P. and S-J. Wei, The bigger they are, the harder they fall : Retail price differences across U.S. Cities, *Journal of International Economics*, 2002, *56*, 21–53.

Papell, D., Searching for Stationarity: Purchasing Power Parity Under the Current Float, *Journal of International Economics*, 1997, *43*, 313–332.

Park, J and M. Shintani, Testing for a Unit Root against Transitional Autoregressive Models, *Manuscript*, 2005.

___ and P. Phillips, Nonlinear Regressions with Integrated Time Series, *Econometrica*, 2001, *69*, 117–161.

Phillips, P. and P. Perron, Testing for unit root in time series regression, *Biometrika*, 1988, *75* (2), 335–346.

Pippenger, M. and G. Goering, Additional results on the Power of Unit root and Cointegration Tests under Threshold Process, *Applied Economics Letters*, 2000, *7* (10), 641–644.

Rogoff, K., The purchasing power parity puzzle, *Journal of Economic Literature*, 1996, *34*, 647–68.

Sercu, P., R. Uppal, and C. Van Hulle, The exchange rate in the presence of transaction costs : implications for tests of purchasing power parity, *The Journal of Finance*, 1995, *50* (4), 1309–19.

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

Taylor, M., D. Peel, and L. Sarno, Non-linear mean reversion in real exchange rates: Towards a solution to the PPP puzzles, *International Economic Review*, 2001, *42*, 1015–1042.

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

Tong, H., *Non-linear Time Series*, Oxford University Press, 1990. Great-Britain.

Uppal, R., A general equilibrium model of international portfolio choice, *Journal of Finance*, 1993, *48*, 529–553.

van Dijk, D., T. Terasvirta, and P. Franses, Smooth transition autoregressive models : a survey of recent developments, *Econometric Reviews*, 2002, *21*, 1–47.

White, H., A Heteroskedasticity-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica*, 1980, *48* (4), 817–838.

Appendix A: Proofs

Proof of Proposition 1. Denote X the $T \times 2$ matrix with rows (y_{t-1}, y_{t-1}^2) and A the 1×2 vector $[0\ 1]$. Let $\hat{\beta}_1$ and $\hat{\sigma}^2$ be the OLS estimators of β_1 and σ^2 in (10). The Wald test is equal to

$$\begin{aligned} WL &= \frac{\hat{\beta}_1^2}{\hat{\sigma}^2 A \left(X'X\right)^{-1} A'} \\ &= \frac{\left\{\left(\sum y_{t-1}^2 \varepsilon_t\right)\left(\sum y_{t-1}^2\right) - \left(\sum y_{t-1}\varepsilon_t\right)\left(\sum y_{t-1}^3\right)\right\}^2}{\hat{\sigma}^2 \left(\sum y_{t-1}^2\right)\left\{\left(\sum y_{t-1}^2\right)\left(\sum y_{t-1}^4\right) - \left(\sum y_{t-1}^3\right)^2\right\}}. \end{aligned}$$

The result follows from the following limits (see Hamilton [1994]):

$$\begin{aligned} &\frac{\sum y_{t-1}^2}{T^2} \xrightarrow{L} \sigma^2 \int_0^1 B(r)^2\, dr, \\ &\frac{\sum y_{t-1}^3}{T^2\sqrt{T}} \xrightarrow{L} \sigma^3 \int_0^1 B(r)^3\, dr, \\ &\frac{\sum y_{t-1}^4}{T^3} \xrightarrow{L} \sigma^4 \int_0^1 B(r)^4\, dr, \\ &\frac{\sum y_{t-1}\varepsilon_t}{T} \xrightarrow{L} \frac{\sigma^2}{2}\left(B(1)^2 - 1\right), \\ &\frac{\sum y_{t-1}^2\varepsilon_t}{T\sqrt{T}} \xrightarrow{L} \frac{\sigma^3}{3} B(1)^3, \\ &\hat{\sigma}^2 \xrightarrow{P} \sigma^2. \end{aligned}$$

■

Proof of Proposition 2.

Denote $\theta = (a_1, ..., a_p, \mu_1, \mu_2, \rho_1, \rho_2)' \equiv (a_1, ..., a_p, \alpha')'$. We want to test $H_0 : \mu_1 = \mu_2 = \rho_1 = \rho_2 = 0$. We will denote $[1 + \exp(-\beta y_{t-d}/\lambda - \beta)]^{-1}$ as G_3 and $[1 + \exp(\beta y_{t-d}/\lambda + \beta)]^{-1}$ as G_1, in the following. The dependence on t is omitted for ease of notation. We are in a case close to Case 2 of Hamilton [1994, page 518]. Using his notation, Model (6) can be rewritten as

$$u_t = x_t'\theta + \varepsilon_t, \tag{19}$$

where $u_t = y_t - y_{t-1}$, $x_t = (u_{t-1}, ..., u_{t-p}, G_3 - G_1, G_2, y_{t-1}(G_1 + G_3), y_{t-1}G_2)'$. We have

$$\hat{\theta} - \theta_0 = \left[\sum x_t x_t'\right]^{-1} \sum x_t \varepsilon_t. \tag{20}$$

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

$$M_{11} = \begin{bmatrix} \sum u_{t-1}^2 & \sum u_{t-1}u_{t-2} & \dots & \sum u_{t-1}u_{t-p} \\ \sum u_{t-2}u_{t-1} & \sum u_{t-2}^2 & \dots & \sum u_{t-2}u_{t-p} \\ \vdots & \vdots & \dots & \vdots \\ \sum u_{t-p}u_{t-1} & \sum u_{t-p}u_{t-2} & \dots & \sum u_{t-p}^2 \end{bmatrix},$$

$$M_{21} = \begin{bmatrix} \sum (G_3 - G_1)\, u_{t-1} & \dots & \sum (G_3 - G_1)\, u_{t-p} \\ \sum G_2 u_{t-1} & \dots & \sum G_2 u_{t-p} \\ \sum y_{t-1}\,(G_1 + G_3)\, u_{t-1} & \dots & \sum y_{t-1}\,(G_1 + G_3)\, u_{t-p} \\ \sum y_{t-1} G_2 u_{t-1} & \dots & \sum y_{t-1} G_2 u_{t-p} \end{bmatrix},$$

and M_{22} is the symmetric matrix such that

$$M_{22} = \begin{bmatrix} \sum (G_3 - G_1)^2 & & & \\ \sum G_2\,(G_3 - G_1) & \sum G_2^2 & & \\ \sum y_{t-1}\left(G_3^2 - G_1^2\right) & \sum y_{t-1} G_2\,(G_1 + G_3) & \sum y_{t-1}^2\,(G_1 + G_3)^2 & \\ \sum y_{t-1} G_2\,(G_3 - G_1) & \sum y_{t-1} G_2^2 & \sum y_{t-1}^2 G_2\,(G_1 + G_3) & \sum y_{t-1}^2 G_2^2 \end{bmatrix},$$

$$\sum x_t \varepsilon_t = \begin{bmatrix} \sum u_{t-1}\varepsilon_t \\ \sum u_{t-2}\varepsilon_t \\ \vdots \\ \sum u_{t-p}\varepsilon_t \\ \sum (G_3 - G_1)\,\varepsilon_t \\ \sum G_2 \varepsilon_t \\ \sum y_{t-1}\,(G_1 + G_3)\,\varepsilon_t \\ \sum y_{t-1} G_2 \varepsilon_t \end{bmatrix}.$$

We use as the scaling matrix the following $(p+4)\times(p+4)$ diagonal matrix Γ_T with diagonal elements $\left(\sqrt{T}, \cdots, \sqrt{T}, T, T\right)$. Premultiplying (20) by Γ_T, we obtain:

$$\Gamma_T\left(\hat{\theta} - \theta_0\right) = \left[\Gamma_T^{-1}\left[\sum x_t x_t'\right]\Gamma_T^{-1}\right]^{-1}\left\{\Gamma_T^{-1}\left[\sum x_t \varepsilon_t\right]\right\}. \tag{21}$$

We have

$$T^{-1}\sum u_{t-i}u_{t-j} \xrightarrow{P} \gamma_{|i-j|}$$

by the Law of Large Numbers where $\gamma_{|i-j|} = E\left[u_{t-i}u_{t-j}\right]$. Under H_0, $y_t = \sum u_{t-j}$ is a random walk and $y_t/\sqrt{T}$ converges to $\delta B(r)$, $r = t/T$, where $B(.)$ is a standard Brownian motion on [0,1]. On the other hand $\frac{1}{\sqrt{T}}\sum_{t=1}^{Tr}\varepsilon_t$ converges to $\sigma B(r)$. Denote

$$\begin{aligned} \tilde{G}_1(r) &= \left[1 + \exp\left(\frac{\beta}{\tilde{\lambda}}\delta B(r) + \beta\right)\right]^{-1}, \\ \tilde{G}_3(r) &= \left[1 + \exp\left(-\frac{\beta}{\tilde{\lambda}}\delta B(r) - \beta\right)\right]^{-1}, \\ \tilde{G}_2(r) &= 1 - \tilde{G}_1(r) - \tilde{G}_3(r). \end{aligned}$$

The continuous transformations of $F\left(x,\beta,\tilde{\lambda}\right)=\left[1+\exp\left(-\frac{\beta}{\tilde{\lambda}}x+\beta\right)\right]^{-1}$ are themselves continuous in x and in $\left(\beta,\tilde{\lambda}\right)\in\Pi$ and therefore regular, we can apply Theorem 3.1. of Park and Phillips [2001]:

$$\begin{aligned}
&\frac{1}{T}\sum\left(G_3-G_1\right)^2\xrightarrow{P}\int_0^1\left(\widetilde{G}_3\left(r\right)-\widetilde{G}_1\left(r\right)\right)^2dr,\\
&\frac{1}{T}\sum G_2^2\xrightarrow{P}\int_0^1\tilde{G}_2\left(r\right)^2dr,\\
&\frac{1}{T}\sum\frac{y_{t-1}^2}{T}\left(G_3-G_1\right)^2\xrightarrow{P}\delta^2\int_0^1B^2(r)\left(\widetilde{G}_3\left(r\right)-\widetilde{G}_1\left(r\right)\right)^2dr,\\
&\frac{1}{T}\sum\frac{y_{t-1}^2}{T}G_2^2\xrightarrow{P}\delta^2\int_0^1B^2(r)\tilde{G}_2\left(r\right)^2dr,\\
&\frac{1}{T}\sum\frac{y_{t-1}}{\sqrt{T}}G_2^2\xrightarrow{P}\delta\int_0^1B(r)\tilde{G}_2\left(r\right)^2dr,
\end{aligned}$$

uniformly in $\pi=\left(\beta,\tilde{\lambda}\right)\in\Pi$. We have also

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum\left(G_3-G_1\right)\varepsilon_t\xrightarrow{L}\sigma\int_0^1\left(\widetilde{G}_3\left(r\right)-\widetilde{G}_1\left(r\right)\right)dB\left(r\right),\\
&\frac{1}{\sqrt{T}}\sum\frac{y_{t-1}}{\sqrt{T}}\left(G_1+G_3\right)\varepsilon_t\xrightarrow{L}\sigma\delta\int_0^1\left(\widetilde{G}_1\left(r\right)+\widetilde{G}_3\left(r\right)\right)B\left(r\right)dB\left(r\right),
\end{aligned}$$

similar results hold for the other terms of M_{22} and $\sum x_t\varepsilon_t$. These limits are also uniform in π, see Theorem 2 of Bec et al. [2008].

Using again Park and Phillips, it can be shown that

$$\begin{aligned}
&T^{-1}\sum G_1u_{t-j}\xrightarrow{P}0,\\
&T^{-1}\sum G_3u_{t-j}\xrightarrow{P}0,\\
&T^{-3/2}\sum y_{t-1}G_1u_{t-j}\xrightarrow{P}0,\\
&T^{-3/2}\sum y_{t-1}G_3u_{t-j}\xrightarrow{P}0.
\end{aligned}$$

Therefore, we have

$$\left[\Gamma_T^{-1}\left[\sum x_tx_t'\right]\Gamma_T^{-1}\right]\xrightarrow{L}\begin{bmatrix}V&0\\0&Q\end{bmatrix}$$

with

$$
\begin{aligned}
V &= \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{bmatrix}, \\
Q &= \Lambda \widetilde{Q \Lambda}, \\
D &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \delta & 0 \\ 0 & 0 & 0 & \delta \end{bmatrix}, \\
\widetilde{Q} &= \begin{bmatrix} Q_{11} & Q_{21}' \\ Q_{21} & Q_{22} \end{bmatrix}, \\
Q_{11} &= \begin{bmatrix} \int_0^1 \left(\widetilde{G}_3(r) - \widetilde{G}_1(r) \right)^2 dr & \int_0^1 \widetilde{G}_2(r) \left(\widetilde{G}_3(r) - \widetilde{G}_1(r) \right) dr \\ \int_0^1 \widetilde{G}_2(r) \left(\widetilde{G}_3(r) - \widetilde{G}_1(r) \right) dr & \int_0^1 \widetilde{G}_2(r)^2 dr \end{bmatrix}, \\
Q_{22} &= \begin{bmatrix} \int_0^1 \left(\widetilde{G}_1(r) + \widetilde{G}_3(r) \right)^2 B^2(r) dr & \int_0^1 \widetilde{G}_2(r) \left(\widetilde{G}_3(r) - \widetilde{G}_1(r) \right) B^2(r) dr \\ \int_0^1 \widetilde{G}_2(r) \left(\widetilde{G}_3(r) - \widetilde{G}_1(r) \right) B^2(r) dr & \int_0^1 \widetilde{G}_2(r)^2 B^2(r) dr \end{bmatrix}, \\
Q_{21} &= \begin{bmatrix} \int_0^1 \left(\widetilde{G}_3(r)^2 - \widetilde{G}_1(r)^2 \right) B(r) dr & \int_0^1 \widetilde{G}_2(r) \left(\widetilde{G}_1(r) + \widetilde{G}_3(r) \right) B(r) d(r) \\ \int_0^1 \widetilde{G}_2(r) \left(\widetilde{G}_3(r) - \widetilde{G}_1(r) \right) B(r) dr & \int_0^1 \widetilde{G}_2(r)^2 B(r) dr \end{bmatrix}.
\end{aligned}
$$

One can decompose $\Gamma_T^{-1} \left[\sum x_t \varepsilon_t \right]$ into two pieces. By Hamilton (page 520), the top part has the following asymptotic distribution

$$
\begin{bmatrix} \frac{1}{\sqrt{T}} \sum u_{t-1} \varepsilon_t \\ \frac{1}{\sqrt{T}} \sum u_{t-2} \varepsilon_t \\ \vdots \\ \frac{1}{\sqrt{T}} \sum u_{t-p} \varepsilon_t \end{bmatrix} \xrightarrow{L} h_1 \sim \mathcal{N}\left(0, \sigma^2 V\right).
$$

And the second part follows asymptotically

$$
\begin{bmatrix} \frac{1}{\sqrt{T}} \sum \left(G_3 - G_1 \right) \varepsilon_t \\ \frac{1}{\sqrt{T}} \sum G_2 \varepsilon_t \\ \frac{1}{T} \sum y_{t-1} \left(G_1 + G_3 \right) \varepsilon_t \\ \frac{1}{T} \sum y_{t-1} G_2 \varepsilon_t \end{bmatrix} \xrightarrow{L} h_2
$$

$$
h_2 = \sigma D \widetilde{h}_2.
$$

with

$$
\widetilde{h}_2 \sim \begin{bmatrix} \int_0^1 \left(\widetilde{G}_3(r) - \widetilde{G}_1(r) \right) dB(r) \\ \int_0^1 \widetilde{G}_2(r) \, dB(r) \\ \int_0^1 \left(\widetilde{G}_1(r) + \widetilde{G}_3(r) \right) B(r) \, dB(r) \\ \int_0^1 \widetilde{G}_2(r) \, B(r) dB(r) \end{bmatrix}.
$$

Substituting into (21), we get

$$\Gamma_T\left(\hat{\theta}-\theta_0\right) \xrightarrow{L} \left[\begin{array}{c} V^{-1}h_1 \\ Q^{-1}h_2 \end{array}\right].$$

H_0 can be rewritten as $A\theta = \alpha = 0$ where A is the appropriate selection matrix and $\alpha = (\mu_1, \mu_2, \rho_1, \rho_2)'$. The Wald test is given by

$$W_T = \left(A\hat{\theta}\right)' \left[\widehat{\sigma}^2 A \left[\sum x_t x_t'\right]^{-1} A'\right]^{-1} A\hat{\theta} = \widehat{\alpha}' \left[\widehat{\sigma}^2 \left[\sum \tilde{x}_t \tilde{x}_t'\right]^{-1}\right]^{-1} \widehat{\alpha},$$

where $\widehat{\alpha}$ is the estimator of α, $\tilde{x}_t$ are the regressors associated with α, and $\widehat{\sigma}^2$ is a consistent estimator of σ^2. Hence the asymptotic distribution of W_T is given by

$$h_2' Q^{-1} h_2/\sigma^2 = \sigma^2 \tilde{h}_2' D \left(D\widetilde{Q}D\right)^{-1} D\tilde{h}_2/\sigma^2 = \tilde{h}_2' \widetilde{Q}^{-1} \tilde{h}_2 \equiv D(k). \tag{22}$$

Note that this distribution is nuisance parameter free. By equivalence between the test statistics, the limiting distribution of LM_T and LR_T is the same as that of W_T.

Under the alternative of a stationary LSTAR model, $\widehat{\alpha}$ converges at the $\sqrt{T}-$rate of convergence to a pseudo true value α_a which is in general not equal to the true α (unless β and λ are the right values). However α_a will be different from 0 and the test statistics diverge. ■

Proof of Proposition 3.

Let us define $\ell_T = \lambda/(\sqrt{T}\delta)$. Note that $k = (\beta, \ell_T)$. Now define

$$\tilde{W}_T(k) = W_T\left(\beta, \lambda\sqrt{T}\delta\right).$$

Hence we have

$$\sup W = \sup_{k \in \mathcal{B} \times \left[\underline{\ell}_T, \bar{\ell}_T\right]} \tilde{W}_T(k).$$

We have shown before

1) $\tilde{W}_T(k) \xrightarrow{L} \mathcal{D}(k)$

2) $\left[\underline{\ell}_T, \bar{\ell}_T\right] \xrightarrow{L} \left[\underline{\ell}, \bar{\ell}\right]$

where the distribution of $\left(\underline{\ell}, \bar{\ell}\right)$ does not depend on any parameter.

We need to show

$$\sup_{k \in \mathcal{B} \times \left[\underline{\ell}_T, \bar{\ell}_T\right]} \tilde{W}_T(k) \xrightarrow{L} \sup_{k \in \mathcal{B} \times \left[\underline{\ell}, \bar{\ell}\right]} \mathcal{D}(k).$$

We can not apply the continuous mapping theorem directly because $\left[\underline{\ell}, \bar{\ell}\right]$ is random and its support is not bounded. This result can be established using a similar proof to that of Theorem 3 in Bec et al. [2004].

Then , it is easy to see that the distribution is nuisance parameter free because $\mathcal{D}(.)$ does not depend on unknown parameters and neither does $\mathcal{B} \times \left[\underline{\ell}, \overline{\ell}\right]$.■

Proof of Proposition 4.

We adopt the same reparametrization as in Kapetanios and al. (2003). Let $\Delta \mathbf{y}_{-j} = (\Delta y_{1-j}, ..., \Delta y_{T-j})'$. Define the $T \times (p-1)$ matrix $\mathbf{Z} = (\Delta \mathbf{y}_{-1}, ..., \Delta \mathbf{y}_{-p+1})$, the $T \times T-$idempotent matrix $\mathbf{M}_T = \mathbf{I}_T - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$, and $\varepsilon = (\varepsilon_1, ..., \varepsilon_T)'$. Let $\mathbf{y}_{-1}^j = \left(y_0^j, ..., y_{T-1}^j\right)'$, $j = 2, 3$. Model (15) can be rewritten as

$$\Delta \mathbf{y} = \sum_{j=1}^{p-1} a_j \Delta \mathbf{y}_{-j} + \beta_1 \mathbf{y}_{-1}^2 + \beta_2 \mathbf{y}_{-1}^3 + \varepsilon, \tag{23}$$

which itself is equivalent to

$$\mathbf{M}_T \Delta \mathbf{y} = \beta_1 \mathbf{M}_T \mathbf{y}_{-1}^2 + \beta_2 \mathbf{M}_T \mathbf{y}_{-1}^3 + \mathbf{M}_T \varepsilon.$$

Let $\theta = (\beta_1, \beta_2)'$. Let X be the $T \times 2-$matrix, $X = \left[\mathbf{y}_{-1}^2 \ \mathbf{y}_{-1}^3\right]$ and

$$\Gamma_T = \begin{bmatrix} T^{3/2} & 0 \\ 0 & T^2 \end{bmatrix}.$$

We have

$$\Gamma_T \hat{\theta} = \left(\Gamma_T^{-1}\left(X'\mathbf{M}_T X\right)\Gamma_T^{-1}\right)\Gamma_T^{-1} X'\mathbf{M}_T \varepsilon.$$

To establish the limiting distribution, we apply Theorem 3.1 of Park and Phillips [2001]:

$$\begin{aligned} \Gamma_T^{-1} X'\mathbf{M}_T \varepsilon &= \begin{pmatrix} \frac{\mathbf{y}_{-1}^{2\prime}\mathbf{M}_T\varepsilon}{T^{3/2}} \\ \frac{\mathbf{y}_{-1}^{3\prime}\mathbf{M}_T\varepsilon}{T^2} \end{pmatrix} = \begin{pmatrix} \frac{\mathbf{y}_{-1}^{2\prime}\varepsilon}{T^{3/2}} \\ \frac{\mathbf{y}_{-1}^{3\prime}\varepsilon}{T^2} \end{pmatrix} + o_p(1) \\ \xrightarrow{L} \sigma \begin{pmatrix} \delta^2 \int_0^1 W(r)^2 dW(r) \\ \delta^3 \int_0^1 W(r)^3 dW(r) \end{pmatrix} &\equiv \sigma \begin{bmatrix} \delta^2 & 0 \\ 0 & \delta^3 \end{bmatrix} h, \end{aligned}$$

where $\delta = \sigma / (1 - a_1 ... - a_{p-1})$.

$$\begin{aligned} \Gamma_T^{-1}\left(X'X\right)\Gamma_T^{-1} &= \begin{pmatrix} \frac{\mathbf{y}_{-1}^{2\prime}\mathbf{M}_T\mathbf{y}_{-1}^2}{T^3} & \frac{\mathbf{y}_{-1}^{2\prime}\mathbf{M}_T\mathbf{y}_{-1}^3}{T^{7/2}} \\ \frac{\mathbf{y}_{-1}^{3\prime}\mathbf{M}_T\mathbf{y}_{-1}^2}{T^{7/2}} & \frac{\mathbf{y}_{-1}^{3\prime}\mathbf{M}_T\mathbf{y}_{-1}^3}{T^4} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\mathbf{y}_{-1}^{2\prime}\mathbf{y}_{-1}^2}{T^3} & \frac{\mathbf{y}_{-1}^{2\prime}\mathbf{y}_{-1}^3}{T^{7/2}} \\ \frac{\mathbf{y}_{-1}^{3\prime}\mathbf{y}_{-1}^2}{T^{7/2}} & \frac{\mathbf{y}_{-1}^{3\prime}\mathbf{y}_{-1}^3}{T^4} \end{pmatrix} + o_p(1) \\ &\xrightarrow{L} \begin{pmatrix} \delta^4 \int_0^1 W(r)^4 dr & \delta^5 \int_0^1 W(r)^5 dr \\ \delta^5 \int_0^1 W(r)^5 dr & \delta^6 \int_0^1 W(r)^6 dr \end{pmatrix} \\ &\equiv \begin{bmatrix} \delta^2 & 0 \\ 0 & \delta^3 \end{bmatrix} Q \begin{bmatrix} \delta^2 & 0 \\ 0 & \delta^3 \end{bmatrix} \end{aligned}$$

with

$$\begin{aligned} h &= \begin{pmatrix} \int_0^1 W(r)^2\, dW(r) \\ \int_0^1 W(r)^3\, dW(r) \end{pmatrix}, \\ Q &= \begin{pmatrix} \int_0^1 W(r)^4\, dr & \int_0^1 W(r)^5\, dr \\ \int_0^1 W(r)^5\, dr & \int_0^1 W(r)^6\, dr \end{pmatrix}. \end{aligned}$$

Let $\hat{\varepsilon}$ be the least-squares estimate of ε in (23), the least-square estimate of σ^2 satisfies

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}' \mathbf{M}_T \hat{\varepsilon}}{T} = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{T} + o_p(1) \xrightarrow{P} \sigma^2.$$

The Wald test statistic is given by

$$\begin{aligned} F_2 &= \hat{\theta}' \left[\hat{\sigma}^2 \left(X' \mathbf{M}_T X \right)^{-1} \right]^{-1} \hat{\theta} \\ &= \frac{1}{\hat{\sigma}^2} \varepsilon' \mathbf{M}_T X \Gamma_T^{-1} \left[\Gamma_T^{-1} \left(X' \mathbf{M}_T X \right) \Gamma_T^{-1} \right]^{-1} \Gamma_T^{-1} X' \mathbf{M}_T \varepsilon \\ &\xrightarrow{L} h' Q^{-1} h. \end{aligned} \tag{24}$$

Note that $h'Q^{-1}h$ does not depend on σ or any other nuisance parameter.■

Appendix B: Expressions of the test statistics

Below, we detail the expression of the heteroskedasticity-robust test statistic LRh. For γ and λ fixed, Model (6) can be rewritten as

$$\Delta y_t = x_t' \theta + \varepsilon_t$$

using the notation of Appendix A. It can be estimated by OLS. Denote $\widetilde{sc}$ the $T \times (p+4)$ matrix with tth element $\widetilde{sc}_t = x_t \left(\Delta y_t - x_t' \theta \right)$, where θ has been replaced by the restricted OLS estimator, $\bar{\widetilde{sc}}$ the $(p+4) \times 1$ vector of the average of the $\widetilde{sc}_t$. The estimator of the covariance matrix is given by

$$\tilde{V}_T = \widetilde{sc}' \widetilde{sc} / T.$$

We call LRh the statistic

$$LRh = T \bar{\widetilde{sc}}' \tilde{V}_T^{-1} \bar{\widetilde{sc}}.$$

This statistic along with others was suggested by Newey and West [1987], it is also a LM type test statistic because when the model is exactly identified, heteroskedastic-robust versions of LR, LM and Wald tests are identical (Newey and West [1987], Section 3).

Appendix C: Estimated G_2 transition functions