THE ACR MODEL: A MULTIVARIATE DYNAMIC MIXTURE AUTOREGRESSION

Frédérique Bec
CREST-LMA, Timbre J360, 15 Boulevard Gabriel Peri, 92245 Malakoff CEDEX
THEMA, University of Cergy-Pontoise, France
bec@ensae.fr

Anders Rahbek
Department of Economics, University of Copenhagen,
Studiestraede 6, DK-1455 Copenhagen K, Denmark
anders.rahbek@econ.ku.dk

Neil Shephard
Oxford-Man Institute and Economics Department, University of Oxford,
Blue Boar Court, Alfred Road, Oxford OX1 4EH, United-Kingdom
neil.shephard@economics.ox.ac.uk

Abstract

In this paper we propose and analyse the Autoregressive Conditional Root (ACR) time series model. It is a multivariate dynamic mixture autoregression which allows for non-stationary epochs. It proves to be an appealing alternative to existing nonlinear models such as e.g. the threshold autoregressive or Markov switching classes of models, which are commonly used to describe nonlinear dynamics as implied by arbitrage in presence of transaction costs. Simple conditions on the parameters of the ACR process and its innovations, are shown to imply geometric ergodicity, stationarity and existence of moments. Furthermore, we establish consistency and asymptotic normality of the maximum likelihood estimators in the ACR model. An application to real exchange rate data illustrates the conclusions and analysis.

Keywords: Dynamic mixture vector autoregressive model; Autoregressive conditional root model; ACR; Regime switching; Stochastic unit root; Threshold autoregression.

1. Introduction

The purpose of this paper is to propose and analyse the Autoregressive Conditional Root (ACR) model. A main feature of this multivariate dynamic mixture vector autoregressive model is that it allows for regime switching between seemingly stationary and non-stationary epochs, where the switching is a function of the magnitude of lagged endogenous variables. This way it allows epochs of seeming non-stationarity, giving the impression that possible long-term relationships, such as e.g. the purchasing power parity, have broken down, before they endogenously collapse back toward their long term relationship.

The kind of dynamics considered here has been increasingly discussed over the past decade. For instance, the general equilibrium models developed by e.g. Dumas (1992), Sercu, Uppal, and Van Hulle (1995), or Berka (2004) imply such dynamics for the real exchange rate in presence of trading costs. The basic underlying idea is that international trade in goods occurs only when the gain expected from the home and foreign price differential is large enough to offset trading costs. Once trade takes place across countries, it induces changes in home and foreign prices which bring the real exchange rate back into the area where international arbitrage is not profitable anymore. The latter area is a non-arbitrage zone where the real exchange rate behaves like a non-stationary process. Nevertheless, since any price differential larger than the trading costs will activate corrective international trade, the real exchange rate process is globally stationary or stable. Another example of the relevance of such kind of non-linear behavior can be found in Anderson (1997), see also Balke and Fomby (1997) for further examples. In Anderson (1997) it is argued that transaction costs translate into two-regime dynamics for the interest rate spread, the switching between an adjusting and a non-adjusting area being defined as a function of the magnitude of the lagged spread value.

The empirical relevance of these theoretical implications has been explored by a large number of studies, using either discontinuous or smooth threshold autoregressive models. For instance, the empirical analyses by Michael, Nobay, and Peel (1997), Obstfeld and Taylor (1997), Kilian and Taylor (2003), Taylor, Peel, and Sarno (2001), or Bec, Ben Salem, and Carrasco (2004) provide support for multiple regime dynamics for real exchange rate data. Regarding the interest rate spread dynamics, similar results are obtained in e.g. Anderson (1997), Enders and Granger (1998) and Enders and Siklos (2001).

The proposed ACR model may be viewed as an appealing alternative to the threshold autoregressive (TAR) class of models retained in the papers cited above, and the Markov switching (MS) autoregressive class of models. As for the TAR and MS autoregressive models, it allows for switching between adjusting and non adjusting regimes, but does so in a different way. By contrast with the TAR models, the ACR model does not require a fixed threshold. And by contrast to MS models, the switching between regimes in the ACR model depends explicitly on lagged endogenous variables, in line with the economic theory outlined above.

The recent mixture autoregressive model (MAR) by Wong and Li (2000) and its extensions in Wong and Li (2001a), Wong and Li (2001b) and Fong, Li, and Wong (2007), as well as the dynamic switching Markov chain model of Gourieroux and Robert (2006) actually share some features similar to our proposed ACR model. Apart from the different dynamic interpretation of the models, our contribution to this literature is two-fold: First, unlike the above mentioned papers, the ACR model and the theory we provide for it are multivariate with any number of lags. Second, using geometric ergodicity results, we provide asymptotic theory for inference in this multivariate framework.

Based on a univariate simple version of the ACR model with one lag only, the ACR(1) model, the mentioned features will be emphasized in Section 2, where the ACR process is also compared with related nonlinear processes in the literature. Despite the epochs of seeming non-stationarity allowed by the ACR model, Section 3 establishes stationarity under simple regularity conditions for the proposed general multivariate ACR model with k lags. The regularity conditions ensure that the collapses regularize the periods of non-stationarity forcing the deviation from the long-term relationship to be globally stationary. Next, Section 4 provides asymptotic theory for the maximum likelihood (ML) estimators of the parameters of the multivariate model and show how the ML estimators can be obtained. In particular, we state conditions under which the ML estimators are consistent and asymptotically normally distributed. These results are illustrated in Section 5 by an empirical analysis of real exchange rate data. Section 6 discusses possible extensions. Finally Section 7 concludes the paper, while the Appendix contains the proofs of the theorems stated in the paper.

Some notation is used throughout: For vectors $a = (a_1, ..., a_k)' \in \mathbb{R}^k$, we use $\|a\|$ to denote some vector norm. Key examples, which we use, include the Euclidean and the L^1 norms, as given by $(a'a)^{1/2}$ and $|a_1| + ... + |a_k|$ respectively. With A a matrix, we use $\|A\|$ to denote the matrix norm as given by $\|A\|^2 = tr\{A'A\}$, and $\rho(A)$ to denote the largest, in absolute value, of the eigenvalues of A. We apply the notation, $dL(A, dA)$ for the differential of the matrix function $L(\cdot)$ with increment dA, see Appendix B.

2. The ACR-like dynamics

This section aims at conveying the flavour of the ACR model. To this end, a simple and univariate version is first presented, and then compared with the threshold autoregressive class of models. Finally, the specific features of the ACR model are further explored in the light of a number of other related models such as the Markov switching.

2.1. Univariate ACR(1) example

To fix ideas, consider initially the simplest version of a univariate autoregression of order one, the ACR(1) model, as given by

$$x_t = \begin{cases} \rho x_{t-1} + \varepsilon_t, & \text{if } s_t = 1 \\ \tilde{\rho} x_{t-1} + \varepsilon_t, & \text{if } s_t = 0. \end{cases} \tag{1}$$

for $t = 1, 2, \ldots, T$, with $\rho, \tilde{\rho}$ scalars, ε_t an i.i.d. $\mathrm{N}(0, \sigma^2)$ sequence and x_0 fixed. For simplicity in the exposition here, set $\tilde{\rho} = 1$ without loss of generality. Then with $\pi = \rho - 1$, the ACR(1) model can be reparametrized as an equilibrium correction model (ECM),

$$\Delta x_t = s_t \pi x_{t-1} + \varepsilon_t, \tag{2}$$

where Δ is the difference operator. The binary variable s_t is allowed to be unobserved, and the switching stochastic rather than deterministic. More precisely, the conditional probability, or the switching probability, that s_t takes the values one or zero is given by

$$\mathrm{P}(s_t = 1 | x_{t-1}, \varepsilon_t) = p(x_{t-1}), \tag{3}$$

where the notation $p(x_{t-1})$ emphasizes dependence solely on x_{t-1}. Vitally if the *regime* s_t is zero the process behaves locally like a random walk, while the case $s_t = 1$ implies it is locally like a stationary autoregression of order one provided $|\rho| = |\pi + 1| < 1$. The essential requirement for the conditional probability $p(x_{t-1})$, is that it tends to one as $|x_{t-1}|$ tends to infinity in addition to it being a function of x_{t-1}. No other condition is needed. A key example is given by the logistic type specification of $p(\cdot)$,

$$\lambda(x_{t-1}) = \log \{p(x_{t-1}) / (1 - p(x_{t-1}))\} = a + bf(x_{t-1}). \tag{4}$$

Here a and b are freely varying reals and $f(\cdot)$ some increasing function in $|x_{t-1}|$. In our empirical illustrations we use the concave function $f(x) = |x|^{1/2}$ and an ACR model with more than one lag.

As emphasized, the ACR(1) process in (1) can have epochs of seeming non-stationarity if $\tilde{\rho} = 1$, while at the same time be globally stable or stationary. More precisely for the case of the simple ACR(1) process with $p(\cdot)$ given by (4), an initial distribution of x_0 exists such that x_t in (1) is stationary and has finite moments of all orders provided that $|\rho| = |1+\pi| < 1$ and $b > 0$. Furthermore, as to estimation of the parameters, which in this case are ρ (or π), $\tilde{\rho}$, σ^2, a and b, the likelihood function can be computed via a prediction decomposition as discussed later. The thereby obtained ML estimators are shown to be consistent and asymptotically Gaussian distributed.

In their recent paper, written independently and concurrently from our paper, Gourieroux and Robert (2006) study in detail a dynamic switching Markov chain model which as mentioned is closely related to the ACR model. Their model may be viewed as the ACR(1) process in the constrained case where there is switching between white noise and a random walk (i.e. the special case of the above process when $\rho = 0$ and $\tilde{\rho} = 1$), and where the switching is governed by the sign of x_{t-1} rather than by its magnitude. In other words, unlike the ACR model, no convergence is assumed regarding $p(x_{t-1})$ as $|x_{t-1}| \to \infty$. Hence, our model is quite different in the dynamic interpretation and well-suited for the real exchange rate application we have in mind. By contrast, the wide ranging paper by Gourieroux and Robert (2006) is motivated by value-at-risk considerations in financial economics. Therefore, it allows a flexible distribution on ε_t and studies specifically the tail behaviour of the marginal distribution of x_t, the distribution of epochs of non-stationary behaviour and discusses stability of x_t in this case. From a methodological point of view, our analysis is complementary: it focuses on estimation and asymptotic inference for use in empirical work in the general, and also multivariate, version of the ACR model. We also note that Bec and Rahbek (2004) apply results from this paper to an analysis of nonlinear adjustments in error correction models.

2.2. ACR and threshold autoregressive models

Clearly, the dynamics of the regimes in the ACR model are determined entirely endogenously and so are similar to the threshold models in Tong (1990) and Enders and Granger (1998). However, now the threshold is allowed to be stochastic rather than only deterministic. In the general formulation of the ACR model, the switching probability $p(\cdot)$ is not bounded away from one, and does allow for deterministic switching by defining $p(\cdot)$ as,

$$p(x_{t-1}) = \begin{cases} 1, & \text{if } |x_{t-1}| > \tau > 0, \\ 0, & \text{otherwise.} \end{cases}$$

This is indeed a Tong (1990) self-exciting threshold autoregressive (SETAR) process, as it implies

$$x_t = \begin{cases} \rho x_{t-1} + \varepsilon_t, & \text{if } |x_{t-1}| > \tau, \\ \tilde{\rho} x_{t-1} + \varepsilon_t, & \text{otherwise.} \end{cases} \tag{5}$$

The implication is that we can view ACR models as softening the thresholds in autoregressive threshold models. This point will be amplified in the next subsection.

Thus a noticeable difference between the ACR and the SETAR models stands in the definition of the binary variable s_t. Contrary to the SETAR model, the ACR(1) model defined by equations (1) and (3) does not require the assumption of a fixed deterministic threshold. While maintained in SETAR models[1], this assumption might be too strong. Indeed, its relevance may be questioned when the threshold reflects e.g. trading costs over several decades as is often the case in empirical analyses. Another difference is that in TAR models (see (5)) there is no uncertainty on the regime conditional of the past values of the series.

Finally, the link between the ACR model and the Smooth Transition Autoregression (STAR) class of models initiated by Chan and Tong (1986) can easily be seen from the conditional expectation of equilibrium correction. For the ACR(1) process with $\tilde{\rho} = 1$, it is given by

$$\mathrm{E}(\Delta x_t | x_{t-1}) = \pi p(x_{t-1}) x_{t-1}.$$

Then, if the conditional probability is given by (4), the conditional expected change is

[1] Note however that the possibility of a Markovian regime for SETAR models has been mentioned by Tong and Lim (1980) and Tong (1983).

given by,

$$\mathrm{E}(\Delta x_t | x_{t-1}) = \pi \left(\frac{\exp\left(a + bf\left(x_{t-1}\right)\right)}{1 + \exp\left(a + bf\left(x_{t-1}\right)\right)} \right) x_{t-1} \tag{6}$$

If we recast this as,

$$\Delta x_t = \pi \left(\frac{\exp\left(a + bf\left(x_{t-1}\right)\right)}{1 + \exp\left(a + bf\left(x_{t-1}\right)\right)} \right) x_{t-1} + \eta_t,$$

where η_t is a martingale difference sequence, then this is a smooth transition autoregression (see Luukkonen, Saikkonen, and Teräsvirta (1988), Tong (1990) and Granger and Teräsvirta (1993, Section 4.2)). Hence the ACR model has many of the features of STAR models. Importantly however, STAR models do not have epochs of nonstationary behaviour – even with $\tilde{\rho} = 1$. Consequently, they do not belong to the class of processes considered in this paper.

2.3. A simulated example

The following simple example allows us to gain a better understanding of the behaviour of the ACR process. Figure 1(a) shows a sample path from the simplest ACR process given by (1), (3) and (4), together with the associated conditional probabilities $p(x_{t-1})$ given in Figure 1(b). Figure 1(c) reports the corresponding expected change in x_t, conditionally to x_{t-1}, as given by $p(x_{t-1})(\rho - 1)x_{t-1}$. The parameters values are $a = -9$, $b = 28$, $\rho = 0.75$, $\tilde{\rho} = 1$, and $\sigma = 0.009$.[2] This simulated process delivers realizations for $p(x_{t-1})$ such that the conditional probability that $s_t = 1$ never exceeds 0.5, which is enough for the x_t series to be stable. The second column of Figure 1 provides the same results based on a SETAR model simulated with $\rho = 0.75$, $\tilde{\rho} = 1$, $\sigma = 0.009$ and $p(x_{t-1}) = s_t = 1(|x_{t-1}| > 0.04)$. Here, the corresponding expected change in x_t, conditionally to x_{t-1} is given by $1(|x_{t-1}| > 0.04)(\rho - 1)x_{t-1}$. Comparing Figures 1(c) and 1(f) makes clear that ACR models may be viewed as softening the SETAR regime switching process.

2.4. Other related models

Apart from the already mentioned threshold class of models, the ACR model is also related to a number of well known models.

The prediction probability defined in equation (3) implies that the ACR(1) model appears similar to a Markov switching autoregressive model. In the Markov switching literature, s_t is usually employed to shift the intercept in a time series model, but it has also been used to make the variance change (Hamilton and Susmel (1994)) delivering a simple stochastic volatility process, and even to make the root of an autoregression move

[2]These values are inspired by the estimates of model ACR-III for French franc/Deutche mark real exchange rate data, reported in section 5. Similarly, the threshold parameter value for the SETAR comes from the estimate obtained by Bec, Ben Salem, and Carrasco (2004) using the same data.

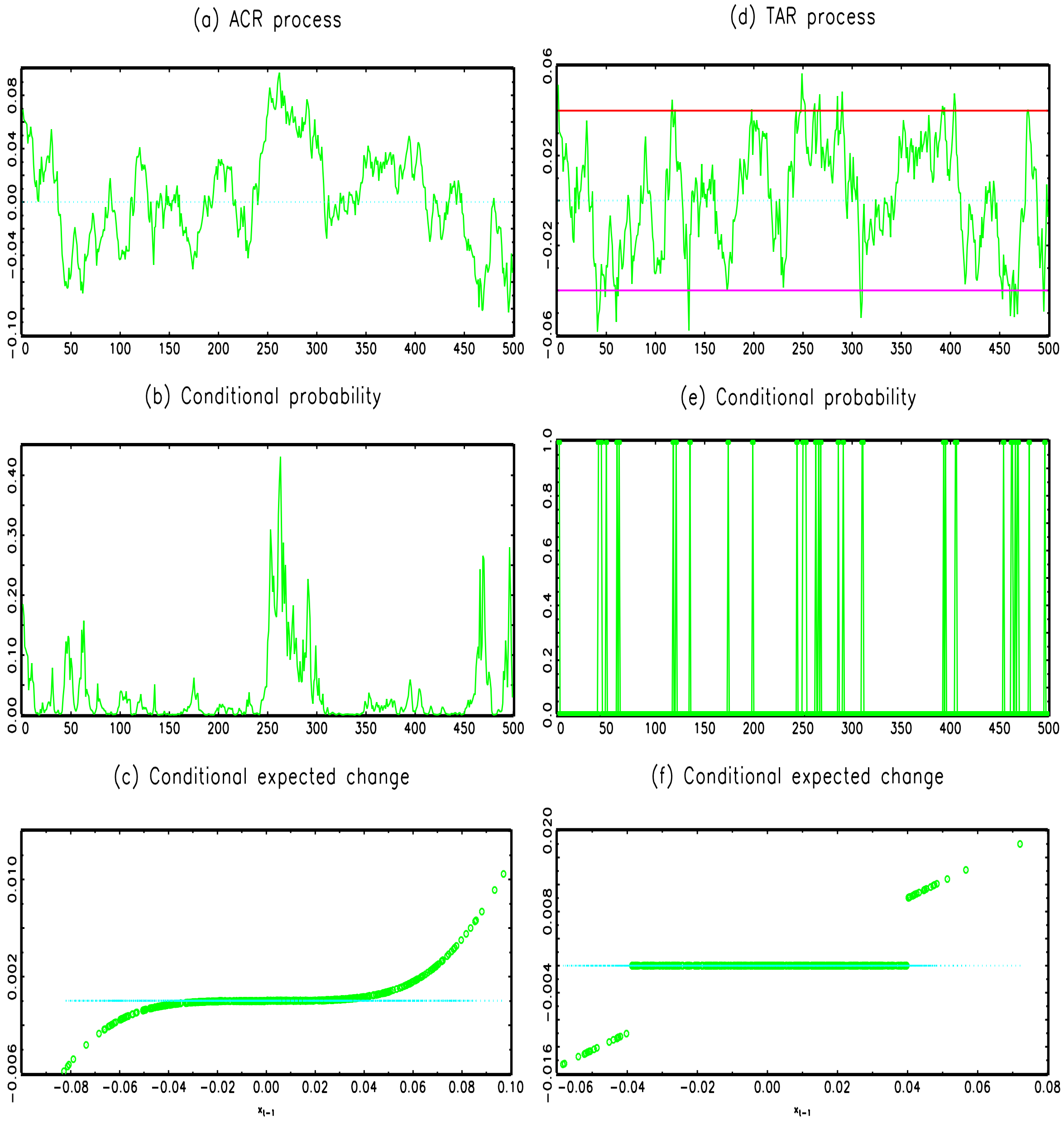

Figure 1: Simulated ACR (first column) and SETAR (second column) models.

between a unit root and a stationary root (Karlsen and Tjøstheim (1990)) or an explosive root (Hall, Psaradakis, and Sola (1999)). However, a fundamental difference between these models and the ACR is that the probability that s_t takes the values one or zero explicitly depends on x_{t-1} in the ACR. This in turn implies that a process defined by equations (1) and (3) is a Markov chain whereas this is not the case for a MS autoregression. This proves important for estimation as well as for the derivation of results for asymptotic inference in the ACR model. In fact, as mentioned, the ML estimators are straightforward to compute and our derived asymptotic theory allows for rigorous inference. This contrasts with the MS autoregressive models where estimation is based on filtering algorithms, and where a full asymptotic theory for inference still needs to be explored, even though much progress in that direction has been made in for example the recent paper by Douc, Moulines, and Ryden (2004).

The ACR model is also related to the stochastic root model introduced by Granger and Swanson (1997) and further studied by Leybourne, McCabe, and Mills (1996). Those papers use (1) but place an exogenous process on the root — allowing stationary, unit and explosive values. An example of this is where the log of the root is specified as being a Gaussian autoregression. These models have many virtues, but the likelihood function cannot usually be computed explicitly. Further, they do not have the clear cut epoch interpretation of the ACR process.

A related approach is the switching regression idea introduced into economics by Goldfeld and Quandt (1973). In our context this builds a model for the regime s_t in (1) which can depend upon explanatory variables and lagged values of the x_t process. A simple example of this is given by defining $\lambda(x_{t-1}) = a + bx_{t-1}$ in (4). This is outside our structure as it does not bound $\lambda(\cdot)$ away from minus infinity and so there is a possibility that the process will indeed be absorbed into the random walk state. The time series setup of $\lambda(x_{t-1}) = a + bx_{t-1}$ was also explicitly studied by Wong and Li (2001a), although its stochastic properties were not derived. Of course this can be generalised to allow $\lambda(x_{t-1})$ to depend upon many lags of x_t or other potentially helpful explanatory variables. Note however that this type of models focus on different dynamics for x positive or negative while in the ACR, the focus is on the magnitude of x, whatever its sign is. Hence these two classes of models have an entirely different interpretation. Note in this respect that the model in Wong and Li (2001a) and our ACR model are extensions of the mentioned static MAR in Wong and Li (2000) where $\lambda(x_{t-1}) = a$, that is, with constant switching probability.

Finally observe that the univariate stochastic permanent breaks model of Engle and Smith (1999) departs from ours as the process is non-stationary and allows for switching between permanent and transitory shocks. In terms of (2), their model can be mimicked by replacing $s_t \pi x_{t-1}$ by $s_t \varepsilon_{t-1}$, thereby introducing a suitable moving average term.

3. ACR(k)

In this section we introduce the general ACR(k) process. Conditions which ensure stationarity of the ACR process, despite epochs of non-stationarity, are discussed. These conditions imply also geometric ergodicity, which, as used in Section 4, again implies that the law of large numbers in Jensen and Rahbek (2007) holds for moment matrices in the asymptotic theory of the ML estimators.

3.1. The ACR(k) process

The ACR(k) process X_t is an immediate extension of the univariate ACR(1) process considered in (1). Switching between two autoregressions of order k, the m-dimensional ACR(k) process X_t is defined by the equation,

$$\begin{aligned} X_t &= s_t\left(A_1 X_{t-1} + ... + A_k X_{t-k}\right) + \left(1 - s_t\right)\left(B_1 X_{t-1} + ... + B_k X_{t-k}\right) + \varepsilon_t \\ &= s_t\left(A_1, ..., A_k\right)\mathbb{X}_{t-1} + (1 - s_t)(B_1, ..., B_k)\mathbb{X}_{t-1} + \varepsilon_t, \end{aligned} \tag{7}$$

for $t = 1, 2, ..., T$, where $\mathbb{X}_{t-1} \equiv (X'_{t-1}, ..., X'_{t-k})'$ and the initial value $\mathbb{X}_0$ is fixed. Furthermore, $(\varepsilon_t)_{t=1,2,...}$ is an i.i.d.$(0, \Omega)$ sequence with $\Omega > 0$, and with ε_t independent of the lagged variables $X_{t-1}, X_{t-2}, ...$. The autoregressive parameters A_i and B_i are $m \times m$ matrices.

Finally, the distribution of the switching variable, s_t, which can take values zero or one, is given by the prediction or switching probability,

$$P(s_t = 1 | \varepsilon_t, \mathbb{X}_{t-1}, \mathbb{X}_{t-2}, ...) = p(\mathbb{X}_{t-1}), \tag{8}$$

where $p(\cdot)$ is a function of $\mathbb{X}_{t-1}$. Note that in particular s_t and ε_t are independent conditional on $\mathbb{X}_{t-1}$, and that an equivalent way of defining s_t, is in terms of the indicator function $1\{\cdot\}$,

$$s_t = 1\left\{\nu_t \geq 1 - p\left(\mathbb{X}_{t-1}\right)\right\},$$

where $(\nu_t)_{t=1,2,...}$ is an i.i.d. sequence, independent of $(\varepsilon_t)_{t=1,2,...}$ and with ν_t uniformly distributed on $[0, 1]$.

We make here the following assumption for the functional form of the switching probability $p(\cdot)$:

Assumption 1. *With* $p : \mathbb{R}^{mk} \to [0, 1]$ *defined in (8), assume that*

$$p(\mathbb{X}) \to 1 \text{ as } \|\mathbb{X}\| \to \infty \tag{9}$$

where $\mathbb{X} \in \mathbb{R}^{mk}$.

As previously emphasized, our focus is on the logistic type specification of $p(\cdot)$ satisfying Assumption 1. The logistic specfication of $p(\cdot)$ is given by,

$$\lambda(\mathbb{X}_{t-1}) = \log\{p(\mathbb{X}_{t-1})/(1-p(\mathbb{X}_{t-1}))\} = a + bf(\mathbb{X}_{t-1}), \tag{10}$$

where a and b are scalar parameters, $b > 0$ and $f(\cdot)$ an increasing function in $\|\mathbb{X}_{t-1}\|$. Trivially, in this case,

$$(1-p(\mathbb{X})) = (1+\exp\lambda(\mathbb{X}))^{-1} \to 0$$

as $\|\mathbb{X}\| \to \infty$ provided $b > 0$. In other words, the probability is such that whatever state the process is in, there is always a non-negative probability that it will (re-)enter the state governed by the A_i parameters in (7). In addition, the structure is such that the further away the process gets from the regime governed by A_i, the more the probability of staying there tends to zero. This mimics closely the economic theory outlined in the introduction and discussed in the references given there.

Thus the generalisation differs from the univariate ACR(1) process in (1) in that we allow for a vector process, a richer lag structure, potentially non-Gaussian errors and additional flexibility in the dynamics by the introduction of the additional autoregressive regime parameters B_i. Specifically, the univariate ACR(1) example in (1) has $m = k = 1$, $A_1 = \rho$ and $B_1 = 1$. Here $A_1 = \rho$ governs the locally stationary regime, while $B_1 = 1$ governs the unit-root regime. In the multivariate extension consider as an example the case of $k = 2$. Choosing, say, $B_2 = I_m - B_1$ introduces m unit-roots in the $s_t = 0$ regime as desired and reflects the flexibility of the dynamics in the current parametrization. Below we demonstrate how the autogressive regime governed by $B_1, ..., B_k$ can have unit roots, even explosive roots, while X_t remains globally stationary provided the other regime corresponding to the A_i parameters has no unit or explosive-roots.

Note that it is straightforward to generalize the switching between two regimes, to switching between any fixed number of regimes. This is not done here in order to avoid unnecessary and complicated notation. To ensure stationarity of the m-dimensional ACR(k) process a further assumption is needed:

Assumption 2. *Assume that*

$$\left|I_m - \varrho A_1 - ... - \varrho^k A_k\right| = 0 \Rightarrow |\varrho| > 1,\ \ \varrho \in \mathbb{C}. \tag{11}$$

Assumption 2 states that the vector autoregressive process corresponding to the $s_t = 1$ regime satisfies the well-known condition for stationarity. Importantly, there are no restrictions on the parameters B_i of the other regime. Hence this regime, may have unit-roots and even explosive roots.

The final assumption addresses the distribution of the innovations ε_t:

Assumption 3. *With $(\varepsilon_t)_{t=1,...,T}$ m-dimensional i.i.d.$(0,\Omega)$, assume that ε_t has a continuous and strictly positive density with respect to the Lebesgue measure on $\mathbb{R}^m$ and that $\mathrm{E}\,\|\varepsilon_t\|^{2n}$ is finite for some $n \geq 1$.*

When discussing ML estimation and inference on the parameters in Section 4, Assumption 3 is particularly satisfied with ε_t Gaussian distributed, in which case also ε_t have finite moments for all $n \geq 1$. The requirement of continuity in Assumption 3 on the density could be replaced by the less strict assumption that for example the density is bounded on compact subsets of $\mathbb{R}^m$.

Theorem 1. *Consider the m-dimensional ACR(k) process X_t defined by (7) in terms of its lagged values in $\mathbb{X}_{t-1} \equiv \left(X'_{t-1}, ..., X'_{t-k}\right)'$ and the switching probability $p\left(\mathbb{X}_{t-1}\right)$ in (8).*

Under Assumptions 1, 2 and 3, the mk-dimensional process $(\mathbb{X}_t)_{t=1,2,...}$ is a geometrically ergodic process. In particular, $\mathbb{X}_0 = \left(X'_0, ..., X'_{-k+1}\right)'$ can be given an initial distribution such that $\mathbb{X}_t$, and hence also the ACR(k) process X_t, are stationary. Moreover, $E\left\|X_t\right\|^{2n} < \infty$.

The proof is given in the appendix.

For the ACR(1) case, the result parallels the result in Gourieroux and Robert (2006), Proposition 7.

As already noted, an important implication of $\mathbb{X}_t$ being geometrically ergodic is that the law of large numbers in Jensen and Rahbek (2007), and hence also a central limit theorem, apply for product moment matrices appearing in the discussion about estimation in the next section. Note furthermore, as emphasized and discussed by Carrasco and Chen (2002), that geometric ergodicity implies that the stationary solution $\mathbb{X}_t$, and hence also the stationary ACR(k) process X_t, will be β-mixing at an exponential decaying rate.

3.2. Switching and Assumption 1

Assumption 1 in Theorem 1 is important as it implies in particular that the switching probability depends on all variables in $\mathbb{X}'_{t-1} = (X'_{t-1}, ..., X'_{t-k})$. Based on existing econometric applications of models with general switching between autoregressions, it is also of interest to allow the switching probability to depend on only one of the lagged variables X_{t-1}, say. This clearly violates Assumption 1. In that case, to ensure stationarity of X_t, while still allowing unit root behaviour in the regime governed by the B_i parameters, the autoregressive parameters of the two regimes must be restricted such that $A_2 = B_2, ..., A_k = B_k$ corresponding to the lags of X_t which do not enter the switching probability. That is, the lag parameters of the variables not entering the switching probability should be identical across the two regimes.

More generally, introduce the known $mk \times q$ dimensional selection matrix η of full rank q, $q \leq mk$ and its orthogonal complement $\eta_\perp$, which is $mk \times (mk - q)$ dimensional of rank $(mk - q)$ and for which $\eta'\eta_\perp = 0$. Now let $\eta'\mathbb{X}_t$ be the q linear combinations of $\mathbb{X}_t$ which enter in the definition of $p\left(\cdot\right)$,

$$P(s_t = 1|\, \varepsilon_t, \mathbb{X}_{t-1}, \mathbb{X}_{t-2}, ...) = p(\eta'\mathbb{X}_{t-1}), \tag{12}$$

while the remaining $(mk-q)$ linear combinations, $\eta'_\perp \mathbb{X}_t$, do not enter. In terms of this notation, replace Assumption 1 by:

Assumption 4. *With η a known $mk \times q$ dimensional matrix of full rank q, $q \leq mk$, and with $p : \mathbb{R}^q \to [0,1]$ defined in (12), assume that:*

$$\begin{array}{ll} (i): & p(\eta'\mathbb{X}) \to 1 \text{ as } \|\eta'\mathbb{X}\| \to \infty \\ (ii): & (A_1 - B_1, ..., A_k - B_k)\, \eta_\perp = 0 \end{array}$$

where $\mathbb{X} \in \mathbb{R}^{mk}$ and A_i and B_i are the autoregressive parameters in (7).

In particular, with $\eta' = (I_m, 0, ..., 0)$ and $\eta'\mathbb{X}_{t-1} = X_{t-1}$, Assumption 4 *(i)* implies that the probability of switching tends to one as the norm of X_{t-1} gets large, independently of the further lagged values, while *(ii)* implies that the autoregressive parameters corresponding to $X_{t-2}, ..., X_{t-k}$ do not switch. With $q = mk$ all elements of $\mathbb{X}_{t-1}$ enter $p(\cdot)$ and all autoregressive parameters switch, while for $q = 0$ the ACR(k) process reduces to the well-known pure vector AR(k) process. The formulation is based on Bec and Rahbek (2004, Theorem 1) where nonstationary ACR(k) processes are studied. Analogous to Theorem 1 we have:

Theorem 2. *Consider the ACR(k) process X_t defined by (7) and the switching probability in (12). Then under Assumptions 2, 3 and 4, the conclusions in Theorem 1 hold.*

The proof is given in the appendix.

4. Likelihood based estimation

In this section we consider estimation and also asymptotic inference for the parameters of the ACR(k) model with Gaussian i.i.d. innovations. The ACR(k) model is defined by equations (7) and (8), or rather (7) and (12), where the switching probability may depend on a few of the lagged variables. Estimation is considered specifically for a logistic prediction probability function which is used in our application. The results have been formulated such that it should be possible to apply them also for other types of switching probability functions. We also discuss briefly how to test hypotheses on the parameters.

4.1. Estimation and inference

We consider here estimation in the general case with switching between AR(k) processes, where the switching probability is logistic and depends on q, $q \leq k$, linear known combinations as given by $\eta'\mathbb{X}_{t-1}$, see Theorem 2. Thereby the cases where switching depends on all lagged X_t in $\mathbb{X}_{t-1}$, or just one X_{t-j}, say, are all covered simultaneously. A convenient way to write the ACR(k) model is then,

$$X_t = s_t A\eta'\mathbb{X}_{t-1} + (1-s_t)B\eta'\mathbb{X}_{t-1} + C\eta'_\perp\mathbb{X}_{t-1} + \varepsilon_t \quad \text{for } t = 1, 2, ..., T \tag{13}$$

with $\mathbb{X}_{t-1} = \left(X'_{t-1}, ..., X'_{t-k}\right)'$, $\mathbb{X}_0$ is fixed and ε_t is an i.i.d.$\mathrm{N}_m(0, \Omega)$ sequence with $\Omega > 0$ and ε_t independent of $X_{t-1}, ..., \mathbb{X}_0$. Here A and B are $m \times q$ dimensional matrices of parameters which switch between the two regimes, while C is the $m \times (mk - q)$ dimensional parameter matrix with non-switching parameters.

The parametrization in terms of A, B and C is a simple reparametrization in terms of $(A_1, ..., A_k)$ and $(B_1, ..., B_k)$ in (7). Specifically, if switching is allowed to depend on all variables, that is $\eta'\mathbb{X}_{t-1} = \mathbb{X}_{t-1}$, then $C = 0$, $A = (A_1, ..., A_k)$ and $B = (B_1, ..., B_k)$. Likewise, $A = A_1$, $B = B_1$, and $C = (A_2, ..., A_k) = (B_2, ..., B_k)$ if $\eta'\mathbb{X}_{t-1} = X_{t-1}$, that is switching is allowed to depend on X_{t-1} alone. Formally, the reparametrization is given by,

$$(A_1, ..., A_k) = (A, C)(\eta, \eta_\perp)' \quad \text{and} \quad (B_1, ..., B_k) = (B, C)(\eta, \eta_\perp)'. \tag{14}$$

The logistic specfication in (10) of the switching probability in (12) is given by,

$$\lambda(p(\eta'\mathbb{X}_{t-1})) = a + bf(\eta'\mathbb{X}_{t-1}), \tag{15}$$

where a and b are scalar parameters, $b > 0$ and $f : \mathbb{R}^q \to \mathbb{R}$ is an increasing function in $\|\eta'\mathbb{X}_{t-1}\|$.

With $\theta \equiv \{A, B, C, a, b, \Omega\}$, the log-likelihood function conditional on $\mathbb{X}_0$ is given by

$$L_T(\theta) = \textstyle\sum_{t=1}^T \ell_t(\theta) = \sum_{t=1}^T \log\left(p_{At}\phi_{At} + p_{Bt}\phi_{Bt}\right), \quad \text{where} \tag{16}$$

$$p_{At} = (1 - p_{Bt}) \equiv p(\eta'\mathbb{X}_{t-1}), \tag{17}$$

and, omitting constants in the Gaussian density,

$$\phi_{Mt} = |\Omega|^{-1/2} \exp\left(-\tfrac{1}{2}\varepsilon_{Mt}\Omega^{-1}\varepsilon'_{Mt}\right), \quad \varepsilon_{Mt} = X_t - M\eta'\mathbb{X}_{t-1} - C\eta'_\perp\mathbb{X}_{t-1}, \tag{18}$$

for $M = A, B$. The likelihood function in (16) is numerically maximised to obtain the maximum likelihood estimator, $\hat{\theta}$, and the following result holds:

Theorem 3. *Consider the ACR model defined by equations (13) and (15). Then under Assumptions 2, 3 and 4, and if $A \neq B$, there exists with probability tending to one as T tends to infinity, a unique ML estimator $\hat{\theta} = \{\hat{A}, \hat{B}, \hat{C}, \hat{a}, \hat{b}, \hat{\Omega}\}$ which satisfies the score equation,*

$$dL_T(\theta, d\theta)|_{\theta = \hat{\theta}} = 0, \tag{19}$$

for all $d\theta$. Moreover, $\hat{\theta} \xrightarrow{\mathrm{P}} \theta$, and $\hat{\theta}_T$ is asymptotically Gaussian,

$$\sqrt{T}\left(\hat{\theta} - \theta\right) \xrightarrow{\mathrm{D}} N(0, \Sigma). \tag{20}$$

The proof of Theorem 3 is based on establishing Cramér type conditions from Jensen and Rahbek (2005, Lemma 1) and is given in Appendix B. When discussing an algorithm to obtain $\hat{\theta}$ below, the explicit form of the score equation (19) is discussed. A consistent estimator of Σ is given in Appendix B, equation (53).

It should be emphasized that the results show that the maximum likelihood estimators are asymptotically Gaussian even if the B_i regime allows unit and even explosive roots, provided that the other has only stationary roots. Thus we provide distribution theory for a model which allows epochs of stationarity and epochs without. As mentioned, we believe this is the first paper providing this kind of result. As discussed below, here ML estimators are straightforward to compute and our derived asymptotic theory allows for rigorous inference. This is illustrated in the empirical application in Section 5, where we also test for the presence of a unit root in the B_i regime by the LR test statistic which by Theorem 3 is asymptotically χ^2 distributed. Note that the imposed restrictions on the parameter space rule out the possibility of a unit root in *both* regimes as well as the possibility of absorption in either of the two regimes. Indeed usual asymptotic expansions in terms of score and information would then be problematic as discussed in general in Andrews and Ploberger (1994), Davies (1987) and Hansen (1996). Related issues have recently been analysed in the context of threshold autoregressive (TAR) models: Based on least squares estimation, Hansen (1997) discusses the theory of Wald type testing for the hypothesis that one of the regimes in a stationary model is an absorbing state. Testing for a unit root in multiple regimes is treated in Caner and Hansen (2001) and Bec, Ben Salem, and Carrasco (2004). Furthermore, cointegrated TAR models are discussed in Hansen and Seo (2002) and Bec and Rahbek (2004).

The results in Theorem 3 are derived specifically for the parametrization and functional choice of a logistic probability in (15). While our derivations do depend on the chosen logistic structure for the probabilities $p(\cdot)$ it is straightforward to modify the results to accommodate alternative specifications of $p(\cdot)$. Specifically, for transparency we have formulated all relevant quantities in terms of the derivative of $\lambda(p(\cdot))$ with respect to the parameters in θ in Lemmas 3, 5 and 6.

4.2. On optimisation of the likelihood

In order to carry out likelihood inference we have to numerically maximise the likelihood function, and an algorithm for this is discussed here. When presenting the algorithm we use notation as in least squares and logistic regression. Note that the algorithm could equivalently be derived as the EM algorithm[3], see e.g. Dempster, Laird, and Rubin (1977) and Ruud (1991).

Define first the weights

$$p^*_{At} = (1 - p^*_{Bt}) = \frac{p_{At}\phi_{At}}{p_{At}\phi_{At} + p_{Bt}\phi_{Bt}}, \tag{21}$$

in terms of the probabilities $p_{At} = (1 - p_{Bt}) = p\left(\eta'\mathbb{X}_{t-1}\right)$ in (17) and the Gaussian densities ϕ_{At} and ϕ_{Bt} in (18). Denote by $\hat{p}_{At}$, the probability p_{At} evaluated at the ML estimator $\hat{\theta}$, and likewise for $\hat{p}_{Bt}$, $\hat{\phi}_{Mt}$ and $\hat{p}^*_{Mt}$ with $M = A, B$.

Next, mimicking least squares regression notation, introduce product moment matrices in terms of the m-dimensional response variable X_t and the q-dimensional explanatory variables $\hat{p}^*_{At}\eta'\mathbb{X}_{t-1}$ and $\hat{p}^*_{Bt}\eta'\mathbb{X}_{t-1}$, as well as $\eta'_\perp\mathbb{X}_{t-1}$ which is $(mk - q)$-dimensional. Define,

$$S_{01} = \textstyle\sum_{t=1}^T \hat{p}^*_{At} X_t \mathbb{X}'_{t-1}\eta, \quad S_{02} = \sum_{t=1}^T \hat{p}^*_{Bt} X_t \mathbb{X}'_{t-1}\eta, \quad \text{and} \quad S_{03} = \sum_{t=1}^T X_t \mathbb{X}'_{t-1}\eta_\perp. \tag{22}$$

Define further the product moments as given by,

$$\begin{array}{ll} S_{11} = \sum_{t=1}^T \hat{p}^*_{At}\eta'\mathbb{X}_{t-1}\mathbb{X}'_{t-1}\eta\ , & S_{13} = S'_{31} = \sum_{t=1}^T \hat{p}^*_{At}\eta'\mathbb{X}_{t-1}\mathbb{X}'_{t-1}\eta_\perp, \\ S_{22} = \sum_{t=1}^T \hat{p}^*_{Bt}\eta'\mathbb{X}_{t-1}\mathbb{X}'_{t-1}\eta\ , & S_{23} = S'_{32} = \sum_{t=1}^T \hat{p}^*_{Bt}\eta'\mathbb{X}_{t-1}\mathbb{X}'_{t-1}\eta_\perp, \\ S_{33} = \sum_{t=1}^T \eta'_\perp\mathbb{X}_{t-1}\mathbb{X}'_{t-1}\eta_\perp & S_{12} = S'_{21} = 0. \end{array} \tag{23}$$

In terms of these, it follows by Lemma 3 in the Appendix, that the components of the ML estimator $\hat{\theta}$ satisfy

$$\left(\begin{array}{ccc} \hat{A} & \hat{B} & \hat{C} \end{array}\right) = \left(\begin{array}{ccc} S_{01} & S_{02} & S_{03} \end{array}\right) S^{-1}, \tag{24}$$

where S is the $(q + mk) \times (q + mk)$ dimensional matrix with entries $\left(S_{ij}\right)_{i,i=1,2,3}$ in (23). Likewise,

$$\hat{\Omega} = \frac{1}{T}\sum_{t=1}^T \left(\hat{p}^*_{At}\hat{\varepsilon}_{At}\hat{\varepsilon}'_{At} + \hat{p}^*_{Bt}\hat{\varepsilon}_{Bt}\hat{\varepsilon}'_{Bt}\right), \tag{25}$$

[3] Specifically, treating $(s_t)_{t=1,2,\ldots,T}$ as observed variables, the log-likelihood function for the EM algorithm is given by,

$$L_T^{EM}(\theta) = \sum_{t=1}^T \log(p_{At}^{s_t} p_{Bt}^{(1-s_t)}\phi_t), \quad \phi_t = |\Omega|^{-1/2}\exp\left(-\tfrac{1}{2}\varepsilon_t\Omega^{-1}\varepsilon'_t\right),$$

and $\varepsilon_t = X_t - (s_t A + (1 - s_t) B)\eta'\mathbb{X}_{t-1} - C\eta'_\perp\mathbb{X}_{t-1}$. By definition, $E\left(s_t|X_t, \mathbb{X}_{t-1}\right) = p_{At}$, and also

$$E\left(s_t|X_t, \mathbb{X}_{t-1}\right) = E\left(s_t^2|\mathbb{X}_T, X_{T-1}, \ldots, X_0\right) = p^*_{At}.$$

Using this, it follows that the updating recursions discussed below are identical to the M, or maximization, step in the EM algorithm.

where $\hat{\varepsilon}_{Mt}$ is ε_{Mt} defined in (18) evaluated at $\hat{\theta}$ for $M = A, B$. Finally, the estimators $\hat{a}$ and $\hat{b}$ for the logistic part satisfy the two equations:

$$\sum_{t=1}^{T} (\hat{p}^*_{At} - \hat{p}_{At}) \left(1 \quad f(\eta' \mathbb{X}_{t-1}) \right) = 0, \tag{26}$$

corresponding to a logistic regression for the 'observations' $\hat{p}^*_{At}$.

In other words, $\hat{\theta}$ satifies equations (24), (25) and (26) which are therefore not in closed form. However, an immediate recursive algorithm is the following. For $M = A, B$, let $\hat{p}^*_{Mt}$ denote p^*_{Mt} evaluated at the previously obtained estimator $\hat{\theta}^{(n-1)}$, say, then the updated estimator $\hat{\theta}^{(n)}$ is obtained by the least squares regression in (24) and (25), and the logistic nonlinear optimization in (26). Convergence is then defined by evaluating the log-likelihood function $L_T(\hat{\theta}^{(n)})$ until convergence.

The algorithm is implemented in the illustration in the next section.

5. An application to real exchange rate data

We illustrate the ACR model by applying it to different real exchange rates, and compare this to the MAR and linear AR models. Also we discuss application of the SETAR model.

The possible nonlinear nature of the dynamics of the real exchange rates has been increasingly discussed, both theoretically and empirically, since the beginning of the nineties. Until then, the so-called Purchasing Power Parity relationship constituted a cornerstone of most open macroeconomic theoretical models. This relationship comes from international arbitrage on goods market under frictionless and costless adjustment assumption. It states that once converted into the same currency, home and foreign general price levels should equalize, thanks to international trade in goods. More formally, the PPP relationship writes $ep^\star = p$, where e denotes the nominal exchange rate, i.e. the price of foreign currency in terms of home currency, p and $p^\star$ are national price levels measured in local currency. As a consequence of this non-arbitrage condition, the real exchange rate, defined as $ep^\star/p$, should be a linear stationary process. Nevertheless, this implication has been challenged by a lot of empirical work[4]. One possible explanation for those results could be the presence of trading costs, or more generally transaction costs including transportation costs, tariffs, information costs, etc...[5] Trading costs imply a nonlinear stationary process for the real exchange rate, as stressed in the theoretical models by Dumas (1992), Sercu, Uppal, and Van Hulle (1995) or Berka (2004), from which it follows that international arbitrage takes place if and only if the international price differential exceeds transaction costs. Hence, price differentials smaller than these costs are not being corrected by international trade. The simplest way to formalize this idea consists in defining two distinct

[4] See e.g. Rogoff (1996) for an overview of this topic.

[5] The crucial role of trading costs is emphasized in Obstfeld and Rogoff (2000).

areas for the real exchange rate process. One is the arbitrage area, concerning relatively large real exchange rate absolute values, where the real exchange rate adjusts towards its long-term equilibrium. The other area is a non arbitrage area, gathering real exchange rate observations which are relatively small in absolute value, where the real exchange rate behaves as if it were non-stationary. A threshold defining these two areas could then be interpreted as the trading costs level.

As underlined in the introduction, evidence of nonlinearity in the real exchange rate process has been found by many authors since a decade, using either discontinuous (SETAR) or smooth (ESTAR, LSTAR) threshold models. The ACR model provides an appealing alternative to model the real exchange rate process, since it does not impose a fixed threshold. So as to illustrate the relevance of this model, let us first consider the logarithm of French franc/Deutsche mark real exchange rate, x_t, defined as $log(e_t) + log(p_t^{DM}) - log(p_t^{FF})$, where e_t is the monthly average of the nominal exchange rate, and p_t^i is the consumption price index of country i. These post- Bretton Woods and pre- Euro data, spanning from 1973:09 to 1998:12, come from Datastream. The centered FF/DM real exchange rate is plotted in figure 2. In order to check the stationarity of

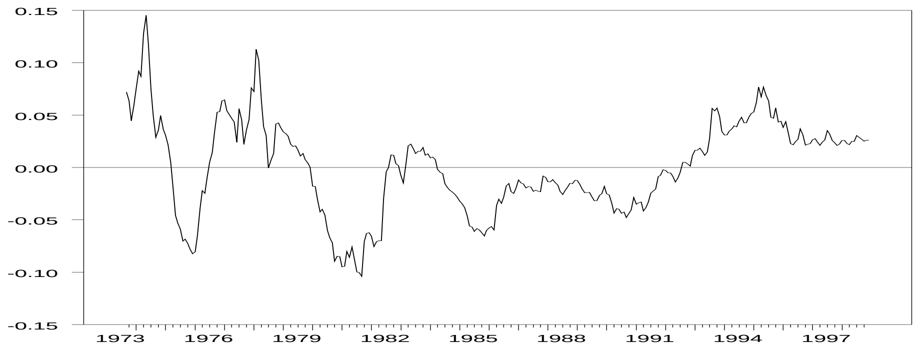

Figure 2: Centered FF/DM real exchange rate.

this series, we apply the $\mathcal{W}_B^{Sup}$ test statistics developed by Bec, Guay, and Guerre (2008). Based on simulation experiments, these authors show that the $\mathcal{W}_B^{Sup}$ unit root test has power against stationary ACR alternatives. For the FF/DM exchange rate data, this statistics reaches 31.81, which is well above the 5% critical value of 13.82 (see Table 1 in Bec, Guay, and Guerre (2008)) and hence allows rejection of the unit-root null hypothesis.

The ACR model considered below is defined by equations (13) and (15), where the function $f(\mathbb{X})$ retained here is: $f(\mathbb{X}_{t-1}) = \sqrt{|x_{t-1}| + |x_{t-2}| + \cdots + |x_{t-k}|}$. To allow for more straightforward inference regarding the existence of non-stationary epochs, we rewrite this ACR model in the following equivalent form:

$$\Delta x_t = s_t(\pi_A x_{t-1} + \sum_{i=1}^{k-1} \gamma_{Ai} \Delta x_{t-i}) + (1 - s_t)(\pi_B x_{t-1} + \sum_{i=1}^{k-1} \gamma_{Bi} \Delta x_{t-i}) + \varepsilon_t \qquad (27)$$

Within this equilibrium correction form of the ACR model, the test for epochs of non-stationarity simply amounts to testing the null $\pi_B = 0$.

The number of lags to include in the ACR model is chosen as the smallest one which succeeds in eliminating residuals autocorrelation according to the LM test. Computing residuals for the ACR model is not so straightforward. We have chosen to compute first the one-step ahead prediction distribution functions

$$\epsilon_t = p\left(\mathbb{X}_{t-1}\right)\Phi\left(\frac{\hat{\varepsilon}_{At}}{\hat{\sigma}_A}\right) + \left(1 - p\left(\mathbb{X}_{t-1}\right)\right)\Phi\left(\frac{\hat{\varepsilon}_{Bt}}{\hat{\sigma}_B}\right),$$

where $\hat{\varepsilon}_{it}$ denotes the maximum likelihood estimators of the ACR residuals in regime i $(i = A, B)$, for instance $\hat{\varepsilon}_{At} = \Delta x_t - \hat{\pi}_A x_{t-1} - \hat{\gamma}_{A1}\Delta x_{t-1} - \cdots - \hat{\gamma}_{Ak-1}\Delta x_{t-k+1}$, while Φ is the distribution function of the standard normal. These $\{\epsilon_t\}$ are approximately standard uniform and i.i.d. if the model is true, ignoring the effect of estimating the parameters. These have been frequently used to define residuals in non-linear time series econometric models (see, for example, Shephard (1994) and Kim, Shephard, and Chib (1998)). We then map these to our residuals for the ACR model by the inverse distribution function, $\varepsilon_t^{ACR} = \Phi^{-1}(\epsilon_t)$.

Using these residuals, the LM test of serial autocorrelation leads to retaining two lags in levels. Then, regarding the ACR model estimation, it is necessary to initialize the parameters in order to use the EM algorithm. All parameters are initialized from the corresponding linear model estimates, obtained by setting $\pi_A = \pi_B$ and $\gamma_{Ai} = \gamma_{Bi}$ $\forall i = 1, \cdots, k-1$ in equation (27). The last issue consists in initializing the logit function parameters. The EM algorithm outcome is in fact quite sensitive to these initial conditions. In order to avoid ending up in a local optimum, we highly recommend choosing them from the plot of the profile likelihood, *i.e.* the likelihood as a function of a and b. The ACR model log-likelihood is estimated using the EM Maximum Likelihood algorithm, considering (a, b) fixed, for a wide range of (a, b) values picked up in a grid consistent with the positiveness requirement for b[6]. For all EM ML estimations presented hereafter, the algorithm is stopped as soon as the log-likelihood increment between two steps is less than 10^{-7}. Moreover, so as to make b approximately scale-free, the logit function is reparameterized by dividing $(|x_{t-1}| + |x_{t-2}|)^{1/2}$ by its sample standard deviation. The plot obtained for the profile log-likelihood of this FF/DM real exchange rate model is in Figure 3, for a grid over $a \in [-60; 10]$, and over $b \in [0.1; 300]$.[7] As can be seen from the graph at the top of Figure 3, the shape of the profile log-likelihood suggests that initializing a from values greater than, say, -20.0, and b from values smaller than 50.0 should allow the EM

[6] The span of the grid should be adapted to the magnitude of the switching variable: for instance, the smaller it is, the larger the maximum of the grid over b.

[7] Extending the grid spans for a and b actually does not change the conclusions, but makes the graph less easy to read. Also note that for values of a greater than 10, the EM ML algorithm failed to estimate the ACR model for b values greater than 200, say, because the variance-covariance matrix of the estimated residuals becomes singular.

algorithm to reach the global maximum. This is confirmed by the graph zooming the profile log-likelihood for these ranges of a and b values, at the bottom of Figure 3.

Table 1 reports linear and ACR models results, where Δx_t is the left-hand side variable. The standard errors reported in parentheses were computed using equation (53) in Appendix. The results corresponding to the ACR model described above are given in column ACR-I. As can be seen from this column, the likelihood of the ACR model is

	Linear	ACR-I	ACR-II	ACR-III
x_{t-1}	**-0.039**	—	—	—
	(0.013)			
$s_t x_{t-1}$	—	**-0.234**	**-0.234**	**-0.258**
		(0.034)	(0.035)	(0.035)
$(1-s_t)x_{t-1}$	—	0.033	0.029	—
		(0.023)	(0.023)	
Δx_{t-1}	**0.304**	—	**0.319**	**0.315**
	(0.055)		(0.056)	(0.054)
$s_t \Delta x_{t-1}$	—	**0.407**	—	—
		(0.104)		
$(1-s_t)\Delta x_{t-1}$	—	**0.227**	—	—
		(0.074)		
a	—	**-6.47**	**-6.49**	**-9.41**
		(2.86)	(2.72)	(3.69)
b	—	**14.32**	**20.09**	**28.25**
		(6.69)	(8.89)	(12.14)
σ_ε	0.010	0.009	0.009	0.009
LM(AR 1-12) [p-value]	[0.14]	[0.18]	[0.18]	[0.18]
log-L	1237.60	1259.91	1258.52	1257.63

Standard errors in parentheses. LM test of no error autocorrelation.

Table 1: Linear and ACR model estimates

higher than the one of the linear model, as also reflected in the smaller standard error of estimated residuals. Moreover, this ACR model points to a sharp contrast between the outer and inner regime dynamics. The outer regime is characterized by a quite strong adjustment with a coefficient of -0.234 for $\hat{\pi}_A$. On the contrary, this model reveals a random walk behavior of the real exchange rate associated with small absolute value of the latter : $\hat{\pi}_B = 0.033$, but it is not significantly different from zero according to its standard error. This provides evidence in favor of the existence of non-stationary epochs, or in other words, the existence of a non-arbitrage area. Hence, the conclusion drawn from the ACR model confirms the findings of numerous empirical studies performed within

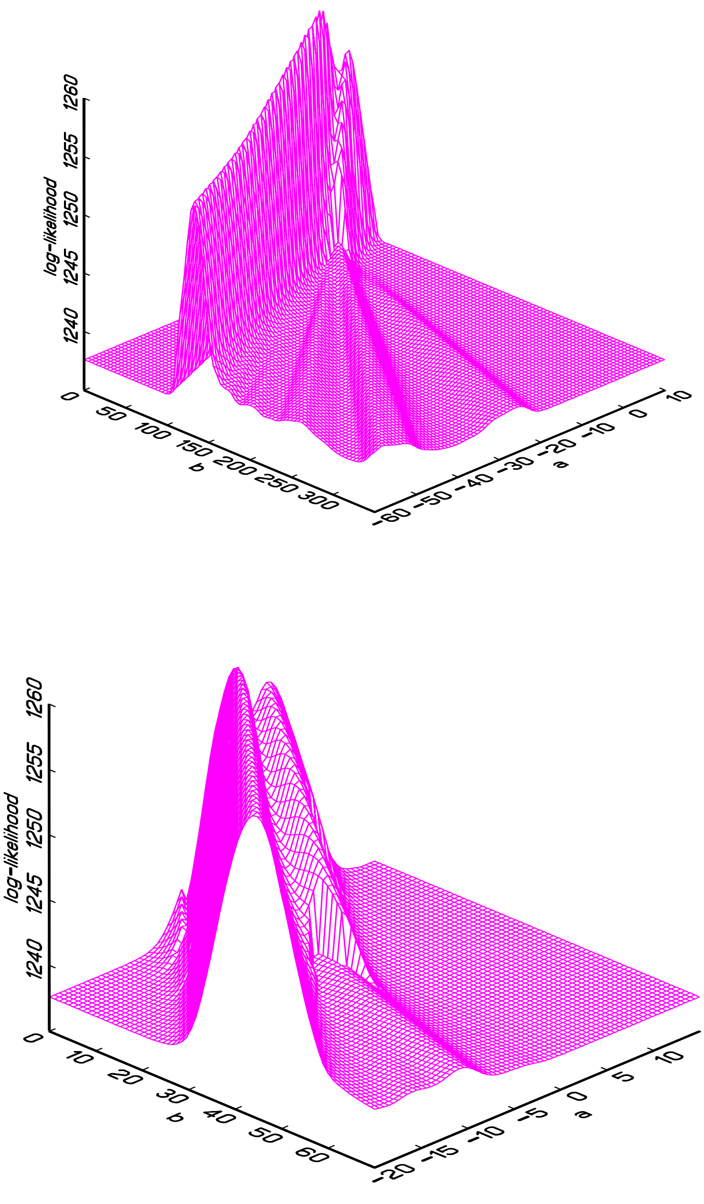

Figure 3: The profile log-likelihood as a function of a and b.

threshold autoregressive models.

So as to make our analysis more comparable with this empirical literature, Theorem 2 above is used to allow the switching probability to depend only on x_{t-1}. Accordingly, the logit function is defined in terms of $|x_{t-1}|^{1/2}$ only, and the parameters of Δx_{t-1} are restricted to be identical across regimes (see Assumption 4 above). The profile log-likelihood obtained in this case (not reported) being very similar to the one plotted in Figure 3, we initialized the EM algorithm with the same values as the ones retained for the ACR-I model's estimation. The resulting ML estimates are reported in column ACR-II of Table 1. They are quite close to their analogues from model ACR-I and clearly point to the same conclusion. Moreover, the decrease in the log-likelihood is very small. This may come from the fact that even though the point estimates of Δx_{t-1} parameters look rather different across regimes in model ACR-I (0.407 and 0.227 in the outer and inner regime respectively), they are not significantly different from each other according to their 5%-confidence intervals. The similarity of the results from models ACR-I and ACR-II also suggests that including x_{t-2} in the switching probability does not convey crucial information about the switches.[8]

Since the parameter associated to $(1 - s_t)x_{t-1}$ is still found not to be significantly different from zero, we also present the results of the estimation of the ACR-II model, imposing that this coefficient is zero (column ACR-III). The log-likelihood is not significantly decreased by this restriction: the LR test does not reject it with a statistic value of 1.78 to be compared to a $\chi^2(1)$ (Theorem 3). Consequently, we will now focus on the restricted ACR-III model. Again, the regime related to large real exchange rate in absolute value is characterized by a quick mean reverting dynamics, with an estimated autoregressive coefficient of -0.258. Overall, these results provide further support to the nonlinear model. When looking at the estimated conditional probability to be in the outer regime, Figure 4,

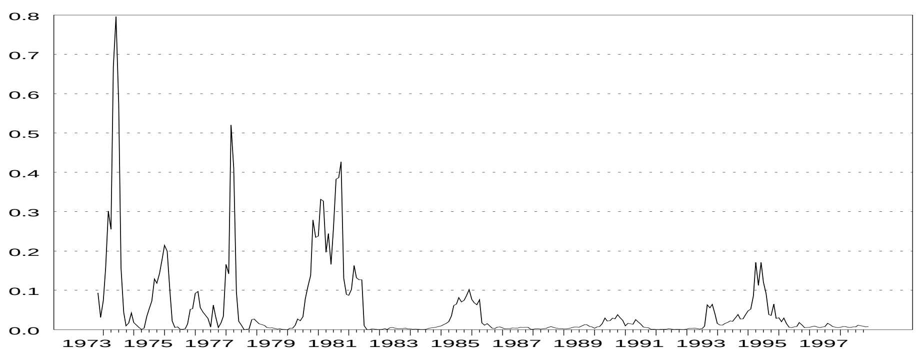

Figure 4: Estimated conditional probability (outer regime).

it appears that it peaks more often over the first half of the sample. The two largest peaks

[8] Testing this hypothesis is not straightforward since the ACR-I and ACR-II models are not nested; this will be addressed in future research.

observed in 1974 and 1978 reflect the sharp widening of the French-German inflation gap after the two oil shocks: the French authorities tried to accommodate the recession by easing monetary policy. The Bundesbank did the same, but to a lesser extent. The smaller peak in between corresponds to the year when France abandoned the European snake system, in 1976. The fourth epoch of increased switching probability also corresponds to a widening of the French and German inflation rates differential. Beyond the high inflation rates inherited from the oil price shocks, the new French government elected in 1981, leaded by Prime Minister Pierre Mauroy and President François Mitterrand, initiated a strong Keynesian policy in order to increase domestic demand. This policy resulted quite quickly in even more inflation and in a sharp weakening of the French franc against the German mark due to a noticeable worsening of the current account. This nominal exchange rate central parity was realigned twice between October 1981 and June 1982. It is worth noting that the conditional switching probability increase precedes the first franc devaluation by roughly one year. Over the second half of the sample, things look quieter than before. The reason for this is twofold. First, the Basle-Nyborg Agreement of September 1987 has probably stabilized the European Monetary System, basically by allowing the (limited) use of EMS credit facilities for intramarginal intervention[9]. Second, this corresponds to the French policy of "franc fort", or "strong franc". Actually, whereas other European countries like United-Kingdom or Italy said they would not defend their exchange rate against the DM when the Bundesbank maintained such high interest rates to finance the German unification, France chose the other way to deal with that issue: the so-called "competitive disinflation". Consequently, the French-German inflation gap decreased sharply, hence contributing to the relative stabilization of the real exchange rate. The last small peaks occur between 1993 and 1995, as a consequence of the speculative attacks against the French franc in July 1993 which caused the widening of the fluctuation bands from $\pm 2.25\%$ to $\pm 15\%$ in September 1993.

Finally, it is worth noting that the conditional switching probability peaks at around 0.80, and that only 1.3% of the sample is associated with a probability larger than 0.5 to switch to the outer regime. By fitting a SETAR to the same data, Bec, Ben Salem, and Carrasco (2004) have found a threshold at 0.0455. In Figure 5, ACR and SETAR estimated probabilities to lie in the outer regime are plotted. The SETAR probabilities closely match their non-zero ACR analogues. However, the SETAR classification looks quite crude compared to the ACR one.

As noted above, the British and Italian exchange rate policies were more independent from the German policy than the French one. As an additional check of the relevance of the ACR model, we now consider these two other real exchange rates series from the ACR model's versions which are close to those commonly used in the empirical literature,

[9]Before this agreement, the use of EMS credit facilities was allowed at the edge of the fluctuation bands only, which weakened the credibility of the EMS.

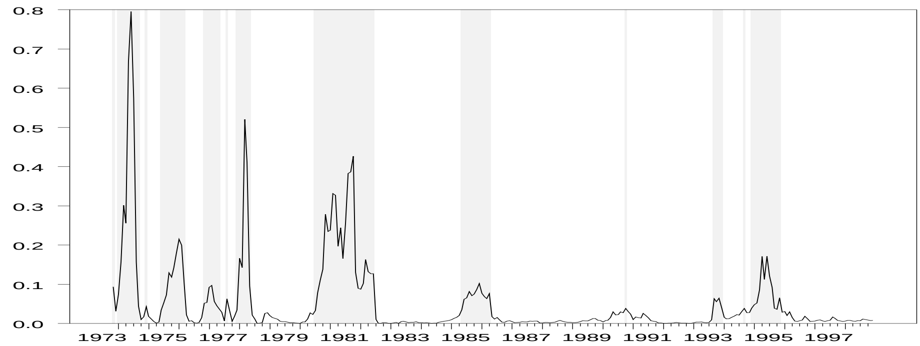

Figure 5: Estimated ACR conditional probability (solid line) and SETAR regimes (shaded area).

namely the ACR-II and ACR-III models. According to the $\mathcal{W}_B^{Sup}$ test statistics, the null of a unit-root is strongly rejected against the stationary alternative, with values of 20.47 and 36.84 for UK and Italy respectively. Table 2 summarizes the ACR-II and ACR-III models estimated for the British and Italian real exchange rates vis-à-vis the DM. The conclusions emerging from these results are quite similar to those obtained for the FF/DM real exchange rate. Actually, large deviations from PPP are associated with strong and significant adjustment coefficients (-0.11 and -0.28 for UK and Italy respectively). In both cases, the null hypothesis that small deviations are not corrected, i.e. $\hat{\pi}_B = 0$, cannot be rejected according to the LR test statistics. The latter equals 0.6 in the British case and 0.06 in the Italian case, and hence is smaller than the $\chi^2(1)$ 5% critical value. The only noticeable difference compared to the French case is the fact that the logit function parameters a and b are less accurately estimated here. Finally, it is possible to compare the ACR model with the Mixture AutoRegressive (MAR) nonlinear model developed by Wong and Li (2000). Indeed, imposing $b = 0$ in the ACR model amounts to assume that the transition probability is constant, in which case the ACR model reduces to the MAR model. Table 3 reports the likelihood ratio tests corresponding to this constraint in the ACR-II model for the three real exchange rates series. In all cases, the MAR restriction is strongly rejected by the data. Hence, this test provides additional empirical support to the ACR model.

6. Potential extensions

It is noted in Wong and Li (2000) that the conditional variance of the static mixture MAR process is non-constant. Likewise for the ACR process in (1) with $\tilde{\rho} = 1$, where straightforward computations give

$$V\left(x_t|x_{t-1}=x\right) = \sigma^2 + \left(p\left(x\right)\rho^2 x^2 + \left(1-p\left(x\right)\right)x^2\right) + \left(p\left(x\right)\rho x + \left(1-p\left(x\right)\right)x\right)^2.$$

	United-Kingdom		Italy	
	ACR-II	ACR-III	ACR-II	ACR-III
$s_t x_{t-1}$	**-0.113**	**-0.111**	**-0.278**	**-0.278**
	(0.031)	(0.030)	(0.036)	(0.036)
$(1-s_t)x_{t-1}$	-0.008	—	-0.002	—
	(0.010)		(0.008)	
Δx_{t-1}	**0.361**	**0.363**	**0.335**	**0.334**
	(0.053)	(0.053)	(0.051)	(0.050)
a	-31.66	-27.38	**-16.65**	**-16.40**
	(18.57)	(15.52)	(7.99)	(7.71)
b	63.97	55.78	31.95	31.40
	(37.39)	(31.74)	(17.91)	(17.33)
σ_ε	0.020	0.020	0.015	0.015
LM(AR 1-12) [p-value]	[0.27]	[0.27]	[0.97]	[0.97]
log-L	1021.29	1220.99	1106.55	1106.52

Standard errors in parentheses. LM test of no error autocorrelation.

Table 2: ACR model estimates for UK and Italy

	France	United-Kingdom	Italy
LR	10.04	9.12	4.16
p-values	0.001	0.002	0.041

Table 3: Comparing ACR and MAR models

Thus the processes indeed allow for non-constant conditional variances (while x_t has a constant unconditional variance). However, note that the form of the conditional variance induced by the ACR is rather restricted as for instance the conditional variance of Δx_t, $V\left(\Delta x_t | x_{t-1}\right)$, is constant.

ACR models could be developed for more sophisticated models of conditional variance[10]. As an example, consider first the traditional financial econometrics model with $x_t|\mathcal{F}_{t-1} \sim N(0, \sigma_t^2)$,where the conditional variance follows a GARCH type recursion (see for a review Bollerslev, Engle, and Nelson (1994)) such as

$$\sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2 + \alpha_2 \sigma_{t-1}^2 = \alpha_0 + \alpha_1 \left(x_{t-1}^2 - \sigma_{t-1}^2\right) + \rho \sigma_{t-1}^2$$

where $\rho = \alpha_1 + \alpha_2$. Here α_0, α_1 and α_2 are non-negative reals and, say, $\mathcal{F}_t = \sigma\left\{x_t, \sigma_t, ...\right\}$.

[10]See also Zhang, Li, and Yuen (2006) and Wong and Li (2001b) for similar extensions of related models.

Although this GARCH model is strictly stationary even if $\rho = 1$, this unit root implies that the process is not covariance stationary and the multistep forecasts of volatility will trend upwards. This is often regarded as being unsatisfactory, however empirically near unit root GARCH models are often estimated. See the discussion in, for example, Bollerslev and Engle (1993) and Engle and Lee (1999).

We can use the ACR structure to construct a GARCH model which behaves mostly like a unit-root process, but which is regularised by periods of stationary GARCH. This is simply achieved by writing $x_t|\mathcal{F}_{t-1}, s_t \sim N(0, \sigma_t^2)$ and changing the conditional variance into

$$\sigma_t^2 = \alpha_0 + \left\{(\alpha_1 + \alpha_2)^{s_t} - \alpha_2\right\} x_{t-1}^2 + \alpha_2 \sigma_{t-1}^2.$$

Now when $s_t = 0$ the GARCH process has a unit root, while when $s_t = 1$, the process is locally covariance stationary. The idea would be to allow, in the simplest case,

$$\lambda(\sigma_{t-1}^2) = \alpha + \gamma \sigma_{t-1}^2,$$

with γ being positive. This would mean that if the conditional variance becomes large the process has a chance to switch to a covariance stationary process, while when the conditional variance is low the process behaves like an integrated GARCH.

7. Conclusion

This paper has proposed a new type of time-series model, an autoregressive conditional root model, which endogenously switches between being stationary and non-stationary. The periods of stationarity regularise the overall properties of the model, implying that although the process has epochs of true non-stationarity, overall the process is both strictly and covariance stationary.

This model was motivated by our desire to reflect the possibility that long-term economic relationships between variables seem to sometimes breakdown over quite prolonged periods, but when the disequilibrium becomes very large there is a tendency for the relationship to reassert itself. This type of behaviour is quite often predicted by economic theory. Now we have a rather flexible time-series model which can test for this type of behaviour within the framework of some established econometric theory. Based on this, cointegration and nonlinear adjustment are discussed for the ACR model in Bec and Rahbek (2004)

8. Acknowledgments

We thank Christian Gourierioux for allowing us to read his unpublished paper on stochastic unit roots, and to Nour Meddahi for pointing out this work to us. Discussions on aspects of the ACR model with Niels R. Hansen, David Hendry, Søren Tolver Jensen, Søren Johansen, Hans Martin Krolzig and Rob Engle have been particularly helpful. Also thanks to comments from participants at ESEM 2002 in Venice and the Thiele II symposium in Copenhagen. Neil Shephard's research is supported by the UK's ESRC through the grant "Econometrics of trade-by-trade price dynamics," which is coded R00023839. Anders Rahbek and Frédérique Bec are grateful to the funding by Danish Social Sciences Research Council (2114-04-0001).

Appendix

The Appendix is divided into two parts: Appendix A is concerned with Markov chain theory used for the proof of geometric ergodicity in Section 2. Appendix B is about asymptotic inference in Markov chain models. This is mostly covered in Section 3 of the paper.

A. Proof of Theorems 1 and 2:

With the m-dimensional ACR(k) process X_t defined by (7) and the switching probability in (12), we show that $\mathbb{X}_t \equiv \left(X'_{t-1}, ..., X'_{t-k}\right)'$ is a Markov chain on $\mathbb{R}^{mk}$ which is geometrically ergodic, see Meyn and Tweedie (1993) and Tong (1990) for an introduction to Markov chain theory and geometric ergodicity. The proof falls in two parts: First it is verified in Lemma 1 that the Markov chain $\mathbb{X}_t$ is irreducible with respect to the Lebesgue measure μ on $\mathbb{R}^{mk}$, it is aperiodic and that compact sets $K \subseteq \mathbb{R}^{mk}$ are small. By Meyn and Tweedie (1993) these regularity conditions imply that if a drift criterion is shown to hold, then $\mathbb{X}_t$ is geometrically ergodic and has finite moments as defined by the drift function. Geometric ergodicity of $\mathbb{X}_t$ implies that $\mathbb{X}_0$ can be given an inital distribution such that $\mathbb{X}_t$, and hence also X_t, are stationary as claimed. This is established in Lemma 2. Thus Theorem 2 holds by Lemmas 1 and 2. Likewise, Theorem 1 holds by setting $\eta = I_{mk}$.

A similar strategy has been used in Bec and Rahbek (2004) and Saikkonen (2005) to establish stationarity of cointegrated relations in nonstationary nonlinear vector autoregressive processes. In particular, Bec and Rahbek (2004, proof of Theorem 1) use the results of Lemma 1 here:

Lemma 1. *Under Assumption 3 and Assumption 4 (i), $(\mathbb{X}_t)_{t=0,1,..}$ is a μ-irreducible, aperiodic Markov chain on $(\mathbb{R}^{mk}, \mathbb{B}^{mk})$, where $\mathbb{B}^{mk}$ is the Borel σ-algebra on $\mathbb{R}^{mk}$. Moreover, compact sets $K \subseteq \mathbb{R}^{mk}$ are small.*

Proof of Lemma 1: By definition of X_t and s_t, X_t conditional on $\mathbb{X}_{t-1}$ has density $f\left(X_t|\mathbb{X}_{t-1}\right)$ given by

$$f\left(X_t|\mathbb{X}_{t-1}\right) = p\left(\eta'\mathbb{X}_{t-1}\right) g\left(X_t - \mathbb{A}\mathbb{X}_{t-1}\right) + (1 - p\left(\eta'\mathbb{X}_{t-1}\right))g\left(X_t - \mathbb{B}\mathbb{X}_{t-1}\right), \tag{28}$$

where $\mathbb{A} = (A_1, ..., A_k)$, $\mathbb{B} = (B_1, ..., B_k)$ and $g(\cdot)$ is the density of ε_t which is well-defined by Assumption 3. Next, by straightforward factorization, $\mathbb{X}_{t+k}$ conditional on $\mathbb{X}_t$ has density,

$$h(\mathbb{X}_{t+k}|\mathbb{X}_t) = \prod_{i=1}^{k} f(X_{t+i}|\mathbb{X}_{t+i-1}). \tag{29}$$

That is, $\mathbb{X}_t$ has a well-defined k-step transition density, which, similar to Tjøstheim (1990), will be exploited in the next step.

Let $\mathrm{P}^n(D|x) = \mathrm{P}(\mathbb{X}_{t+n} \in D|\mathbb{X}_t = x)$ denote the n-step transition probabilities for the Markov chain $\mathbb{X}_t$, where $x \in \mathbb{R}^{mk}$ and $D \in \mathbb{B}^{mk}$. Then irreducibility with respect to μ follows by Meyn and Tweedie (1993, Proposition 4.2.1 (ii)), by noting that for all $x \in \mathbb{R}^{mk}$ and $D \in \mathbb{B}^{mk}$, with $\mu(A) > 0$,

$$\textstyle\sum_{n=1}^{\infty} \mathrm{P}^n(D|x) \geq \mathrm{P}^k(D|x) = \int_D h(y|x)\,dy > 0, \tag{30}$$

which holds by (29) and Assumption 3.

Likewise, with $K \subseteq \mathbb{R}^{mk}$ a compact set, and $(x, y) \in K \times K$, $h(y|x) \geq \delta$ for some $\delta > 0$ by Assumptions 4 *(i)* and 3. Then for any $x \in K$ and any $D \in \mathbb{B}^p$,

$$\mathrm{P}^k(D \mid x) \geq \mathrm{P}^k(D \cap K \mid x) = \textstyle\int_{D\cap K} h(y|x)dy \geq \delta\mu(D \cap K).$$

Hence for all $x \in K$, $\mathrm{P}^k(\cdot \mid x)$ is minorized by $\mu(\cdot \cap K)$ and the compact set K by definition is small, cf. Meyn and Tweedie (1993, p. 106).

Finally, an irreducible chain is periodic if it has period $d > 1$ and aperiodic if $d = 1$. If $\mathbb{X}_t$ has period $d > 1$, then by Meyn and Tweedie (1993, Theorem 5.4.4) there exist disjoint sets $D_0, D_1, ..., D_{d-1}$ in $\mathbb{B}^{mk}$ such that

$$\mathrm{P}^1(D_{i+1}|x) = 1 \text{ for } x \in D_i \text{ and } i = 0, 1, .., d-1 \pmod{d}$$

and furthermore, $\psi(\bigcup_{i=0}^{d-1} D_i)^c = 0$, where ψ is a maximal irreducibility measure. By Meyn and Tweedie (1993, Proposition 4.2.2 (ii)) μ is absolutely continuous with respect to ψ and therefore also $\mu(\bigcup_{i=1}^{d} D_{i-1})^c = 0$. For this to hold at least one of the sets D_1, say, must have $\mu(D_1) > 0$, which implies $\mathrm{P}^k(D_1|x) > 0$ for all x as in (30).

Iterating k times one gets for some j, the contradiction,

$$\mathrm{P}^k(D_1|x) = 0 \text{ with } x \in \textstyle\bigcup_{i\neq j} D_i.$$

Hence $\mathbb{X}_t$ has period $d = 1$ and is aperiodic. □

Lemma 2. *Under Assumptions 3 and 2, and Assumption 4 (i) and (ii), $(\mathbb{X}_t)_{t=0,1,..}$ satisfies a drift criterion such that $\mathbb{X}_t$ is geometrically ergodic and has finite $2n'th$ order moments.*

Proof of Lemma 2: By Lemma 1 $\mathbb{X}_t$ is a Markov chain for which we can apply the drift criterion as stated in *e.g.* Meyn and Tweedie (1993, Theorem 15.0.1 (iii)). As to choice of drift function $d(\mathbb{X}_t) \geq 1$ and calculation of $E(d(\mathbb{X}_t)|\mathbb{X}_{t-1} = x)$ the arguments mimic Bec and Rahbek (2004, proof of Theorem 1). Specifically, a drift function in Feigin and Tweedie (1985) implying finite second-order moments is given by

$$d(x) = 1 + x'Vx, V = \sum_{i=0}^{\infty} \mathbf{A}^{i'}\mathbf{A}^{i},$$

where $\mathbf{A}$ is the $mk \times mk$ dimensional matrix given by,

$$\mathbf{A} \equiv \begin{pmatrix} A_1 & A_2 & \cdots & A_k \\ I_m & 0 & \cdots & 0 \\ & \ddots & \ddots & \vdots \\ 0 & & I_m & 0 \end{pmatrix}.$$

Note that $\mathbf{A}'V\mathbf{A} = V - I$. Assumption 2 is equivalent to the assumption that $\rho(\mathbf{A}) < 1$, and therefore $d(\cdot)$ is well-defined. Defining $\mathbf{B}$ similarly in terms of the B_i coefficients, it follows that

$$E(d(\mathbb{X}_t)|\mathbb{X}_{t-1} = x) = \left(1 - \frac{x'x - tr(\Omega V) - (1 - p(\eta' x))h(x)}{d(x)}\right) d(x), \text{ with} \tag{31}$$
$$h(x) = x'(\mathbf{A} - \mathbf{B})'D(\mathbf{A} - \mathbf{B})x - 2x'\mathbf{A}'D(\mathbf{A} - \mathbf{B})x,$$

For some $\lambda > 1$, define the compact set $K \equiv \{ x |\ x'Vx \leq \lambda\}$. Note initially, that on its complement K^c it holds by definition that

$$d(x) = 1 + x'Vx \leq x'Vx \left(1 + \tfrac{1}{\lambda}\right) \leq 2x'Vx.$$

Hence for λ large enough,

$$\begin{aligned} \tfrac{x'x - tr(\Omega V) - (1-p(\eta' x))h(x)}{d(x)} &\geq \tfrac{x'x}{2x'Vx} - \frac{tr(\Omega V) + (1 - p(\eta' x))h(x)}{d(x)} \\ &\geq \tfrac{1}{2\rho(V)} - \frac{tr(\Omega V) + (1 - p(\eta' x))h(x)}{d(x)} \end{aligned}$$

Write $x = \eta\bar{\eta}'x + \eta_\perp \bar{\eta}_\perp' x$, where e.g. $\bar{\eta} = \eta(\eta'\eta)^{-1}$. Assumption 4 *(ii)* implies that $h(x)$ can be written as:

$$h(x) = (x'\eta)\bar{\eta}'(\mathbf{A} - \mathbf{B})'D(\mathbf{A} - \mathbf{B})\bar{\eta}(\eta' x) - 2x'\mathbf{A}'D(\mathbf{A} - \mathbf{B})\bar{\eta}(\eta' x) = h_1(x) - h_2(x)$$

Note that as $\|x\|^2 \to \infty$, either *(a)* $\|\eta' x\| \to \infty$ or *(b)* $\|\eta_\perp' x\| \to \infty$. In case of *(a)*, as $h_1(x)$ and $h_2(x)$ are $O(1)$, clearly $(1 - p(\eta' x))h(x)/d(x) \to 0$ by assumption on $p(\cdot)$. In case *(b)* $h_1(x) \to 0$ and also $h_2(x) \to 0$ since $h_2(x) = O(\|\eta' x\| \|x\|)$ and again $(1 - p(\eta' x))h(x)/d(x) \to 0$ holds since $1 - p(\cdot)$ is bounded. We can therefore conclude that

$$\frac{tr(\Omega V) + (1 - p(\eta' x))h(x)}{d(x)} \to 0$$

Summarizing, for λ large enough,

$$E(\, d(\mathbb{X}_t)|\, \mathbb{X}_{t-1} = x) \leq (1-\delta)d(x)$$

for $x \in K^c$ and some $\delta > 0$. On K, $E(\, d(\mathbb{X}_t)|\, \mathbb{X}_{t-1} = x)$ given by (31) which is continuous and hence bounded on the compact set. $\square$

B. Proof of Theorem 3:

Theorem 3 holds by establishing the regularity conditions (A.1), (A.2) and (A.3) in Jensen and Rahbek (2004, Lemma 1), which are classical Cramér type conditions addressing first, second and third order differentials of the log-likelihood function. These hold by Lemma 5 and Lemma 6 below.

We apply notation as in Magnus and Neudecker (1988) for derivatives of matrix functions: With k, l, m and n integers, the mapping G, $G : \mathbb{R}^{k\times l} \to \mathbb{R}^{m\times n}$, G is differentiable of order three in $X \in \Xi \subset \mathbb{R}^{k\times l}$ if

$$G(X + dX) = G(X) + dG(X, dX) + d^2G(X, dX, dX) + d^3G(X, dX, dX, dX) + o(\|dX\|^3)$$

as $\|dX\| \to 0$. Here, say, $dG(X, dX)$ is the differential of G at X with increment $dX \in \mathbb{R}^{k\times l}$, where $X + dX$ is in the interior of Ξ. The Jacobian, $\frac{\partial}{\partial vec(X)} vec\{G(X)\}$, and the differential are connected through the vec-operator by the identity,

$$vec\,\{dG(X, dX)\} = \left[\tfrac{\partial vec\{G(X)\}}{\partial\{vec(X)\}'}\right]' vec(dX). \tag{32}$$

Likewise for the second order derivative or Hessian, see Magnus and Neudecker (1988).

B.1. First and second order differentials

We start by listing the first and second order differentials, or score and observed information. In both case, the differentials have been stated such that it is possible to accommodate different choices of the logistic specification in (15).

Lemma 3. *With $p(\cdot)$ on the logistic form in (15), the first-order differential for the* log-*likelihood function in (16) is given by*

$$d\ell_t(\theta, d\theta) = (p^*_{At} - p_{At})\, d\lambda(\theta, d\theta) + \{p^*_{At} d\log\phi_{At}(\theta, d\theta) + p^*_{Bt} d\log\phi_{Bt}(\theta, d\theta)\}\,, \tag{33}$$

such that with p_{Mt}, p^*_{Mt}, ϕ_{Mt} and ε_{Mt} for $M = A, B$ defined in (17), (21) and (18)

$$d\ell_t\,(\theta, dA) = tr\{\Omega^{-1}p^*_{At}\varepsilon_{At}\mathbb{X}'_{t-1}\eta dA'\}, \quad d\ell_t\,(\theta, dB) = tr\{\Omega^{-1}p^*_{Bt}\varepsilon_{Bt}\mathbb{X}'_{t-1}\eta dB'\} \tag{34}$$

$$d\ell_t\,(\theta, dC) = tr\{\Omega^{-1}\,(p^*_{At}\varepsilon_{At}{+}p^*_{Bt}\varepsilon_{Bt})\,\mathbb{X}'_{t-1}\eta_\perp dC'\}, \tag{35}$$

$$d\ell_t\,(\theta, d(a, b)') = (p^*_{At} - p_{At})\, d(a, b)v_t, \qquad v_t = (1,\; f\,(\eta'\mathbb{X}_{t-1}))'. \tag{36}$$

$$d\ell_t\,(\theta, d\Omega) = \tfrac{1}{2}tr\{\Omega^{-1}d\Omega[\Omega^{-1}(p^*_{At}\varepsilon_{At}\varepsilon'_{At} + p^*_{Bt}\varepsilon_{Bt}\varepsilon'_{Bt}) - I_m]\}. \tag{37}$$

Proof of Lemma 3: The result follows by direct differentiation of the log likelihood function in (16) combined with the identity (21). □

Lemma 4. *With the notation from Lemma 3,*

$$d^2\ell_t(\theta, d\theta, d\theta) = p^*_{At}p^*_{Bt}\left\{d\lambda(\theta, d\theta) + d\log\phi_{At}(\theta, d\theta) - d\log\phi_{Bt}(\theta, d\theta)\right\}^2 + \left\{p^*_{At}d^2\log\phi_{At}(\theta, d\theta, d\theta) + p^*_{Bt}d^2\log\phi_{Bt}(\theta, d\theta, d\theta)\right\} - p_{At}p_{Bt}\left\{d\lambda(\theta, d\theta)\right\}^2. \tag{38}$$

The second-order differentials for the autoregressive parameters are given by,

$$\begin{aligned}
d^2\ell_t(\theta, dA, dA) &= p^*_{At}p^*_{Bt}(tr\{\Omega^{-1}\varepsilon_{At}\mathbb{X}'_{t-1}\eta dA'\})^2 - p^*_{At}tr\{\Omega^{-1}dA\eta'\mathbb{X}_{t-1}\mathbb{X}'_{t-1}\eta dA'\} \\
d^2\ell_t(\theta, dA, dB) &= -p^*_{At}p^*_{Bt}tr\{\Omega^{-1}\varepsilon_{At}\mathbb{X}'_{t-1}\eta dA'\}tr\{\Omega^{-1}\varepsilon_{Bt}\mathbb{X}'_{t-1}\eta dB'\} \\
d^2\ell_t(\theta, dA, dC) &= -p^*_{At}tr\{\Omega^{-1}dC\eta'_\perp\mathbb{X}_{t-1}\mathbb{X}'_{t-1}\eta dA'\} + \\
&\quad p^*_{At}p^*_{Bt}tr\{\Omega^{-1}\varepsilon_{At}\mathbb{X}'_{t-1}\eta dA'\}tr\{\Omega^{-1}[\varepsilon_{At} - \varepsilon_{Bt}]\mathbb{X}'_{t-1}\eta_\perp dC'\} \\
d^2\ell_t(\theta, dB, dB) &= p^*_{At}p^*_{Bt}(tr\{\Omega^{-1}\varepsilon_{Bt}\mathbb{X}'_{t-1}\eta dB'\})^2 - p^*_{Bt}tr\{\Omega^{-1}dB\eta'\mathbb{X}_{t-1}\mathbb{X}'_{t-1}\eta dB'\} \\
d^2\ell_t(\theta, dB, dC) &= -p^*_{Bt}tr\{\Omega^{-1}dC\eta'_\perp\mathbb{X}_{t-1}\mathbb{X}'_{t-1}\eta dB'\} - \\
&\quad p^*_{At}p^*_{Bt}tr\{\Omega^{-1}\varepsilon_{Bt}\mathbb{X}'_{t-1}\eta dB'\}tr\{\Omega^{-1}[\varepsilon_{At} - \varepsilon_{Bt}]\mathbb{X}'_{t-1}\eta_\perp dC'\} \\
d^2\ell_t(\theta, dC, dC) &= -tr\{\Omega^{-1}dC\eta'_\perp\mathbb{X}_{t-1}\mathbb{X}'_{t-1}\eta_\perp dC'\} + \\
&\quad p^*_{At}p^*_{Bt}(tr\{\Omega^{-1}[\varepsilon_{At} - \varepsilon_{Bt}]\mathbb{X}'_{t-1}\eta_\perp dC'\})^2
\end{aligned} \tag{39}$$

Next, for the logistic parameters,

$$\begin{aligned}
d^2\ell_t(\theta, d(a,b)', d(a,b)) &= (p^*_{At}p^*_{Bt} - p_{At}p_{Bt})\left\{d(a,b)v_t\right\}^2 \\
d^2\ell_t(\theta, d(a,b)', dA) &= p^*_{At}p^*_{Bt}tr\{\Omega^{-1}\varepsilon_{At}\mathbb{X}'_{t-1}\eta dA'\}d(a,b)v_t \\
d^2\ell_t(\theta, d(a,b)', dB) &= -p^*_{At}p^*_{Bt}tr\{\Omega^{-1}\varepsilon_{Bt}\mathbb{X}'_{t-1}\eta dB'\}d(a,b)v_t \\
d^2\ell_t(\theta, d(a,b)', dC) &= p^*_{At}p^*_{Bt}tr\{\Omega^{-1}[\varepsilon_{At} - \varepsilon_{Bt}]\mathbb{X}'_{t-1}\eta_\perp dC'\}d(a,b)v_t
\end{aligned} \tag{40}$$

And finally, for the covariance,

$$\begin{aligned}
d^2\ell_t(\theta, d\Omega, d\Omega) &= tr\{(\tfrac{1}{2}I_m - \Omega^{-1}[p^*_{At}\varepsilon_{At}\varepsilon'_{At} + p^*_{Bt}\varepsilon_{Bt}\varepsilon'_{Bt}])\Omega^{-1}d\Omega\Omega^{-1}d\Omega\} + \\
&\quad p^*_{At}p^*_{Bt}(tr\{\tfrac{1}{2}\Omega^{-1}d\Omega\Omega^{-1}[\varepsilon_{At}\varepsilon'_{At} - \varepsilon_{Bt}\varepsilon'_{Bt}]\})^2 \\
d^2\ell_t(\theta, d\Omega, dA) &= p^*_{At}p^*_{Bt}tr\{\tfrac{1}{2}\Omega^{-1}d\Omega\Omega^{-1}[\varepsilon_{At}\varepsilon'_{At} - \varepsilon_{Bt}\varepsilon'_{Bt}]\}tr\{\Omega^{-1}\varepsilon_{At}\mathbb{X}'_{t-1}\eta dA'\} - \\
&\quad p^*_{At}tr\{\Omega^{-1}d\Omega\Omega^{-1}\varepsilon_{At}\mathbb{X}'_{t-1}\eta dA'\} \\
d^2\ell_t(\theta, d\Omega, dB) &= -p^*_{At}p^*_{Bt}tr\{\tfrac{1}{2}\Omega^{-1}d\Omega\Omega^{-1}[\varepsilon_{At}\varepsilon'_{At} - \varepsilon_{Bt}\varepsilon'_{Bt}]\}tr\{\Omega^{-1}\varepsilon_{Bt}\mathbb{X}'_{t-1}\eta dB'\} - \\
&\quad p^*_{Bt}tr\{\Omega^{-1}d\Omega\Omega^{-1}\varepsilon_{Bt}\mathbb{X}'_{t-1}\eta dB'\} \\
d^2\ell_t(\theta, d\Omega, dC) &= -tr\{\Omega^{-1}d\Omega\Omega^{-1}(p^*_{At}\varepsilon_{At} + p^*_{Bt}\varepsilon_{Bt})\mathbb{X}'_{t-1}\eta_\perp dC'\} + \\
&\quad p^*_{At}p^*_{Bt}tr\{\tfrac{1}{2}\Omega^{-1}d\Omega\Omega^{-1}[\varepsilon_{At}\varepsilon'_{At} - \varepsilon_{Bt}\varepsilon'_{Bt}]\}tr\{\Omega^{-1}[\varepsilon_{At} - \varepsilon_{Bt}]\mathbb{X}'_{t-1}\eta_\perp dC'\} \\
d^2\ell_t(\theta, d\Omega, d(a,b)) &= p^*_{At}p^*_{Bt}tr\{\tfrac{1}{2}\Omega^{-1}d\Omega\Omega^{-1}[\varepsilon_{At}\varepsilon'_{At} - \varepsilon_{Bt}\varepsilon'_{Bt}]\}d(a,b)v_t
\end{aligned} \tag{41}$$

Proof of Lemma 4: Differentiation in (33) gives,

$$\begin{aligned}d^2\ell_t(\theta,d\theta,d\theta) &= p^*_{At}d^2\log\phi_{At}(\theta,d\theta,d\theta)+p^*_{Bt}d^2\log\phi_{Bt}(\theta,d\theta,d\theta)\\ &\quad+\left(p^*_{At}-p_{At}\right)d^2\lambda(\theta,d\theta,d\theta)+\left[dp^*_{At}(\theta,d\theta)-dp_{At}(\theta,d\theta)\right]d\lambda(\theta,d\theta)\\ &\quad+dp^*_{At}(\theta,d\theta)\left(d\phi_{At}(\theta,d\theta)-d\phi_{Bt}(\theta,d\theta)\right)\end{aligned}$$

which equals (38) using the identity $dp_{At}(\theta,d\theta)=p_{At}p_{Bt}d\lambda(\theta,d\theta)$, the identity

$$dp^*_{At}(\theta,d\theta)=p^*_{At}p^*_{Bt}(d\lambda(\theta,d\theta)+d\log\phi_{At}(\theta,d\theta)-d\log\phi_{Bt}(\theta,d\theta))$$

and that $d^2\lambda(\theta,d\theta,d\theta)=0$. The results in (39)-(41) hold by using the identities,

$$\begin{aligned}d\log\phi_{At}\left(\theta,dA\right) &= tr\{\Omega^{-1}\varepsilon_{At}\mathbb{X}'_{t-1}\eta dA'\},\\ d\log\phi_{At}\left(\theta,dC\right) &= tr\{\Omega^{-1}\varepsilon_{At}\mathbb{X}'_{t-1}\eta_\perp dC'\}\\ d^2\log\phi_{At}(\theta,dA,dA) &= tr\{\Omega^{-1}dA\eta'\mathbb{X}_{t-1}\mathbb{X}'_{t-1}\eta dA'\}\\ d^2\log\phi_{At}(\theta,dA,dC) &= tr\{\Omega^{-1}dA\eta'\mathbb{X}_{t-1}\mathbb{X}'_{t-1}\eta_\perp dC'\}\end{aligned}\tag{42}$$

and similarly for dB as well as standard matrix calculus results such as $d\log|\Omega| = tr\left\{\Omega^{-1}d\Omega\right\}$. □

B.2. Regularity conditions

Next, we verify that the information equality holds, positive definiteness of the information and that the third-order differential is bounded:

Lemma 5. *Under the Assumptions of Theorem 3 it holds that*

$$E(d\ell_t(\theta,d\theta))^2=-E(d^2\ell_t(\theta,d\theta,d\theta))>0.\tag{43}$$

Furthermore, for each θ there is a neighborhood $N(\theta)$ of θ such that

$$E\sup_{\tilde\theta\in N(\theta)}\left|d^3\ell_t\left(\theta,d\theta,d\theta,d\theta\right)\right|<\infty.$$

Proof of Lemma 5:

To see that e.g. $E\left(d\ell_t(\theta,dA)^2\right)=-E\left(d^2\ell_t(\theta,dA,dA)\right)$ for all $m\times q$ matrices dA we use the conditional independence of s_t and ε_t given $\mathbb{X}_{t-1}$: First note that

$$s_t\varepsilon_{At}=s_t(X_t-A\eta'\mathbb{X}_{t-1}-C\eta'_\perp\mathbb{X}_{t-1})=s_t\varepsilon_t$$

and using (21),

$$E\left(p^*_{At}\varepsilon_{At}\right|\mathbb{X}_{t-1}\left.\right)=E\left(E\left(s_t\right|X_t,\mathbb{X}_{t-1}\left.\right)\varepsilon_{At}\right|\mathbb{X}_{t-1}\left.\right)=E\left(s_t\varepsilon_t\right|\mathbb{X}_{t-1}\left.\right)=0.\tag{44}$$

By definition,

$$E\left(p^*_{At}\right|\mathbb{X}_{t-1}\left.\right)=E\left(s_t\right|\eta'\mathbb{X}_{t-1}\left.\right)=p_{At}.\tag{45}$$

Now,

$$\left(d\ell_t(\theta,dA)^2\right)+d^2\ell_t(\theta,dA,dA)=p^*_{At}[tr\left\{\Omega^{-1}\varepsilon_{At}\mathbb{X}'_{t-1}\eta dA'\right\}^2-tr\left\{\Omega^{-1}dA\eta'\mathbb{X}_{t-1}\mathbb{X}'_{t-1}\eta dA'\right\}],$$

and it holds that

$$E\left(\left(d\ell_t(\theta,dA)^2\right)+d^2\ell_t(\theta,dA,dA)\right|\mathbb{X}_{t-1}\right)=0.$$

as desired. Likewise for the remaining terms in (43) the results follow by repeated use of the additional identities

$$(1-s_t)\varepsilon_{Bt}=(1-s_t)\varepsilon_t,\quad E\left(\left(p^*_{Bt}\varepsilon_{Bt}\right|\mathbb{X}_{t-1}\right)=0 \tag{46}$$

$$E\left(p^*_{At}\varepsilon_{At}\varepsilon'_{At}+p^*_{Bt}\varepsilon_{Bt}\varepsilon'_{Bt}\right|\mathbb{X}_{t-1}\right)=s_t\Omega+(1-s_t)\Omega=\Omega \tag{47}$$

$$Cov\left(tr\left\{\varepsilon_t\varepsilon'_tP\right\},tr\left\{\varepsilon_t\varepsilon'_tQ\right\}\right)=2tr\left\{P\Omega Q\Omega\right\} \tag{48}$$

for P,Q symmetric $p\times p$ matrices. For instance, using (48) together with (47) and (46) it follows that

$$\begin{aligned}&E\left(d\ell_t(\theta,d\Omega)\right)^2+E\left(d^2\ell_t(\theta,d\Omega,d\Omega)\right)\\&=\tfrac{1}{4}E\left[tr\left\{\varepsilon_t\varepsilon'_t\Omega^{-1}d\Omega\Omega^{-1}\right\}\right]^2-\left[\tfrac{1}{4}tr\left\{\Omega^{-1}d\Omega\right\}^2+\tfrac{1}{2}tr\left\{\left[\Omega^{-1}d\Omega\right]^2\right\}\right]=0.\end{aligned}$$

Next, observe that $E(d\ell_t(\theta,d\theta))^2>0$ for all $d\theta$, is equivalent to linear independence of the first-order differentials or simply,

$$d\ell_t(\theta,dA)+d\ell_t(\theta,dB)+d\ell_t\left(\theta,dC\right)+d\ell_t(\theta,d(a,b)'+d\ell_t(\theta,d\Omega)=0$$

implies $dA=dB=dC=d(a,b)=d\Omega=0$. Note initially that by the definition of p^*_{At} in (21) then

$$p^*_{At}-p_{At}=p_{At}p_{Bt}(\phi_{At}-\phi_{Bt}) \tag{49}$$

Thus if $A=B$ then by (49) $p^*_{At}=p_{At}$ and the claimed implication does not hold. More precisely, conditioning on $\mathbb{X}_{t-1}$ and choosing $dA=\rho dB\neq 0$ for some real ρ, $d\ell_t(\theta,dA)+d\ell_t(\theta,dB)=0$. This is a consequence of the fact that conditional on $\mathbb{X}_{t-1}$, and with a and b known, the considerations simplify to the well-known for mixed normal models, see e.g. Titterington, Smith, and Makov (1985). Therefore we focus on the non-singularity of the derivative with respect to $(a,b)'$,

$$\begin{aligned}d\ell_t(\theta,d(a,b)'&=d\ell_t\left(\theta,d(a,b)'\right)=\left(p^*_{At}-p_{At}\right)d(a,b)v_t\\&=\left(p^*_{At}-p_{At}\right)\left(da+f\left(\eta'\mathbb{X}_{t-1}\right)db\right)\end{aligned}$$

By (49) and Assumption 1, $(p^*_{At}-p_{At})\neq 0$ almost surely (as $b>0$). Next, the proof of geometric ergodicity of $\mathbb{X}_t$ implies that the Markov chain has the Lebesgue measure as irreducibility measure. This again implies, by the Lebesgue decomposition, that the invariant measure has a component which has a strictly positive density w.r.t. Lebesgue measure and hence that, $\Pr(f(\eta'\mathbb{X}_{t-1}) \neq \text{constant}) > 0$ and therefore $d\ell_t(\theta, d(a,b)' \neq 0$ almost surely.

For the third-order differential, use Lemma 4 and note that with

$$w_t^M = d\log\phi_t^M = tr\left\{\Omega^{-1}\left[X_t\mathbb{X}_{t-1}'\eta - M\eta'\mathbb{X}_{t-1}\mathbb{X}_{t-1}'\eta\right]dM\right\} \tag{50}$$

for $M = A, B$, cf. (42), then

$$\left|w_t^M\right| \leq \kappa_1 \left\|\mathbb{X}_{t-1}X_t'\right\| + \kappa_2 \left\|\mathbb{X}_{t-1}\mathbb{X}_{t-1}'\right\|$$

for $\tilde{\theta} \in N(\theta)$ and some constants κ_i, i=1,2. Consider first the direction of A,

$$\left|d^3\ell_t(\theta, dA, dA, dA)\right| = \left|(p_{At}^* p_{Bt}^* w_t^A tr\left\{\Omega^{-1}dA\eta'\mathbb{X}_{t-1}\mathbb{X}_{t-1}'\eta dA'\right\} + (1-2p_{At}^*)p_{At}^*\left(w_t^A\right)^3\right|$$
$$\leq \tilde{\kappa}_1\left|w_t^A\right|\left\|\mathbb{X}_t\right\|^2 + \tilde{\kappa}_2\left|w_t^A\right|^3$$

for some constants $\tilde{\kappa}_i$, i=1,2. Hence $E_\theta \sup_{\tilde{\theta}\in N(\theta)}\left|d^3\ell_t(\theta, dA, dA, dA)\right|$ is finite by existence of second-order moments of $\mathbb{X}_t$. Apart from tedious calculus, similar results hold for the remaining third-order differentials. □

Lemma 6. *Under the assumptions of Theorem 3, then as $T \to \infty$:*
Provided $\phi(\cdot,\cdot)$ is measurable and $E\left\|\phi(X_t, \mathbb{X}_{t-1})\right\| < \infty$, then for each θ

$$\tfrac{1}{T}\Sigma_{t=1}^T d^2\ell_t(\theta, d\theta, d\theta) \xrightarrow{\text{P}} E(d^2\ell_t(\theta, d\theta, d\theta)). \tag{51}$$

Furthermore,

$$\tfrac{1}{T}\Sigma_{t=1}^T d\ell_t(\theta, d\theta) \xrightarrow{\text{D}} N\left(0, E\left[d\ell_t(\theta, d\theta)\right]^2\right), \tag{52}$$

where $E\left[d\ell_t(\theta, d\theta)\right]^2$ satisfies (43).

Proof of Lemma 6: By the law of large numbers in Jensen and Rahbek (2007),

$$\tfrac{1}{T}\Sigma_{t=1}^T\phi(X_t, \mathbb{X}_{t-1}) \xrightarrow{\text{P}} E(\phi(X_t, \mathbb{X}_{t-1})),$$

for all $\phi(\cdot,\cdot)$ measurable and $E\left\|\phi(X_t, \mathbb{X}_{t-1})\right\| < \infty$ as $\mathbb{X}_t$ is geometrically ergodic. Using the expressions in Lemma 4 for the second-order differential, the convergence in (51) holds as all moments are finite. Next, note that $d\ell_t(\theta, d\theta)$ is a Martingale difference sequence with respect to $\mathcal{F}_t = \sigma(X_t, X_{t-1}, ...)$. Specifically,

$$E(\,d\ell_t(\theta, d\theta)|\,\mathcal{F}_{t-1}) = E(\,d\ell_t(\theta, d\theta)|\,\mathbb{X}_{t-1}) = 0$$

using the expression for the differentials in Lemma 3 together with the identities (44), (45), (46) and (47) applied in the proof of Lemma 4. Again the established geometric ergodicity and existence of moments imply that

$$\tfrac{1}{T}\sum_{t=1}^T E(\left[d\ell_t(\theta, d\theta)\right]^2\big|\,\mathbb{X}_{t-1})$$

converges in probability by the law of large numbers. Furthermore, the Lindeberg condition in Brown (1971) applies and the claimed asymptotic normality of the first-order differential follows. □

B.3. Information

We end this section by stating the observed information, that is minus the second-order derivative of the log-likelihood function, which is used in the application as a consistent estimator for Σ^{-1} in Theorem 3. Consistency of the observed information, or Hessian, evaluated at $\hat{\theta}$, holds by Lemma 5 and Lemma 6 as the third-order derivative is uniformly bounded in mean in a neigborhood of θ, that the Hessian evaluated at θ is consistent, and finally that $\hat{\theta}$ is consistent by Theorem 3.

Using Lemma 4, the observed information can be represented as follows setting

$$\text{vec}\theta = \left((\text{vec}A)', (\text{vec}B)', (\text{vec}C)', (\text{vec}\Omega)', a, b, \right)'.$$

The corresponding consistent estimator of the covariance matrix is given by Σ^{-1} evaluated at $\hat{\theta}$ where Σ^{-1} is given by

$$\Sigma^{-1} = \Sigma_1 + \Sigma_2 = \frac{1}{T}\sum_{t=}^{T}(\text{blockdiag}\left(\Upsilon_t \otimes \Omega^{-1}, 0_{2\times 2}\right) + \Psi_t) \tag{53}$$

Here Υ_t is the $(q+m(k+1))\times$ $(q+m(k+1))$ dimensional matrix defined by,

$$\begin{pmatrix} p^*_{At}\eta'\mathbb{X}_{t-1}\mathbb{X}'_{t-1}\eta & 0 & p^*_{At}\eta'\mathbb{X}_{t-1}\mathbb{X}'_{t-1}\eta_\perp & p^*_{At}\eta'\mathbb{X}_{t-1}\varepsilon'_{At}\Omega^{-1} \\ 0 & p^*_{Bt}\eta'\mathbb{X}_{t-1}\mathbb{X}'_{t-1}\eta & p^*_{Bt}\eta'\mathbb{X}_{t-1}\mathbb{X}'_{t-1}\eta_\perp & p^*_{Bt}\eta'\mathbb{X}_{t-1}\varepsilon'_{Bt}\Omega^{-1} \\ p^*_{At}\eta'_\perp\mathbb{X}_{t-1}\mathbb{X}'_{t-1}\eta & p^*_{Bt}\eta'_\perp\mathbb{X}_{t-1}\mathbb{X}'_{t-1}\eta & \eta'_\perp\mathbb{X}_{t-1}\mathbb{X}'_{t-1}\eta_\perp & \eta'_\perp\mathbb{X}_{t-1}\sum_{M=A,B}p^*_{Mt}\varepsilon'_{Mt}\Omega^{-1} \\ p^*_{At}\Omega^{-1}\varepsilon_{At}\mathbb{X}'_{t-1}\eta & p^*_{Bt}\Omega^{-1}\varepsilon_{Bt}\mathbb{X}'_{t-1}\eta & \Omega^{-1}\sum_{M=A,B}p^*_{Mt}\varepsilon_{Mt}\mathbb{X}'_{t-1}\eta_\perp & \Omega^{-1}(\sum_{M=A,B}p^*_{Mt}\varepsilon_{Mt}\varepsilon'_{Mt} - \frac{1}{2}\Omega)\Omega^{-1} \end{pmatrix}$$

Next, Ψ_t is the $(q+m(k+1)+2)\times$ $(q+m(k+1)+2)$ dimensional matrix given by

$$\Psi_t = p_{At}p_{Bt}\psi_t\psi'_t - p^*_{At}p^*_{Bt}\psi^*_t\psi^{*\prime}_t$$

where

$$\psi^{*\prime}_t = \left[\begin{array}{cccc} vec\left(\Omega^{-1}\varepsilon_{At}\mathbb{X}'_{t-.1}\eta\right)', & -vec\left(\Omega^{-1}\varepsilon_{Bt}\mathbb{X}'_{t-1}\eta\right)', & vec\left(\frac{1}{2}\Omega^{-1}Q_t\Omega^{-1}\right)', & v'_t \end{array}\right],$$

$$\psi'_t = \left[\begin{array}{cccc} 0, & 0, & 0, & v'_t \end{array}\right], \quad Q_t = [\varepsilon_{At}\varepsilon'_{At} - \varepsilon_{Bt}\varepsilon'_{Bt}] \quad \text{and } v'_t = (1, f(\eta'\mathbb{X}_{t-1})).$$

References

Anderson, H. (1997). Transaction costs and non-linear adjustment towards equilibrium in the U.S. treasury bill market. *Oxford Bulletin of Economics and Statistics 59*, 465–484.

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

Balke, N. and T. Fomby (1997). Threshold cointegration. *International Economic Review 38*, 627–465.

Bec, F., M. Ben Salem, and M. Carrasco (2004). Tests for unit-root versus threshold specification with an application to the PPP. *Journal of Business and Economic Statistics 22*, 382–395.

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

Bec, F. and A. Rahbek (2004). Vector equilibrium correction models with non-linear discontinuous adjustments. *Econometrics Journal 7*, 628–651.

Berka, M. (2004). General equilibrium model of arbitrage trade and real exchange rate persistence. Unpublished paper: University of British Columbia.

Bollerslev, T. and R. F. Engle (1993). Common persistence in conditional variances. *Econometrica 61*, 167–186.

Bollerslev, T., R. F. Engle, and D. B. Nelson (1994). ARCH models. In R. F. Engle and D. McFadden (Eds.), *The Handbook of Econometrics, Volume 4*, pp. 2959–3038. Amsterdam: North-Holland.

Brown, B. M. (1971). Martingale central limit theorems. *Annals of Mathematical Statistics 49*, 59–66.

Caner, M. and B. Hansen (2001). Threshold autoregression with a unit root. *Econometrica 69*, 1555–1697.

Carrasco, M. and X. Chen (2002). Mixing and moment properties of various garch and stochastic volatility models. *Econometric Theory 18*, 17–39.

Chan, K. S. and H. Tong (1986). On estimating thresholds in autoregressive models. *JTSA 7*, 179–190.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika 74*, 33–43.

Dempster, A. P., N. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B 39*, 1–38.

Douc, R., E. Moulines, and T. Ryden (2004). Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regimes. *Annals of Statistics 32*, 2254–2304.

Dumas, B. (1992). Dynamic equilibrium and the real exchange rate in a spatially separated world. *Review of Financial Studies 5*, 153–180.

Enders, W. and C. Granger (1998). Unit-root tests and asymmetric adjustment with an example using the terms structure of interest rates. *Journal of Business and Economic Statistics 16*, 304–311.

Enders, W. and P. Siklos (2001). Cointegration and threshold adjustment. *Journal of Business and Economic Statistics 19*, 166–176.

Engle, R. F. and G. G. J. Lee (1999). A permanent and transitory component model of stock return volatility. In R. F. Engle and H. White (Eds.), *Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, pp. 475–497. Oxford: Oxford University Press.

Engle, R. F. and A. Smith (1999). Stochastic permanent breaks. *The Review of Economics and Statistics 81*, 553–574.

Feigin, P. and R. Tweedie (1985). Random coefficient autoregressive processes: A Markov chain analysis of stationarity and finiteness of moments. *Journal of Time Series Analysis 6*, 1–14.

Fong, P., W. Li, and C. Wong (2007). On a mixture vector autoregressive model. *The Canadian Journal of Statistics 35*, 135–150.

Goldfeld, S. M. and R. E. Quandt (1973). A Markov model for switching regressions. *Journal of Econometrics 1*, 3–15.

Gourieroux, C. and C. Y. Robert (2006). Stochastic unit root models. *Econometric Theory 22*, 1052–1090.

Granger, C. W. J. and N. R. Swanson (1997). An introduction to stochastic unit-root processes. *Journal of Econometrics 80*, 35–62.

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

Hall, S. G., Z. Psaradakis, and M. Sola (1999). Detecting periodically collapsing bubbles: a Markov-switching unit root test. *Journal of Applied Econometrics 14*, 143–154.

Hamilton, J. D. and R. Susmel (1994). Autoregressive conditional heteroskedasticity and changes in regimes. *Journal of Econometrics 64*, 307–333.

Hansen, B. and B. Seo (2002). Testing for two-regime threshold cointegration in vector error-correction models. *Journal of Econometrics 110*, 293–318.

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

Hansen, B. E. (1997). Inference in TAR models. *Studies in Nonlinear Dynamics and Econometrics 2*, 1–14.

Jensen, S. T. and A. Rahbek (2007). On the law of large numbers for (geometrically) ergodic markov chains. *Econometric Theory 23*, 761–766.

Karlsen, H. A. and D. Tjøstheim (1990). Autoregressive segmentation of signal traces with application to geological dipmeter measurements. *IEEE Transactions on Geoscience and Remote Sensing 28*, 171–181.

Kilian, L. and M. Taylor (2003). Why is it so difficult to beat the random walk forecast of exchange rates ? *Journal of International Economics 60*, 85–107.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies 65*, 361–393.

Leybourne, S. J., B. McCabe, and T. C. Mills (1996). Randomised unit root processes for modelling and forecasting financial time series: theory and applications. *Journal of Forecasting 15*, 253–70.

Luukkonen, R., P. Saikkonen, and T. Teräsvirta (1988). Testing linearity against smooth transition autoregressive models. *Biometrika 75*, 491–99.

Magnus, J. R. and H. Neudecker (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York: Wiley.

Meyn, S. P. and R. L. Tweedie (1993). *Markov Chains and Stochastic Stability*. London: Springer-Verlag.

Michael, P., A. Nobay, and D. Peel (1997). Transactions costs and nonlinear adjustment in real exchange rates : An empirical investigation. *Journal of Political Economy 105*, 862–879.

Obstfeld, M. and K. Rogoff (2000). The six major puzzles in international macroeconomics : Is there a common cause ? *NBER Macroeconomics Annual*, 339–389.

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

Rogoff, K. (1996). The purchasing power parity puzzle. *Journal of Economic Literature 34*, 647–668.

Ruud, P. (1991). Extensions of estimation methods using the EM algorithm. *Journal of Econometrics 49*, 305–341.

Saikkonen, P. (2005). Stability results for nonlinear error correction models. *Journal Of Econometrics 127*, 69–81.

Sercu, P., R. Uppal, and C. Van Hulle (1995). The exchange rate in the presence of transaction costs : Implications for tests of purchasing power parity. *The Journal of Finance 50*, 1309–1319.

Shephard, N. (1994). Partial non-Gaussian state space. *Biometrika 81*, 115–131.

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

Titterington, D. M., A. F. M. Smith, and U. E. Makov (1985). *Statistical Analysis of Finite Mixture Distributions*. Chichester: Wiley.

Tjøstheim, D. (1990). Non-linear time series and Markov chains. *Advances in Applied Probability 22*, 587–611.

Tong, H. (1983). *Threshold models in non-linear time series analysis*. Heidelberg: Springer-Verlag.

Tong, H. (1990). *Non-linear Time Series*. Oxford: Oxford University Press.

Tong, H. and K. S. Lim (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society, Series B 42*, 245–292.

Wong, C. S. and W. K. Li (2000). On a mixture autoregressive model. *Journal of the Royal Statistical Society, Series B 62*, 95–115.

Wong, C. S. and W. K. Li (2001a). On a logistic mixture autoregressive model. *Biometrika 88*, 833–846.

Wong, C. S. and W. K. Li (2001b). On a mixture autoregressive conditional heteroscedastic model. *Journal of the American Statistical Association 96*, 982–995.

Zhang, Z., W. K. Li, and K. C. Yuen (2006). On a mixture garch time-series model. *Journal of Time Series Analysis 27*, 577–597.