INFERENCE AND ERGODICITY IN THE AUTOREGRESSIVE CONDITIONAL ROOT MODEL

Anders Rahbek
Department of Statistics and Operations Research, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø , Denmark
rahbek@stat.ku.dk
www.stat.ku.dk/~rahbek/

Neil Shephard
Nuffield College, University of Oxford, New Road, Oxford OX1 1NF, U.K.
neil.shephard@nuf.ox.ac.uk
www.nuff.ox.ac.uk/users/shephard/

This version November 2002.

Abstract

In this paper we develop a vector time series model which allows long-term disequilibriums to have *epochs* of non-stationarity, giving the impression that long term relationships between economic variables have temporarily broken down, before they *endogenously* collapse back towards their long term relationship. The autoregressive conditional root (ACR) process is shown to be geometrically ergodic, stationary and posess all moments under simple conditions. Furthermore, we establish consistency and asymptotic normality of the maximum likelihood estimators in the ACR model.

Keywords: Cointegration; Equilibrium correction model; GARCH; Hidden Markov model; Likelihood; Regime switching; STAR model; Stochastic break; Stochastic unit root; Switching regression; Threshold autoregression; Unit root hypothesis.

1 Introduction

1.1 The ACR model

Much of macroeconomic theory is concerned with long term relationships between variables such as the quantity theory of money and purchasing power parity (PPP). The variables are frequently considered non-stationary and modern econometrics analyses long term relationships between them using the cointegration framework developed by Engle and Granger (1987). In terms of dynamics, cointegration allows short term deviations from long term relationships by imposing stationarity on the transitory disequilibriums.

With the disequilibriums in mind, we introduce a general multivariate non-linear time series model which allows the disequilibriums to have *epochs* of seeming non-stationarity, giving the impression that the long term relationships have temporarily broken down, before they *endogenously* collapse back towards their long term relationship. We show that despite the epochs of seeming non-stationarity the time series is indeed stationary under suitable regularity conditions. Stated differently, the collapses regularise the periods of non-stationarity forcing the disequilibrium to be globally stationary. This type of behavior is reflected in, for example, an economic theory model developed by Bec, Ben Salem, and Carrasco (2001) where it is shown how trading costs in a two-country stochastic general equilibrium model create a region of no trade where the PPP does not hold, while stationarity holds outside this region.

Our analysis of transitory disequilibriums will be based around an autoregression whose autoregressive root(s) switches endogenously and stochastically between being possibly unity and being stable. We call this model an *autoregressive conditional root* (ACR) model. In addition to establishing stationarity we also show that the maximum likelihood (ML) estimator of the parameters in the model are consistent and asymptotically normal. We believe this is the first paper which provides the distribution theory of the maximum likelihood estimator of a model which shifts between epochs of being stationary and non-stationary.

1.2 Univariate ACR(1) example

To fix ideas in this introduction we will focus solely on the simplest fully parametric form corresponding to a univariate autoregression of order one, the ACR(1) process, which is given by the equation

$$x_t = \rho^{s_t} x_{t-1} + \varepsilon_t, \quad \text{for} \quad t = 1, 2, \ldots, T. \tag{1}$$

Here s_t is binary, ρ is a real number and ε_t is an *i.i.d.*N(0,σ^2) sequence. In the next Section the general vector version, with k lags and no assumed parametric distribution on ε_t will be spelt out. Here our aim is to convey the flavour of the model. 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. We assume the *prediction probability* has the form

$$\Pr(s_t = 1 | x_{t-1}, \varepsilon_t) = p_{x_{t-1}}, \tag{3}$$

where the notation $p_{x_{t-1}}$ reflects dependence solely on x_{t-1}. By assumption s_t and ε_t are independent conditionally 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 $|\rho| = |\pi + 1| < 1$.

Clearly, the dynamics of the regime are determined entirely endogenously and so are similar to the threshold models of Tong (1990) and Enders and Granger (1998). However, now the threshold is actually stochastic rather than deterministic. The essential requirement for the conditional probability $p_{x_{t-1}}$, will be that it tends to one as $|x_{t-1}|$ tends to infinity in addition to it being measurable with respect to x_{t-1}. Importantly we do not bound $p_{\cdot}$ away from one. In particular, we can allow processes of the type

$$p_{x_{t-1}} = \begin{cases} 1, & \text{if } |x_{t-1}| > c > 0, \\ 0, & \text{otherwise.} \end{cases}$$

which is a Tong (1990) threshold autoregressive process, for it implies

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

The implication is that we can view ACR models as softening the thresholds in autoregressive threshold models.

We show that an initial distribution exists such that x_t is indeed strictly stationary and possess all of its moments provided the regime corresponding to $s_t = 1$ is stationary, or equivalently, $|\rho| = |1 + \pi| < 1$ in (1) and (2) respectively. No other condition is needed apart from the mentioned convergence of the probability $p_{\cdot}$. As emphasized this means that the model can have epochs of seeming non-stationarity, but at the same time be globally stable or stationary. Estimation is straightforward for the likelihood function can be computed via a prediction decomposition. We argue that inference is regular while we briefly mention aspects of inference in the case of x_t being a random walk for the entire period.

In a recent paper, written independently and concurrently from our paper, Gourieroux and Robert (2001) have studied in detail the above process in the case where there is switching between white noise and a random walk (i.e. the case of the above process when $\rho = 0$). Their wide ranging paper, motivated by value-at-risk considerations in financial economics, 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 discuss geometric ergodicity of x_t in this case. Our analysis will be complementary, focusing on estimation and asymptotic inference for use in empirical work in the general and also multivariate version of the ACR model.

The following simple example allows us to gain a better understanding of the behaviour of this process. We have chosen to write the dynamics in terms of the logistic transform

$$\begin{aligned} \lambda(x_{t-1}) &= \log\left\{p_{x_{t-1}} / \left(1 - p_{x_{t-1}}\right)\right\} \\ &= \alpha + \beta x_{t-1}^2, \end{aligned} \tag{4}$$

with α and β being freely varying reals. So long as β is non-negative and α and β are finite, $\lambda(x_{t-1})$ will be bounded and so the process x_t in (1) will be stationary as demonstrated below.

Example 1 *Figure 1(a) shows a sample path from the simplest Markov ACR process (4), together with the associated conditional probabilities $p_{x_{t-1}}$ given in Figure 1(b). The parameter*

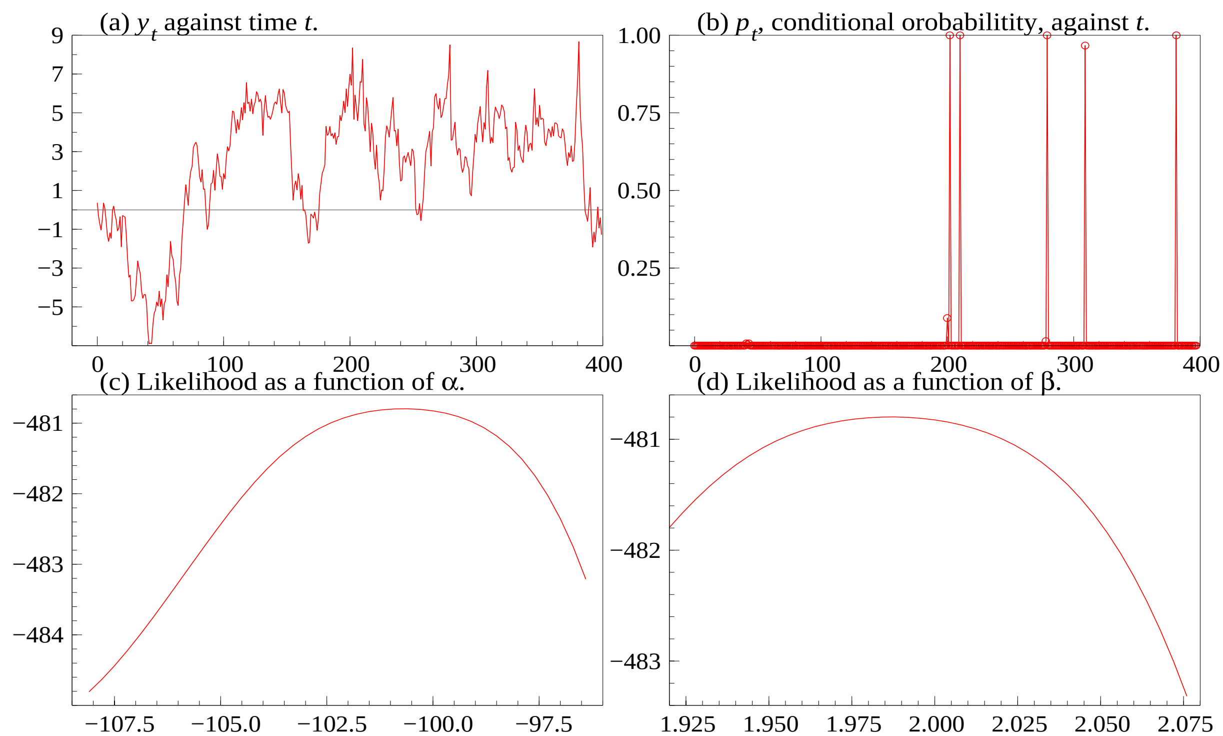

Figure 1: *(a) Simulations from the Markov model y_t. (b) corresponding p_t. (c) Likelihood function as a function of α. (d) Likelihood function as a function of β. Code:* `regime_sim.ox`*.*

values are $\alpha = -100$, $\beta = 2.0$, $\rho = 0.5$, and $\sigma = 0.8$. This process delivers a jagged realisations for $p_{x_{t-1}}$, which never spends substantial consecutive periods close to one. This is enough however for the x_t series to be stable, never going much above ten in absolute value. The likelihood function for this model, as a function of α and β, is drawn in Figures 1(c) and 1(d) respectively. The likelihood function will be detailed in Section 3 of the paper. It is drawn here by varying only one parameter at a time, fixing the others at their true parameter values.

1.3 Related models

The ACR model seems new. However, it is related to a number of well known models. Apart from the already mentioned threshold class of models, perhaps the closest is 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.

In the very large Markov switching literature in economics, following Hamilton (1989), the regime s_t is regarded as a latent variable which follows a Markov process specified by the probability $\Pr\left(s_t = 1 | s_{t-1}, x_{t-1}\right) = \Pr\left(s_t = 1 | s_{t-1}\right)$. Thus in contrast to the ACR process the prediction probability is determined implicitly. In the Markov switching literature s_t is usually employed to shift the intercept in a time series model, but it has been used to make the variance to change (Hamilton and Susmel (1994)) delivering a simple stochastic volatility process and even to make the root of an autoregression move between a unit root and a stationary root (Karlsen and Tjøstheim (1990)) or an explosive root (Hall, Psaradakis, and Sola (1999)). In this framework the regimes are an exogenous process with the observable x_t not feeding back into the regime. The likelihood function for this model can be computed via a relatively simple filtering argument so long as the model has an autoregressive structure of finite order. This model can be generalised in a number of ways, allowing explanatory variables to influence the probabilities which govern the switching between the regimes. Two papers which carries this out in some detail in the context of macroeconomics are Diebold, Lee, and Weinbach (1994) and Durland and McCurdy (1994). In statistics and engineering the above model is often called the hidden Markov model (HMM) and is a special case of a state space or parameter driven model (e.g. Harvey (1989) and Cox (1981)). An early important reference in the HMM literature is Baum, Petrie, Soules, and Weiss (1970).

A related approach is the switching regression idea introduced into economics by Goldfeld and Quandt (1973). In our context this would build 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 $\lambda(x_{t-1}) = \alpha + \beta x_{t-1}$ 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. Hence this model has an entirely different interpretation than the ACR model. The time series setup of $\lambda(x_{t-1}) = \alpha + \beta x_{t-1}$ was explicitly studied recently by Wong and Li (2001), 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.

The conditional expectation of equilibrium correction form of the ACR model is

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

Suppose we again define $\lambda(x_{t-1}) = \alpha + \beta x_{t-1}^2$ then

$$\mathrm{E}(\Delta x_t | x_{t-1}) = \pi \frac{\exp\left(\alpha + \beta x_{t-1}^2\right)}{1 + \exp\left(\alpha + \beta x_{t-1}^2\right)} x_{t-1}.$$

If we recast this as

$$\Delta x_t = \pi \frac{\exp\left(\alpha + \beta x_{t-1}^2\right)}{1 + \exp\left(\alpha + \beta x_{t-1}^2\right)} x_{t-1} + \eta_t,$$

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.

Finally, recently Engle and Smith (1999) have proposed an interesting stochastic break model which has some of the above features. They write, in their simplest model

$$\Delta x_t = q_t \varepsilon_t, \qquad \varepsilon_t | \mathcal{F}_{t-1} \sim N(0, \sigma^2)$$

and q_t is a deterministic function of ε_t, bounded below by zero and above by one. Further, $\partial q_t / \partial \left|\varepsilon_t\right|$ is assumed to be finite and strictly negative. A simple example of this is where

$$q_t = \frac{\varepsilon_t^2}{\gamma + \varepsilon_t^2}, \qquad \gamma > 0.$$

This model has shocks which are all permanent but of varying magnitude. It is quite different from the model we desire, which moves between stationary and non-stationary behaviour, but is globally stationary. Our model is more in the stochastic root tradition.

1.4 Outline of the paper

This paper has three other main sections. In Section 2 we extend the model to the multivariate case, including more lags, and derive the stochastic properties of the ACR process including the existence of moments and stationary solutions. Section 3 gives an asymptotic likelihood analysis for the multivariate model. In particular, we state conditions under which the maximum likelihood estimators are consistent and asymptotically normally distributed. We also briefly discuss the use of various testing procedures to look at special cases of the model structure. Section 4 discusses possible extensions and finally Section 5 concludes the paper, while in the Appendix we prove the theorems stated in the paper.

Some notation is used throughout the paper: With A a matrix, $\rho(A)$ denotes the spectral radius or equivalently, the maximal modulus of the eigenvalues of A. The matrix norm $\|A\|$ is given by $\|A\|^2 = tr\left\{A'A\right\}$ and similarly for a vector a, $\|a\|^2 = a'a$. We use the notational definition $A^{\otimes 2} = AA'$. Moreover we apply the notation, $dL(A, dA)$ for the differential of the matrix function $L(\cdot)$ with increment dA; see also Appendix B.

2 Stability and existence of moments for the vector ACR(k) process

In this section we introduce the general multivariate ACR(k) process and discuss some of its properties. In particular we study conditions under which it is stationary and all of its moments exist.

2.1 The vector ACR(k) process

Our general m-dimensional vector ACR(k) process is defined by the equation

$$\begin{aligned} X_t &= s_t\left(A_1 X_{t-1} + ... + A_k X_{t-k}\right) + (1-s_t)\left(B_1 X_{t-1} + ... + B_k X_{t-k}\right) + \varepsilon_t \\ &= s_t A \mathbf{X}_{t-1} + (1-s_t) B \mathbf{X}_{t-1} + \varepsilon_t \end{aligned} \tag{5}$$

together with the prediction probability equation

$$P(s_t = 1 | \mathbf{X}_{t-1}, \varepsilon_t) = 1 - P(s_t = 0 | \mathbf{X}_{t-1}, \varepsilon_t) = p_{x_{t-1}}, \tag{6}$$

for $t = 1, 2, ..., T$. We have used the notational convention,

$$\mathbf{X}_{t-1} = (X'_{t-1}, ..., X'_{t-k})' \tag{7}$$

$$A = (A_1, ..., A_k) \text{ and } B = (B_1, ..., B_k). \tag{8}$$

Here A_i and B_i are $m \times m$ matrices and ε_t is an *i.i.d.* mean-zero sequence with variance Ω positive definite. Furthermore ε_t is assumed to have a positive and continuous density with respect to the Lebesgue measure. The two autoregressive regimes are governed by the $m \times mk$ dimensional matrices A and B respectively. As previously noted the notation $p_{x_{t-1}}$ for the prediction probability for switches between the two regimes, indicates that it is a function of $\mathbf{X}_{t-1}$ alone so that in particular s_t and ε_t are independent conditional on $\mathbf{X}_{t-1}$.

The generalisation differs from the univariate ACR(1) process in (1) in that we allow for a vector process (of dimension m), a richer lag structure (of order k), potentially non-Gaussian errors and additional flexibility in the dynamics by the introduction of the additional autoregressive regime parameter B. Specifically, the univariate ACR(1) example in (1) has $m = 1$, $k = 1$, $A = \rho$ and $B = 1$. Here A governs the locally stationary regime, while $B = 1$ governs the unit-root regime. In the multivariate extension consider as an example the case of $k = 2$. Choosing, say, $B = (B_1, I_m - B_1)$ introduces (at least) m unit-roots in the B regime as desired and which reflects the flexibility in the dynamics in the current parametrization.

Below we demonstrate how the regime governed by B can have unit and even explosive-roots while X_t is globally stationary as a process provided the A-regime has no unit or explosive-roots.

2.2 Stationarity and ergodicity

In order to address stationarity and ergodicity of the ACR(k) process in (5) we consider the companion form of X_t as given by

$$\mathbf{X}_t = s_t \mathbf{A} \mathbf{X}_{t-1} + (1 - s_t) \mathbf{B} \mathbf{X}_{t-1} + \mathbf{e}_t \tag{9}$$

$$P(s_t = 1 | \mathbf{X}_{t-1}, \mathbf{e}_t) = 1 - P(s_t = 0 | \mathbf{X}_{t-1}, \mathbf{e}_t) = p_{x_{t-1}} \tag{10}$$

with $\mathbf{A}$, $\mathbf{B}$ the $mk \times mk$ matrices defined as

$$\mathbf{A} = \begin{pmatrix} A_1 & A_2 & \cdots & A_k \\ I_m & 0 & \cdots & 0 \\ & \ddots & \ddots & \vdots \\ 0 & & I_m & 0 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} B_1 & B_2 & \cdots & B_k \\ I_m & 0 & \cdots & 0 \\ & \ddots & \ddots & \vdots \\ 0 & & I_m & 0 \end{pmatrix}$$

and likewise $\mathbf{e}_t = (\varepsilon_t', 0, ..., 0)'$. By definition $\mathbf{X}_t$ is a Markov chain on $\mathbb{R}^p$, $p = mk$ and below we state the regularity conditions for which this chain is geometrically ergodic and has moments. The geometric ergodicity implies in particular that a stationary version exists.

We formulate the conditions more generally than the specific choice of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{e}_t$ corresponding to the companion form of X_t in (5) is treated. Specifically, we show that if the eigenvalues of $\mathbf{A}$ in absolute value are smaller than one and under suitable conditions on the functional form of $p_{x_{t-1}}$ the process $\mathbf{X}_t$ has appropriate moments and is stable in the sense that an invariant distribution exists for $\mathbf{X}_t$. We note as a corollary that this implies that the same holds for X_t. First, we turn to regularity conditions on the transition probabilities:

Assumption 1 *Consider the Markov chain $\mathbf{X}_t$ defined by (9) with $\mathbf{e}_t$ an i.i.d. mean zero sequence with finite variance, $E \|\mathbf{e}_t\|^2 < \infty$. With $\mathbb{B}^p$ being the Borel σ-algebra on $\mathbb{R}^p$, assume that for all sets $A \in \mathbb{B}^p$ and for some integer $k \geq 1$, that the k step transition density with respect to the Lebesgue measure, $f(\cdot|\cdot)$ as defined by*

$$P(\mathbf{X}_t \in A | \mathbf{X}_{t-k} = x) = \int_A f(y|x) dy$$

is strictly positive and continuous in both arguments.

Assumption 1 is, in particular, satisfied if the sequence of the form $\mathbf{e}_t = (\varepsilon_t', 0, ..., 0)'$ is ε_t is $i.i.d.N_m(0, \Omega)$ with a positive definite covariance matrix, $\Omega > 0$, see Corollary 1. Note also that we could replace the requirement of continuity by lower semi-continuity.

Next, turn to the assumption on the autoregression matrices $\mathbf{A}$ and $\mathbf{B}$:

Assumption 2 *The spectral radius of $\mathbf{A}$ is smaller than one,*

$$\rho(\mathbf{A}) < 1.$$

In other words, in the regime governed by $\mathbf{A}$ the process behaves – locally – as a stationary vector autoregressive process. There are no restrictions on the matrix $\mathbf{B}$ whatsoever and therefore regime switches between stationary and non-stationary, even explosive regimes are allowed for.

Finally, we turn to the assumption on the functional form of the state probability $p_.$.

Assumption 3 *The prediction probability function p_x is continuous in x and converges to one as $\|x\|$ tends to infinity. Equivalently,*

$$(1 - p_x) = o(1) \tag{11}$$

as $\|x\| \to \infty$.

As emphasized in the introductory section, of particular interest for us is the case of a logistic function where

$$\log \{p_x / (1 - p_x)\} = \lambda(x) = \alpha + \beta \|x\|^2 \tag{12}$$

which is a continuous function in (the norm of) x. Trivially,

$$(1 - p_x) = (1 + \exp \lambda(x))^{-1} = o(1)$$

provided $\beta > 0$. In other words, the probability is such that whatever state the process is in, there is always a non-negative probability that we will (re-)enter the state governed by the $\mathbf{A}$ matrix in (9) corresponding to a stationary regime. In addition, the structure is such that the further away the process gets from the regime governed by $\mathbf{A}$ the probability of staying there tends to zero.

Theorem 1 *Consider the p-dimensional ACR process $\mathbf{X}_t$ defined by (9). Then under Assumptions 1, 2 and 3, $\mathbf{X}_t$ is geometrically ergodic and X_0 can be given an initial distribution such that $\mathbf{X}_t$ is stationary. Furthermore, $\mathbf{X}_t$ has finite second order moments.*

Remark 1 *In the special case where $\mathbf{B}$ represents a pure random walk regime, Assumption 3 can be replaced by the requirement that $p_x \geq \delta > 0$ for all x, where δ is some constant. For the particular choice of a logistic function in (12) this implies that $\beta \geq 0$ as opposed to $\beta > 0$ which was required for general $\mathbf{B}$.*

The results on the existence of higher order moments are generalised in the following way:

Theorem 2 *Consider the p-dimensional ACR process $\mathbf{X}_t$ defined by (9). Then under Assumptions 1, 2 and 3, $\mathbf{X}_t$ is geometrically ergodic and has finite $2m$'th order moments, $E\|\mathbf{X}_t\|^{2m} < \infty$ provided $E\|\mathbf{e}_t\|^{2m} < \infty$.*

Remark 2 *Note that in the case of* $\mathbf{e}_t$ *being Gaussian trivially* $E\left\|\mathbf{e}_t\right\|^{2m}<\infty$ *for all* m.

Finally, we are now in position to state the corollary which establishes the properties of the vector ACR(k) process:

Corollary 1 *The multivariate ACR*(k) *process* X_t *defined by (5) is geometrically ergodic if* $p_{x_{t-1}}$ *satisfies Assumption 3 and if*

$$\left|I_m - A_1 z - \ldots - A_k z^k\right| = 0 \Rightarrow |z| > 1,\ z \in \mathbb{C}. \tag{13}$$

In particular, $(X_0, \ldots, X_{-k+1}) = \mathbf{X}_0$ *can be given an initial distribution such that* X_t *is stationary. Furthermore,* $E\left\|X_t\right\|^{2m}<\infty$ *provided* $E\left\|\varepsilon_t\right\|^{2m}<\infty$.

3 Inference

3.1 Likelihood based estimation and testing

In this section we consider asymptotic inference for the multivariate ACR model defined by (5) and (6) under the specialised assumptions of a logistic prediction probability function (12) as well as normality of the innovations ε_t. We study how to test hypotheses which leave the epochs or mixing structure intact. In addition, we discuss inference for hypotheses, such as the unit root hypothesis, which do not allow epochs of either mean-reversion or random-walk type behaviour.

3.2 Distribution of the ML estimator

The ACR model is given by (5) as

$$X_t = s_t A \mathbf{X}_{t-1} + (1 - s_t) B \mathbf{X}_{t-1} + \varepsilon_t \tag{14}$$

for $t = 1, \ldots, T$ and with A, B and $\mathbf{X}_t$ defined in (8). For the statistical analysis we assume that ε_t is $i.i.d.N(0, \Omega)$ where $\Omega > 0$. Furthermore the logistic prediction probability function $P(\, s_t = 1 |\, \mathbf{X}_{t-1}, \varepsilon_t)$ is given by (12),

$$\log\left\{p_{x_{t-1}}/(1 - p_{x_{t-1}})\right\} = \lambda(\mathbf{X}_{t-1}) = \alpha + \beta \left\|\mathbf{X}_{t-1}\right\|^2. \tag{15}$$

The freely varying parameters A, B are $mk \times mk$ matrices, α, β are scalars and $\Omega > 0$.

Denote by $\theta = \{A, B, \alpha, \beta, \Omega\}$, then the log-likelihood function (given the initial observation $\mathbf{X}_0$) can be written as

$$L_T(\theta) = \textstyle\sum_{t=1}^T \log f_\theta(X_t | \mathbf{X}_{t-1}) = \sum_{t=1}^T \ell_t(\theta), \tag{16}$$

where

$$\ell_t(\theta) = \log\left\{p_{x_{t-1}} \phi_t^A + (1 - p_{x_{t-1}}) \phi_t^B\right\}. \tag{17}$$

Here the convenient notation

$$\phi_t^M = |\Omega|^{-1/2} \exp\left[-\tfrac{1}{2}tr\left\{\Omega^{-1}(X_t - M\mathbf{X}_{t-1})^{\otimes 2}\right\}\right], \tag{18}$$

for $M = A$, B has been used for the Gaussian density, deliberately ignoring constants. The likelihood function in (16) is numerically maximised to obtain the maximum likelihood estimator, $\hat{\theta}$. Numerical aspects of this tasked are addressed in the next subsection.

Here we state the asymptotic behavior of $\hat{\theta}$ restricting the parameter space appropriately. It should be emphasized that the results show that the maximum likelihood estimators are asymptotically Gaussian even if the regime goverened by B allows unit (even explosive) roots, provided of course that the A regime 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: In particular in existing literature on e.g. TAR models, both regimes need stationary roots for distributional theory, or the estimation is not based on maximum likelihood. We comment further on the results and assumptions immediately after the theorem.

Theorem 3 *Consider the ACR model defined by equations (9), (10) and (12). Then if the assumption about the characteristic roots in (13) holds for the A-regime and $A \neq B$ and finally Assumption 3 holds, there exists with probability tending to one as T tends to infinity a sequence of $\hat{\theta} = \{\hat{A}, \hat{B}, \hat{\alpha}, \hat{\beta}, \hat{\Omega}\}$ which satisfies the score equation,*

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

for all $d\theta$. The sequence is consistent,

$$\hat{\theta} \xrightarrow{P} \theta,$$

and furthermore asymptotically Gaussian,

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

Here Σ^{-1} is defined in Appendix B, Lemma 3. A consistent estimator of Σ^{-1} can be derived from equation (31) corresponding to the observed information. Finally, likelihood ratio tests for simple hypotheses on θ are asymptotically χ^2 distributed.

The proof is based on establishing the standard Cramér type conditions for consistency and asymptotic normality and is given in Appendix B. Note that also likelihood ratio tests for composite hypotheses on θ as given by $h(\phi) = \theta$, say, are asymptotically χ^2 under the usual (rank) regularity conditions for the mapping $h(\cdot)$.

The results are derived specifically for the parametrization and functional choice of

$$\lambda(x) = \alpha + \beta \|x\|^2$$

in (4). However, while our derivations do depend on the chosen logistic structure for the probabilities p_x, it is straightforward to modify the results to accommodate alternative specifications of $\lambda(\cdot)$. Specifically, for transparency we have formulated all sufficient and relevant quantities in terms of the derivative of $\lambda(\cdot)$ with respect to the parameters in θ in Lemmas 2, 4 and 5. In addition recall that the results on ergodicity have been formulated for flexible choices of $\lambda(\cdot)$.

Clearly 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. Consider the univariate case with x_t given by (1) or (2). The null of x_t being a random walk without any epochs of stationarity can be parametrized by $\pi = 0$ and likewise the null of an autoregressive process can be reparametrized in the ACR(1) model as $\gamma = 0$ where, say, $\gamma = \exp(-\alpha)$. While both hypotheses are simple, the parameter β vanishes under the null (as does α for the unit root hypothesis) and usual asymptotic expansions in terms of score and information are therefore problematic as discussed in general in Andrews and Ploberger (1994), Davies (1987) and Hansen (1996).

Some related issues have been recently analysed in the context of threshold autoregressive (TAR) models. 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. His theory is based upon the least squares method. Testing for a unit root in both regimes is treated in Caner and Hansen (2001) and Bec, Ben Salem, and Carrasco (2001).

3.3 On optimisation of the likelihood

In order to carry out likelihood inference we have to numerically maximise the likelihood function. Experimenting has lead us to favour maximising the likelihood function via the EM algorithm (e.g. Dempster, Laird, and Rubin (1977) and Ruud (1991)). With $X = (X_1, ..., X_T)$ this regards the indicators $s = (s_1, ..., s_T)$ as missing data, iteratively computing the conditional mean $E\{\log f(X, s)|X_T, X_{T-1}, \ldots, \mathbf{X}_0\}$ and then maximising it with respect to the parameters of the model. As is well-known repeated switching between these the expectation (E) and the maximisation (M) steps leads to an algorithm which maximises $\log f(X)$.

More precisely for the EM algorithm consider the log-likelihood function for $s = (s_1, ..., s_T)$

and $X = (X_1, ..., X_T)$ conditional on $\mathbf{X}_0$ as given by

$$\log f(X, s) = -\frac{T}{2} \log |\Omega| - \frac{1}{2}\sum_{t=1}^{T} tr \left[\Omega^{-1} \left\{X_t - (s_t A + (1 - s_t)B)\mathbf{X}_{t-1}\right\}^{\otimes 2}\right]$$
$$+ s_t \log \left\{\frac{p_{x_{t-1}}}{(1-p_{x_{t-1}})}\right\} + \log\left(1 - p_{x_{t-1}}\right),$$

see also (16). Conditioning on $X_T, X_{T-1}, \ldots, \mathbf{X}_0$ and taking expectations, one finds that

$$E\left(s_t | X_T, X_{T-1}, \ldots, \mathbf{X}_0\right) = E\left(s_t | X_t, \mathbf{X}_{t-1}\right) = E\left(s_t^2 | \mathbf{X}_T, X_{T-1}, \ldots, X_0\right) = p^*_{x_t}, \tag{19}$$

where

$$p^*_{x_t} = \frac{p_{x_{t-1}} \phi_t^A}{p_{x_{t-1}} \phi_t^A + (1 - p_{x_{t-1}}) \phi_t^B}, \tag{20}$$

with ϕ_t^A and ϕ_t^B defined in (18). It immediately follows that $E\left\{\log f(X, s) | X_T, X_{T-1}, \ldots, \mathbf{X}_0\right\}$ equals

$$-\frac{1}{2}\sum_{t=1}^{T} tr \left\{\Omega^{-1}\left[X_t^{\otimes 2} + p^*_{x_t}(A\mathbf{X}_{t-1})^{\otimes 2} + (1 - p^*_{x_t})(B\mathbf{X}_{t-1})^{\otimes 2} - 2(p^*_{x_t}A - (1 - p^*_{x_t})B)\mathbf{X}_{t-1}X_t'\right]\right\}$$
$$-\frac{T}{2} \log |\Omega| + \sum_{t=1}^{T} \left[p^*_{x_t} \log\left\{\frac{p_{x_{t-1}}}{(1 - p_{x_{t-1}})}\right\} + \log(1 - p_{x_{t-1}})\right]. \tag{21}$$

The M-step maximises this with respect to the parameters, holding $p^*_{x_t}$ fixed. Importantly the first two components depend only on A, B and Ω, while the last is a function of only the parameters which determine $p_{x_{t-1}}$. Hence these two parts can be maximised separately.

An attractive feature of this problem is that the maximisation can be carried out analytically. In particular, maximizing the first two parts in (21) gives,

$$\hat{A} = \sum_{t=1}^{T} p^*_{x_{t-1}} X_t \mathbf{X}'_{t-1} \left(\sum_{t=1}^{T} p^*_{x_{t-1}} \mathbf{X}_{t-1}\mathbf{X}'_{t-1}\right)^{-1},$$
$$\hat{B} = \sum_{t=1}^{T} (1 - p^*_{x_{t-1}}) X_t \mathbf{X}'_{t-1} \left(\sum_{t=1}^{T} p^*_{x_{t-1}} \mathbf{X}_{t-1}\mathbf{X}'_{t-1}\right)^{-1},$$

while

$$\hat{\Omega} = \frac{1}{T}\sum_{t=1}^{T} \left[X_t^{\otimes 2} + p^*_{x_t}(\hat{A}\mathbf{X}_{t-1})^{\otimes 2} + (1 - p^*_{x_t})(\hat{B}\mathbf{X}_{t-1})^{\otimes 2}\right.$$
$$\left. -(p^*_{x_t}\hat{A} - (1 - p^*_{x_t})\hat{B})\mathbf{X}_{t-1}X_t' - X_t\mathbf{X}'_{t-1}(p^*_{x_t}\hat{A}' - (1 - p^*_{x_t})\hat{B}')\right]$$

The last component of (21) can be rewritten as

$$\sum_{t=1}^{T} \left[p^*_{x_t}\lambda(\mathbf{X}_{t-1}) - \log\left\{1 + \exp\left(\lambda(\mathbf{X}_{t-1})\right)\right\}\right], \tag{22}$$

which has to be optimised numerically. In cases where $\lambda(\mathbf{X}_{t-1})$ is a linear function of past data, such as in the pure autoregressive scheme

$$\lambda(\mathbf{X}_{t-1}) = \alpha + \beta g(\mathbf{X}_{t-1}),$$

where $g(x) = x'x$ or some other known function, then (22) takes on the form of a logistic regression for the "observations" $p^*_{x_t}$. As a result the likelihood function is concave, a property which extends to any dynamic model where $\lambda(\cdot)$ is linear in functions of lagged data. For more general model structures this is not the case which implies the M-step in the EM algorithm has to be carried out using automatic numerical optimisation algorithms. Having completed the M-step we have to return to the E-step, to perform another iteration of the algorithm.

4 Potential extensions

In this Section we discuss briefly discuss some natural extensions of the ACR model structure, focusing on cointegration, deterministic components and conditional heteroskedasticity.

4.1 ACR based cointegration models

At the start of this paper we motivated the development of the ACR model as a way to formalise the idea that a long term equilibrium or cointegration between variables breaks down yielding a disequilibrium which is a random walk. As the size of the equilibrium grows so the chance the long-term relationship reasserts itself increases. Thus in the very long-term the disequilibrium is stationary.

A generalization of the univariate model to the multivariate case would furthermore allow for analysis of not only the real exchange rate, but also potentially include, say, money and bonds markets and in particular interest rate parities — see Taylor (1995) for an overview and Frydman and Goldberg (2002) for a recent discussion with non-linear type dynamics.

In econometrics there already exists a substantial literature on cointegration models where the cointegrating relationships change through time. These are usually phrased in terms of threshold models and leading references include Enders and Granger (1998), Tsay (1989) and Tsay (1998).

To incorporate the ACR kind of dynamics in cointegration consider the first order m-dimensional canonical equilibrium correction model as given by

$$\Delta X_t = \gamma \delta' X_{t-1} + \varepsilon_t,$$

where ε_t is an i.i.d.$N(0,\Omega)$ sequence. Leading references to this model structure include Hendry (1995), Engle and Granger (1987) and Johansen (1995). The *disequilibrium term* $\delta' X_t$ here measures the size of the *out of equilibrium.* Then suppose that the PPP, say, is given by the univariate process $\delta' X_t$. It immediately follows that the vector ACR process

$$\Delta X_t = \gamma s_t \delta' X_{t-1} + \varepsilon_t.$$

has some of the desired features: In particular, the process $\delta' x_t$,

$$\Delta\delta' X_t = s_t(\delta'\gamma)\delta' X_{t-1} + \delta'\varepsilon_t$$

is a univariate ACR process and so is strictly stationary using the results discussed above provided $|\delta'\gamma| < 1$. Likewise with $\gamma_\perp$ denoting the $m \times m - 1$ dimensional matrix of full rank $m - 1$ and with $\gamma'\gamma_\perp = 0$,

$$\Delta\gamma_\perp' X_t = \gamma_\perp' \varepsilon_t$$

there are $m - 1$ common trends. In epochs where s_t is zero the series has no cointegrating relationships even though they exist in the long run.

4.2 Deterministic components

It is straightforward to extend the basic ACR model to handle deterministic components. This can be carried out without removing the important properties of ergodicity and existence of moments. Consider, say, the case where the model for the observed X_t allows for a constant level through the parameter μ which is given by

$$X_t = \mu + Y_t$$

where Y_t follows an ACR(1) model. The probabilistic properties of X_t can be directly derived from those of Y_t — in particular the X_t process has a level μ in both regimes. With respect to estimation, we can write X_t as the solution to

$$\Delta X_t = s_t \pi (X_{t-1} - \mu) + \varepsilon_t$$

which we reparameterise as

$$\Delta X_t = s_t(\pi X_{t-1} - \gamma) + \varepsilon_t$$

with π and γ freely varying. This line of argument generalises to flexibly parameterised trend components. A detailed discussion of these types of issues is given in Nielsen and Rahbek (2000).

4.3 Conditional heteroskedasticity

ACR models could also be developed for models of conditional variance, which is a commonly used concept in financial econometrics. Consider first the traditional 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, $\mathcal{F}_t = \sigma\left\{x_t, \sigma_t, ...\right\}$. 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 then we change 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 then the conditional variance is low the process behaves like an integrated GARCH.

5 Conclusions

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.

6 Acknowledgments

We thank Christian Gourieroux 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 Frederique Bec, 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 is grateful to the Danish Social Sciences Research Council and CAF for support. Computations reported in this paper were made using version 3 of the matrix programming language Ox (Doornik (2001)).

Appendix

The Appendix is divided into two parts: Appendix A is concerned with Markov chain theory which focuses on the essential elements for the proof of geometric ergodicity developed in Section 2. Appendix B is about asymptotic inference in Markov chain models. This is mostly covered in Section 3 of the paper. Both parts include a brief introduction to the relevant material as well as the proofs needed in the paper.

A Drift Criteria

A.1 Introduction

To address geometric ergodicity and stationarity of x_t Markov chain theory will be used, see e.g. Meyn and Tweedie (1993) and Chan and Tong (1985) for further details. Let $(\mathbf{X}_t)_{t=0,1,2,...}$ be a general p-dimensional time homogenous Markov chain on $(\mathbb{R}^p, \mathbb{B}^p)$ which satisfies Assumption 1. By Lemma 1 below we show that the chain is *irreducible* with respect to the Lebesgue measure μ, it is *aperiodic* and compact sets $C \subset \mathbb{R}^p$ are *small*, see also An, Chen, and Huang (1997, Lemma 2.5) and Hansen and Rahbek (1998) for similar results in different settings.

Lemma 1 *Under Assumption 1 the homogenous Markov chain $(\mathbf{X}_t)_{t=0,1,..}$ on $(\mathbb{R}^p, \mathbb{B}^p)$ is μ-irreducible, aperiodic and compact sets $C \subset \mathbb{R}^p$ are small.*

Proof of Lemma 1: First note that for any n, the n-step transition probabilities can be defined recursively as follows, $P^1(A|\, x) = P(\mathbf{X}_1 \in A|\, \mathbf{X}_0 = x)$ and

$$P^n(A|\, x) = \int_{\mathbb{R}^p} P^{n-1}(A|\, y)\, dP^1(y|\, x) \quad \text{for } n \geq 2,\ x \in \mathbb{R}^p \text{ and } A \in \mathbb{B}^p$$

with P the chains kernel. Assumption 1, states that for some $n = k$,

$$P^k\left(A|\,x\right) = \textstyle\int_A f(y|x)dy,$$

with f positive and continuous. Next turn to irreducibility, aperiodicity and smallness of compact sets:

(i): Irreducibility with respect to μ follows by Meyn and Tweedie (1993, Proposition 4.2.1 (ii)) by simply noting that

$$\textstyle\sum_{n=1}^{\infty} P^n\left(A|x\right) \geq P^k\left(A|x\right) = \int_A f\left(y|x\right)dy > 0$$

by Assumption 1 and the result follows.

(ii): An irreducible chain is *periodic* if it has period $d > 1$ and *aperiodic* if it has period $d = 1$. If it has period $d > 1$,then by Meyn and Tweedie (1993, Theorem 5.4.4)) there exists disjoint sets $D_0, D_1, ..., D_{d-1}$ in $\mathbb{B}^p$ such that

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

and furthermore

$$\psi(\textstyle\bigcup_{i=1}^{d} D_{i-1})^c = 0,$$

where ψ is a *maximal irreducibility measure.* Now, by Meyn and Tweedie (1993, Proposition 4.2.2 (ii)) the Lebesgue measure μ is absolutely continuous with respect to ψ and therefore also

$$\mu(\textstyle\bigcup_{i=1}^{d} D_{i-1})^c = 0.$$

For this to hold at least one of the sets D_1, say, must have $\mu(D_1) > 0$ which by Assumption 1 again implies $P^k\left(D_1|x\right) > 0$ for *all* x. Iterating k times one gets for some j the contradiction,

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

Hence the chain has period $d = 1$ and is therefore *aperiodic.*

(iii): If C is a compact set, $f\left(\cdot|\cdot\right)$ attains its minimum on $C \times C$ which is strictly positive since $f > 0$. In other words,

$$f\left(y|x\right) \geq \delta$$

for some $\delta > 0$ and $(x, y) \in C \times C$. For any $x \in C$ and any $A \in \mathbb{B}^p$,

$$P^k(A \mid x) \geq P^k(A \cap C \mid x) = \textstyle\int_{A \cap C} f(y|x)dy \geq \delta\mu\left(A \cap C\right).$$

Hence for all $x \in C$, $P^k(\cdot \mid x)$ is minorized by $\mu(\cdot \cap C\,)$ and therefore C is by definition small, cf. Meyn and Tweedie (1993, p. 106). $\square$

A.2 Proof of Theorem 1

Proof of Theorem 1: By Lemma 1 $\mathbf{X}_t$ 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 the choice of drift function we use

$$g(x) = 1 + x'Vx \geq 1, \quad V = \textstyle\sum_0^\infty \mathbf{A}^{i'}\mathbf{A}^i$$

corresponding to the second order moment of $\mathbf{X}_t$. The drift function is well-defined by Assumption 2 as $\rho(\mathbf{A} \otimes \mathbf{A}) < 1$. An equivalent choice of drift function appears in Feigin and Tweedie (1985) for the analysis of the class of random coefficients autoregressive processes (which does not include the ACR). It follows that

$$\begin{aligned} E(g(\mathbf{X}_t)|\mathbf{X}_{t-1} = x) &= 1 + tr\{V\Omega\} + p_x x'\mathbf{A}'V\mathbf{A}x + (1-p_x)x'\mathbf{B}'V\mathbf{B}x \\ &= 1 + tr\{V\Omega\} + x'\mathbf{A}'V\mathbf{A}x + (1-p_x)h(x), \end{aligned} \tag{23}$$

where

$$h(x) = x'(\mathbf{B}'V\mathbf{B} - \mathbf{A}'V\mathbf{A})x.$$

It follows that,

$$\begin{aligned} E(g(\mathbf{X}_t)|\,\mathbf{X}_{t-1} = x) &= 1 + tr(\Omega V) + x'\mathbf{A}'V\mathbf{A}x + (1-p_x)h(x) \\ &= \left\{\frac{1 + \varepsilon + tr(\Omega V) + x'Vx - x'x + (1-p_x)h(x)}{g(x)}\right\} g(x) \\ &= \left\{1 - \frac{x'x - tr(\Omega V) - (1-p_x)h(x)}{g(x)}\right\} g(x). \end{aligned}$$

Next, define for some $\lambda_c > 0$ the compact set

$$C = \left\{\, x \in \mathbb{R}^p |\ \ x'Vx \leq \gamma_c \right\}.$$

On C^c it holds by definition that

$$g(x) = 1 + x'Vx \leq x'Vx\left(1 + \tfrac{1}{\gamma_c}\right) \leq 2x'Vx$$

and therefore for γ_c large enough,

$$\begin{aligned} \tfrac{x'x - tr(\Omega V) - (1-p_x)h(x)}{1+x'Vx} &\geq \tfrac{x'x}{2x'Vx} - \frac{tr(\Omega V) + (1-p_x)h(x)}{1+x'Vx} \\ &\geq \tfrac{1}{2\rho(V)} - \frac{tr(\Omega V) + (1-p_x)h(x)}{1+x'Vx} \geq \eta > 0 \end{aligned}$$

using Assumption 3 which implies

$$\frac{tr(\Omega V) + (1-p_x)h(x)}{1+x'Vx} \to 0$$

as $\|x\|^2 \to \infty$. In short,

$$E(g(\mathbf{X}_t)|\,\mathbf{X}_{t-1}=x) \le (1-\eta)g(x)$$

for $x \in C^c$. On C, $E(g(\mathbf{X}_t)|\,\mathbf{X}_{t-1}=x)$ given by (23) which is continuous and hence bounded on the compact set. $\square$

A.3 Proof of Corollary

Proof of Corollary 1: The process $\mathbf{X}_t$ is a mk-dimensional ACR(1) Markov chain for which the k-step transition density is continuous and positive. To see this note that the density of $\mathbf{X}_{t+k}$ conditional on $\mathbf{X}_t$ is given by,

$$\begin{aligned} f(\mathbf{X}_{t+k}|\mathbf{X}_t) &= f\left(X_{t+k},\ldots,X_{t+1}|X_t,\ldots,X_{t-k+1}\right) = \frac{f\left(X_{t+k},\ldots,X_t,\ldots,X_{t-k+1}\right)}{f\left(X_t,\ldots,X_{t-k+1}\right)} \\ &= \textstyle\prod_{i=1}^k f\left(X_{t+i}|X_{t-1+i},\ldots,X_{t-k+i}\right), \end{aligned}$$

where the last line follows by k-dependence in X_t. Hence by definition $f(\mathbf{X}_{t+k}|\mathbf{X}_t)$ as each term in the product is continuous and positive and Theorem 1 gives the result.

A.4 Proof of Theorem 2:

Proof of Theorem 2: Apart from modifications, the proof structure is similar to the proof of Theorem 1 above and the proof of Feigin and Tweedie (1985, Theorem 5). We illustrate the proof for $m=2$ or fourth order moments. Define

$$\bar{\mathbf{X}}_t = (\mathbf{X}_t \otimes \mathbf{X}_t),\ \bar{\mathbf{A}} = (\mathbf{A}\otimes\mathbf{A}) \text{ and likewise } \bar{\mathbf{B}} = (\mathbf{B}\otimes\mathbf{B}).$$

Next note that under Assumption 2, $\rho(\bar{\mathbf{A}}\otimes\bar{\mathbf{A}}) < 1$ and therefore the $p^2\times p^2$ positive definite matrix $\bar{V}$ as well as the drift function $\bar{g}$ are well-defined, where

$$\bar{V} = \textstyle\sum_{i=0}^{\infty}\bar{\mathbf{A}}^{i\prime}\bar{\mathbf{A}}^{i} \text{ and } \bar{g}(x) = 1+\bar{x}'\bar{V}\bar{x} \text{ with } \bar{x} = (x\otimes x).$$

Similarly to the proof of Theorem 1, one finds

$$E(\bar{g}(\mathbf{X}_t)|\mathbf{X}_{t-1}=x) = 1+\bar{x}'\bar{\mathbf{A}}'\bar{V}\bar{\mathbf{A}}\bar{x} + K(\mathbf{A},\bar{V},x) + (1-p_x)\bar{h}(x). \tag{24}$$

Here

$$\bar{h}(x) = \bar{x}'\{\bar{\mathbf{B}}'\bar{V}\bar{\mathbf{B}} - \bar{\mathbf{A}}'\bar{V}\bar{\mathbf{A}}\}\bar{x} + K(\mathbf{B},\bar{V},x) - K(\mathbf{A},\bar{V},x) = O(\|x\|^4)$$

and $K(\mathbf{M},\bar{V},x)$ for $M=\mathbf{A},\mathbf{B}$ is the sum of terms which apart from x, $\mathbf{M}$ and $\bar{V}$ involves 4^{th} order moments of ε_t, 3^{rd} order moments (4 terms) and second order moments (6 terms). Similar to Feigin and Tweedie (1985, p.13) it follows that

$$\begin{aligned} &K(\mathbf{M},\bar{V},x) \le \qquad (25) \\ &\rho(\bar{V})\left\{6\rho(\mathbf{M}'\mathbf{M})tr(\Omega)\left\|\bar{x}\right\| + 4\rho(\mathbf{M}'\mathbf{M})^{1/2}\left\|\bar{x}\right\|^{1/2}E\left\|\varepsilon_t\right\|^3 + E\left\|\varepsilon_t\right\|^4\right\} = O(\|x\|^2). \end{aligned}$$

Now,

$$E(\,g(\mathbf{X}_t)|\,\mathbf{X}_{t-1}=x) = 1+\bar{x}'\bar{\mathbf{A}}'\bar{V}\bar{\mathbf{A}}\bar{x}+K(\mathbf{A},\bar{V},x)+(1-p_x)\bar{h}(x)$$
$$= \left\{1-\tfrac{\bar{x}'\bar{x}-K(\mathbf{A},\bar{V},x)-(1-p_x)\bar{h}(x)}{1+\bar{x}'\bar{V}\bar{x}}\right\}\bar{g}(x)\leq(1-\eta)\bar{g}(x)$$

for $\eta>0$, provided that for some constants κ_i, using (25),

$$\|\bar{x}\|^2\geq\rho(\bar{V})\left\{\eta\,\|\bar{x}\|^2+\kappa_1\,\|\bar{x}\|+\kappa_2\,\|\bar{x}\|^{1/2}+\kappa_3\right\}$$

which holds if $\eta\rho(\bar{V})<1$ and $\|\bar{x}\|^2$ large enough since $\bar{g}(x)=\|x\|^4$. The rest now follows as in the proof of Theorem 1. □

B Regularity conditions for the asymptotic inference

As noted the proof establishes standard Cramér or Wald type conditions which appear in various forms in the literature. We apply the formulation in Billingsley (1961) Theorems 2.1 and 2.2. The results therein are formulated for a Markov chain dependence structure which by Basawa, Feigin, and Heyde (1976) apply also to the k-dependence structure as in the ACR(k) model. The regularity Conditions 1.1 and 1.2 in Billingsley (1961) corresponds to (B1)-(B7) Basawa, Feigin, and Heyde (1976) and these we restate as Conditions 1 and 2 below for the current case.

Some notation is needed in order to handle derivatives of functions of matrices, see Magnus and Neudecker (1988) for a general introduction to matrix differential calculus. Consider the mapping G,

$$G:\mathbb{R}^{k\times l}\to\mathbb{R}^{m\times n}$$

where k,l,m and n are integers. Then G is differentiable of order three in $X\in\Xi\subset\mathbb{R}^{k\times l}$ if

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

as $\|H\|\to 0$. Here, say, $dG(X,H)$ is the differential of G at X with increment H and $X+H$ is in the interior of the set Ξ. To ease the presentation we use the notation $dG(X,dX)$ below. Although we shall *not* use this here, we note that alternatively one can work with the Jacobian,

$$\frac{\partial}{\partial vec(X)}vec\left\{G(X)\right\}$$

and that the differential and Jacobian are connected through the *vec*-operator by the identity,

$$vec\left\{dG(X,H)\right\}=\left[\frac{\partial vec\left\{G(X)\right\}}{\partial\left\{vec(X)\right\}'}\right]'vec(H)$$

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

B.1 First regularity condition

Condition 1 *For the transition density $f_\theta(Y|\mathbf{X})$*

(i): For all X the set of Y for which $f_\theta(Y|\mathbf{X}) > 0$ does not depend on θ.

(ii): For all $Y, \mathbf{X}$ the log-*likelihood $\ell_t(\theta) = \log f_\theta$ is well-defined except for a set of measure zero with respect to $f_\theta(Y|X)$. Also f_θ is three times continuously differentiable in θ.*

(iii): $Ed\ell_t(\theta, d\theta)^2 < \infty$ and

$$E\left(d\ell_t(\theta, d\theta)^2\right) = -E\left(d^2\ell_t(\theta, d\theta, d\theta)\right) > 0 \tag{26}$$

for all $d\theta$.

(iv): For each θ there exists a neighbourhood $N(\theta)$ of θ such that

$$E \sup_{\tilde{\theta}\in N(\theta)} \left|d^3\ell_t(\theta, d\theta, d\theta, d\theta)\right| < \infty \tag{27}$$

Together the Lemmas 2, 3 and 4 in the following show that Condition 1 applies to the ACR model.

Lemma 2 *With $p_{x_{t-1}}$ on the logistic form in (15) the first order differential for the model in (9) is given by*

$$d\ell_t(\theta, d\theta) = \left(p^*_{x_t} - p_{x_{t-1}}\right) d\lambda(\theta, d\theta) + \left\{p^*_{x_t} d\log\phi^A_t(\theta, d\theta) + \left(1 - p^*_{x_t}\right) d\log\phi^B_t(\theta, d\theta)\right\} \tag{28}$$

such that

$$\begin{aligned}
d\ell_t(\theta, dA) &= p^*_{x_t} tr\left(\Omega^{-1} e_{At}\mathbf{X}'_{t-1} dA'\right) \\
d\ell_t(\theta, dB) &= (1 - p^*_{x_t}) tr\left(\Omega^{-1} e_{Bt}\mathbf{X}'_{t-1} dB'\right) \\
d\ell_t\left(\theta, d(\alpha, \beta)'\right) &= \left(p^*_{x_t} - p_{x_{t-1}}\right) d(\alpha, \beta) z_t \\
d\ell_t(\theta, d\Omega) &= -\tfrac{1}{2} tr\left(\Omega^{-1} d\Omega\right) + \tfrac{1}{2} tr\left[\left\{p^*_{x_t} e_{At}e'_{At} + (1 - p^*_{x_t}) e_{Bt}e'_{Bt}\right\}\Omega^{-1} d\Omega\Omega^{-1}\right].
\end{aligned}$$

Here $p^*_{x_t}$ is defined in (20), ϕ^A_t, ϕ^B_t are given by (18) and finally,

$$e_{Mt} = X_t - M\mathbf{X}_{t-1} \text{ for } M = A, B \tag{29}$$

$$z_t = \left(1, \|\mathbf{X}_{t-1}\|^2\right)'. \tag{30}$$

Proof of Lemma 2: The result follows by direct differentiation of the log likelihood function in (16) combined with the identity (20). □

Next turn to the second order differential.

Lemma 3 *Under the assumptions of Theorem 3 then*

$$
\begin{aligned}
&d^2\ell_t(\theta, d\theta, d\theta) \qquad (31)\\
&= \left\{p^*_{x_t} d^2 \log \phi^A_t(\theta, d\theta, d\theta) + \left(1 - p^*_{x_{t-1}}\right) d^2 \log \phi^B_t(\theta, d\theta, d\theta)\right\} - \left(1 - p_{x_{t-1}}\right) p_{x_{t-1}} \left\{d\lambda(\theta, d\theta)\right\}^2\\
&+ \left(1 - p^*_{x_{t-1}}\right) p^*_{x_{t-1}} \left\{d\lambda(\theta, d\theta) + d\log \phi^A_t(\theta, d\theta) - d\log \phi^B_t(\theta, d\theta)\right\}^2.
\end{aligned}
$$

Note that,

$$
d\log \phi^A_t = tr\left(\Omega^{-1} e_{At}\mathbf{X}'_{t-1} dA'\right) \qquad (32)
$$

$$
d^2 \log \phi^A_t(\theta, dA, dA) = tr\left(\Omega^{-1} dA\mathbf{X}_{t-1}\mathbf{X}'_{t-1} dA'\right) \qquad (33)
$$

such that in particular,

$$
\begin{aligned}
d^2\ell_t(\theta, dA, dA) &= -p^*_{x_t} tr\left(\Omega^{-1} dA\mathbf{X}_{t-1}\mathbf{X}'_{t-1} dA'\right) \qquad (34)\\
&\quad + (1 - p^*_{x_t})p^*_{x_t}\left\{tr\left(\Omega^{-1} e_{At}\mathbf{X}'_{t-1} dA'\right)\right\}^2\\
d^2\ell_t(\theta, d(\alpha, \beta)', d(\alpha, \beta)) &= \left\{(1 - p^*_{x_t})p^*_{x_t} - (1 - p_{x_{t-1}})p_{x_{t-1}}\right\}\left\{d(\alpha, \beta) z_t\right\}^2 \qquad (35)\\
d^2\ell_t(\theta, d\Omega, d\Omega) &= tr\left\{\left(\tfrac{1}{2} - \left[p^*_{x_t} e_{At}e'_{At} + (1 - p^*_{x_t})e_{Bt}e'_{Bt}\right]\Omega^{-1}\right)\left(\Omega^{-1} d\Omega\Omega^{-1} d\Omega\right)\right\}\\
&\quad + (1 - p^*_{x_t})p^*_{x_t}\left[\tfrac{1}{2} tr\left\{(e_{At}e'_{At} - e_{Bt}e'_{Bt})\Omega^{-1} d\Omega\Omega^{-1}\right\}\right]^2 \qquad (36)
\end{aligned}
$$

Moreover, (26) in Condition 1 holds.

Note that the expressions for the score and 'information' can alternatively be derived by using the EM algorithm and treating s_t as unobserved, see Louis (1982) and Ruud (1991).

Proof of Lemma 3: It follows directly that

$$
\begin{aligned}
d^2\ell_t(\theta, d\theta, d\theta) &= p^*_{x_t} d^2 \log \phi^A_t(\theta, d\theta, d\theta) + \left(1 - p^*_{x_t}\right) d\log \phi^B_t(\theta, d\theta, d\theta)\\
&\quad + \left(p^*_{x_t} - p_{x_{t-1}}\right) d^2\lambda(\theta, d\theta, d\theta) + \left[dp^*_{x_t}(\theta, d\theta) - dp_{x_{t-1}}(\theta, d\theta)\right] d\lambda(\theta, d\theta)\\
&\quad + dp^*_{x_t}(\theta, d\theta)\left(d\phi^A_t(\theta, d\theta) - d\phi^B_t(\theta, d\theta)\right)
\end{aligned}
$$

which equals (31) using the identities

$$
\begin{aligned}
dp_{x_{t-1}}(\theta, d\theta) &= \left(1 - p_{x_{t-1}}\right) p_{x_{t-1}} d\lambda(\theta, d\theta)\\
dp^*_{x_t}(\theta, d\theta) &= \left(1 - p^*_{x_t}\right) p^*_{x_t} d\log\left(\frac{p^*_{x_t}}{1 - p^*_{x_t}}\right)\\
&= \left(1 - p^*_{x_t}\right) p^*_{x_t}\left(d\lambda(\theta, d\theta) + d\log \phi^A_t(\theta, d\theta) - d\log \phi^B_t(\theta, d\theta)\right)
\end{aligned}
$$

and the fact that $d^2\lambda(\theta, d\theta, d\theta) = 0$. In particular, we find (34)-(36) by using (32) and (33) as well as standard matrix calculus results such as $d\log|\Omega| = tr\left\{\Omega^{-1} d\Omega\right\}$.

Next 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 $p \times p$ matrices dA we use the conditional independence of s_t and ε_t given $\mathbf{X}_{t-1}$: First note that

$$s_t e_{At} = s_t(X_t - A\mathbf{X}_{t-1}) = s_t\left\{(s_t - 1)[A - B]\mathbf{X}_{t-1} + \varepsilon_t\right\} = s_t\varepsilon_t$$

and hence using (20),

$$E\left(p^*_{x_t} e_{At}\middle|\, \mathbf{X}_{t-1}\right) = E\left(E\left(s_t\middle|\, X_t, \mathbf{X}_{t-1}\right) e_{At}\middle|\, \mathbf{X}_{t-1}\right) = E\left(s_t\varepsilon_t\middle|\, \mathbf{X}_{t-1}\right) = 0. \tag{37}$$

Recall that by definition

$$E\left(p^*_{x_t}\middle|\, \mathbf{X}_{t-1}\right) = E\left(s_t\middle|\, \mathbf{X}_{t-1}\right) = p_{x_{t-1}}. \tag{38}$$

Therefore

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

and it holds that

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

as desired. Likewise for the remaining terms in (26) the results follow by repeated use of the additional identities

$$(1 - s_t)(e_{Bt} - \varepsilon_t) = 0 \tag{39}$$

$$E\left((1 - p^*_{x_t})e_{Bt}\middle|\, \mathbf{X}_{t-1}\right) = 0 \tag{40}$$

$$E\left(p^*_{x_t}e_{At}e'_{At} + (1 - p^*_{x_t})e_{Bt}e'_{Bt}\middle|\, \mathbf{X}_{t-1}\right) = s_t\Omega + (1 - s_t)\Omega = \Omega \tag{41}$$

$$Cov\left(tr\left\{\varepsilon_t\varepsilon'_t P\right\}, tr\left\{\varepsilon_t\varepsilon'_t Q\right\}\right) = 2tr\left\{P\Omega Q\Omega\right\} \tag{42}$$

for P, Q symmetric $p \times p$ matrices. For instance, using (42) together with (41) and (40) 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,

$$d\ell_t(\theta, dA) + d\ell_t(\theta, dB) + d\ell_t(\theta, d(\alpha, \beta)' + d\ell_t(\theta, d\Omega) = 0$$

implies $dA = dB = d(\alpha, \beta) = d\Omega = 0$. Note initially that by the definition of $p^*_{x_t}$ in (20) then

$$p^*_{x_t} - p_{x_{t-1}} = p_{x_t}(1 - p_{x_t})(\phi^A_t - \phi^B_t) \tag{43}$$

Thus if $A = B$ then by (43) $p^*_{x_t} = p_{x_{t-1}}$and the claimed implication does not hold. More precisely, conditioning on $\mathbf{X}_{t-1}$ 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 $\mathbf{X}_{t-1}$, and with α and β 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 $(\alpha, \beta)'$,

$$\begin{aligned} d\ell_t(\theta, d(\alpha,\beta)' &= d\ell_t\left(\theta, d(\alpha,\beta)'\right) = \left(p^*_{x_t} - p_{x_{t-1}}\right) d(\alpha,\beta) z_t \\ &= \left(p^*_{x_t} - p_{x_{t-1}}\right)\left(d\alpha + \|\mathbf{X}_{t-1}\|^2 d\beta\right) \end{aligned}$$

By (43) and Assumption 3, $\left(p^*_{x_t} - p_{x_{t-1}}\right) \neq 0$ almost surely (as $\beta > 0$). Next, the proof of geometric ergodicity of $\mathbf{X}_t$ implies that the Markov chain has the Lebesgue measure as a 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,

$$\Pr\left(\|\mathbf{X}_{t-1}\|^2 \neq \text{ constant}\right) > 0.$$

and therefore $d\ell_t(\theta, d(\alpha,\beta)' \neq 0$ almost surely. $\square$

Lemma 4 *Under the assumptions of Theorem 3 then (27) in Condition 1 holds.*

Proof of Lemma 4: The result is shown by using Lemma 3 and noting that with

$$v^M_t = d\log\phi^M_t = tr\left\{\Omega^{-1}\left[X_t\mathbf{X}'_{t-1} - M\mathbf{X}_{t-1}\mathbf{X}'_{t-1}\right]dM\right\} \tag{44}$$

for $M = A, B$, cf. (32), then

$$\left|v^M_t\right| \leq \kappa_1\left\|\mathbf{X}_{t-1}X'_t\right\| + \kappa_2\left\|\mathbf{X}_{t-1}\mathbf{X}'_{t-1}\right\|$$

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

$$\begin{aligned} &\left|d^3\ell_t(\theta, dA, dA, dA)\right| \\ &= \left|(p^*_{x_t} - 1)p^*_{x_t} v^A_t tr\left\{\Omega^{-1}dA\mathbf{X}_{t-1}\mathbf{X}'_{t-1}dA'\right\} + (1 - 2p^*_{x_t})p^*_{x_t}\left(v^A_t\right)^3\right| \\ &\leq \tilde{\kappa}_1\left|v^A_t\right|\|X_t\|^{2k} + \tilde{\kappa}_2\left|v^A_t\right|^3 \end{aligned}$$

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 all moments of X_t. Apart from tedious calculus similar results hold for the remaining third order differentials. $\square$

B.2 Second regularity condition

Condition 2 *Independently of choice of initial distribution and as $T \to \infty$:*
(v): Provided $\phi(\cdot,\cdot)$ is measurable and $E \left\| \phi\left(X_t, \mathbf{X}_{t-1}\right) \right\| < \infty$, then for each θ

$$\tfrac{1}{T}\Sigma_{t=1}^{T}\phi\left(X_t, \mathbf{X}_{t-1}\right) \overset{P}{\to} E\phi\left(X_t, \mathbf{X}_{t-1}\right).$$

(vi): Furthermore,

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

where $E\left[d\ell_t(\theta, d\theta)\right]^2$ satisfies (26).

Lemma 5 *Condition 2 holds for the ACR model under the assumptions of Theorem 3.*

Proof of Lemma 5: Note that $d\ell_t(\theta, d\theta)$ is a Martingale difference sequence with respect to $\mathcal{F}_t = \sigma\left(X_t, X_{t-1}, ...\right)$. Specifically,

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

using the expression for the differentials in Lemma 2 together with the identities (37), (38), (40) and (41) applied in the proof of Lemma 3. The established geometric ergodicity and existence of moments imply that

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

converges in probability by the law of large numbers in Meyn and Tweedie (1993, Theorem 17.0.1). Furthermore, the Lindeberg condition in Brown (1971) applies and the claimed asymptotic normality of the first order differential follows by Brown (1971). The chain defined by $\tilde{X}_t = (X_t, X_{t-1})$, $t = 1, 2, ..$ is geometrically ergodic and Condition 2 *(v)* using follows again by Meyn and Tweedie (1993, Theorem 17.0.1). $\square$

References

An, H., M. Chen, and F. Huang (1997). The geometric ergodicity and existence of moments for a class of non-linear time series models. *Statistics and Probability Letters 31*, 213–224.

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

Basawa, I. V., P. D. Feigin, and C. C. Heyde (1976). Asymptotic properties of maximum likelihood estimators for stochastic processes. *Sankhya: The Indian Journal of Statistics, Series A 38*, 259–270.

Baum, L. E., T. Petrie, G. Soules, and N. Weiss (1970). A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics 41*, 164–171.

Bec, F., M. Ben Salem, and M. Carrasco (2001). Tests for unit-root versus threshold specification with an application to the PPP. Unpublished paper: CREST-ENSAE, Universite de Poitiers and University of Rochester.

Billingsley, P. (1961). *Statistical Inference for Markov Processes*. Chicago: Chicago University Press.

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.

Chan, K. S. and H. Tong (1985). On the use of deterministic Lyaponov function for the ergodicity of stochastic differential equations. *Advances in Applied Probability 17*, 666–678.

Cox, D. R. (1981). Statistical analysis of time series: some recent developments. *Scand. J. Statist. 8*, 93–115.

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.

Diebold, F. X., J.-H. Lee, and G. C. Weinbach (1994). Regime switching with time-varying transition probabilities. In C. Hargreaves (Ed.), *Nonstationary Time Series Analysis and Cointegration*, pp. 283–302. Oxford: Oxford University Press.

Doornik, J. A. (2001). *Ox: Object Oriented Matrix Programming, 3.0.* London: Timberlake Consultants Press.

Durland, J. M. and T. H. McCurdy (1994). Duration dependence in a Markov model of U.S. GDP growth. *Journal of Business and Economic Statistics 12*, 279–88.

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.

Engle, R. F. and C. W. J. Granger (1987). Co-integration and error correction: representation, estimation and testing. *Econometrica 55*, 251–276.

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*, Chapter 20, 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.

Frydman, R. and M. Goldberg (2002). Imperfect knowledge expectations, uncertainty adjusted UIP and exchange rate dynamics. In P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford (Eds.), *Knowledge, Information and Expectations in Modern Macroeconomics: In Honour of Edmund S Phelps*. Princeton: Princeton University Press. forthcoming.

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 (2001). Tails and extremal behaviour of stochastic unit root models. Unpublished paper: CREST, CEPREMAP, University of Toronto and University Paris VII.

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. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*, 357–384.

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

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.

Hansen, E. and A. Rahbek (1998). Stationarity and asymptotics of multivariate ARCH time series with an application to robustness and cointegration analysis. Unpublished paper: preprint number 12, Department of Theoretical Statistics.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge: Cambridge University Press.

Hendry, D. F. (1995). *Dynamic Econometrics.* Oxford: Oxford University Press.

Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Auto-regressive Models.* Oxford: Oxford University Press.

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.

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.

Louis, T. A. (1982). Finding observed information using the EM algorithm. *Journal of the Royal Statistical Society, Series B 44*, 98–103.

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.

Nielsen, B. and A. Rahbek (2000). Similiarity issues in cointegrated models. *Oxford Bulletin of Economics and Statistics 62*, 5–22.

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

Taylor, M. (1995). Economics of exchange rates. *Journal of Economic Literature 33*, 13–47.

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

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

Tsay, R. S. (1989). Testing and modelling threshold autoregressive processes. *Journal of the American Statistical Association 84*, 231–240.

Tsay, R. S. (1998). Testing and modeling multivariate threshold models. *Journal of the American Statistical Association 93*, 1188–1202.

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