Autoregressive conditional root model

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

NEIL SHEPHARD
Nuffield College, Oxford OX1 1NF, UK
neil.shephard@nuf.ox.ac.uk
www.nuff.ox.ac.uk/users/shephard/

First draft April 2001
This version February 2002.

Abstract

In this paper we develop a 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. This autoregressive root model is shown to be ergodic and covariance stationary under some rather general conditions. We study how this model can be estimated and tested, developing appropriate asymptotic theory for this task. Finally we apply the model to assess the purchasing power parity relationship.

Keywords: Cointegration; Equilibrium correction model; GARCH; Hidden Markov model; Likelihood; Regime switching; STAR model; Stochastic break; Stochastic unit root; Switching regression; Real Exchange Rate; PPP; Unit root hypothesis.

1 Introduction

1.1 The model

Much of macroeconomic theory is concerned with long term relationships between variables. Examples of this include the quantity theory of money and purchasing power parity (PPP). When analysing non-stationary processes, modern econometrics formalises this using the concept of cointegration as in Engle and Granger (1987). Cointegration allows only short term deviations from long term relationships by imposing stationarity on the transitory disequilibriums and the present paper follows that classic cointegration tradition. However, we introduce a non-linear time series model which allow the disequilibriums to have *epochs* of true non-stationarity, giving the impression that the long term relationships have temporarily broken down, before they *endogenously* collapse back towards their long term relationship. The collapses regularise the periods of non-stationarity forcing the disequilibrium to be globally stationary. This type of behaviour is reflected in an economic theory model developed by Bec, 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.

We formalise the modelling of the time series of transitory disequilibriums y_t by a combination of non-stationarity and collapse. Our analysis will be based around a first order autoregression whose root switches endogenously and stochastically between being exactly unity and being stable. This can be thought of as a softening of threshold autoregressive models, which are often used in economics, where the autoregressive parameter is a deterministic function of past data. We call this model an *autoregressive conditional root* (ACR) model.

In its simplest form corresponding to an autoregression of order one, the ACR model is given by the equation

$$y_t = \rho^{s_t} y_{t-1} + \varepsilon_t, \quad t = 1, 2, \ldots, T, \tag{1}$$

where s_t is binary, ρ is a real number and ε_t is an *i.i.d.*N(0,σ^2) sequence. Alternatively this model can be reparametrized as an equilibrium correction model (ECM)

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

where Δ is the difference operator.

With $\mathcal{F}_t = \sigma(y_t, y_{t-1}, \ldots, y_0)$ the information up to time t we write the *prediction probability*

$$p_t = \Pr(s_t = 1 | \mathcal{F}_{t-1}, \varepsilon_t) = p(y_{t-1}, \ldots, y_{t-p}), \tag{3}$$

where the notation $p(y_{t-1}, \ldots, y_{t-p})$ reflects that p_t depends only on $y_{t-1}, \ldots, y_{t-p}$. Note that by assumption s_t and ε_t are independent conditionally on $\mathcal{F}_{t-1}$. We will allow $p(\cdot)$ to be unconstrained except that $p(\cdot)$ is measurable with respect to $\mathcal{F}_{t-1}$ and is bounded awap from zero.

We show that an initial distribution exists such that y_t is strictly stationary and possess all of its moments provided $\rho \in (-1, 1)$ or $\pi \in (-2, 0)$ in (1) and (2) respectively. No other condition is needed apart from the mentioned boundedness of the probability $p(\cdot)$. This means that the model can have epochs of non-stationarity, but is globally stable. Estimation is straightforward as for instance the likelihood function can be computed via a prediction decomposition. We argue that inference is regular within the parameter region just outline above while we briefly mention aspects of inference in the case of y_t being a random walk for the entire period.

In this model 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 an autoregression of order one. Note that in the ACR model the dynamics of the regime are determined entirely endogenously similar to threshold models with an endogenous threshold variable as applied in Enders and Granger (1998). However, now the threshold is actually stochastic rather than deterministic. In a recent paper, written independently and concurrently from our paper, Gourieroux and Robert (2001) have studied in detail the above model in the case where there is switching between white noise and a random walk (ie. the case of the above model when $\rho = 0$) with $p_t = p(y_{t-1})$. 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 y_t, the distribution of epochs of non-stationary behaviour and shows that y_t is geometrically ergodic. Our analysis will be complementary, focusing on estimation and asymptotic inference for use in empirical work.

The following simple examples allows us to gain a better understanding of the behaviour of this process. The first is the Markov case which is written in terms of the logistic transform

$$\begin{aligned} \lambda_t &= \log \{p_t / (1 - p_t)\} && (4) \\ &= \lambda(y_{t-1}) && (5) \\ &= \alpha + \beta y_{t-1}^2, && (6) \end{aligned}$$

with α and β being freely varying reals. So long as β is non-negative and α and β are finite, λ_t will be bounded and so the process y_t in (1) will be stationary as demonstrated below.

Example 1 *We give in Figure 1 a sample path from the simplest Markov ACR processes. In the Markov model (6) the parameter values will be taken to be $\alpha = -100$, $\beta = 1.1$, $\rho = 0.9$, and $\sigma = 0.8$. This process delivers a jagged realisations for p_t, which never spend substantial consecutive periods close to one. This is enough however to for the y_t series to be stable, never going much above ten in absolute value.*

1.2 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). These 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.

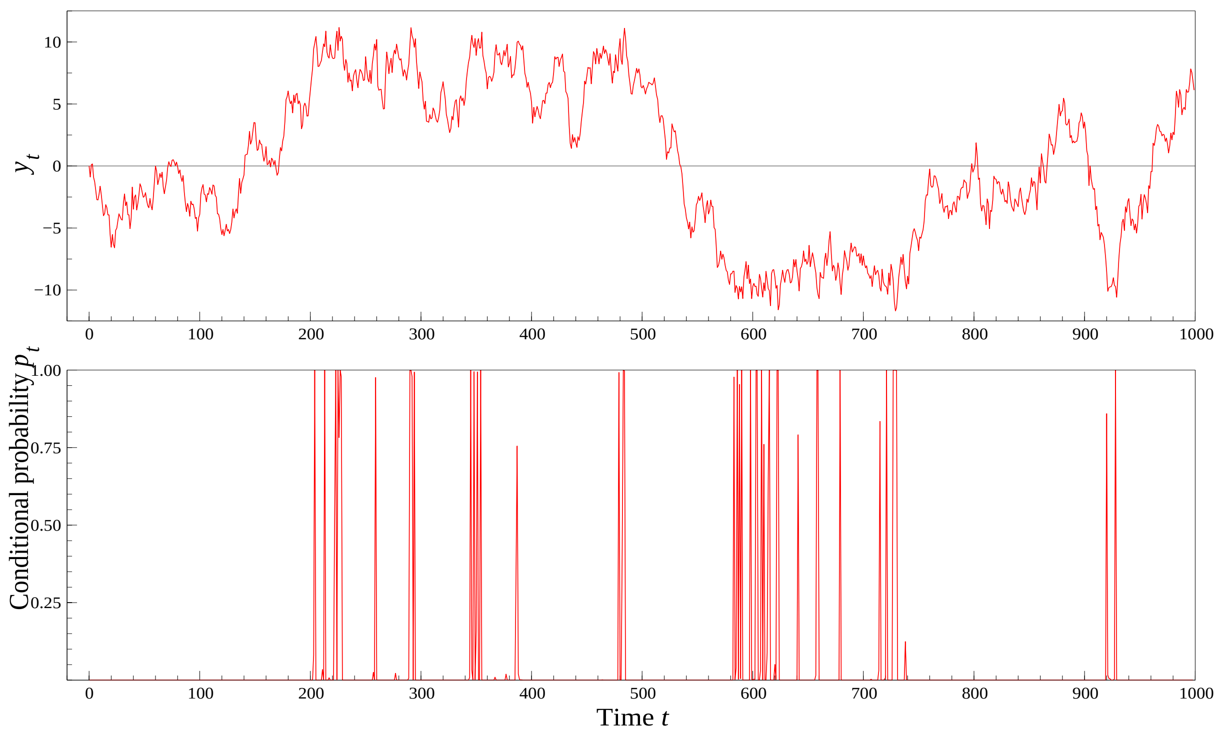

Figure 1: *Simulations from the Markov model, with corresponding conditional probabilisties* p_t *given below. Code: regime_sim.ox.*

In the Markov switching literature in economics, following Hamilton (1989), the regime is regarded as a latent variable and follows a Markov process governed by $\Pr(s_t|s_{t-1})$. It 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 y_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 y_t process. A simple example of this is given by defining $\lambda_t = \alpha + \beta y_{t-1}$ in (6). This is outside our structure as it does not bound λ_t 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 $\lambda_t = \alpha + \beta y_{t-1}$ was explicitly studied recently by Wong and Li (2001), although its stochastic properties were not derived. Of course this can be generalised to allow λ_t to depend upon many lags of y_t or other potentially helpful explanatory variables.

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

$$\mathrm{E}(\Delta y_t | y_{t-1}) = \pi p_t y_{t-1}.$$

Suppose we again define $\lambda_t = \alpha + \beta y_{t-1}^2$ then

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

If we reinvent this model as

$$\Delta y_t = \pi \frac{\exp\left(\alpha + \beta y_{t-1}^2\right)}{1 + \exp\left(\alpha + \beta y_{t-1}^2\right)} y_{t-1} + \varepsilon_t,$$

then this is a simple 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. 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 features of the above setup. They write, in their simplest model

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

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.

1.3 Outline of the paper

This paper has three other main sections. In Section 2 we derive the stochastic properties of the ACR model, deriving the conditions needed for the model to be strictly stationary. In Section 3 the likelihood function for the model is derived and we give conditions under which the maximum likelihood estimator is asymptotically normally distributed. We also discuss the use of various testing procedures to look at special cases of the model structure. In Section 4 of the paper we illustrate the model on simulated and real data. The paper has two other sections. Section 5 dicsusses possible extensions and finally Section 6 concludes the paper, while the Appendix proves theorems stated in the paper.

2 Stability of the ACR(p) process

In this section we formalise the discussion given above and study some stochastic properties of the ACR process. In the first stage of this discussion we will deal with models with the prediction probability p_t is allowed to depend flexibly on p lagged values of the process $y_{t-1}, ..., y_{t-p}$.

We maintain the process as given by (1), (3) and (4) but assume that λ_t is some measurable function of y_{t-1},y_{t-2},...,y_{t-p}:

$$\lambda_t = \lambda\left(y_{t-1}, y_{t-2}, ..., y_{t-p}\right). \tag{7}$$

This is denoted a p-th order autoregressive conditional root, or ACR(p), process. The case of $p = 1$ is methodologically interesting for then the y_t process is a Markov chain while for the general case $(y_t, ..., y_{t-p})$ is a Markov chain.

Throughout we will assume λ satisfies:

Assumption 2. *$\lambda(\cdot)$ is continuous in x and there exists a constant $\gamma > 0$, such that $\lambda(x) \geq -\gamma > -\infty$ for all $x \in R^p$.*

It is important to remark that we could have equivalently written the ACR model directly in terms of p_t, flexibly parameterising it in terms of y_{t-1},y_{t-2},...,y_{t-p}. The use of the logistic transformation is not constraining, while Assumption 2 would imply that we are assuming that the p_t is always bounded away from zero. This is important, for it means that whatever state the process is in, there is always a non-negative probability that we will enter the locally stationary regime. For the specific case in (6) the assumption would imply that β should be non-negative. Finally, we should note that the assumption does not bound $\lambda(\cdot)$ from above, so we are not removing the possibility that the process will spend its entire time in the $s_t = 1$ regime.

For the ACR(p) process the following holds:

Theorem 1 *Under Assumption 2 then y_t in (1) with λ_t given by (7) is geometrically ergodic and has moments of any order if*

$$|\rho| = |\pi + 1| < 1.$$

In particular, y_t is stationary and $\beta-$mixing with exponential decay if initiated from the invariant distribution.

The proof is based on Markov chain theory and uses the concept of a drift function. It is located in the appendix.

The theorem has a number of important features. The theorem requires only very weak assumptions on $\lambda(\cdot)$ to continue to hold. This is sufficient to imply the process is stationary and ergodic. Further, all the moments of the process will exist which in particular implies, for example, that the process is also covariance stationary. Geometric ergodicity also has the implication that the law of large numbers as well as a central limit theorem holds irrespectively of the choice of initial distribution of the process, see Meyn and Tweedie (1993). These features are used in the next section where inference is considered.

Note that the tail properties of an ACR(1) type process is studied in Gourieroux and Robert (2001) in the case where $\rho = 0$, $\lambda_t = \lambda(y_{t-1})$ and The ε_t is an *i.i.d.* sequence with exponential type distributions.

3 Inference

3.1 Likelihood based estimation and testing

In this section we consider asymptotic inference in the basic ACR(1) model as defined by (1), (3) and (4). The focus is on hypotheses which leave the epochs or mixing structure intact. In addition, we briefly mention inference for hypotheses, such as the unit root hypothesis, which does not allow epochs of either mean-reversion or random-walk type behaviour.

3.2 Distribution of the ML estimator

The specific prediction probability $p_t(\cdot)$ of interest here is given by

$$p_t = \tfrac{\exp \lambda_t}{1+\exp \lambda_t}, \qquad \lambda_t = \alpha + \beta y_{t-1}^2 \tag{8}$$

in which case the parameters of the model ACR(1) are given by

$$\theta = (\pi, \alpha, \beta, \sigma^2)'$$

where $\sigma^2 > 0$ and π, α and β are freely varying. Conditional on the initial observation y_0 the log-likelihood function to be maximized is found from

$$\log f_\theta(y_1, ..., y_T|y_0) = \sum_{t=1}^{T} \log f_\theta(y_t|y_{t-1}) = \sum_{t=1}^{T} l_t(\theta) \tag{9}$$

with

$$l_t(\theta) = \log\left\{p_t\phi_t^\pi + (1-p_t)\phi_t^0\right\}.$$

Here the convenient notation

$$\phi_t^\pi = \tfrac{1}{\sqrt{\sigma^2}} \exp\left\{-\tfrac{1}{2\sigma^2}\left(\Delta y_t - \pi y_{t-1}\right)^2\right\} \tag{10}$$

has been used for the Gaussian density apart from a constant. The likelihood function in (9) is numerically maximised to obtain the maximum likelihood estimator, $\hat{\theta}$. Full expressions for the score and observed information are given in equations (18) and (22) of Appendix B, respectively.

In the following it is shown that provided π is in the interval $]-2,0[$ or equivalently ρ in (1) is smaller than one in absolute value, then $\hat{\theta}$ is (locally) consistent, asymptotically Gaussian and likelihood ratio tests for simple hypotheses on θ, $\theta = \theta_0$ are asymptotically χ^2 distributed. More precisely the following Theorem holds:

Theorem 2 *With the ACR(1) model defined by equations (1), (2), (3) and (4), then if $\pi \in]-2,0[$ ($|\rho| < 1$) there exists with probability tending to one as T tends to infinity a sequence of $\hat{\theta} = \left(\hat{\pi}, \hat{\alpha}, \hat{\beta}, \hat{\sigma}^2\right)$ which satisfies the score equation,*

$$\left.\frac{\partial l_t}{\partial \theta}(\theta)\right|_{\theta=\hat{\theta}} = 0$$

and is consistent,

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

Furthermore,

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

with

$$\Sigma = E\left(\frac{\partial l_t}{\partial \theta}(\theta)\frac{\partial l_t}{\partial \theta'}(\theta)\right) = -E\left(\frac{\partial^2 \ell_t(\theta)}{\partial \theta \partial \theta'}\right) > 0.$$

Finally, likelihood ratio tests for simple hypotheses on θ are asymptotically χ^2 distributed.

The proof is given in Appendix B and is an application of Billingsley (1961) Theorems 2.1 and 2.2 which hold under regularity Conditions 1.1. and 1.2 therein. However, while regularity Condition 1.1 remain unaltered the Condition 1.2 is modified based on Markov chain theory for geometrically ergodic processes. The regularity conditions which are used are stated as Conditions 5 and 6 respectively in Appendix B. Note also that Σ is consistently estimated by the observed information, see also Lemma 5 in Appendix B.

3.3 Testing for no epochs

Our main concern has been so far the class of hypotheses which leave the epochs or mixing structure intact under the null. However, two alternative kinds of null hypotheses are also of interest when analyzing data such as the PPP data briefly considered. The first is the null of y_t being a random walk process without any epochs of stationarity or mean-reverting behaviour and a test for this hypothesis may be viewed as a misspecification test, see also Bec, Salem, and Carrasco (2001). The second hypothesis is the null of y_t being a Gaussian autoregressive process, i.e. in this case a process without any epochs of random walk type behaviour. The null of a random walk is implied by the simple restriction,

$$\pi = 0.$$

On the other hand the null of an autoregressive process can be characterized in the reparametrized ACR(1) model with $\gamma := \exp(-\alpha)$ as

$$\gamma = 0.$$

However, even though both hypotheses are simple, in both cases the parameter β vanish under the null (as do α for the random walk hypothesis). Therefore the usual asymptotic expansion of the likelihood ratio statistic in terms of score and information is problematic similar to the situation discussed in Davies (1987) and Andrews and Ploberger (1994).

4 Empirical illustrations

In this section we will illustrate our analysis by applying our ACR model to purchasing power parity data analyzed in Bec, Salem, and Carrasco (2001) by a threshold autoregressive model. To carry this through we need to be able to numerically maximise the likelihood for our model and we develop a simple algorithm for this in this subsection, before going on to look at a class of conveniently parameterised models for λ_t. We explain why and discuss if our model has a better and a more meaningful fit than conventional linear models to the PPP series between the French Franc and Italian Lira. The final subsection will give other examples, based on a number of other series.

4.1 Numerical 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)). This regards the indicators $s_1, ..., s_n$ as missing data.

More precisely for the EM algorithm consider the likelihood function for $s = (s_1, ..., s_n)$ and $y = (y_1, ..., y_T)$ conditional on y_0 as given by

$$\begin{aligned}\log f(y,s) &= -\frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(\Delta y_t - \pi s_t y_{t-1}\right)^2 \\ &\quad + s_t \log\left\{\frac{p_t}{(1-p_t)}\right\} + \log(1-p_t).\end{aligned}$$

see also (9). Taking conditional expectation of this, conditioning on $\mathcal{F}_T = \sigma\left(y_T, y_{T-1}, \ldots, y_0\right)$ and using

$$E\left(s_t|\mathcal{F}_T\right) = p_t^* \qquad \text{and} \qquad E\left(s_t^2|\mathcal{F}_T\right) = p_t^*,$$

see Lemma 4 and equation (19), we immediately find

$$\begin{aligned}E\left\{\log f(y,s)|\mathcal{F}_T\right\} &= -\frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left\{(\Delta y_t)^2 - 2\pi p_t^* y_{t-1}\Delta y_t + \pi^2 p_t^* y_{t-1}^2\right\} \\ &\quad + \sum_{t=1}^{T}\left[p_t^*\log\left\{\frac{p_t}{(1-p_t)}\right\} + \log(1-p_t)\right].\end{aligned}$$

We maximise this with respect to the parameters.

Some of it is is in close form:

$$\widehat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\left\{(\Delta y_t)^2 - 2\widehat{\pi} p_t^* y_{t-1}\Delta y_t + \widehat{\pi}^2 p_t^* y_{t-1}^2\right\},$$

and

$$\widehat{\pi} = \sum_{t=1}^{T} p_t^* y_{t-1}\Delta y_t \Bigg/ \sum_{t=1}^{T} p_t^* y_{t-1}^2 .$$

The other parameters in the model effect only

$$\begin{aligned}L &= \sum_{t=1}^{T}\left[p_t^*\log\left\{\frac{p_t}{(1-p_t)}\right\} + \log(1-p_t)\right] && (11) \\ &= \sum_{t=1}^{T}\left[p_t^*\lambda_t - \log\left\{1+\exp\left(\lambda_t\right)\right\}\right]. && (12)\end{aligned}$$

This has to be optimised numerically. In cases where λ_t is a linear function of past data, such as in the pure autoregressive scheme

$$\lambda_t = \alpha + \beta g(y_{t-1}),$$

where $g(x) = x^2$ or some other known function, then (12) takes on the form of a logistic regression for the "observations" p_t^*. In this case this part of the likelihood function is concave, a result which extends to any dynamic model where λ_t 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.

4.2 Finite memory, smooth models

As well as fitting simple Markov type models for λ_t, we would like to fit models with smoother evolutions for the λ_t, depending not just on a last few values of the lagged dependent variable. Following the work on GARCH and ACD models by Bollerslev (1986) and Engle and Russell (1998), it would seem sensible to allow structures with, for example,

$$\lambda_t = \alpha + \beta g(y_{t-1}) + \gamma \lambda_{t-1}, \qquad 0 \leq \lambda < 1, \tag{13}$$

and $\beta > 0$. Although this is straightforward to fit, our theory does not directly cover this case for we have only proved stationarity of the model when

$$\lambda_t = \lambda(y_{t-1}, ..., y_{t-p}),$$

for any integer $p \geq 1$, a finite memory process. We have not been able to prove the stationarity result covers the (13) model, although our firm conjecture is that it holds so long as $0 \leq \gamma < 1$. As a result of this deficiency we use finite memory models of the form

$$\lambda_t = \alpha + \beta \sum_{j=1}^{p} \gamma^{j-1} g(y_{t-j}), \qquad 0 \leq \lambda < 1.$$

These models certainly fall inside the compass of the theory that we have proved. Of course we can rewrite this model in the computationally convenient form

$$\begin{aligned}
\lambda_t &= \alpha + \beta y_{t-1}^2 + \beta \sum_{j=1}^{p-1} \gamma^j g(y_{t-j-1}) \\
&= \alpha + \beta y_{t-1}^2 + \beta \sum_{j=1}^{p} \gamma^j g(y_{t-j-1}) - \beta \gamma^{p-1} g(y_{t-p}) \\
&= \alpha(1-\gamma) + \beta \left\{ g(y_{t-1}) - \gamma^{p-1} g(y_{t-p}) \right\} + \alpha\gamma + \beta \sum_{j=1}^{p} \gamma^j g(y_{t-j-1}) \\
&= \alpha(1-\gamma) + \beta \left\{ g(y_{t-1}) - \gamma^{p-1} g(y_{t-p}) \right\} + \gamma \lambda_{t-1}.
\end{aligned}$$

In practice it will be necessary to set λ_0, as well as $y_0, y_{-1}, ...$ in order to use the above recursion. In our paper we have used the ad hoc choice of putting $\lambda_0 = \log \{0.1/(1-0.1)\}$ with $y_j = 0$ for $j = 0, -1, -2, ...$ This means that the process is started with only a moderate probability of having a stationary root.

4.3 Purchasing power parity example

Of the vast amount of real exchange rates analyzed in Bec, Salem, and Carrasco (2001) we focus here on one real exchange rate which is the French Franc versus Italian Lira for the period

α	β	γ	ρ	σ	log-L
-14.96	33.43	0.4056	0.7360	0.01367	920.88
-16.41	34.49		0.7278	0.01364	919.45
			0.9872	0.01538	896.61
			1	0.01533	895.58
-3.262	9.352	0.9788		0.01473	906.28
-5.605	1.440			0.01470	902.16

Table 1: *Fit of ACR models, using a logistic model* $\lambda_t = \alpha(1-\gamma) + \beta\left\{g(y_{t-1}) - \gamma^{p-1} g(y_{t-p})\right\} + \gamma\lambda_{t-1}$*, to the ZFRIT series. Here* $p = 100$*. Results below the line enforced the autoregressive root to be exactly zero, so the series moves between a unit root and white noise. Results were not sensitive to the power used on the absolute value. Model was initialised with* $\lambda_1 = \log\{0.1/(1-0.1)\}$.

September 1973 to September 2000. The monthly data are from *Datastream* and are based on nominal exhange rates which are monthly averages and on consumer price indeces.

Figure 2 displays the real exchange rate series. It indicates a series which moves around zero, with large movements away from zero seemingly being forced eventually back to zero. There are three periods when the series became large in absolute value: around 1976, 1993 and 1995 observations. Table 1 shows the results from fitting a Gaussian random walk and a Gaussian autoregressive model to the data. The difference in the fit is modest but important in terms of the log-likelihoods. We experimented with different functional forms for $\lambda(y_{t-1})$, studying the empirical impact of changing the power we raise the absolute value of the lagged variable. However, throughout the impact of varying this effect was small.

The fitted ACR models have very much smaller values of ρ. This means that the PPP series is stationary: it behaves like a random walk but when it gets a long way from equilibrium the series will quickly collapse backwards corresponding to the small value of ρ. This is much in accordance with the results for this series by the threshold AR modelling in Bec, Salem, and Carrasco (2001) where the null of non-stationarity is strongly rejected based on formal testing.

The pure autoregressive model has a different meaning. It has a high, but less than one, value for ρ. This means at all points the PPP has a small tendency to go back to zero, but this is uniform over its sample space. Hence we would not expect, for example, the AR model to have residuals which are large when the absolute value of the process is large.

Table 1 has two other main results. First the extension to allow for smoother effects in the probability of collapse is not vital here, with a small increase in the likelihood function. Second, the special case of when ρ is set to zero appears not to be supported by the data. This simple model is better than a simple autoregression, however the series seems to prefer a fast but not instant collapse in the series.

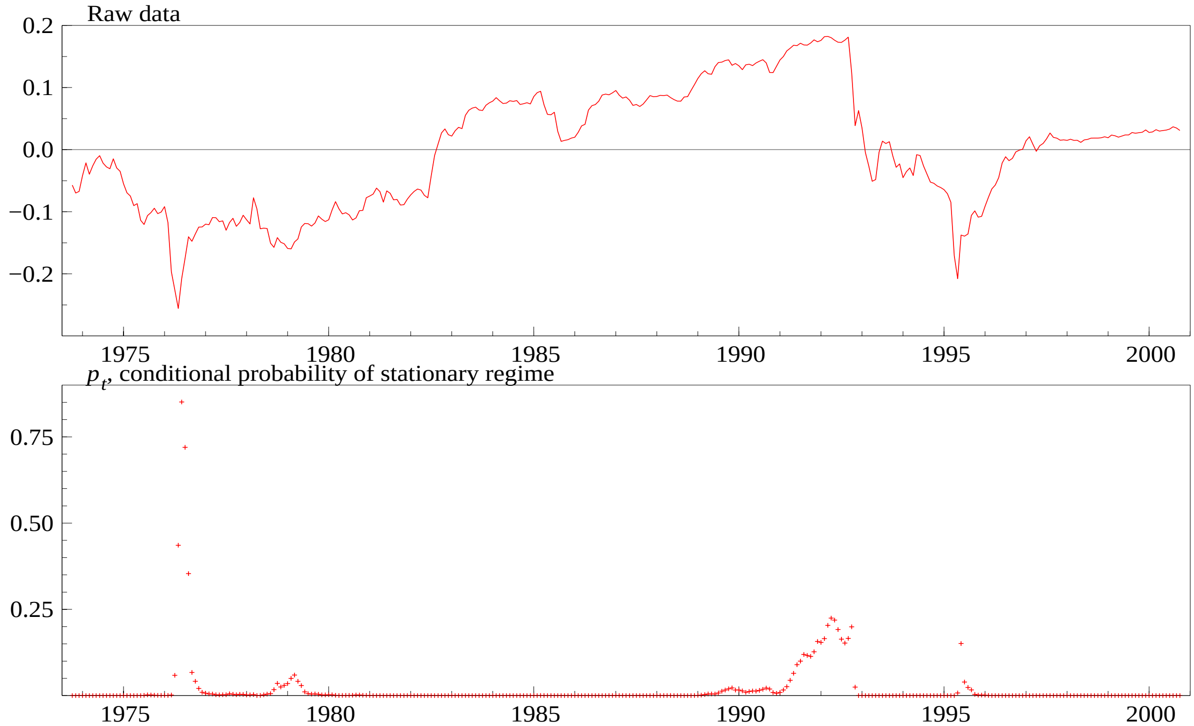

Figure 2: *Top graph is the ZFRIT series against time. Bottom time is the implied* $p_t = \exp(\lambda_t)/\{1+\exp(\lambda_t)\}$, *the conditional probability of a stationary regime. Code:* `regime.ox`.

Figure 2 plots p_t against t for the fitted ACR model, given by the first line of Table 1. The fit of the model given in the second line is very similar. The important point is that the probability of moving into the stationary regime is highest around the 1976 observation, when the original series is furthest from zero. It has three other important times when the probability is away from zero. However, none of them reach the level of the earlier peak. Interesting when we force $\rho = 0$ the corresponding figure, not reported here, is very different for now if just has a single spike around the 1995 observation. This is because this model structure makes the series return to zero instantly, and this feature seems to occur only once in the data.

The top right graph of Figure 3 repeats the picture of the conditional probability p_t against t. In addition Figure 3 shows the standardised residuals from the fitted ACR model and the corresponding residuals from the standard autoregressive models. The standard residuals from the AR model are computed as

$$u_t^{AR} = \frac{y_t - \widehat{\rho} y_{t-1}}{\widehat{\sigma}},$$

where $\widehat{\rho}$ and $\widehat{\sigma}$ denote the maximum likelihood estimators of the AR parameters. Computing equivalent residuals for the ACR model is not so straightforward. We have chosen to compute

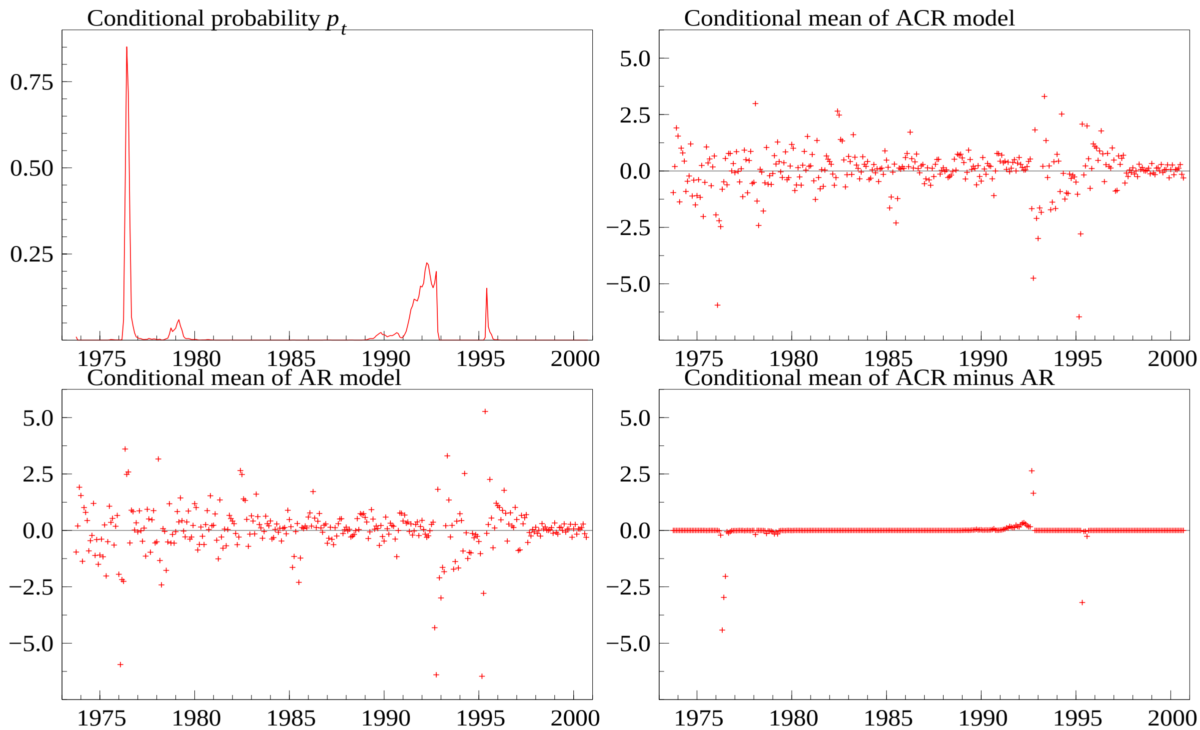

Figure 3: *Top graph left is p_t. Top right are the residuals from the ACR model. Bottom left are the residuals from the AR model. Bottom right is the difference between the residuals from the ACR and AR models. Code:* `regime.ox`*.*

first the one-step ahead prediction distribution functions

$$v_t = p_t \Phi\left(\frac{y_t - \widetilde{\rho} y_{t-1}}{\widetilde{\sigma}}\right) + (1 - p_t)\,\Phi\left(\frac{y_t - y_{t-1}}{\widetilde{\sigma}}\right),$$

where $\widetilde{\rho}$ and $\widetilde{\sigma}$ denote the maximum likelihood estimators of the ACR parameters, while Φ is the distribution function of the standard normal. These $\{v_t\}$ 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

$$u_t^{ACR} = \Phi^{-1}(v_t).$$

The other three graphs are time series plots of the residuals from the models. The plots have been drawn so that the plots are all on the same scale, with the top left picture being the residuals from the ACR model. This shows large failures in the models with residuals which are even five standard deviations from zero. The bottom left picture shows similar effects from the autoregression, although the number of poor residuals is actually higher. The bottom right plot shows the differences in the residuals, it shows that the difference in fit between the two models is focused on 3 periods.

Both models have a large negative residual in 1976, but the ACR model has taken out a number of positive residuals which follow the unwinding of the large negative equilibrium. That is the fitted AR process regards the sharp move back towards zero in 1976 as surprising, while the ACR model did not. Neither is able to predict the preceding sharp move away from equilibrium, hence the shared negative residuals.

The next difference between the two fitted models occur in 1993 when the ACR model has smaller negative residuals. This is the flip side of the above discussion. Now the PPP relationship was misbalanced the other way and the ACR process correctly allowed for a rapid decline in the PPP series, while this was outside the scope of the AR model. The third large discrepancy repeats the 1976 episode in 1995.

Overall the ACR model has only improved upon the AR model in a modest way, really fitting three occurrences in the PPP series. However, the model seems to allow the model to accord more closely with what we would expect: When the adjustment back to equilibrium happens then it is likely to happen quickly.

5 Potential extensions

5.1 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 interst rate parities — see Taylor (1995) for an overview and Frydman and Goldberg (2002) for a recent discussion with non-linear type dynamics.

In econometrics already there 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 encorporate the ACR kind of dynamics in cointegration consider the first order m-

dimensional canonical equilibrium correction model as given by

$$\Delta Y_t = \alpha\beta' Y_{t-1} + \varepsilon_t,$$

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* $\beta' Y_t$ here measures the size of the *out of equilibrium.*

Then suppose that the PPP, say, is given by the univariate process $\beta' Y_t$. It immediately follows that the vector ACR process (VACR)

$$\Delta Y_t = \alpha s_t \beta' Y_{t-1} + \varepsilon_t.$$

has some of the desired features: In particular, the process $\beta' Y_t$,

$$\Delta\beta' Y_t = s_t(\beta'\alpha)\beta' Y_{t-1} + \beta'\varepsilon_t$$

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

$$\Delta\alpha'_\perp Y_t = \alpha'_\perp \varepsilon_t$$

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.

5.2 More general autoregressive models

Some natural extensions of the model are not discussed in the paper. First of all the first order nature of the autoregression in (1) of the basic ACR model can be relaxed. By using higher order autoregressions we produce AR(k)-ACR models. Of particular interest is parameterising the AR part. We prefer to work with the equilibrium correction form used extensively in Hendry (1995). In the AR(2) case the ACR model takes on the form

$$\Delta y_t = s_t \pi y_{t-1} + \gamma \Delta y_{t-1} + \varepsilon_t$$

It seems natural to expect that provided the characteristic polynomial given by

$$A(z) = (1-z) - \pi z - \gamma(1-z)z$$

has roots with absolute value greater than one the y_t process will be stationary. However, the proof of this conjucture is challanging.

5.3 Conditional heteroskedasticity

ACR models could also be developed for models of conditional volatility, which is a commonly used concept in financial econometrics. Consider first the traditional model with

$$y_t|\mathcal{F}_{t-1} \sim N(0, \sigma_t^2),$$

where the conditional variance follows a GARCH type recursion (see for a review Bollerslev, Engle, and Nelson (1994)) such as

$$\begin{aligned}\sigma_t^2 &= \alpha_0 + \alpha_1 y_{t-1}^2 + \alpha_2 \sigma_{t-1}^2 \\ &= \alpha_0 + \alpha_1 y_{t-1}^2 + \alpha_2 \sigma_{t-1}^2 \\ &= \alpha_0 + \alpha_1 \left(y_{t-1}^2 - \sigma_{t-1}^2\right) + \rho \sigma_{t-1}^2\end{aligned}$$

where

$$\rho = \alpha_1 + \alpha_2.$$

Here α_0, α_1 and α_2 are non-negative reals. 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

$$y_t|\mathcal{F}_{t-1}, s_t \sim N(0, \sigma_t^2)$$

and then we change the conditional variance into

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

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,

$$\lambda_t = \alpha + \gamma \sigma_{t-1}^2,$$

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.

6 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 from 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.

7 Acknowledgements

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. 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 for support. Discussions on this topic with Frederique Bec, David Hendry, Søren Tolver Jensen, Hans Martin Krolzig and Rob Engle have been particularly helpful. All the 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 with focus on essentials 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 from Markov Chain Theory

A.1 Introduction

To address geometric ergodicity and stataionarity Markov chain theory will be used. A brief introduction is given to drift criteria from Markov chain theory on general state spaces, see also Meyn and Tweedie (1993).

Let $(X_t)_{t=0,1,2,...}$ be a time homogenous Markov chain on $(\mathcal{X},\mathcal{E}) = (\mathbb{R}^p,\mathcal{B}^p)$ for some p and where $\mathcal{B}^p$ is the Borel σ-algebra on $\mathbb{R}^p$. The kth step transition probability for $k \geq 1$ is denoted $P^k\left(A|\, x\right)$, that is

$$P^k\left(A|\, x\right) = P\left(X_k \in A|\, X_0 = x\right) = P\left(X_{m+k} \in A|\, X_m = x\right),$$

$x \in \mathcal{X}$ and $A \in \mathcal{E}$ and all $m \geq 0$. The following regularity condition will be imposed:

Assumption A.1. *For some $k \geq 1$, the kth step transition probability has a strictly positive and continuous density with respect to the Lebesque measure, i.e.*

$$P\left(X_{t+k} \in A|\, X_t = x\right) = \int_A f(y|\, x)dy$$

for all n and all $x \in X$.

A Markov process satisfying Assumption A.1, is by Lemma 1 below *irreducible* with respect to the Lebesque measure μ, it is *aperiodic* and compact sets $C \subset \mathcal{X}$ are *small.* For a discussion of these concepts see e.g. Chan and Tong (1985) and Meyn and Tweedie (1993).

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

Proof of Lemma 1: First note that the n-step transition probabilities can be defined recursively as follows, $P^1\left(A|\, x\right) = P\left(X_1 \in A|\, X_0 = x\right)$ and

$$P^n\left(A|\, x\right) = \int_{\mathcal{X}} P^{n-1}\left(A|\, y\right) dP^1\left(y|\, x\right) \quad \text{for } n \geq 2,\ x \in \mathcal{X} \ \text{ and } A \in \mathcal{E}$$

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

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

with f positive and continuous. Then the lemma has three parts: irreducibility, aperiodicity and smallness of compact sets.

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

$$\sum_{n=1}^{\infty} P^n\left(A|x\right) > 0 \text{ for all } x \in \mathcal{R}^p \text{ and } A \in \mathcal{B}^p \text{ with } \mu\left(A\right) > 0.$$

Note that,

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

by Assumption A.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 $\mathcal{B}^k$ 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

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

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

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

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

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

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 $A \in \mathcal{B}^k$,

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

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). □

To derive geometric ergodicity define a drift function, which is any function $V : \mathcal{X} \to [1, \infty]$ not identically ∞. For a drift function V consider the m-step conditional expectation $E^m\{V(x)\}$, defined as

$$E^m V(x) := E(V(X_{t+m}) | X_t = x).$$

Definition 1 *A drift function $V : X \to [1, \infty]$ satisfies an m-step geometric drift criterion (relative to the given Markov chain) if there is a compact set $C \subset X$, and constants $\beta \in (0, 1)$, $b > 0$, such that*

$$E^m\{V(x)\} \leq \beta V(x) + b\,1_C(x) \qquad \textit{for all} \quad x \in \mathcal{X}.$$

An important consequence of the m-step geometric drift criterion is the following theorem which can be obtained from Tjøstheim (1990) or Hansen and Rahbek (1998):

Theorem 3 *Let $(X_t)_{t=0,1,\ldots}$ be a time homogenous Markov chain on $(\mathcal{X}, \mathcal{E})$ which is irreducible, aperiodic and for which compact sets are small. Then if it satisfies a m-step geometric drift criterion for some drift function V, the process is geometrically ergodic and there exists an invariant measure for the process. If X_t is initiated at the invariant distribution then the Markov chain is stationary and ergodic. Finally $EV(X_t) < \infty$.*

Remark. If X_t satisfies the drift criterion, not only is the process stationary and ergodic, but as V is integrable, any moments of X_t which are bounded by V exist.

Remark. *Geometric ergodicity* is defined by

$$\lim_{n \to \infty} \gamma^{-n} \|P^n(\cdot | x) - \pi\| = 0$$

for all x, with $\|g\| = \sup\left\{ \left|\int_{\mathcal{X}} f(x) dg(x)\right| \, |f(x)| \leq 1 \right\}$ and $0 < \gamma < 1$. Similarly, *ergodicity by* $\lim_{n \to \infty} \|P^n(\cdot | x) - \pi\| = 0$.

A.2 ACR(1): Geometric Ergodicity

As mentioned y_t is a Markov chain with one-step transition density given by

$$f(y_t | y_{t-1}) = (1 - p_t)\,\phi(\Delta y_t / \sigma) + p_t \phi((\Delta y_t - \pi y_{t-1}) / \sigma)$$

Lemma 2 *$f(\cdot|\cdot)$ satisfies Assumption A.1 if $\lambda(\cdot)$ satisfies Assumption 2*

Proof of Lemma 2: Using the notation $p_x = \frac{\exp(\lambda(x))}{1+\exp(\lambda(x))}$ the density can be written as

$$f(y | x) = \frac{1}{\sqrt{2\pi\sigma^2}} \left[p_x \exp\left\{ -\frac{(y - x)^2}{2\sigma^2} \right\} + (1 - p_x) \exp\left\{ -\frac{(y - \rho x)^2}{2\sigma^2} \right\} \right]$$

which is positive and continuous as desired. □

Theorem 4 *Under Assumption 2 then y_t is geometrically ergodic and has moments of any order if $|\rho| = |\pi + 1| < 1$.*

Proof of Theorem 4: Consider the drift function given by $V(x) = 1+x^2$. Use iterated expectations to see that

$$\begin{aligned} E\left(y_t^2|y_{t-1}=x\right) &= (1-p_x)\left(x^2+\sigma^2\right)+p_x\left(\sigma^2+\rho^2x^2\right) \\ &= \sigma^2+x^2\beta(\lambda(x)) \end{aligned}$$

where

$$\beta(\lambda(x)) = \tfrac{1+\rho^2\exp\lambda(x)}{1+\exp\lambda(x)}.$$

Next, for there to be a constant $\beta < 1$, such that $\beta\left(\lambda(x)\right) < \beta$ for all x it is necessary that $\rho^2 < 1$. As $\lambda\left(x\right)$ is continuous in x and furthermore bounded below by $-\gamma > -\infty$ define $\beta = \beta\left(-\gamma\right)$ which is smaller than one if $\rho^2 < 1$. For moments of order $2k$, simply use the drift function $V_k(x) = 1 + x^{2k}$. □

A.3 ACR(p): Geometric Ergodicity

Consider the process given by (1) but with

$$\lambda_t = \lambda\left(y_{t-1}, .., y_{t-p}\right).$$

The p-dimensional Markov process is defined by

$$X_t = \left(y_t, ..., y_{t-p+1}\right)' \tag{14}$$

and we initially note that:

Lemma 3 *With $(X_t)_{t=1,2,...}$ given by (14), Assumption A.1 holds.*

Proof of Lemma 3: Note that the density of X_{t+p} conditional on X_t is given by,

$$\begin{aligned} f(X_{t+p}|X_t) &= f\left(y_{t+p},\ldots,y_{t+1}|y_t,\ldots,y_{t-p+1}\right) \\ &= \frac{f\left(y_{t+p},\ldots,y_t,\ldots,y_{t-p+1}\right)}{f\left(y_t,\ldots,y_{t-p+1}\right)} = \prod_{i=1}^{p} f\left(y_{t+i}|y_{t-1+i},\ldots,y_{t-p+i}\right) > 0 \end{aligned}$$

where the last line follows by p-dependence in y_t. Hence by definition $f(X_{t+p}|X_t)$ is continuous and positive. □

Next the drift criterion can be applied with, say,

$$V(x) = 1 + \left(c'x\right)^2 \tag{15}$$

with $c' = (1, 0, 0, ..., 0)$.

Proof of Theorem 1: Applying the drift function in (15), one finds

$$E(V(X_t)|X_{t-1}) = 1 + E(y_t^2|X_{t-1}) = 1 + \sigma^2 + \beta(\lambda(X_{t-1}))V(X_{t-1})$$

with $\beta(\cdot)$ defined in (A.2). This shows the result for second order moments. For the general case, consider the drift function $V_k(x) = 1 + (c'x)^{2k}$. $\square$

B Regularity conditions for the asymptotic inference

As mentioned the proof of Theorem 2 is an application of Billingsley (1961) Theorems 2.1 and 2.2 which hold under regularity Conditions 1.1. and 1.2 therein. Regularity Condition 1.2 is modified below based on Markov chain theory for geometrically ergodic processes, while Condition 1.1 remain unaltered.

We show that geometric ergodicity imply that the ACR(1) model satsify the regularity conditions, stated in their present form as Conditions 5 and 6.

B.1 Regularity Condition 1.1

Regularity Condition 1.1 of Billingsley (1961) is here restated as follows:

Condition 5

(i): For all x the set of y for which $f_\theta(y|x) > 0$ does not depend on θ.

(ii): For all y, x the log-*likelihood $l_t(\theta) = \log f_\theta$ is well-defined except for a set of measure zero with respect to the one-step transition probability*[1]*. Also the derivatives*

$$\frac{\partial}{\partial\theta_i} l_t(\theta), \quad \frac{\partial^2}{\partial\theta_i \partial\theta_j} l_t(\theta), \quad \frac{\partial^2}{\partial\theta_i \partial\theta_j \partial\theta_k} l_t(\theta)$$

exist and are continuous in θ

(iii): $E\left|\frac{\partial}{\partial\theta_i} \log f_\theta\right|^2 < \infty$ and

$$\Sigma = E\left(\frac{\partial}{\partial\theta} l_t(\theta)\right)\left(\frac{\partial}{\partial\theta} l_t(\theta)\right)' = -E\left(\frac{\partial^2}{\partial\theta\partial\theta'} l_t(\theta)\right) > 0 \tag{16}$$

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

$$E \sup_{\tilde{\theta} \in N} \left|\frac{\partial^3}{\partial\theta_i \partial\theta_j \partial\theta_k} l_t(\theta)\right| < \infty \tag{17}$$

Together the Lemmas 4, 5 and 6 in the following show that Condition 5 applies to the ACR model.

[1] *See Appendix A*

Lemma 4 *With p_t given by (8) the score for the model in (1) is given by*

$$\frac{\partial}{\partial\theta} l_t(\theta) = (p_t^* - p_t)\frac{\partial\lambda_t}{\partial\theta} + \left\{ p_t^* \frac{\partial \log \phi_t^\pi}{\partial\theta} + (1 - p_t^*)\frac{\partial \log \phi_t^0}{\partial\theta} \right\} \tag{18}$$

such that

$$\begin{aligned} \frac{\partial}{\partial\pi} l_t(\theta) &= p_t^* e_t \frac{1}{\sigma^2} y_{t-1} \frac{\partial}{\partial(\alpha,\beta)'} l_t(\theta) \\ &= (p_t^* - p_t) z_t \frac{\partial}{\partial\sigma^2} l_t(\theta) \\ &= -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}\left(p_t^* e_t^2 + (1 - p_t^*)\Delta y_t^2\right) \end{aligned}$$

Here

$$p_t^* = E(s_t | y_t, y_{t-1}) = \frac{p_t\phi_t^\pi}{p_t\phi_t^\pi + (1-p_t)\phi_t^0} \tag{19}$$

$$e_t = \Delta y_t - \pi y_{t-1} \tag{20}$$

$$z_t = \left(1, y_{t-1}^2\right)' \tag{21}$$

and ϕ_t^π, ϕ_t^0 are given by (10).

Proof of Lemma 4: The result follows by direct differentiation of the log likelihood function in (9) combined with the identity (19). □

Next turn to the information matrix.

Lemma 5 *The observed information is given by*

$$\begin{aligned} \frac{\partial^2}{\partial\theta\partial\theta'} l_t(\theta) = & \left\{ p_t^* \frac{\partial^2 \log \phi_t^\pi}{\partial\theta\partial\theta'} + (1 - p_t^*)\frac{\partial^2 \log \phi_t^0}{\partial\theta\partial\theta'} \right\} - (1 - p_t)p_t \frac{\partial\lambda_t}{\partial\theta}\frac{\partial\lambda_t}{\partial\theta'} \\ & + (1 - p_t^*)p_t^* \left(\frac{\partial\lambda_t}{\partial\theta} + \frac{\partial \log \phi_t^\pi}{\partial\theta} - \frac{\partial \log \phi_t^0}{\partial\theta}\right)\left(\frac{\partial\lambda_t}{\partial\theta'} + \frac{\partial \log \phi_t^\pi}{\partial\theta'} - \frac{\partial \log \phi_t^0}{\partial\theta'}\right) \end{aligned} \tag{22}$$

Moreover, (16) 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 5: Using the identities

$$\frac{\partial}{\partial\theta} p_t = (1 - p_t)p_t \frac{\partial\lambda_t}{\partial\theta} \tag{23}$$

$$\begin{aligned} \frac{\partial}{\partial\theta} p_t^* &= (1 - p_t^*)p_t^* \frac{\partial}{\partial\theta}\log\left(\frac{p_t^*}{1-p_t^*}\right) \\ &= (1 - p_t^*)p_t^* \left(\frac{\partial\lambda_t}{\partial\theta} + \frac{\partial \log \phi_t^\pi}{\partial\theta} - \frac{\partial \log \phi_t^0}{\partial\theta}\right) \end{aligned} \tag{24}$$

it follows directly that

$$\begin{aligned}
\tfrac{\partial^2}{\partial\theta\partial\theta'} l_t(\theta) &= \left(p_t^* - p_t\right) \tfrac{\partial^2 \lambda_t}{\partial\theta\partial\theta'} + \left\{ p_t^* \tfrac{\partial^2 \log \phi_t^\pi}{\partial\theta\partial\theta'} + \left(1 - p_t^*\right) \tfrac{\partial^2 \log \phi_t^0}{\partial\theta\partial\theta'} \right\} \\
&\quad + \left(\tfrac{\partial}{\partial\theta} p_t^* - \tfrac{\partial}{\partial\theta} p_t \right) \tfrac{\partial \lambda_t}{\partial\theta'} + \tfrac{\partial}{\partial\theta} p_t^* \left(\tfrac{\partial \log \phi_t^\pi}{\partial\theta'} - \tfrac{\partial \log \phi_t^0}{\partial\theta'} \right) \\
&= \left(p_t^* - p_t\right) \tfrac{\partial^2 \lambda_t}{\partial\theta\partial\theta'} + \left\{ p_t^* \tfrac{\partial^2 \log \phi_t^\pi}{\partial\theta\partial\theta'} + \left(1 - p_t^*\right) \tfrac{\partial^2 \log \phi_t^0}{\partial\theta\partial\theta'} \right\} - \left(1 - p_t\right) p_t \tfrac{\partial \lambda_t}{\partial\theta} \tfrac{\partial \lambda_t}{\partial\theta'} \\
&\quad + \left(1 - p_t^*\right) p_t^* \left(\tfrac{\partial \lambda_t}{\partial\theta} + \tfrac{\partial \log \phi_t^\pi}{\partial\theta} - \tfrac{\partial \log \phi_t^0}{\partial\theta} \right) \left(\tfrac{\partial \lambda_t}{\partial\theta'} + \tfrac{\partial \log \phi_t^\pi}{\partial\theta'} - \tfrac{\partial \log \phi_t^0}{\partial\theta'} \right).
\end{aligned}$$

In particular,

$$\begin{aligned}
\tfrac{\partial^2}{\partial\pi^2} l_t(\theta) &= -p_t^* \tfrac{1}{\sigma^2} y_{t-1}^2 + \left(1 - p_t^*\right) p_t^* \left\{ \tfrac{1}{\sigma^2} y_{t-1} \left(\Delta y_t - \pi y_{t-1}\right) \right\}^2 \\
&= \left\{ \left(1 - p_t^*\right) p_t^* e_t^2 \tfrac{1}{\sigma^2} - p_t^* \right\} \tfrac{1}{\sigma^2} y_{t-1}^2 \tfrac{\partial^2}{\partial(\alpha,\beta)'\partial(\alpha,\beta)} l_t(\theta) \\
&= \left\{ \left(1 - p_t^*\right) p_t^* - \left(1 - p_t\right) p_t \right\} z_t z_t'
\end{aligned}$$

such that for example

$$-E \tfrac{\partial^2}{\partial\pi^2} l_t(\theta) = E \left(p_t^* e_t \tfrac{1}{\sigma^2} y_{t-1} \right)^2 = E \left(\tfrac{\partial}{\partial\pi} l_t\left(\theta\right) \right)^2$$

by the conditional independence of s_t and ε_t given y_{t-1}. Likewise for the remaining terms in (16) the results follow by repeated use of the identities

$$\begin{aligned}
E \left(p_t^* e_t^k \middle| y_{t-1} \right) &= E \left(s_t \varepsilon_t^k \middle| y_{t-1} \right) = p_t E \left(\varepsilon_t^k \right) \\
\pi^2 E \left(p_t y_{t-1}^2 \right) + \sigma^2 &= E \left(\Delta y_t^2 \right) \\
\pi^4 E \left(p_t y_{t-1}^4 \right) &= E \left(\Delta y_t^4 \right) - 3\sigma^2 \left\{ 2\sigma^2 - 2E \left(\Delta y_t^2 \right) \right\}
\end{aligned} \tag{25}$$

for all positive integers k. □

Lemma 6 *With $(\theta_1, \theta_2, \theta_3, \theta_4) = \left(\pi, \alpha, \beta, \sigma^2\right)$ then (17) holds.*

Proof of Lemma 6: The result is shown by using Lemma 5 and noting that with

$$v_t = \tfrac{\partial \log \phi_t^\pi}{\partial\pi} = \tfrac{1}{\sigma^2} y_{t-1} e_t = \tfrac{1}{\sigma^2} y_{t-1} \left(\Delta y_t - \pi y_{t-1}\right) \tag{26}$$

then

$$|v_t| \leq \kappa_1 \left| y_{t-1} \Delta y_t \right| + \kappa_2 y_{t-1}^2$$

for $\tilde{\theta} \in N\left(\theta\right)$. One finds

$$\begin{aligned}
\left| \tfrac{\partial^3}{\partial\pi^3} l_t(\theta) \right| &= \left| -p_t^* \left(1 - p_t^*\right) v_t \tfrac{1}{\sigma^2} y_{t-1}^2 + \left(1 - 2p_t^*\right) p_t^* v_t^3 \right| \\
&\leq |v_t| \tfrac{1}{\sigma^2} y_{t-1}^2 + |v_t|^3
\end{aligned}$$

and hence $E_\theta \sup_{\tilde{\theta}\in N(\theta)} \left| \frac{\partial^3}{\partial\pi^3} l_t(\theta) \right|$ is finite by existence of all moments of y_t (in particular $Ey_t^6 < \infty$). Next,

$$\begin{aligned} \left| \tfrac{\partial^3}{\partial\alpha^3} l_t(\theta) \right| &= \left| (1-2p_t^*)(1-p_t^*) p_t^* - (1-2p_t)(1-p_t) p_t \right| \\ &\leq 2 \end{aligned}$$

for all $\tilde{\theta}$ and the condition is trivially satisfied. Similarly for the derivatives $\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k} l_t(\theta)$, $i,j,k = 2,3$ using here the existence of sixth order moments as well. As to the derivatives $\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\pi} l_t(\theta)$ for $i,j=2,3$ then e.g.

$$\begin{aligned} \left| \tfrac{\partial^3}{\partial\alpha^2\partial\pi} l_t(\theta) \right| &= \left| (1-2p_t^*)(1-p_t^*) p_t^* v_t \right| \\ &\leq |v_t| \end{aligned}$$

and hence it is bounded by an integrable function. Similar arguments can be used for the remaining derivatives. □

B.2 Regularity Condition 1.2

As noted regularity Condition 1.2 of Billingsley (1961) is modified here reflecting that we use Markov chain theory on general state spaces.

Condition 6 *Independently of choice of initial distribution and as $T \to \infty$:*
(v): Provided $\phi(\cdot,\cdot)$ is measurable and $E\left|\phi(y_t, y_{t-1})\right| < \infty$, then for each θ

$$\tfrac{1}{T}\Sigma_{t=1}^T \phi(y_t, y_{t-1}) \overset{\text{a.s.}}{\to} E\phi(y_t, y_{t-1})$$

(vi): Furthermore

$$\tfrac{1}{T}\Sigma_{t=1}^T \tfrac{\partial}{\partial\theta} l_t(\theta) \overset{\text{D}}{\to} N(0, \Sigma)$$

with Σ defined in (16).

Lemma 7 *If $|\rho| = |\pi+1| < 1$ then Condition 6 holds for the ACR model*

Proof of Lemma 7: With $\mathcal{F}_t = \sigma(y_t, y_{t-1}, ...)$ then

$$\begin{aligned} E\left\{ \tfrac{\partial}{\partial\theta} l_t(\theta) \middle| \mathcal{F}_{t-1} \right\} &= E\left\{ \tfrac{\partial}{\partial\theta} l_t(\theta) \middle| y_{t-1} \right\} \\ &= E\left(p_t^* v_t \middle| y_{t-1} \right) \\ &= 0 \end{aligned}$$

by conditional independence, (26) and (25). Moreover, $E\left\{ \frac{\partial}{\partial(\alpha,\beta)'} l_t(\theta) \middle| y_{t-1} \right\}$ and $E\left\{ \frac{\partial}{\partial \sigma^2} l_t(\theta) \middle| y_{t-1} \right\}$ are identical to zero by the identities

$$E\left\{ (p_t^* - p_t) \middle| y_{t-1} \right\} = 0 \qquad E\left\{ (1 - p_t^*)\, \Delta y_t^2 \middle| y_{t-1} \right\} = (1 - p_t)\, \sigma^2.$$

Hence $\frac{\partial}{\partial\theta} l_t(\theta)$ is a martingale difference sequence with respect to the filtration $\mathcal{F}_t$.

Suppose next that (y_t) is initiated by the invariant distribution. Then the law of large numbers in *(v)* holds by the existence of all moments together with the demonstrated geometric ergodicity. In particular, the average of the conditional second order moments

$$\tfrac{1}{T} \Sigma_{t=1}^{T} E\left\{ \tfrac{\partial}{\partial\theta} l_t(\theta) \tfrac{\partial}{\partial\theta'} l_t(\theta) \middle| \mathcal{F}_{t-1} \right\}$$

will converge. Also the Lindeberg condition in Brown (1971) applies and the claimed asymptotic normality of the score follows as well.

The choice of initial distribution can be relaxed by using the law of large numbers and central limit theorem in Theorem 17.0.1 of Meyn and Tweedie (1993) as it can be shown that the chain defined by $X_t = (y_t, y_{t-1})$, $t = 1, 2, ..$ is geometrically ergodic by using the results in the appendix used for showing that the (y_t) chain was geometrically ergodic. □

References

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

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. B. 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. (1986). Generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics 51*, 307–327.

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.

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. *Scandinavian Journal of Statistics 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 J. R. Russell (1998). Forecasting transaction rates: the autoregressive conditional duration model. *Econometrica 66*, 1127–1162.

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

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.

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, 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.

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.

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.

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

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

Shephard, N. (1994). Partial non-Gaussian state space. *Biometrika 81*, 115–31.

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

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

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.