A Gaussian mixture autoregressive model for univariate time series*

Leena Kalliovirta
University of Helsinki

Mika Meitz
Koç University

Pentti Saikkonen
University of Helsinki

April 2012

[Preliminary and incomplete version. Do not quote.]

Abstract

This paper presents a general formulation for the univariate nonlinear autoregressive model discussed by Glasbey [*Journal of the Royal Statistical Society: Series C*, 50(2001), 143–154] in the first order case, and provides a more thorough treatment of its theoretical properties and practical usefulness. The model belongs to the family of mixture autoregressive models but it differs from its previous alternatives in several advantageous ways. A major theoretical advantage is that, by the definition of the model, conditions for stationarity and ergodicity are always met and these properties are much more straightforward to establish than is common in nonlinear autoregressive models. Moreover, for a pth order model an explicit expression of the $(p+1)$–dimensional stationary distribution is known and given by a mixture of Gaussian distributions with constant mixing weights. Lower dimensional stationary distributions have a similar form whereas the conditional distribution given the past observations is a Gaussian mixture with time varying mixing weights that depend on p lagged values of the series in a natural way. Due to the known stationary distribution exact maximum likelihood estimation is feasible, and one can assess the applicability of the model in advance by using a nonparametric estimate of the density function. An empirical example with interest rate series illustrates the practical usefulness of the model.

*The first and third authors thank the Academy of Finland and the OP-Pohjola Group Research Foundation for financial support. Contact addresses: Leena Kalliovirta, Department of Political and Economic Studies, University of Helsinki, P. O. Box 17, FI–00014 University of Helsinki, Finland; e-mail: leena kalliovirta@helsinki.fi. Mika Meitz, Department of Economics, Koç University, Rumelifeneri Yolu, 34450 Sariyer, Istanbul, Turkey; e-mail: mmeitz@ku.edu.tr. Pentti Saikkonen, Department of Mathematics and Statistics, University of Helsinki, P. O. Box 68, FI–00014 University of Helsinki, Finland; e-mail: pentti.saikkonen@helsinki.fi.

1 Introduction

During the past two or three decades various nonlinear autoregressive (AR) models have been proposed to model time series data. This paper is confined to univariate parametric models although multivariate models and nonparametric models have also attracted interest. Tong (1990) and Granger and Teräsvirta (1993) provide comprehensive accounts of the early stages of threshold autoregressive (TAR) models and smooth transition autoregressive (STAR) models which have become perhaps the most popular nonlinear AR models. An up-to-date discussion of TAR and STAR models, as well as other nonlinear time series models, can be found in Teräsvirta, Tjøstheim, and Granger (2010). TAR and STAR models are distinctively models for the conditional expectation of a time series given its past history although they may also include a time varying conditional variance. The conditional expectation is specified as a convex combination of conditional expectations of two or more linear AR models and similarly for the conditional variance if it is assumed time varying. The weights of these convex combinations (typically) depend on a past value of the time series so that different models are obtained by different specifications of the weights.

The specification of TAR and STAR models is focused on the conditional expectation (and possibly conditional variance) and not so much on the conditional distribution which in parameter estimation is assumed to be Gaussian. In so-called mixture AR models the focus is more on the specification of the whole conditional distribution. In these models the conditional distribution, not only the conditional expectation (and possibly conditional variance) is specified as a convex combination of (typically) Gaussian conditional distributions of linear AR models. Thus, the conditional distribution is a mixture of Gaussian distributions and, similarly to TAR and STAR models, different models are obtained by different specifications of the mixing weights, often assumed to be functions of past values of the series. Models of this kind were introduced by Le, Martin, and Raftery (1996) and further developed by Wong and Li (2000, 2001a,b). Further references include Glasbey (2001), Lanne and Saikkonen (2003), Gourieroux and Robert (2006), Dueker, Sola, and Spagnolo (2007), and Bec, Rahbek, and Shephard (2008) (for reasons to be discussed in Section 2.3 we treat the model of Dueker, Sola, and Spagnolo (2007) as a mixture model although the authors call it a STAR model). Markov switching AR models (see, e.g., Hamilton (1994; Ch. 22)) could also be included in mixture AR models but, as their structure is rather different from the mixture AR models we are interested in, we shall not discuss them any further in this paper.

A property that makes the stationary linear Gaussian AR model different from most, if not all, of its nonlinear AR alternatives is that the probability structure of the underlying stochastic process is fully known. In particular, the joint distribution of any finite realization is Gaussian with mean and covariance matrix being simple functions of the parameters of the conditional distribution used to parameterize the model. In nonlinear AR models the situation is typically very different. The conditional distribution is known by construction but what is usually known beyond that is only the existence of a stationary distribution and finiteness of some of its moments. No explicit expression for the stationary distribution or its density function is known and conditions under which the stationary distribution exists may not be fully known. A notable exception is the mixture AR model discussed by Glasbey (2001, Section 3). In his paper Glasbey (2001) explicitly considers the model only in the first order case and applies it to solar radiation data. In this paper, we extend this model to the general pth order case and provide a more detailed discussion of its properties.

In the considered mixture AR model the mixing weights are defined in a specific way which turns out to have very convenient implications from both theoretical and practical point of view. A theoretical consequence is that stationarity of the underlying stochastic process is a simple consequence of the definition of the model and ergodicity can also be established straightforwardly without imposing any additional restrictions on the parameter space of the model. Moreover, in the pth order case, the $(p+1)$–dimensional stationary distribution is known to be a mixture of Gaussian distributions with constant mixing weights and known structure for the mean and covariance matrix of the component distributions. Consequently, all lower dimensional stationary distributions are of the same type. From the specification of the mixing weights it also follows that the conditional distribution is a Gaussian mixture with time varying mixing weights that depend on p lagged values of the series in a way that has a natural interpretation. Thus, similarly to the linear Gaussian AR process, and contrary to other nonlinear AR models, the structure of stationary marginal distributions of order $p+1$ or smaller is fully known. Stationary marginal distributions of order higher than $p+1$ are not Gaussian mixtures and for them no explicit expressions are available. This need not be a drawback, however, because a process with all finite dimensional distributions being Gaussian mixtures (with constant mixing weights) cannot be ergodic, as we shall demonstrate in the paper. Despite this fact, the formulation of the model is based on the assumption of Gaussianity, and therefore we call the model a Gaussian Mixture AR (GMAR) model.

A practical convenience of having an explicit expression for the stationary marginal

density is that one can use a nonparametric density estimate to examine the suitability of the GMAR model in advance and, after fitting a GMAR model to data, assess the fit by comparing the density implied by the model with the nonparametric estimate. Because the p–dimensional stationary distribution of the process is known the exact likelihood function can be constructed and used to obtain exact maximum likelihood (ML) estimates. An advantage, which also stems from the formulation of the model, is the specific form of the time varying mixing weights which appears very flexible. As demonstrated by our empirical example, we are able to capture features in the data that appear difficult to capture with previous mixture AR models or with usual TAR and STAR models where the transition functions are counterparts of the mixing weights.

The rest of the paper is organized as follows. After discussing general Gaussian mixture AR models Section 2 presents the GMAR model along with a discussion of its properties. Section 3 describes how the parameters of the GMAR model can be estimated by the method of maximum likelihood. Section 4 presents an empirical example with interest rated data, and Section 5 concludes. An appendix provides some technical derivations.

2 Models

2.1 Mixture autoregressive models

Let y_t $(t = 1, 2, \ldots)$ be the real-valued time series of interest, and let $\mathcal{F}_{t-1}$ denote the σ–algebra generated by $\{y_{t-j},\ j > 0\}$. We consider an AR mixture model in which the conditional density function of y_t given its past, $f(\cdot \mid \mathcal{F}_{t-1})$, is of the form

$$f(y_t \mid \mathcal{F}_{t-1}) = \sum_{m=1}^{M} \alpha_{m,t} \frac{1}{\sigma_m} \phi\left(\frac{y_t - \mu_{m,t}}{\sigma_m}\right). \tag{1}$$

Here the (positive) mixing weights $\alpha_{m,t}$ are $\mathcal{F}_{t-1}$–measurable and satisfy $\sum_{m=1}^{M} \alpha_{m,t} = 1$ (for all t). Furthermore, $\phi(\cdot)$ denotes the density function of a standard normal random variable, $\mu_{m,t}$ is defined by

$$\mu_{m,t} = \varphi_{m,0} + \sum_{i=1}^{p} \varphi_{m,i} y_{t-i}, \qquad m = 1, \ldots, M, \tag{2}$$

and $\boldsymbol{\vartheta}_m = (\varphi_{m,0}, \boldsymbol{\varphi}_m, \sigma_m^2)$ with $\boldsymbol{\varphi}_m = (\varphi_{m,1}, \ldots, \varphi_{m,p})$ and $\sigma_m^2 > 0$ $(m = 1, \ldots, M)$ contain the unknown parameters introduced in the above equations. (By replacing p in (2) with p_m, the autoregressive orders in the component models could be allowed to vary; on the

other hand, this can also be achieved by restricting some of the $\varphi_{m,i}$–coefficients in (2) to be zero.) As equation (2) indicates, the definition of the model also requires a specification of the initial values $y_{-p+1}, \ldots, y_0$. Different mixture AR models are obtained by different specifications of the mixing weights. Section 2.3 provides a more detailed discussion of the various specifications proposed in the literature.

For further intuition we express the model (1)–(2) in a different format. Let $P_{t-1}(\cdot)$ signify the conditional probability of the indicated event given $\mathcal{F}_{t-1}$, and let ε_t be a sequence of independent standard normal random variables ($\varepsilon_t \sim NID(0,1)$) such that ε_t is independent of $\{y_{t-j},\ j>0\}$. Furthermore, let $\boldsymbol{s}_t = (s_{t,1}, \ldots, s_{t,M})$ $(t = 1, 2, \ldots)$ be a sequence of (unobserved) M–dimensional random vectors such that, conditional on $\mathcal{F}_{t-1}$, $\boldsymbol{s}_t$ and ε_t are independent. The components of $\boldsymbol{s}_t$ are such that, for each t, exactly one of them takes the value one and others are equal to zero, with conditional probabilities $P_{t-1}(s_{t,m} = 1) = \alpha_{m,t}$, $m = 1, \ldots, M$. Now y_t can be expressed as

$$y_t = \sum_{m=1}^{M} s_{t,m}(\mu_{m,t} + \sigma_m \varepsilon_t) = \sum_{m=1}^{M} s_{t,m} \left(\varphi_{m,0} + \sum_{i=1}^{p} \varphi_{m,i} y_{t-i} + \sigma_m \varepsilon_t \right). \tag{3}$$

This formulation suggests that the mixing weights $\alpha_{m,t}$ can be thought of as probabilities that determine which one of the M autoregressive components of the mixture generates the next observation.

From (1)–(2) or (3) one immediately finds that the conditional mean and variance of y_t given $\mathcal{F}_{t-1}$ are

$$E[y_t \mid \mathcal{F}_{t-1}] = \sum_{m=1}^{M} \alpha_{m,t}\mu_{m,t} = \sum_{m=1}^{M} \alpha_{m,t} \left(\varphi_{m,0} + \sum_{i=1}^{p} \varphi_{m,i} y_{t-i} \right) \tag{4}$$

and

$$Var[y_t \mid \mathcal{F}_{t-1}] = \sum_{m=1}^{M} \alpha_{m,t}\sigma_m^2 + \sum_{m=1}^{M} \alpha_{m,t} \left(\mu_{m,t} - \left(\sum_{m=1}^{M} \alpha_{m,t}\mu_{m,t} \right) \right)^2. \tag{5}$$

These expressions apply for any specification of the mixing weights $\alpha_{m,t}$. The conditional mean is a weighted average of the conditional means of the M autoregressive components with weights generally depending on the past history of the process. The conditional variance also contains a similar weighted average of the conditional (constant) variances of the M autoregressive components but there is an additional additive term which depends on the variability of the conditional means of the component processes. This additional term makes the conditional variance nonconstant even if the mixing weights are nonrandom and constant over time.

2.2 The Gaussian Mixture Autoregressive (GMAR) model

The mixture autoregressive model considered in this paper is based on a particular choice of the mixing weights in (1). Using the parameters $\varphi_{m,0}$, $\boldsymbol{\varphi}_m = (\varphi_{m,1}, \ldots, \varphi_{m,p})$ and σ_m (see equation (1) or (3)) we first define the M auxiliary Gaussian AR(p) processes

$$\nu_{m,t} = \varphi_{m,0} + \sum_{i=1}^{p} \varphi_{m,i}\nu_{m,t-i} + \sigma_m \varepsilon_t, \qquad m = 1, \ldots, M,$$

where the autoregressive coefficients $\boldsymbol{\varphi}_m$ are assumed to satisfy

$$\varphi_m(z) = 1 - \sum_{i=1}^{p} \varphi_{m,i} z^i \neq 0 \quad \text{for } |z| \leq 1, \qquad m = 1, \ldots, M. \tag{6}$$

This condition implies that the processes $\nu_{m,t}$ are stationary and also that each of the component models in (3) satisfies the usual stationarity condition of the conventional linear AR(p) model. Set $\boldsymbol{\nu}_{m,t} = (\nu_{m,t}, \ldots, \nu_{m,t-p+1})$ and $\mathbf{1}_p = (1, \ldots, 1)$ ($p \times 1$), and let $\mu_m \mathbf{1}_p$ and $\boldsymbol{\Sigma}_{m,p}$ signify the mean vector and covariance matrix of $\boldsymbol{\nu}_{m,t}$ ($m = 1, \ldots, M$). Here $\mu_m = \varphi_{m,0}/\varphi_m(1)$ and each $\boldsymbol{\Sigma}_{m,p}$, $m = 1, \ldots, M$, has the familiar form of being a $p \times p$ symmetric Toeplitz matrix with $\gamma_{m,0} = Cov[\nu_{m,t}, \nu_{m,t}]$ along the main diagonal, and $\gamma_{m,i} = Cov[\nu_{m,t}, \nu_{m,t-i}]$, $i = 1, \ldots, p-1$, on the diagonals above and below the main diagonal. For the dependence of the covariance matrix $\boldsymbol{\Sigma}_{m,p}$ on the parameters $\boldsymbol{\varphi}_m$ and σ_m, see Reinsel (1997, Sec. 2.2.3). The random vector $\boldsymbol{\nu}_{m,t}$ follows the p–dimensional multivariate normal distribution with density

$$\mathsf{n}_p(\boldsymbol{\nu}_{m,t}; \boldsymbol{\vartheta}_m) = (2\pi)^{-p/2} \det(\boldsymbol{\Sigma}_{m,p})^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\nu}_{m,t} - \mu_m \mathbf{1}_p)' \boldsymbol{\Sigma}_{m,p}^{-1} (\boldsymbol{\nu}_{m,t} - \mu_m \mathbf{1}_p)\right\}. \tag{7}$$

Now set $\boldsymbol{y}_{t-1} = (y_{t-1}, \ldots, y_{t-p})$, and define the mixing weights $\alpha_{m,t}$ as

$$\alpha_{m,t} = \frac{\alpha_m \mathsf{n}_p(\boldsymbol{y}_{t-1}; \boldsymbol{\vartheta}_m)}{\sum_{n=1}^{M} \alpha_n \mathsf{n}_p(\boldsymbol{y}_{t-1}; \boldsymbol{\vartheta}_n)}, \tag{8}$$

where the $\alpha_m \in (0,1)$, $m = 1, \ldots, M$, are unknown parameters satisfying $\sum_{m=1}^{M} \alpha_m = 1$. (Clearly, the coefficients $\alpha_{m,t}$ are measurable functions of $\boldsymbol{y}_{t-1} = (y_{t-1}, \ldots, y_{t-p})$ and satisfy $\sum_{m=1}^{M} \alpha_{m,t} = 1$ for all t.) We collect the unknown parameters to be estimated in the vector $\boldsymbol{\theta} = (\boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_M, \alpha_1, \ldots, \alpha_{M-1})$ ($(M(p+3)-1) \times 1$); the coefficient α_M is not included due to the restriction $\sum_{m=1}^{M} \alpha_m = 1$. Equations (1), (2), and (8) (or (3) and (8)) define the Gaussian Mixture Autoregressive model or the GMAR model. We use the abbreviation GMAR(p, M) when the autoregressive order and number of component models need to be emphasized.

A major motivation for specifying the mixing weights as in (8) is theoretical attractiveness. We shall discuss this point briefly before providing an intuition behind this particular choice of the mixing weights. First note that the conditional distribution of $\boldsymbol{y}_t$ given $\mathcal{F}_{t-1}$ only depends on $\boldsymbol{y}_{t-1}$, implying that the process $\boldsymbol{y}_t$ is Markovian. This fact is formally stated in the following theorem which shows that there exists a choice of initial values $\boldsymbol{y}_0$ such that $\boldsymbol{y}_t$ is a stationary and ergodic Markov chain. An explicit expression for the stationary distribution is also provided. As will be discussed in more detail shortly, it is quite exceptional among mixture autoregressive models or other related nonlinear autoregressive models such as TAR models or STAR models that the stationary distribution is fully known. As our empirical examples demonstrate, this result is also practically very convenient.

The proof of the following theorem can be found in the Appendix.

Theorem 1. *Consider the GMAR process y_t generated by (1), (2), and (8) (or, equivalently, (3) and (8)) with condition (6) satisfied. Then $\boldsymbol{y}_t = (y_t, \ldots, y_{t-p+1})$ $(t = 1, 2, \ldots)$ is a Markov chain on $\mathbb{R}^p$ with a stationary distribution characterized by the density*

$$f(\boldsymbol{y}; \boldsymbol{\theta}) = \sum_{m=1}^{M} \alpha_m \mathsf{n}_p \left(\boldsymbol{y}; \boldsymbol{\vartheta}_m\right). \tag{9}$$

Moreover, $\boldsymbol{y}_t$ is ergodic.

Thus, the stationary distribution of $\boldsymbol{y}_t$ is a mixture of M multinormal distributions with constant mixing weights α_m that appear in the time varying mixing weights $\alpha_{m,t}$ defined in (8). An immediate consequence of this result is that all moments of the stationary distribution exist and are finite. In the proof of Theorem 1 it is also demonstrated that the stationary distribution of the $(p+1)$–dimensional random vector $\left(y_t, \boldsymbol{y}_{t-1}\right)$ is a Gaussian mixture with density of the same form as in (9) or, specifically, $\sum_{m=1}^{M} \alpha_m \mathsf{n}_{p+1} \left((y, \boldsymbol{y}); \boldsymbol{\vartheta}_m\right)$ with an explicit form of the density function $\mathsf{n}_{p+1} \left((y, \boldsymbol{y}); \boldsymbol{\vartheta}_m\right)$ given in the proof of Theorem 1. It is straightforward to check that the marginal distributions of this Gaussian mixture belong to the same family (this can be seen by integrating the relevant components of $(y, \boldsymbol{y})$ out of the density). It may be worth noting, however, that this does not hold for higher dimensional realizations so that the stationary distribution of $\left(y_{t+1}, y_t, \boldsymbol{y}_{t-1}\right)$, for example, is not a Gaussian mixture. This fact was already pointed out by Glasbey (2001) who considered a first order version of the same model by using a slightly different formulation. Glasbey (2001) did not discuss higher order models explicitly and he did not establish ergodicity obtained in Theorem 1. Interestingly, in the discussion section of his

paper he mentions that a drawback of his model is that joint and conditional distributions in higher dimensions are not Gaussian mixtures. It would undoubtedly be convenient in many respects if all finite dimensional distributions of a process were Gaussian mixtures (with constant mixing weights) but an undesirable implication would then be that ergodicity could not hold true. We demonstrate this in the Appendix by using a simple special case.

A property that makes our GMAR model different from most, if not all, previous nonlinear autoregressive models is that its stationary distribution obtained in Theorem 1 is fully known (for a discussion, see Teräsvirta, Tjøstheim, and Granger (2010, Sec. 3.4.1)). As illustrated in Section 4 a nonparametric estimate of the stationary density of y_t can thus be used (as one tool) to assess the need of a mixture model and the fit of a specified GMAR model. It is also worth noting that in order to prove Theorem 1 we are not forced to restrict the parameter space over what is used to define the model and the parameter space is defined by familiar conditions that can readily be checked. This is in contrast with similar previous results where conditions for stationarity and ergodicity are only sufficient and restrict the parameter space or, if sharp, cannot be verified without resorting to simulation or numerical methods (see, e.g., Cline (2007, JTSA)). It is also worth noting that Theorem 1 can be proved in a much more straightforward manner than most of its previous counterparts. In particular, we do not need to apply the so-called drift criterion which has been a standard tool in previous similar proofs (see, e.g., Bec, Rahbek, and Shephard (2008), Meyn and Tweedie (2009)). On the other hand, our GMAR model assumes that the components of the mixture satisfy the usual stationarity condition of a linear AR(p) model which is not required in all previous models. For instance, Bec, Rahbek, and Shephard (2008) prove an analog of Theorem 1 with $M = 2$ without any restrictions on the autoregressive parameters of one of the regimes (see also Cline (2007)).

Unless otherwise stated, the rest of this section assumes the stationary version of the process. According to Theorem 1, the parameter α_m $(m = 1, \ldots, M)$ then has an immediate interpretation as the unconditional probability of the random vector $\boldsymbol{y}_t = (y_t, \ldots, y_{t-p+1})$ being generated from a distribution with density $\mathsf{n}_p(\boldsymbol{y}; \boldsymbol{\vartheta}_m)$, that is, from the mth component of the Gaussian mixture characterized in (9). As a direct consequence, α_m $(m = 1, \ldots, M)$ also represents the unconditional probability of the component y_t being generated from a distribution with density $\mathsf{n}_1(y; \boldsymbol{\vartheta}_m)$ which is the mth component of the (univariate) Gaussian mixture density $\sum_{m=1}^{M} \alpha_m \mathsf{n}_1(y; \boldsymbol{\vartheta}_m)$ where $\mathsf{n}_1(y; \boldsymbol{\vartheta}_m)$ is the density function of a normal random variable with mean μ_m and variance $\gamma_{m,0}$. Furthermore, it is straightforward to check that α_m also represents the unconditional probability of (the

scalar) y_t being generated from the mth autoregressive component in (3) whereas $\alpha_{m,t}$ represents the corresponding conditional probability $P_{t-1}(\boldsymbol{s}_{t,m}=1)=\alpha_{m,t}$. This conditional probability depends on the (relative) size of the product $\alpha_m \mathsf{n}_p\left(\boldsymbol{y}_{t-1};\boldsymbol{\vartheta}_m\right)$, the numerator of the expression defining $\alpha_{m,t}$ (see (8)). The latter factor of this product, $\mathsf{n}_p\left(\boldsymbol{y}_{t-1};\boldsymbol{\vartheta}_m\right)$, can be interpreted as the likelihood of the mth autoregressive component in (3) based on the observation $\boldsymbol{y}_{t-1}$. Thus, the larger this likelihood is the more likely it is to observe y_t from the mth autoregressive component. However, the product $\alpha_m \mathsf{n}_p\left(\boldsymbol{y}_{t-1};\boldsymbol{\vartheta}_m\right)$ is also affected by the former factor α_m or the weight of $\mathsf{n}_p\left(\boldsymbol{y}_{t-1};\boldsymbol{\vartheta}_m\right)$ in the stationary mixture distribution of $\boldsymbol{y}_{t-1}$ (evaluated at $\boldsymbol{y}_{t-1}$; see (9)). Specifically, even though the likelihood of the mth autoregressive component in (3) is large (small) a small (large) value of α_m attenuates (amplifies) its effect so that the likelihood of observing y_t from the mth autoregressive component can be small (large). This seems intuitively natural because a small (large) weight of $\mathsf{n}_p\left(\boldsymbol{y}_{t-1};\boldsymbol{\vartheta}_m\right)$ in the stationary mixture distribution of $\boldsymbol{y}_{t-1}$ means that observations cannot be generated by the mth autoregressive component too frequently (too infrequently).

It may also be noted that the probabilities $\alpha_{m,t}$ are formally similar to posterior model probabilities commonly used in Bayesian statistics (see, e.g., Sisson (2005) or Del Negro and Schorfheide (2011). An obvious difference is that in our model the parameters $\boldsymbol{\vartheta}_1,\ldots,\boldsymbol{\vartheta}_M$ are treated as fixed so that no prior distributions are specified for them. Therefore, the marginal likelihood used in the Bayesian setting equals the density $\mathsf{n}_p(\boldsymbol{y};\boldsymbol{\vartheta}_m)$ associated with the mth model. However, as α_m only requires knowledge of the stationary distribution of the process, not observed data, it can be thought of as one's prior probability of the observation y_t being generated from the mth autoregressive component in (3). When observed data $\mathcal{F}_{t-1}$ (or $\boldsymbol{y}_{t-1}$) are available one can compute $\alpha_{m,t}$, an analog of the corresponding posterior probability, which provides more accurate information about the likelihood of observing y_t from the mth autoregressive component in (3). Other things being equal a decrease (increase) in the value of α_m decreases (increases) the value of $\alpha_{m,t}$. That the stationary distribution of the process explicitly affects the conditional probability of observing y_t from the mth autoregressive component appears intuitively natural regardless of whether one interprets α_m as a prior probability or a mixing weight in the stationary distribution.

Using the facts that the density of $\left(y_t,\boldsymbol{y}_{t-1}\right)$ is $\sum_{m=1}^{M}\alpha_m \mathsf{n}_{p+1}\left((y_t,\boldsymbol{y}_{t-1});\boldsymbol{\vartheta}_m\right)$ and that of y_t is $\sum_{m=1}^{M}\alpha_m \mathsf{n}_1\left(y;\boldsymbol{\vartheta}_m\right)$ we can obtain explicit expressions for the mean, variance, and first p autocovariances of the process y_t. With the notation introduced in equation (7) we

can express the mean as

$$\mu \overset{def}{=} E\left[y_t\right] = \sum_{m=1}^{M} \alpha_m \mu_m$$

and the variance and first p autocovariances as

$$\gamma_j \overset{def}{=} Cov\left(y_t, y_{t-j}\right) = \sum_{m=1}^{M} \alpha_m \gamma_{m,j} + \sum_{m=1}^{M} \alpha_m \left(\mu_m - \mu\right)^2, \qquad j = 0, 1, \ldots, p.$$

Using these autocovariances and Yule-Walker equations (see, e.g., Box, Jenkins, and Reinsel (2008, p. 59)) one can derive the parameters of the linear AR(p) process that best approximates a GMAR(p, M) process. As higher dimensional stationary distributions are not Gaussian mixtures and appear difficult to handle no simple expressions are available for autocovariances at lags larger than p.

2.3 Discussion of models

In this section, we discuss the GMAR model in relation to other nonlinear autoregressive models introduced in the literature. If the mixing weights are assumed constant over time the general mixture autoregressive model (1) reduces to the MAR model studied by Wong and Li (2000). The MAR model, in turn, is a generalization of a model considered by Le, Martin, and Raftery (1996). Wong and Li (2001b) considered a model with time-varying mixing weights. In their Logistic MAR (LMAR) model, only two regimes are allowed, with a logistic transformation of the two mixing weights, $\log(\alpha_{1,t}/\alpha_{2,t})$, being a linear function of past observed variables. Related two-regime mixture models with time-varying mixing weights were also considered by Gourieroux and Robert (2006) and Bec, Rahbek, and Shephard (2008). Lanne and Saikkonen (2003) considered a mixture AR model in which multiple regimes are allowed. In their model, the mixing weights are specified as

$$\alpha_{m,t} = \begin{cases} 1 - \Phi((y_{t-d} - c_1)/\sigma_\eta), & m = 1, \\ \Phi((y_{t-d} - c_{m-1})/\sigma_\eta) - \Phi((y_{t-d} - c_m)/\sigma_\eta), & m = 2, \ldots, M-1, \\ \Phi((y_{t-d} - c_{M-1})/\sigma_\eta), & m = M, \end{cases} \tag{10}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of a standard normal random variable, $d \in \mathbb{Z}_+$ is a delay parameter, and the real constants $c_1 < \cdots < c_{M-1}$ are location parameters. In their model, the probabilities determining which of the M autoregressive components the next observation is generated from depend on the location of y_{t-d} relative to the location parameters $c_1 < \cdots < c_{M-1}$. Thus, when $p = d = 1$ a similarity between

the mixing weights in the model of Lanne and Saikkonen (2003) and in the GMAR model is that the value of y_{t-1} gives indication concerning which regime will generate the next observation. However, even in this case the functional forms of the mixing weights and their interpretation are rather different.

An interesting two-regime mixture model with time-varying mixing weights was recently introduced by Dueker, Sola, and Spagnolo (2007) (see also Dueker, Psaradakis, Sola, and Spagnolo (2011) for a multivariate extension).[1] In their model the mixing weights are defined by

$$\alpha_{1,t} = \frac{\Phi\left((c_1 - \varphi_{1,0} - \boldsymbol{\varphi}_1'\boldsymbol{y}_{t-1})/\sigma_1\right)}{\Phi\left((c_1 - \varphi_{1,0} - \boldsymbol{\varphi}_1'\boldsymbol{y}_{t-1})/\sigma_1\right) + \left[1 - \Phi\left((c_1 - \varphi_{2,0} - \boldsymbol{\varphi}_2'\boldsymbol{y}_{t-1})/\sigma_2\right)\right]} \tag{11}$$

and $\alpha_{2,t} = 1 - \alpha_{1,t}$. Here c_1 is interpreted as a location parameter similar to that in the model of Lanne and Saikkonen (2003). However, similarly to our model the mixing weights are determined by p lagged values of the observed series and the autoregressive parameters of the component models. This makes the interpretation of the mixing weights closer to that of our GMAR model than is the case for the model of Lanne and Saikkonen (2003). The probability that the next observation is generated from the first or second regime is determined by the locations of the conditional means of the two autoregressive components from the location parameter c_1 whereas in the GMAR model this probability is determined by the stationary densities of the two component models and their weights in the stationary mixture distribution. The functional form of the mixing weights of Dueker, Sola, and Spagnolo (2007) is also similar to ours except that instead of the Gaussian density function used in our GMAR model Dueker, Sola, and Spagnolo (2007) have the Gaussian cumulative distribution function.

The GMAR model is also related to threshold and smooth transition type nonlinear models. In particular, the conditional mean function $E[y_t \mid \mathcal{F}_{t-1}]$ of our GMAR model is similar to those of a TAR or a STAR model (see, e.g., Tong (1990) and Teräsvirta (1994)). In a basic two-regime TAR model, whether a threshold variable (a lagged value

[1] According to Dueker et al. (2007, 2011) their model belongs to the family of STAR models. However, it is not clear to us whether this model should be regarded as a STAR model or a mixture model. Initially Dueker, Sola, and Spagnolo (2007) use equations (1)–(4) of their paper to define the model and these equations lend support to the interpretation of the model as a STAR model. However, treating the model as a mixture model also appears reasonable because the likelihood function used to fit the model to data is determined by conditional density functions that are of the mixture form (1) (with $\phi(\cdot)$ a possibly nonnormal density). These conditional density functions are given in equation (7) of Dueker, Sola, and Spagnolo (2007) but but their connection to the aforementioned equations (1)–(4) is not clear to us.

of y_t) exceeds a certain threshold or not determines which of the two component models describes the generating mechanism of the next observation. The threshold and threshold variable are analogous to the location parameter c_1 and the variable y_{t-d} in the mixing weights used in the two-regime ($M = 2$) mixture model of Lanne and Saikkonen (2003) (see (10)). In a STAR model, one gradually moves from one component model to the other as the threshold (or transition) variable changes its value. In a GMAR model, the mixing weights follow similar smooth patterns. A difference to STAR models is that while the mixing weights of the GMAR model vary smoothly, the next observation is generated from one particular AR component whose choice is governed by these mixing weights. In a STAR model, the generating mechanism of the next observation is described by a convex combination of the two component models. This difference is related to the fact that the conditional distribution of the GMAR model is of a different type than the conditional distribution of the STAR (or TAR) model which is not a mixture distribution. This difference is also reflected in differences between the conditional variances associated with the GMAR model and STAR (or TAR) models.

Figure 1 illustrates the preceding discussion. In the top panels, we plot the mixing weight $\alpha_{1,t}$ of the GMAR model as a function of $y_{t-1} = y$ in the case $M = 2$, $p = 1$, with certain parameter combinations. The bottom left panel shows $\alpha_{1,t}$ in some cases for the LMAR model of Wong and Li (2001b); in the model of Lanne and Saikkonen (2003) $\alpha_{1,t}$ behaves in a comparable way (no picture presented). The two pictures on the left illustrate that the three models can produce mixing weights of similar monotonically increasing patterns. The figure in the top left panel also illustrates the previously mentioned fact about the mixing weights of the GMAR model that, other things being equal, a decrease in the value of α_m decreases the value of $\alpha_{m,t}$. In the conditional expectation of a basic logistic two-regime STAR model, referred to as the LSTAR1 model in Teräsvirta, Tjøstheim, and Granger (2010, Sec. 3.4.1), the transition function, which is the counterpart of the mixing weight $\alpha_{1,t}$, also behaves in a similar monotonically increasing way. Given these observations it is interesting that with suitable parameter values our GMAR model can produce nonmonotonic mixing weights even in the case $M = 2$. The top right panel illustrates this. The models of Wong and Li (2001b) and Lanne and Saikkonen (2003) can produce mixing weights of this form only when $M > 2$. Similarly, in LSTAR models a transition function of this form cannot be obtained with a LSTAR1 model. For that one needs an LSTAR2 model (see Teräsvirta, Tjøstheim, and Granger (2010, Sec. 3.4.1)) or some other similar model such as the exponential autoregressive model of Haggan and Ozaki (1981). Thus, once the number of component models is specified our GMAR model appears more

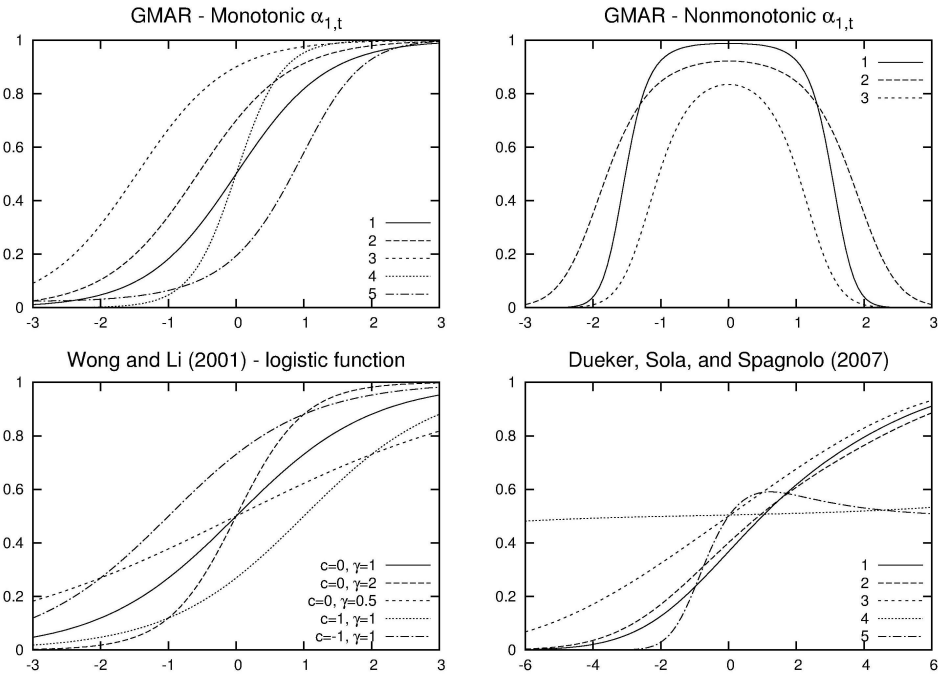

Figure 1: **Top left panel:** $\alpha_{1,t}$ in the GMAR model ($p = 1$) as a function of y_{t-1}. Parameter values used: model 1: $\varphi_{1,0} = 0.5$, $\varphi_{2,0} = -0.5$, $\varphi_{1,1} = \varphi_{2,1} = 0.5$, $\sigma_1^2 = \sigma_2^2 = 1$, $\alpha_1 = 0.5$; model 2: same as model 1 but $\alpha_1 = 0.7$; model 3: same as model 1 but $\alpha_1 = 0.9$; model 4: $\varphi_{1,0} = 1$, $\varphi_{2,0} = -1$, $\varphi_{1,1} = \varphi_{2,1} = 0.5$, $\sigma_1^2 = \sigma_2^2 = 1$, $\alpha_1 = 0.5$; model 5: $\varphi_{1,0} = \varphi_{2,0} = 0.5$, $\varphi_{1,1}0.75$, $\varphi_{2,1} = 0.25$, $\sigma_1^2 = \sigma_2^2 = 1$, $\alpha_1 = 0.5$. **Top rigth panel:** $\alpha_{1,t}$ in the GMAR model ($p = 1$) as a function of y_{t-1}. Parameter values used: model 1: $\varphi_{1,0} = \varphi_{2,0} = 0$, $\varphi_{1,1} = 0.2$, $\varphi_{2,1} = 0.9$, $\sigma_1^2 = 0.25$, $\sigma_2^2 = 4$, $\alpha_1 = 0.5$; model 2: same as model 1 but $\sigma_1^2 = \sigma_2^2 = 0.5$, $\alpha_1 = 0.7$; model 3: same as model 1 but $\sigma_2^2 = 0.25$. **Bottom left panel:** $\alpha_{1,t}$ in the model of Wong and Li (2001) as a function of y_{t-1}. Logistic equation assumed to be of the form $\log(\alpha_{1,t}/\alpha_{2,t}) = \gamma(y_{t-1} - c)$, or equivalently, $\alpha_{1,t} = \frac{1}{1+e^{-\gamma(y_{t-1}-c)}}$. Note that this is exactly the standard form of the logistic function. Curves correspond to different values of c and γ. **Bottom right panel:** $\alpha_{2,t} = 1-\alpha_{1,t}$ in the model of Dueker, Sola, and Spagnolo (2007) as a function of y_{t-1}. Parameter values used: model 1: $c_1 = 1$, $\varphi_{1,0} = -0.5$, $\varphi_{2,0} = 0.5$, $\varphi_{1,1} = \varphi_{2,1} = 0.9$, $\sigma_1 = 3$, $\sigma_2 = 2$; model 2: $c_1 = 1$, $\varphi_{1,0} = -1$, $\varphi_{2,0} = 1$, $\varphi_{1,1} = \varphi_{2,1} = 0.9$, $\sigma_1 = 3$, $\sigma_2 = 2$; model 3: $c_1 = 0$, $\varphi_{1,0} = -1$, $\varphi_{2,0} = 1$, $\varphi_{1,1} = \varphi_{2,1} = 0.9$, $\sigma_1 = \sigma_2 = 3$; model 4: $c_1 = 0$, $\varphi_{1,0} = -10$, $\varphi_{2,0} = 10$, $\varphi_{1,1} = \varphi_{2,1} = 0.7$, $\sigma_1 = 5$, $\sigma_2 = 4$; model 5: $c_1 = 0$, $\varphi_{1,0} = \varphi_{2,0} = 0$, $\varphi_{1,1} = -0.3$, $\varphi_{2,1} = 0.3$, $\sigma_1 = 1$, $\sigma_2 = 0.25$.

flexible in terms of the form of mixing weights than the aforementioned previous mixture models and the same is true when the mixing weights of our GMAR model are compared to the transition functions of LSTAR models. As far as the mixing weights of the model of Dueker, Sola, and Spagnolo (2007) are concerned they can be nonmonotonic, as illustrated in the bottom right panel of Figure 1. After trying a number of different parameter combinations it seems, however, that (at least in the case $p = 1$) nonmonotonic mixing weights are rather special for this model. The first four (monotonic) graphs in the bottom right panel correspond to parameter configurations in Table 2 of Dueker, Sola, and Spagnolo (2007). The fourth one is interesting in that it produces a nearly constant graph (the graph would be constant if the values of the standard deviations σ_1 and σ_2 were changed to be equal). Finally, note that a common convenience of the GMAR model as well as of the models of Wong and Li (2001b) and Dueker, Sola, and Spagnolo (2007) is that there is no need to choose a threshold variable such as y_{t-d} in the model of Lanne and Saikkonen (2003) or in TAR and STAR models.

3 Estimation

The parameters of the GMAR model can be estimated by the method of maximum likelihood (ML). As the stationary distribution of the GMAR process is known it is even possible to make use of initial values and construct the exact likelihood function and obtain exact ML estimates, as already discussed by Glasbey (2001) in the first order case. Assuming the observed data $y_{-p+1}, ..., y_0, y_1, ..., y_T$ and stationary initial values the log-likelihood function takes the form

$$\begin{aligned} l_T(\boldsymbol{\theta}) &= \log\left(\sum_{m=1}^{M}\alpha_m \mathsf{n}_p(\boldsymbol{y}_0;\boldsymbol{\vartheta}_m)\right) \\ &\quad + \sum_{t=1}^{T}\log\left(\sum_{m=1}^{M}\alpha_{m,t}(\boldsymbol{\theta})\left(2\pi\sigma_m^2\right)^{-1/2}\exp\left(-\frac{\left(y_t-\mu_{m,t}(\boldsymbol{\vartheta}_m)\right)^2}{2\sigma_m^2}\right)\right), \end{aligned} \tag{12}$$

where dependence of the mixing weights $\alpha_{m,t}$ and the conditional expectations $\mu_{m,t}$ of the component models on the parameters is made explicit (see (8) and (2)). Maximizing the log-likelihood function $l_T(\boldsymbol{\theta})$ with respect to the parameter vector $\boldsymbol{\theta}$ yields the ML estimate denoted by $\hat{\boldsymbol{\theta}}$ (a similar notation is used for components of $\hat{\boldsymbol{\theta}}$). Here we have assumed that the initial values in the vector $\boldsymbol{y}_0$ are generated by the stationary distribution. If this assumption seems inappropriate one can condition on initial values and drop the first term on the right hand side of (12). For reasons of identification

the inequality restrictions $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_M$ are imposed on the parameters $\alpha_{m,t}$ $\left(m = 1, ..., M,\ \alpha_M = 1 - \sum_{m=1}^{M} \alpha_m\right)$. In our empirical examples we have used the optimization algorithms in the cmlMT library of Gauss to maximize the likelihood function or its conditional version. Especially the Newton-Raphson algorithm in that library seemed to work quite well but one could alternatively follow Wong and Li (2001b) and use the EM algorithm.

Dueker, Sola, and Spagnolo (2007) show that, under appropriate regularity conditions, the usual results of consistency and asymptotic normality of the ML estimator hold in their mixture model. The conditions they use are of general nature and using the ergodicity result of Theorem 1 along with similar "high level" conditions it is undoubtedly possible to show the consistency and asymptotic normality of the ML estimator in our GMAR model as well. However, we prefer to leave a detailed analysis of this issue for future work. In our empirical examples we treat the ML estimator $\hat{\boldsymbol{\theta}}$ as approximately normally distributed with mean vector $\boldsymbol{\theta}$ and covariance matrix the inverse of the Fisher information matrix $E\left[-\partial^2 l_T\left(\boldsymbol{\theta}\right)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'\right]$ that can be estimated by inverting the observed information matrix $-\partial^2 l_T(\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$. It is worth noting that the aforementioned results require a correct specification of the number of autoregressive components. In particular, standard likelihood-based tests are not applicable if the number of component models is chosen too large because then some parameters of the model are not identified. This particularly happens when one tests for the number of component models. For further discussion of this issue, see Dueker et al. (2007, 2011) and the references therein.

4 Empirical example

To illustrate how the GMAR model works in practice we present an example with interest rate data. Interest rate series are typically highly persistent and exhibit behavior possibly due to regime switching dynamics. Consequently, various regime switching models have previously been used in modelling interest rate series. Examples can be found in Hamilton (1988), Garcia and Perron (1996), Enders and Granger (1998), Banzal and Zhou (2002), Dueker, Sola, and Spagnolo (2007), and Dueker, Psaradakis, Sola, and Spagnolo (2011). The data set we use differs from those employed by these authors in that it also contains the recent turbulences of the financial crisis since 2008. Before empirical analyses we provide a brief discussion on some general aspects of building a GMAR model.

As mentioned in the previous section, standard likelihood-based tests cannot be used to test for the number of component models. Instead of trying to develop proper test

procedures for this purpose we use residual based diagnostics and information criteria (AIC and BIC) to select a model. As discussed in Kalliovirta (2012), conventional Pearson residuals are unsuitable for mixture models and, therefore, we use quantile residuals and related diagnostic tests developed in that paper. These tests correctly allow for the uncertainty caused by parameter estimation so that, under correct specification, the obtained p-values are asymptotically valid. Note also that for the linear Gaussian AR model the quantile residuals are identical to conventional (Pearson) residuals. We used cmlMT library in GAUSS Windows Version 10.0 to estimate the parameters of the model. The estimation codes are available upon request from the first author.

We start the model building by finding a conventional linear Gaussian AR model that adequately describes the autocorrelation structure of the series. If the residual diagnostics show signs of non-Gaussianity and possibly also conditional heteroskedasticity we consider the GMAR model. At this point also the graph of the series and a nonparametric estimate of the density function may be useful. To reduce the risk of estimating an unidentified model we first consider a two component model, that is GMAR$(p, 2)$. Our first guess for the order p bases on the order chosen for the linear AR model. However, in our first attempts with the GMAR model we also consider smaller alternatives if the order selected for the linear AR model appears large. If an adequate two component model is not found we try a three component model and, if needed, we consider even more components. After finding an adequate GMAR model we examine possible simplifications obtained by parameter restrictions. For instance, some of the parameters may be restricted to be equal in each component.

We consider the difference between the U.S. and Euro Area long-term government bond yields using monthly observations from January 1989 to December 2009.[2] This series, also referred to as *interest-rate differential*, is depicted in the left panel of Figure 2 (solid line). The autocorrelation at the first lag, 0.987, indicates that the series is highly persistent and around 1997 the level of the series has dropped.

For this series AIC and BIC suggested linear AR(2) and AR(5) models when the considered maximum order was eight (parameters of these models were estimated by least squares). There seemed to be some autocorrelation in the residuals of the AR(2) model which were largely eliminated by the AR(5) model. However, in terms of residual diag-

[2]The data was retrieved from the OECD Statistics. The Euro area data prior to 2001 refer to EU11 (Belgium, Germany, Ireland, Spain, France, Italy, Luxembourg, the Netherlands, Austria, Portugal and Finland), from 2001 to 2006 refer to EU12 (EU11 plus Greece), and from January 2007 data refer to EU13 (EU12 plus Slovenia).

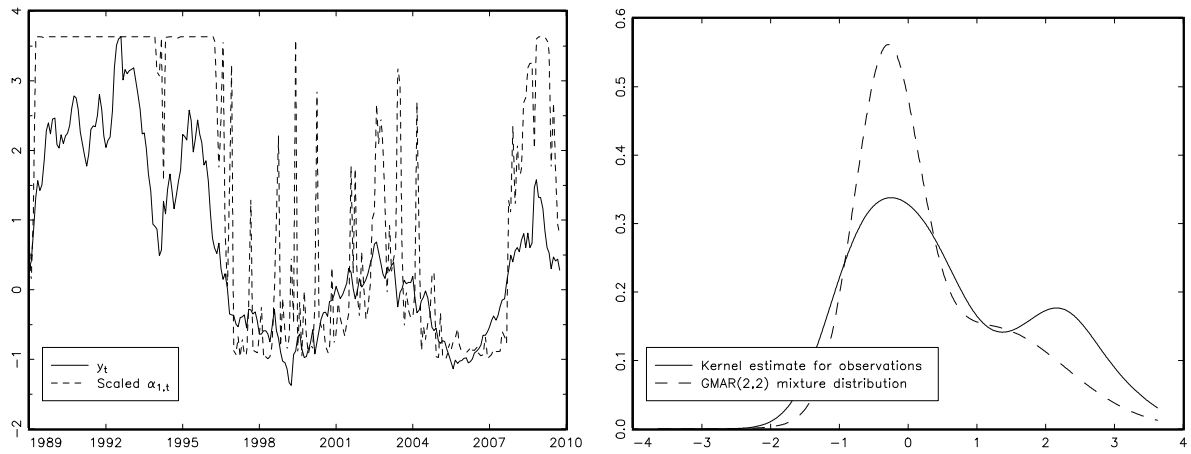

Figure 2: **Left panel:** Interest rate differential between US and Euro area, and scaled mixing weights based on the estimates of the restricted GMAR(2,2) model in Table 2. The scaling is such that $\hat{\alpha}_{1,t} = \max y_t$, when $\hat{\alpha}_{1,t} = 1$, and $\hat{\alpha}_{1,t} = \min y_t$, when $\hat{\alpha}_{1,t} = 0$. **Right panel:** A kernel density estimate of the observations and mixture density implied by the same GMAR(2,2) model as in the left panel.

nostics the AR(4) model appeared equally good and was chosen as a more parsimonious alternative. Although the AR(4) model behaved reasonably well in terms of residual autocorrelations the squared residuals were correlated and the distribution of the residuals was clearly nonnormal (these conclusions also apply to the AR(5) model). Table 1 reports the values of AIC and BIC for the linear AR(4) model along with the (quantile) residual based tests of normality, autocorrelation, and conditional heteroskedasticity. The results show that the tests for normality and conditional heteroskedasticity clearly reject the AR(4) model. A nonparametric kernel density estimate (not shown) indicates that compared to the normal distribution the error distribution is heavy tailed. The kernel density estimate of the original series depicted in the right panel of Figure 2 similarly suggests clear departures from normality. The estimate is bimodal with the mode 0.33 at -0.18 and a local mode 0.18 at 2.2.

Of the GMAR models we first tried an unrestricted GMAR$(2, 2)$ specification. Two AR components seem to match with the graph of the series, where two major levels can be seen, as well as with the bimodal expression of the kernel density estimate (see Figure 2). For this series all estimation and test results based on the exact likelihood and conditional likelihood were practically the same, so we only present those based on the exact likelihood. We used the formula $\mu_m = \varphi_{m,0}/\varphi_m(1)$ with $\varphi_m(1) = 0.1$ and μ_1 and

Table 1: P-values of tests based on quantile residuals and exact likelihood

	N	A_1	A_4	H_1	H_4	$\max l_T(\theta)$	AIC	BIC
AR(4)	0	0.36	0.27	0.003	0	58.3	-107	-89
GMAR(2,2)	0.81	0.97	0.12	0.98	0.77	79.7	-145	-121

NOTE: The test statistic for normality, N, is based on moments of quantile residuals and the test statistics for autocorrelation, A_k, and conditional heteroskedasticity, H_k, are based on the first k autocovariances and squared autocovariances of quantile residuals, respectively. Under correct specification, test statistic N is approximately distributed as χ^2_2 (AR(4)) or χ^2_3 (GMAR(2,2)) and test statistics A_k and H_k are approximately distributed as χ^2_k. A p-value < 0.001 is denoted by 0. For the Gaussian AR(4) model quantile residuals are identical to conventional (Pearson) residuals. GMAR(2,2) refers to a restricted model with a common AR polynomials (see Table 2).

μ_2 chosen to match with the locations of the mode and local mode, respectively, to compute the initial values for $\varphi_{1,0}$ and $\varphi_{2,0}$. Our experience on the estimation of GMAR models indicates that good initial values for these intercept terms are important for numerical optimization of the likelihood function whereas the precision of initial values used for the AR coefficients and error variances appears less critical.

According to quantile residual diagnostics (not reported) the unrestricted GMAR$(2, 2)$ specification turned out to be adequate but, as the AR polynomials in the two components seemed to be very close to each other, we restricted them to be the same. This restriction was not rejected by the LR test (p-value 0.61) and, as Table 1 shows, diagnostic tests based on quantile residuals also lend support to the restricted GMAR$(2, 2)$ model. To get a wider picture of the properties of quantile residuals Figure 3 depicts the time series and QQ-plot of quantile residuals as well as the first ten standardized sample autocovariances of quantile residuals and their squares (the employed standardization is such that, under correct specification, the distribution of the sample autocovariances is approximately standard normal). The time series of quantile residuals computed from a correctly specified model should resemble a realization from an independent standard normal sequence. The graph of quantile residuals and related QQ-plot give no obvious reason to suspect this, although some large positive quantile residuals occur. According to the approximate 99% critical bounds only two somewhat larger autocovariances are seen but even they are found at larger lags (we use 99% critical bounds because, from the viewpoint of statistical testing, several tests are performed). It is particularly encouraging that the GMAR model has been able to allow for the conditional heteroskedasticity observed in the considered linear AR

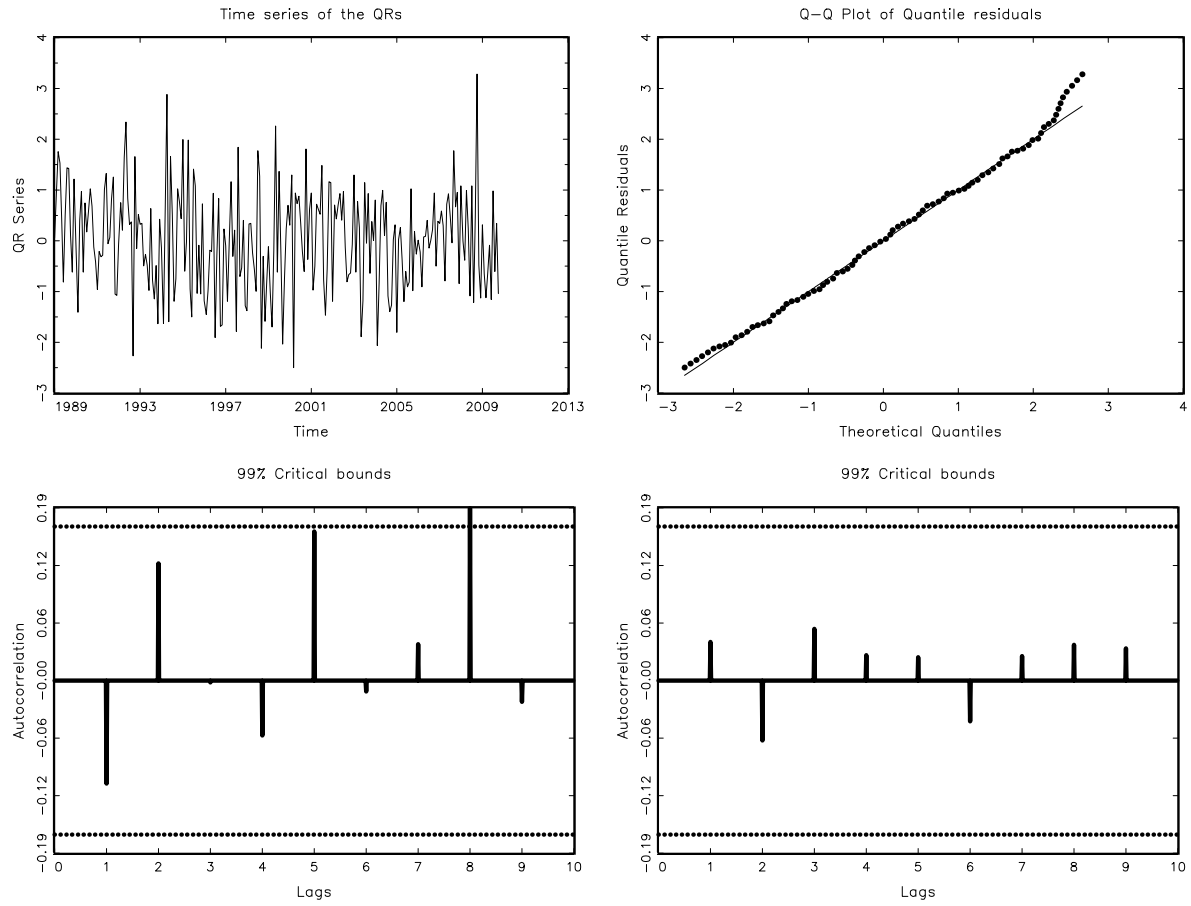

Figure 3: Diagnostics of the restricted GMAR(2,2) model described in Table 2. **Top left panel**: Time series of quantile residuals. **Top right panel**: QQ-plot of quantile residuals. **Bottom left panel**: Ten first scaled autocovariances of quantile residuals. **Bottom right panel**: Ten first scaled autocovariances of squared quantile residuals. The lineas in the bottom panels show approximate 99% critical bounds.

models (see the bottom right panel of Figure 3 and the first panel of Table 1). Thus, unlike the linear AR models the GMAR$(2, 2)$ model seems to provide an adequate description for the interest rate series and according to AIC and BIC it also outperforms the chosen linear AR(4) model by a wide margin (this also holds for the more parsimonious linear AR(2) model suggested by BIC).

Table 2 displays estimates obtained for the parameters of the restricted GMAR$(2, 2)$ model as well as estimates derived for the expectations μ_m and elements of the covariance matrix $\Sigma_{m,2}$ (see Section 2.2). The estimated sum of the AR coefficients is 0.966 which is slightly less than the corresponding sum 0.982 obtained in the linear AR(4) model. The

Table 2: Parameter estimates of the restricted GMAR(2,2) model based on exact likelihood

$\max l_T(\theta)$	φ_{01}	φ_{02}	φ_{11}	φ_{21}	α_1	σ_1^2	σ_2^2
79.7	0.041 (0.022)	−0.011 (0.006)	1.262 (0.064)	−0.296 (0.065)	0.586 (0.174)	0.058 (0.008)	0.011 (0.002)
μ_1	μ_2	$\gamma_{1,0}$	$\gamma_{1,1}$	$\gamma_{1,1}/\gamma_{1,0}$	$\gamma_{2,0}$	$\gamma_{2,1}$	$\gamma_{2,1}/\gamma_{2,0}$
1.175	-0.321	1.214	1.182	0.973	0.221	0.215	0.973

NOTE: Approximate standard errors obtained by taking square roots of the diagonal elements of the matrix $(-\partial^2 l_T(\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}')^{-1}$ in parentheses.

reduction is presumably related to the differences in the intercept terms of the AR components which is directly reflected in different means of the two regimes with point estimates 1.175 and -0.321. The estimated error variances of the AR components are also very different and, consequently, the same is true for the variances of the two regimes with point estimates 1.214 and 0.221. This feature is of course related to the above-mentioned fact that the model has been able to remove the conditional heteroskedasticity observed in linear modeling. According to the approximate standard errors in Table 2 the estimation accuracy appears quite reasonable except for the parameter α_1, the weight of the first component in the stationary distribution of the GMAR$(2,2)$ process. The point estimate of this parameter is 0.586 with approximate standard error 0.174. A possible explanation for this rather unreliable estimation is that the series is not sufficiently long to reveal the nature of the stationary distribution to which the parameter α_1 is directly related. To put this another way, with the estimated parameter values the process may generate realizations that look rather different from our series. On the other hand, this does not show in differences between estimates based on exact likelihood and conditional likelihood, a potential reason being that the observed series starts rather close to its mean.

Based on the estimates of Table 2, Figure 4 shows the estimate of the two dimensional stationary mixture density $\sum_{m=1}^{2}\alpha_m \mathsf{n}_2(\boldsymbol{y};\boldsymbol{\vartheta}_m)$ along with a related contour plot. A figure of the one dimensional mixture density $\sum_{m=1}^{2}\alpha_m \mathsf{n}_1(y;\boldsymbol{\vartheta}_m)$ and its two components is also included. These figures clearly illustrate the large differences between the shapes of the two component densities already apparent in the estimates of Table 2. The one dimensional mixture density is also drawn in the right panel of Figure 2 (dashed line) and, as can be seen, there are rather large departures between the density implied by the model and the nonparametric kernel density estimate. The density implied by the model is more peaked and more concentrated than the kernel density estimate. The kernel

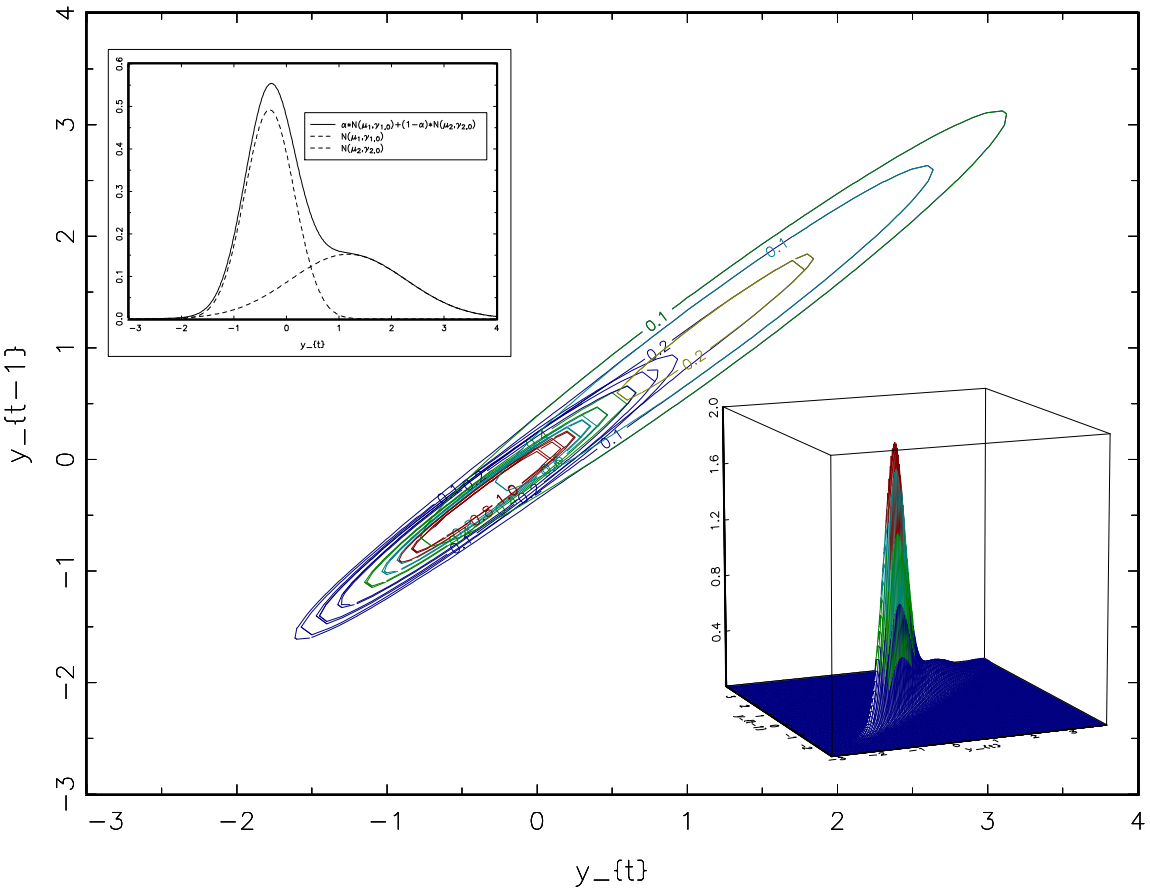

Figure 4: Estimate of the two dimensional stationary mixture density implied by the GMAR(2,2) model described in Table 3 (below), its contour plots (middle), and the corresponding one dimensional marginal density and its two components (above).

density estimate may not be too reliable, however, because in some parts of the empirical distribution the number of observations seems to be rather small and the choice of the bandwidth parameter has a noticeable effect on the shape of the kernel density.

Figure 2 (left panel) depicts the time series of the estimated mixing weight $\hat{\alpha}_{1,t}$ scaled so that $\hat{\alpha}_{1,t} = \max y_t$, when $\hat{\alpha}_{1,t} = 1$, and $\hat{\alpha}_{1,t} = \min y_t$, when $\hat{\alpha}_{1,t} = 0$ (dashed line). During the period before 1996 or 1997 the first regime is clearly dominating. Except for only a few exceptional months the mixing weights $\hat{\alpha}_{1,t}$ are practically unity. This period corresponds to a high level regime or regime where U.S. bond yields are smaller than Euro Area bond yields. After this period a low level regime, where U.S. bond yields are larger than Euro Area bound yields, prevails until 2008 or the early stages of the financial crisis. Interestingly, the period roughly between 1997 and 2004 is "restless" in that several narrow

spikes in the estimated mixing weights occur. Because no marked increases appear in the level of the series it seems reasonable to relate these spikes to the rather large differences between the variances in the two regimes. Although the second regime is here dominating observations are occasionally generated by the first AR component whose estimated error variance is six times the estimated error variance of the first AR component. However, despite these large shocks from the first AR component the level of the series has remained rather low. To discuss this point in more detail, recall that the mixing weights $\alpha_{1,t}$ and $\alpha_{2,t}$ depend on the density functions $\mathsf{n}_2(\boldsymbol{y}_{t-1};\boldsymbol{\vartheta}_1)$ and $\mathsf{n}_2(\boldsymbol{y}_{t-1};\boldsymbol{\vartheta}_2)$ where $\boldsymbol{y}_{t-1} = (y_{t-1}, y_{t-2})$. As Figure 4 indicates, the density function $\mathsf{n}_2(\boldsymbol{y}_{t-1};\boldsymbol{\vartheta}_2)$ is concentrated on the lower tail of $\mathsf{n}_2(\boldsymbol{y}_{t-1};\boldsymbol{\vartheta}_1)$ (see also the estimates in the lower panel of Table 2). This means that it is possible for the process to be on both tails of these two dimensional distributions and at the same time not far from the mean of $\mathsf{n}_2(\boldsymbol{y}_{t-1};\boldsymbol{\vartheta}_2)$. Switching from the second regime to the first one can then happen without much increase in the level of the series. This seems to have happened between 1997 and 2004 when the series has evolved in the second regime and the lower tail of the first regime. Based on the time series of estimated mixing weights $\hat{\alpha}_{1,t}$ a reasonable interpretation is that the second regime has been dominating and the process has only occasionally paid short visits to the (lower tail of the) first regime. The domination of the second regime has been clearer from 2005 until the early stages of the 2008 financial crisis when the first regime becomes dominating. During the last couple of years the estimated mixing weights $\hat{\alpha}_{1,t}$ have part of the time been very high but the level of the series has still remained rather moderate. Again, it seems reasonable to think that the dominance of the first regime is mainly caused by its large variance. This interpretation, as well as the one related to the narrow spikes between 1997 and 2004, is supported by the time series graph of the conditional variance implied by the estimated model. Without showing this graph we just note that its shape is more or less indistinguishable from the time series graph of the estimated mixing weight $\hat{\alpha}_{1,t}$ in the left panel of Figure 2.

To gain further insight into the preceding discussion Figure 5 depicts the estimated mixing weight $\hat{\alpha}_{1,t}$ as a function of y_{t-1} and y_{t-2}. The functional form is similar to an overturned version of the estimated density function $\mathsf{n}_2(\boldsymbol{y}_{t-1};\hat{\boldsymbol{\vartheta}}_2)$. Outside an ellipse, roughly corresponding to an ellipse where the estimated density $\mathsf{n}_2(\boldsymbol{y}_{t-1};\hat{\boldsymbol{\vartheta}}_2)$ has nonnegligible mass, the estimated mixing weight $\hat{\alpha}_{1,t}$ is nearly unity and in the center of this ellipse, or close to the point where $y_{t-1} = y_{t-2} \approx -0.5$, the estimated mixing weight $\hat{\alpha}_{1,t}$ attains its minimum value. The closer the series is to this minimum point the clearer it evolves in the lower regime and when it approaches the border of the aforementioned ellipse the

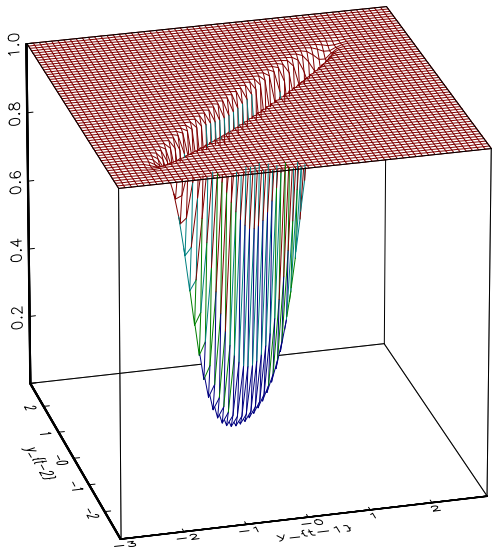

Figure 5: Estimated mixing weight $\hat{\alpha}_{1,t}$ of the restricted GMAR(2,2) model described in Table 2.

probability of switching to the upper regime increases. The spikes in the time series graph of $\hat{\alpha}_{1,t}$ in Figure 2 (left panel) between 1997 and 2004 have apparently occurred when the series has been close to the border of this ellipse. It is interesting to note that the spikes before 2001 have occurred when the level of the series is quite low so that the series has evolved in a way which has increased the (conditional) probability of obtaining an observation from the upper regime but without much increase in the level of the series. As Figure 5 illustrates, this is possible. A feature like this may be difficult to capture by previous mixture AR models as well as by TAR and STAR models in which regime switches are basically determined by the level of the series. For instance, in the model of Lanne and Saikkonen (2003) the probability of a regime switch is determined by the level of a single lagged value of the series and similarly for TAR and STAR models (see Tong (1990), Teräsvirta (1994), and Teräsvirta, Tjøstheim, and Granger (2010)). In the models of Wong and Li (2001b) and Dueker, Sola, and Spagnolo (2007) regime switches are determined by the level of a linear combination of a few past values of the series but this hardly makes any major difference at least for persistent series like the one we consider.

5 Conclusion

This paper provides a more detailed discussion of the mixture AR morel considered by Glasbey (2001) in the first order case. This model, referred to as the GMAR model, has several attractive properties. Unlike other nonlinear AR models the GMAR model has a clear probability structure which translates into simple conditions for stationarity and ergodicity. These theoretical features are due to the definition of the mixing weights which have a natural interpretation. As our empirical example demonstrates, the GMAR model also appears flexible in applications being able to describe features in the data that may be difficult to capture by alternative (nonlinear) AR models.

In this paper we have only considered a univariate version of the GMAR model. In the future we plan to explore a multivariate extension. Providing a detailed analyses of the asymptotic theory of estimation and statistical inference is another topic left for future work. In this context, the problem of developing statistical tests that can be used to test for the number of AR components is of special interest. Due to its nonstandard nature this testing problem may be quite challenging, however. Finally, applications of the GMAR model to different data sets will also be presented.

6 Technical Appendix

Proof of Theorem 1. We first note some properties of the stationary auxiliary autoregressions $\nu_{m,t}$. Denoting $\boldsymbol{\nu}_{m,t}^{+} = (\nu_{m,t}, \boldsymbol{\nu}_{m,t-1})$, it is seen that $\boldsymbol{\nu}_{m,t}^{+}$ follows the $(p+1)$–dimensional multivariate normal distribution with density

$$\begin{aligned} \mathsf{n}_{p+1}\left(\boldsymbol{\nu}_{m,t}^{+}; \boldsymbol{\vartheta}_m\right) &= (2\pi)^{-(p+1)/2} \det(\boldsymbol{\Sigma}_{m,p+1})^{-1/2} \\ &\quad \times \exp\left\{-\frac{1}{2}\left(\boldsymbol{\nu}_{m,t}^{+} - \mu_m \mathbf{1}_{p+1}\right)' \boldsymbol{\Sigma}_{m,p+1}^{-1}\left(\boldsymbol{\nu}_{m,t}^{+} - \mu_m \mathbf{1}_{p+1}\right)\right\}, \end{aligned}$$

where $\mathbf{1}_{p+1} = (1, \ldots, 1)$ $((p+1) \times 1)$ and the matrices $\boldsymbol{\Sigma}_{m,p+1}$, $m = 1, \ldots, M$, have the usual symmetric Toeplitz form similar to their counterparts in (7) with each $\boldsymbol{\Sigma}_{m,p+1}$ depending on the parameters $\boldsymbol{\varphi}_m$ and σ_m (see, e.g., Reinsel (1997, Sec. 2.2.3)). This joint density can be decomposed as

$$\mathsf{n}_{p+1}\left(\boldsymbol{\nu}_{m,t}^{+}; \boldsymbol{\vartheta}_m\right) = \mathsf{n}_1\left(\nu_{m,t} \mid \boldsymbol{\nu}_{m,t-1}; \boldsymbol{\vartheta}_m\right) \mathsf{n}_p\left(\boldsymbol{\nu}_{m,t-1}; \boldsymbol{\vartheta}_m\right), \tag{13}$$

where the normality of the two densities on the right-hand side follows from properties of the multivariate normal distribution (see, e.g., Anderson (2003, Theorems 2.4.3 and

2.5.1)). Moreover, $\mathsf{n}_p(\cdot;\boldsymbol{\vartheta}_m)$ clearly has the form given in (7), and making use of the Yule-Walker equations (see, e.g., Box, Jenkins, and Reinsel (2008, p. 59)), it can be seen that

$$\begin{aligned}\mathsf{n}_1(\nu_{m,t} \mid \boldsymbol{\nu}_{m,t-1};\boldsymbol{\vartheta}_m) &= \left(2\pi\sigma_m^2\right)^{-1/2}\exp\left\{-\frac{1}{2\sigma_m^2}(\nu_{m,t}-\mu_m-\boldsymbol{\varphi}_m'(\boldsymbol{\nu}_{m,t-1}-\mu_m\mathbf{1}_p))^2\right\}\\ &= \left(2\pi\sigma_m^2\right)^{-1/2}\exp\left\{-\frac{1}{2\sigma_m^2}(\nu_{m,t}-\varphi_{m,0}-\boldsymbol{\varphi}_m'\boldsymbol{\nu}_{m,t-1})^2\right\}. \qquad (14)\end{aligned}$$

The rest of the proof makes use of the theory of Markov chains (for the employed concepts, see Meyn and Tweedie (2009)). To make the Markov chain representation of $\boldsymbol{y}_t$ explicit we denote $\imath_p = (1,0,\ldots,0)$ $(p\times 1)$, and for $m=1,\ldots,M$,

$$\Phi_m = \begin{bmatrix} \varphi_{m,1} & \cdots & \cdots & \varphi_{m,p} \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}.$$

Then (3) can be written as

$$\boldsymbol{y}_t = \sum_{m=1}^{M} \boldsymbol{s}_{t,m}\left(\varphi_{m,0}\imath_p + \Phi_m\boldsymbol{y}_{t-1} + \sigma_m\varepsilon_t\imath_p\right),$$

making clear that $\boldsymbol{y}_t$ is a Markov chain on $\mathbb{R}^p$.

Now, let $\boldsymbol{y}_0 = (y_0,\ldots,y_{-p+1})$ be a random vector whose distribution has the density

$$f(\boldsymbol{y}_0;\boldsymbol{\theta}) = \sum_{m=1}^{M}\alpha_m\mathsf{n}_p(\boldsymbol{y}_0;\boldsymbol{\vartheta}_m).$$

According to (1) and (14), the conditional density of y_1 given $\boldsymbol{y}_0$ is

$$\begin{aligned} f(y_1 \mid \boldsymbol{y}_0;\boldsymbol{\theta}) &= \sum_{m=1}^{M}\alpha_{m,1}\mathsf{n}_1(y_1 \mid \boldsymbol{y}_0;\boldsymbol{\vartheta}_m)\\ &= \sum_{m=1}^{M}\frac{\alpha_m}{\sum_{n=1}^{M}\alpha_n\mathsf{n}_p(\boldsymbol{y}_0;\boldsymbol{\vartheta}_n)}\mathsf{n}_p(\boldsymbol{y}_0;\boldsymbol{\vartheta}_m)\,\mathsf{n}_1(y_1 \mid \boldsymbol{y}_0;\boldsymbol{\vartheta}_m)\\ &= \sum_{m=1}^{M}\frac{\alpha_m}{\sum_{n=1}^{M}\alpha_n\mathsf{n}_p(\boldsymbol{y}_0;\boldsymbol{\vartheta}_n)}\mathsf{n}_{p+1}((y_1,\boldsymbol{y}_0);\boldsymbol{\vartheta}_m), \end{aligned}$$

where the second and third equalities are due to (8) and (13). It thus follows that the density of $(y_1,\boldsymbol{y}_0) = (y_1,y_0,\ldots,y_{-p+1})$ is

$$f((y_1,\boldsymbol{y}_0);\boldsymbol{\theta}) = \sum_{m=1}^{M}\alpha_m\mathsf{n}_{p+1}((y_1,\boldsymbol{y}_0);\boldsymbol{\vartheta}_m).$$

Integrating y_{-p+1} out it follows that the density of $\boldsymbol{y}_1$ is

$$f(\boldsymbol{y}_1;\boldsymbol{\theta}) = \sum_{m=1}^{M} \alpha_m \mathsf{n}_p\left(\boldsymbol{y}_1;\boldsymbol{\vartheta}_m\right)$$

Therefore, $\boldsymbol{y}_0$ and $\boldsymbol{y}_1$ are identically distributed. As $\{\boldsymbol{y}_t\}_{t=1}^{\infty}$ is a (time homogeneous) Markov chain, we can thus conclude that $\{\boldsymbol{y}_t\}_{t=1}^{\infty}$ has a stationary distribution $\pi_{\boldsymbol{y}}(\cdot)$, say, characterized by the density

$$f(\cdot;\boldsymbol{\theta}) = \sum_{m=1}^{M} \alpha_m \mathsf{n}_p\left(\cdot;\boldsymbol{\vartheta}_m\right)$$

(cf. Meyn and Tweedie (2009, pp. 230-231)). Being a mixture of multivariate normal distributions, all moments of the stationary distribution are finite.

It remains to establish ergodicity. To this end, let $P_{\boldsymbol{y}}^{p}(\boldsymbol{y},\cdot) = \Pr(\boldsymbol{y}_p \mid \boldsymbol{y}_0 = \boldsymbol{y})$ signify the p-step transition probability measure of $\boldsymbol{y}_t$. It is straightforward to check that $P_{\boldsymbol{y}}^{p}(\boldsymbol{y},\cdot)$ has a density given by

$$f(\boldsymbol{y}_p \mid \boldsymbol{y}_0;\boldsymbol{\theta}) = \prod_{t=1}^{p} f(y_t \mid \boldsymbol{y}_{t-1};\boldsymbol{\theta}) = \prod_{t=1}^{p}\sum_{m=1}^{M} \alpha_{m,t}\mathsf{n}_1\left(y_t \mid \boldsymbol{y}_{t-1};\boldsymbol{\vartheta}_m\right).$$

The last expression makes clear that $f(\boldsymbol{y}_p \mid \boldsymbol{y}_0;\boldsymbol{\theta}) > 0$ for all $\boldsymbol{y}_p \in \mathbb{R}^p$ and all $\boldsymbol{y}_0 \in \mathbb{R}^p$ so that, from every initial state $\boldsymbol{y}_0 = \boldsymbol{y}$ $(\in \mathbb{R}^p)$, the chain $\boldsymbol{y}_t$ can in p steps reach any set of the state space $\mathbb{R}^p$ with positive Lebesgue measure. Using the definitions of irreducibility and aperiodicity we can therefore conclude that the chain $\boldsymbol{y}_t$ is irreducible and aperiodic (see Meyn and Tweedie (2009, Chapters 4.3 and 5.4)). Moreover, also the p-step transition probability measure $P_{\boldsymbol{y}}^{p}(\boldsymbol{y},\cdot)$ is irreducible, aperiodic, and has $\pi_{\boldsymbol{y}}$ as its invariant distribution (Meyn and Tweedie, 2009, Theorem 10.4.5).

A further consequence of the preceding discussion is that the p-step transition probability measure $P_{\boldsymbol{y}}^{p}(\boldsymbol{y},\cdot)$ is equivalent to the Lebesgue measure on $\mathbb{R}^p$ for all $\boldsymbol{y} \in \mathbb{R}^p$. As the stationary probability measure $\pi_{\boldsymbol{y}}(\cdot)$ also has a (Lebesgue) density positive everywhere in $\mathbb{R}^p$ it is likewise equivalent with the Lebesgue measure on $\mathbb{R}^p$. Consequently, the p-step transition probability measure $P_{\boldsymbol{y}}^{p}(\boldsymbol{y},\cdot)$ is absolutely continuous with respect to the stationary probability measure $\pi_{\boldsymbol{y}}(\cdot)$ for all $\boldsymbol{y} \in \mathbb{R}^p$.

To complete the proof, we now use the preceding facts and conclude from Theorem 1 and Corollary 1 of Tierney (1994) that $\left\|P_{\boldsymbol{y}}^{pn}(\boldsymbol{y},\cdot) - \pi_{\boldsymbol{y}}(\cdot)\right\| \to 0$ as $n \to \infty$ for all $\boldsymbol{y} \in \mathbb{R}^p$, where $\|\cdot\|$ signifies the total variation norm of probability measures. Now, by Proposition 13.3.2 of Meyn and Tweedie (2009), also $\left\|P_{\boldsymbol{y}}^{n}(\boldsymbol{y},\cdot) - \pi_{\boldsymbol{y}}(\cdot)\right\| \to 0$ as $n \to \infty$ for all $\boldsymbol{y} \in \mathbb{R}^p$

(as the total variation norm is non-increasing in n). Hence, $\boldsymbol{y}_t$ is ergodic in the sense of Meyn and Tweedie (2009, Ch. 13). ■

Remark. In the discussion following Theorem 1 it was pointed out that non-ergodicity would be an undesirable implication for a process having all finite dimensional distributions being Gaussian mixtures. To see that this holds in a particular special case, suppose all finite dimensional distributions of a process x_t, say, are Gaussian mixtures of the form (9) so that, for any $T \geq 1$, the distribution of a realization $(x_1, ..., x_T)$ is

$$f(\boldsymbol{x}; \boldsymbol{\theta}) = \sum_{m=1}^{M} \alpha_m \mathsf{n}_T (\boldsymbol{x}; \boldsymbol{\vartheta}_m),$$

where the density function $\mathsf{n}_T (\boldsymbol{x}; \boldsymbol{\vartheta}_m)$ is a T–dimensional analog of that in (7). The process x_t is clearly stationary. For simplicity, consider the special case where $M = 2$, $\varphi_{m,i} = 0$ $(i = 1, \ldots, p,\ m = 1, 2)$, $\sigma_1 = \sigma_2 = \sigma$, and $\varphi_{1,0} \neq \varphi_{2,0}$. Then $\mathsf{n}_T (\boldsymbol{x}; \boldsymbol{\vartheta}_i)$ is the joint density of T independent Gaussian random variables with mean $\varphi_{i,0}$ and variance σ^2 $(i = 1, 2)$. This means that, for every T,

$$(x_1, ..., x_T) \sim \begin{cases} \mathsf{n}_T (\boldsymbol{x}; \boldsymbol{\vartheta}_1), & \text{with probability } \alpha_1 \\ \mathsf{n}_T (\boldsymbol{x}; \boldsymbol{\vartheta}_2), & \text{with probability } 1 - \alpha_1. \end{cases}$$

This implies that, for every T, the sample mean $\bar{X}_T = T^{-1} \sum_{t=1}^{T} x_t$ is distributed as $N(\varphi_{1,0}, \sigma^2/T)$ with probability α_1 and as $N(\varphi_{2,0}, \sigma^2/T)$ with probability $1 - \alpha_1$. As $\varphi_{1,0} \neq \varphi_{2,0}$ and $0 < \alpha_1 < 1$ is assumed, it is therefore immediate that no law of large numbers holds, and consequently the process x_t cannot be ergodic. Indeed, it is not difficult to check that $\bar{X}_T$ converges in distribution to a random variable taking the values $\varphi_{1,0}$ and $\varphi_{2,0}$ with probability α_1 and $1 - \alpha_1$, respectively.

References

[1]

Anderson, T. W. (2003): *An Introduction to Multivariate Statistical Analysis*, 3rd edn. Wiley, Hoboken NJ.

Banzal, R., and H. Zhou (2002): "Term Structure of Interest Rates with Regime Shifts," *The Journal of Finance*, 57, 1997–2043.

Bec, F., A. Rahbek, and N. Shephard (2008): "The ACR model: a multivariate dynamic mixture autoregression," *Oxford Bulletin of Economics and Statistics*, 70, 583–618.

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel (2008): *Time Series Analysis: Forecasting and Control*, 4th edn. Wiley, Hoboken NJ.

Del Negro, M. and F. Schorfheide (2011): Bayesian Macroeconometrics, Chapter 7 in the *Handbook of Bayesian Econometrics*, ed. by J.F. Geweke, G. Koop, and H.K. van Dijk, Oxford University Press, Oxford.

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

Dueker, M. J., Z. Psaradakis, M. Sola and F. Spagnolo (2011): "Multivariate contemporaneous-threshold autoregressive models," *Journal of Econometrics*, 160, 311–325.

Dueker, M. J., M. Sola and F. Spagnolo (2007): "Contemporaneous threshold autoregressive models: estimation, testing and forecasting," *Journal of Econometrics*, 141, 517–547.

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

Garcia, R., and P. Perron (1996): "An Analysis of the Real Interest Rate Under Regime Shifts," *The Review of Economics and Statistics*, 78, 111–125.

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

Glasbey, C. A. (2001): "Non-linear autoregressive time series with multivariate Gaussian mixtures as marginal distributions," *Journal of the Royal Statistical Society: Series C*, 50, 143–154.

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

Haggan, V., and T. Ozaki (1981): "Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model," *Biometrika*, 68, 189–196.

Hamilton, J. (1988): "Rational-expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates," *Journal of Economic Dynamics and Control*, 12, 385–423.

Hamilton, J, D. (1995): *Time Series Analysis*, Princeton University Press, Princeton.

Kalliovirta, L. (2012): "Misspecification tests based on quantile residuals," *The Econometrics Journal*, forthcoming.

Lanne, M., and P. Saikkonen (2003): "Modeling the U.S. short-term interest rate by mixture autoregressive processes," *Journal of Financial Econometrics*, 1, 96–125.

Le, N. D., R. D. Martin, and A. E. Raftery (1996): "Modeling flat stretches, bursts, and outliers in time series using mixture transition distribution models," *Journal of the American Statistical Association*, 91, 1504–1515.

McLachlan, G. J., and T. Krishnan (2008): *The EM Algorithm and Extensions*, 2nd edn. Wiley, Hoboken NJ.

Meyn, S., and R. L. Tweedie (2009): *Markov Chains and Stochastic Stability*, 2nd edn. Cambridge University Press, Cambridge.

Reinsel, G. C. (1997): *Elements of Multivariate Time Series Analysis*, 2nd edn. Springer, New York.

Sisson, S. A. (2005): Trans-dimensional Markov chains: A decade of progress and future perspectives, *Journal of the American Statistical Association*, 100, 1077–1089.

Teräsvirta, T. (1994): "Specification, estimation, and evaluation of smooth transition autoregressive models," *Journal of the American Statistical Association*, 89, 208–218.

Teräsvirta, T., D. Tjøstheim and C. W. J. Granger (2010): *Modelling Nonlinear Economic Time Series*. Oxford University Press, Oxford.

Tierney, L. (1994): "Markov chains for exploring posterior distributions," *Annals of Statistics*, 22, 1701–1762.

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

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

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

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

Wu, C. F. J (1983): "On the convergence properties of the EM algorithm," *Annals of Statistics*, 11, 95–103.