Liuc Papers n. 169, Serie Metodi quantitativi 16, maggio 2005

MIXTURE OF AUTOREGRESSIVE COMPONENTS FOR MODELING FINANCIAL MARKET VOLATILITY

Stefano Sampietro

Introduction

This paper presents a Bayesian analysis of a mixture model for dependent data. In particular, we deal with a mixture of normal distributions whose means are linear functions of the past values of the observed variables. Since the component densities of the mixture can be viewed as the conditional distributions of different Gaussian autoregressive models, the model will be referred as mixture of autoregressive components.

Our model can also be regarded as a generalization of the standard "mixture of normals" model, since it allows dependence between the observable variables.

Wong and Li (2000) introduced the mixture of autoregressive models, performing a numerical maximum likelihood estimation. In order to choose among competing models, they consider both the AIC (Akaike 1973) and the BIC (Schwarz 1978) criterion, although they point out that these selection procedures are not particularly appropriate in this context.

We use a fully Bayesian approach relying on Reversible Jump sampler (Green 1995) and MCMC methods for the evaluation of marginal likelihoods (Chib and Jeliazkov 2001).

Our analysis takes into account the stationarity conditions on the autoregressive parameters by means of a reparametrization in terms of partial autocorrelations (Barndorff-Nielsen and Schou 1973; Barbieri and O'Hagan 1997). Usually, it is difficult to consider prior setting compatible with such constraints and Bayesian approaches often ignore them (see e.g. Zellner 1971; Broemeling and Shaaraway 1988).

Finally, we consider an application to the return volatility of financial markets. The assumptions of the model are indeed consistent with some stylized facts about volatility which are recognized in the financial literature, like persistence, clustering effects, nonsymmetrical dependencies (see e.g. Poterba and Summers 1986; Schwert 1987; French, Schwert and Stambaugh 1987; Hsieh 1991; Jones, Lamont and Lumsdaine 1998). A special advantage of our

model consists in taking into account model uncertainty which is particulary important in a financial context. Influence of this kind of uncertainty on financial models has been investigated by some recent papers (Barberis 2000; MacKinley and Pastor 2000; Pastor 2000; Pastor and Stambaugh 2000; Cremers 2002).

Mixture of autoregressive components

Let y_ts be the collection of observable variables. The mixture of autoregressive components can be defined by

$$y_t|\ldots \sim \sum_{j=1}^{k} w_j \, \mathrm{N}(y_t|\nu_{j,t}, \sigma_j^2), \tag{1}$$

where "$|\ldots$" is used to denote conditioning on the past observations and on all other variables, $\mathrm{N}(x|a,b)$ stands for the univariate Normal distribution with mean a and variance b, and

$$\nu_{j,t} = \mu_j + \phi_{1,j}(y_{t-1} - \mu_j) + \cdots + \phi_{\rho_j,j}(y_{t-\rho_j} - \mu_j), \tag{2}$$

for $j = 1, \ldots, k$.

Note that equation (2) is the conditional mean of a stationary autoregressive model, with order ρ_j, stationary mean μ_j and autoregressive coefficients $\phi_j = (\phi_{1,j}, \ldots, \phi_{\rho_j,j})$. If the number k of the mixture components is equal to 1, the model (1) reduces to a simple autoregressive model with Gaussian error term.

As previously mentioned, another special case of the model is the mixture of independent normal distributions, obtained simply by letting $\rho_j = 0$, so that $\nu_{j,t} = \mu_j$ for $j = 1, \ldots, k$.

The mixing weights w_j satisfy the constraints:

$$w_j > 0, \quad j = 1, \ldots, k; \qquad w_1 + \cdots + w_k = 1.$$

The conditional expectation of y_t is

$$E(y_t|\ldots) = \sum_{j=1}^{k} w_j \nu_{j,t},$$

while the conditional variance depends on the conditional means and is given by

$$VAR(y_t|\ldots) = \sum_{j=1}^{k} w_j \sigma_j^2 + \sum_{j=1}^{k} w_j \nu_{j,t}^2 - \left(\sum_{j=1}^{k} w_j \nu_{j,t} \right)^2. \tag{3}$$

Expression (3) shows that the model allows a time-varying conditional variance.

In this paper we focus on the so-called *direct application* of finite mixture models (Titterington, Smith and Makov 1985): the existence of k underlying categories is assumed and each

of the observed variables belongs to only one of these categories. From this point of view, the mixing weight w_j represents the probabilities that an observation comes from the j-th component.

Clearly, we do not observe the component of y_t directly. As a consequence, it is natural to consider a latent variable z_t which is a "component label": for every t, $z_t = j$ if the t-th observation comes from the j-th component. Probabilistically, we assume that the z_is are discrete random variables independently drawn from the discrete distribution

$$Pr(z_t = j|k, w_1, \ldots, w_k) = w_j, \qquad j = 1, \ldots, k.$$

Note that, conditionally on z_t, the observations are drawn from their respective individual subpopulation, i.e.

$$y_t|z_t = j, \cdots \sim \mathrm{N}(y_t|\nu_{j,t}, \sigma_j^2). \tag{4}$$

Prior distributions and stationary conditions

For fixed k, let $w = (w_1, \ldots, w_k)$, $\mu = (\mu_1, \ldots, \mu_k)$, $\sigma^2 = (\sigma_1^2, \ldots, \sigma_k^2)$. We set the following prior distributions for the parameters w, μ and σ^2

$$w \sim \mathrm{Di}(w|\delta_1, \delta_2, \ldots, \delta_k), \tag{5}$$

$$\mu_j \overset{iid}{\sim} \mathrm{N}(\mu_j|\mu_0, \tau^2), \qquad j = 1, \ldots, k, \tag{6}$$

$$\sigma_j^2 \overset{iid}{\sim} \mathrm{Ig}(\sigma^2|\alpha, \beta), \qquad j = 1, \ldots, k, \tag{7}$$

where $\mathrm{Di}(w|\delta_1, \delta_1, \ldots, \delta_k)$ denotes the Dirichlet distribution with $E(w_j) = \delta_j / \sum_{l=1}^{k} \delta_l$, for $j = 1, \ldots, k$, while $\mathrm{Ig}(\sigma_j^2|\alpha, \beta)$ denotes the Inverted-Gamma distribution with $E(\sigma_j^2) = \beta/(\alpha - 1)$, for $\alpha > 1$.

These prior distributions are conventional choices for a mixture model (Diebolt and Robert 1994; Richardson and Green 1997). Unfortunately, noninformative priors do not lead to proper posterior distributions in a mixture context (Diebolt and Robert 1994).

The parameters $\delta_1, \ldots, \delta_k$, μ_0, τ^2, α and β are assumed to be fixed. We take the prior for μ_j to be essentially flat over the range of the observed data R_y, where $R_y = \max(y) - \min(y)$, and where y is the vector of the observations. Following Richardson and Green (1997), we choose $\mu_0 = \min(y) + R_y/2$ and τ equal to a small multiple of R_y, a typical choice being $\tau = R_y$.

We also use the knowledge of the range of the data in setting the hyperparameters of the distribution of σ^2, in particular, β will be proportional to $1/R^2$ (see below for details). It is also possible to consider an additional hierarchical level (for instance, β could have a gamma distribution, as suggested by Richardson and Green 1997).

In this analysis we consider the existence of the stationarity regions for the autoregressive coefficients ϕ. Apart from the cases of small orders, it is well known that the form of these regions for an autoregressive model is very complex. As a consequence, a Bayesian analysis can be difficult in terms of prior specification and these constraints have been often ignored (Zellner 1971; Broemeling and Shaaraway 1988).

Let Φ_j be the stationarity region for the j-th autoregressive component and let $\pi_{h,j}$ be the partial autocorrelation coefficient at lag h for the j-th model. Clearly $-1 < \pi_{h,j} < 1$, for $h = 1, \ldots, \rho_j$, and $\pi_{h,j} = 0$, for $h > \rho_j$. In addition, $\pi_{1,j}$ equals the first autocorrelation coefficient.

Barndorff-Nielsen and Schou (1973) proposed a useful reparametrization in terms of the partial autocorrelation coefficients, $\phi_j \to \pi_j$, where $\phi_j = (\phi_{1,j}, \ldots, \phi_{\rho_j,j})$ and $\pi_j = (\pi_{1,j}, \ldots, \pi_{\rho_j,j})$, with

$$\phi_j \in \Phi_j \Longleftrightarrow |\pi_{h,j}| < 1, \quad h = 1, \ldots, \rho_j.$$

That is, the stationarity region for the j-th model in terms of π_j is simply the hypercube $(-1,1)^{\rho_j}$. For details see Barndorff-Nielsen and Schou (1973).

We can specify a prior distribution for $\phi = (\phi_1, \ldots, \phi_k)$ because of the following result (Jones 1987)

$$\phi_j \sim \text{Uniform on } \Phi_j \Longleftrightarrow \pi_{i,j} \overset{ind}{\sim} \text{Be}_{(-1,+1)}\left(\pi_{i,j} \middle| \left[\frac{i+1}{2}\right], \left[\frac{i}{2}\right] + 1\right),$$

for $i = 1, \ldots, \rho_j$ and $j = 1, \ldots, k$, where $\text{Be}_{(-1,+1)}(x|a,b)$ denotes a generalized beta distribution (see e.g. Bernardo and Smith 1994, pp. 116-117) defined on $(-1,+1)$ and where $[a]$ means "integer part of a". In other words, we put a uniform prior distribution for the original parameters on the complicated stationarity region Φ_j, simply choosing a generalized beta prior for the $\pi_{i,j}$s on $(-1,+1)$.

The last part concerns the parameters k (the number of the autoregressive components) and $\rho^k = (\rho_1, \ldots, \rho_k)$ (the orders of the autoregressive components). Clearly, the dimension of ρ^k depends on k but, for notational simplicity, we henceforth write $\rho^k = \rho$. Notice that k and ρ uniquely identifies a specific model and we stress that they are random quantities, representing the distinctive feature of our analysis.

The priors we set are

$$k \sim \mathrm{Un}(k|1, k_{max}), \tag{8}$$

$$\rho_j|k \overset{iid}{\sim} \mathrm{Un}(\rho_j|0, \rho_{max}), \qquad j = 1, \ldots, k, \tag{9}$$

where $\mathrm{Un}(x|a, b)$ denotes the discrete uniform distribution for $x = a, a+1, \ldots, b$ and where ρ_{max} and k_{max} are fixed (details on their values will be given in section 6).

Dealing with mixture models involves the so-called label switching problem (see Marin, Mengersen and Robert 2004), which derives from the symmetry in the likelihood of the parameter labels. In a Bayesian analysis, if we have no prior information that distinguishes between the components of the mixture (that is, the joint prior distribution is invariant to permutations of the parameters), then the posterior distribution will be similarly symmetric (Frühwirth-Schnatter 2001). As a result, the posterior distribution shows artificial multimodality, which poses obvious problems in terms of parameter estimations.

The usual solution of the label switching problem consists in imposing an *identifiability constraint* on the parameter space, such as $\mu_1 < \mu_2 < \ldots, < \mu_k$. This kind of constraints can be satisfied by only one permutation of the parameter label and this breaks the symmetry of the prior.

Parameter estimation

Let $m = (k, \rho)$ be the model index, and denote with $\theta = (w, \mu, \sigma^2, \phi)$ the corresponding vector of parameters (a more rigorous notation would employ θ_m instead of θ, in order to emphasize the dependence of the parameters on the model). In this section we discuss parameter estimation for a given model, while model determination will be explained in the next one.

In order to calculate the conditional posterior distribution $p(\theta|y, m)$, we implemented a standard *MCMC* componentwise algorithm. The vector θ is split into four blocks corresponding to the parameter (w, μ, σ^2, ϕ) and each block is updated separately. In addition, it is necessary to consider a further move for the allocation variable z.

The set of moves is

i. Updating the weights w,

ii. Updating the means μ,

iii. Updating the partial autocorrelations π,

iv. Updating the variances σ^2,

v. Updating the allocation variable z.

Moves (i), (ii), (iv) and (v) are Gibbs type move. The full conditionals are respectively

$$w|\cdots \sim \mathrm{Di}(w|\delta_1 + n_1, \ldots, \delta_k + n_k), \tag{10}$$

$$\mu_j|\cdots \sim \mathrm{N}\left(\mu_j \left| \frac{n_j\,\bar{v}_j\,B_j\,\tau^2 + \sigma_j^2\mu_0}{n_j\,B_j^2\tau^2 + \sigma_j^2}, \frac{\sigma_j^2\tau^2}{n_j\,B_j^2\tau^2 + \sigma_j^2}\right.\right), \tag{11}$$

$$\sigma_j^2|\cdots \sim \mathrm{Ig}\left(\sigma_j^2 \left| \alpha + \frac{1}{2}n_j, \beta + \frac{1}{2}\sum_{t:z_t=j}(y_t - \nu_{j,t})^2\right.\right), \tag{12}$$

$$Pr(z_t = j|\ldots) \propto \frac{w_j}{\sigma_j}\ \exp\left\{-\frac{1}{2}\frac{(y_t - \nu_{j,t})^2}{\sigma_j^2}\right\}. \tag{13}$$

where $n_j = \sum_{t:z_t=j} 1$, $\bar{v}_j = \frac{1}{n_j}\sum_{t:z_t=j} v_{tj}$, with $v_{tj} = y_t - \phi_{j,1}y_{t-1} - \cdots - \phi_{j,\rho_j}y_{t-\rho_j}$, and where $B_j = 1 - \phi_{j,1} - \cdots - \phi_{j,\rho_j}$.

Move (iii) updates π_j, for $j = 1, \ldots, k$, by a Metropolis-Hastings mechanism. A candidate $\pi_{j,i}^*$ is generated by a normal density truncated to $(-1, +1)$ and centered in the current state of the chain $\pi_{j,i}$. The variance σ_q^2 is tuned so that a satisfactory acceptance rate can be reached (possibly 20% or more).

Let π_j^* be the proposal vector for the partial autocorrelations: using π_j^*, the corresponding parameters $\phi_j^* = (\phi_{j,1}^*, \ldots, \phi_{j,\rho_j}^*)$ are derived trough the transformation previously illustrated.

The acceptance probability is $\min(1, R)$, where R is given by

$$\begin{aligned} R =& \exp\left\{-\frac{1}{2\sigma_j^2}\sum_{t:z_t=j}\left[(y_t - \nu_{j,t}^*)^2 - (y_t - \nu_{j,t})^2\right]\right\} \\ &\times \prod_{i=2}^{\rho_j}\frac{(\pi_{j,i}^* + 1)^{[(i-1)/2]}(1 - \pi_{j,i}^*)^{[i/2]}}{(\pi_{j,i} + 1)^{[(i-1)/2]}(1 - \pi_{j,i})^{[i/2]}} \\ &\times \prod_{i=1}^{\rho_j}\frac{F_N(1|\pi_{j,i}, \sigma_q^2) - F_N(-1|\pi_{j,i}, \sigma_q^2)}{F_N(1|\pi_{j,i}^*, \sigma_q^2) - F_N(-1|\pi_{j,i}^*, \sigma_q^2)}, \end{aligned} \tag{14}$$

where $\nu_{j,t}^* = \mu_j + \phi_{1,j}^*(y_{t-1} - \mu_j) + \cdots + \phi_{\rho_j,j}^*(y_{t-\rho_j} - \mu_j)$ and where $F_N(x|a, b)$ is the cumulative distribution function of the $N(x|a, b)$ distribution evaluated in x. Note that the first line of formula (14) corresponds to the likelihood ratio, the second one to the prior ratio and the third one to the proposal ratio.

Model determination

The model posterior distribution can be factorized as $p(m|y) = p(\rho|k, y)p(k|y)$. To update the orders, for given k, a reversible jump type move is added to the MCMC sequence described in the previous section. The derivation of the marginal posterior $p(k|y)$ involves method by Chib and Jeliazkov (2001).

Order of the autoregressive components

The reversible jump sampler (Green 1995) can be viewed as a Metropolis-Hastings method (Metropolis, Rosenbluth, Rosenbluth and Teller 1953; Hastings 1970) adapted for general state spaces. Suppose to denote the dimension of a parameter vector θ with $d(\theta)$. Let (θ, m) be the current value of the chain and let $g_{m,m'}$ be an invertible function. At each iteration, a candidate model m' is generated from $q(m'|m)$. Note that θ depends on m, so $d(\theta)$ can change. In order to have a dimension matching, a random vector u is generated from a proposal density $q(u|\theta, m, m')$. Set $(\theta', u') = g_{m,m'}(\theta, u)$, with $d(u') = d(\theta) + d(u) - d(\theta')$. The proposed value (θ', m') is accepted with probability

$$\alpha[(\theta, m), (\theta', m')] = \min\left(1, \frac{h(\theta', m')}{h(\theta, m)} \times \frac{q(m|m')q(u'|\theta', m', m)}{q(m'|m)q(u|\theta, m, m')} \times \left|\frac{\partial g_{m,m'}(\theta, u)}{\partial(\theta, u)}\right|\right),$$

where h is the distribution of interest. The final term is a Jacobian arising from the change of variable from (θ, u) to (θ', u').

The move which updates the orders $\rho = (\rho_1, \ldots, \rho_k)$ starts by selecting a component, say j^*, randomly chosen in $\{1, \ldots, k\}$. Then ρ_{j^*} increases by one with probability $b(\rho_{j^*})$ and decreases by one with probability $d(\rho_{j^*})$, where $b(\rho_j) = 1 - d(\rho_j)$, for $j = 1, \ldots, k$, $d(1) = 0$ and $b(\rho_{max}) = 0$.

It is now necessary to change the partial autocorrelation coefficients. Following Barbieri and O'Hagan (1997), if the order is decreased, the last partial autocorrelation is simply discarded. Otherwise, we need a new parameter $\pi^*_{\rho^*_{j^*},j^*}$, which is generated from the beta prior.

That is, letting $\pi^*_{j^*}$ be the proposal vector of the partial autocorrelations

- If $\rho^*_{j^*} = \rho_{j^*} - 1$, $\quad \pi^*_{j^*} = (\pi_{1,j^*}, \ldots, \pi_{\rho^*_{j^*},j^*})$
- If $\rho^*_{j^*} = \rho_{j^*} + 1$, $\quad \pi^*_{j^*} = (\pi_{1,j^*}, \ldots, \pi_{\rho_{j^*},j^*}, \pi^*_{\rho^*_{j^*},j^*})$

with $\pi^*_{\rho^*_{j^*},j^*} \sim \mathrm{Gb}\left(\pi_{i,j}|\left[\frac{i+1}{2}\right], \left[\frac{i}{2}\right] + 1\right)$.

Note that in both cases all the autoregressive parameters are updated because of the choosen reparametrization.

If $\rho^*_{j^*} = \rho_{j^*} - 1$, the acceptance probability ratio is $\min(1, R)$, where R is given by

$$R = \exp\left\{-\frac{1}{2\sigma_j^2}\sum_{t:z_t=j}\left[(y_t - \nu_{j,t})^2 - (y_t - \nu^*_{j,t})^2\right]\right\}\frac{b(\rho^*_{j^*})}{d(\rho_{j^*})} \tag{15}$$

where $\nu^*_{j^*,t} = \mu_{j^*} + \phi^*_{1,j^*}(y_{t-1} - \mu_{j^*}) + \cdots + \phi^*_{\rho_{j^*},j^*}(y_{t-\rho_{j^*}} - \mu_{j^*})$.

On the other hand, if $\rho^*_{j^*} = \rho_{j^*} + 1$

$$R = \exp\left\{-\frac{1}{2\sigma_j^2}\sum_{t:z_t=j}\left[(y_t - \nu_{j,t})^2 - (y_t - \nu^*_{j,t})^2\right]\right\}\frac{d(\rho^*_{j^*})}{b(\rho_{j^*})}. \tag{16}$$

The equations (15) and (16) are quite simple because of some cancellations between prior and proposal ratios. Furthermore, the Jacobian is one because the matrix of derivatives of the transformation g is the identity matrix.

Finally $p(\rho|k, y)$ is simply estimated by the proportions of every possible value for ρ in the sample obtained by the previous complete MCMC algorithm.

Number of the autoregressive components

Through Bayes' theorem, the marginal posterior distribution of k is $p(k|y) \propto p(k)f(y|k)$, where $p(k)$ is the prior on k and $f(y|k)$ is the marginal likelihood

$$f(y|k) = \sum_{\rho} \int f(y|\theta, \rho, k)p(\theta, \rho|k)\, d\theta. \tag{17}$$

The starting idea of the method by Chib and Jeliazkov (2001) arises from the basic marginal likelihood identity: suppressing for notational convenience the model index k, we can write

$$\begin{aligned} f(y) &= \frac{f(y|\theta^*, \rho^*)p(\theta^*, \rho^*)}{p(\theta^*, \rho^*|y)} \\ &= \frac{f(y|\theta^*, \rho^*)p(\theta^*|\rho^*)p(\rho^*)}{p(\theta^*|\rho^*, y)p(\rho^*|y)}, \end{aligned} \tag{18}$$

for any fixed point (θ^*, ρ^*). Note that the only quantity in (18) that is not already available is $p(\theta^*|\rho^*, y)$. Chib and Jeliazkov (2001) proposes an efficient method to produce the estimate $\bar{p}(\theta^*|\rho^*, y)$ using the output from a MCMC simulation with fixed ρ^*.

First of all, $\bar{p}(\theta^*|\rho^*, y)$ is factorized as

$$\begin{aligned} \bar{p}(\theta^*|\rho^*,y) = \bar{p}(\pi^*|\rho^*, y) \times \bar{p}(\mu^*|\pi^*, \rho^*, y)\times \\ \times \bar{p}(\sigma^{2*}|\mu^*, \pi^*, \rho^*, y) \times \bar{p}(w^*|\sigma^{2*}, \mu^*, \pi^*, \rho^*, y). \end{aligned} \tag{19}$$

Suppose to have a sample $\{\theta^{(i)}, z^{(i)}\}$, for $i = 1, \ldots, N_1$, from the MCMC algorithm for a given ρ^* (i.e. a sample from $p(\theta|\rho^*)$). Let $\eta_{j-1} = (\rho, \pi_1, \ldots, \pi_{j-1})$ and $\eta^{j+1} = (\pi_{j+1}, \ldots, \pi_k, \mu, \sigma^2, w)$. The terms of the (19) are estimated by the following steps

1. Sample $\{\tilde{\eta}^{j+1,(i)}, \tilde{z}^{(i)}\}$, for $i = 1, \ldots, N_{j+1}$, from a reduced MCMC algorithm with distribution of interest $p(\eta^{j+1}, z|\eta_j^*, y)$. Also draw $\tilde{\pi}_j^{(i)}$ from $q_p(\pi_j^*, \pi_j) = \prod_{s=1}^{\rho_j} q(\pi_{s,j}^*, \pi_{s,j})$, where $q(.,.)$ was previously defined.

   Set

$$\bar{p}(\pi_j^*|\rho^*, \pi_1^*, \ldots, \pi_{j-1}^*) = \frac{N_j^{-1} \sum_{i=1}^{N_j} \alpha(\pi_j^{(i)}, \pi_j^*)q_p(\pi_j^{(i)}, \pi_j^*)}{N_{j+1}^{-1} \sum_{i=1}^{N_{j+1}} \alpha(\pi_j^*, \tilde{\pi}_j^{(i)})},$$

   where $\alpha(.,.) = \min(1, R)$ with R defined in equation (14).

   Set $\eta^{j+1,(i)} = \tilde{\eta}^{j+1,(i)}$ and $z^{(i)} = \tilde{z}^{(i)}$, for $i = 1, \ldots, N_{j+1}$.

Repeat this step for $j = 1, \ldots, k$ and finally set

$$\bar{p}(\pi^*|\rho^*, y) = \prod_{j=1}^{k} \bar{p}(\pi_j^*|\rho^*, \pi_1^*, \ldots, \pi_{j-1}^*).$$

2. The second term is

$$\bar{p}(\mu^*|\pi^*, \rho^*, y) = N_{k+1}^{-1} \sum_{i=1}^{N_{k+1}} \prod_{j=1}^{k} p(\mu_j^*|\pi^*, \sigma^{2(i)}, z^{(i)}, \rho^*, y),$$

where $(\sigma^{2(i)}, z^{(i)})$ are draws from the last iteration of the previous step (thus they are marginally from $p(\sigma^2, w, z|\pi^*, \rho^*, y)$) and $p(\mu_j^*|\pi^*, \sigma^{2(i)}, z^{(i)}, \rho^*, y)$ is given by equation(11).

3. Sample $\{\sigma^{2(s)}, w^{(s)}, z^{(s)}\}$, for $s = 1, \ldots, S$, from a reduced MCMC algorithm with distribution of interest $p(\sigma^2, w, z|\pi^*, \mu^*, \rho^*, y)$ and set

$$\bar{p}(\sigma^{2*}|\pi^*, \mu^*, \rho^*, y) = S^{-1} \sum_{s=1}^{S} \prod_{j=1}^{k} p(\sigma_j^{2*}|\pi^*, \mu^*, z^{(s)}, \rho^*, y),$$

where $p(\sigma_j^{2*}|\pi^*, \mu^*, z^{(i)}, \rho^*, y)$ is given by equation (12).

4. Sample $\{w^{(v)}, z^{(v)}\}$, for $v = 1, \ldots, V$, from a reduced MCMC algorithm with distribution of interest $p(w, z|\pi^*, \mu^*, \sigma^{2*}, \rho^*, y)$ and set

$$\bar{p}(w^*|\pi^*, \mu^*, \sigma^{2*}, \rho^*, y) = V^{-1} \sum_{v=1}^{V} p(w^*|z^{(v)}, \rho^*, y),$$

where $p(w^*|z^{(v)}, \rho^*, y)$ is given by equation (10)

Return volatility

Modelling and forecasting return volatility is one of the most important tasks in financial markets. Within a rich literature (for a recent review see Poon and Granger 2003), several stylized facts have been recognized.

First of all, volatility is persistent (see e.g. Poterba and Summers 1986; Schwert 1987; French *et al.* 1987 and Hsieh 1991) and it can have long memory properties (see e.g. Ding, Granger and Engle 1993; Bollerslev and Mikkelsen 1996).

Second, observations of financial time series reveal volatility clustering. Autoregressive conditional heteroskedasticity (ARCH) models and stochastic volatility (SV) models (for a survey, see Bollerslev, Chou and Kroner 1992 and Ghysels *at al.* 1996 respectively) are well-known instruments proposed in literature and they are essentially built to mimic this volatility feature.

Moreover, volatility shows threshold effects, nonsymmetrical dependencies and mean reversion. In particular, it has been argued that volatility adjustments follow a twin-speed process: low volatility state is more persistent with respect to high state (Longing 1987; Jones *et al.* 1998). In order to consider changing volatility persistence, models in a regime switching framework have been proposed (Hamilton 1989; Cao and Tsay 1992; Gray 1996). Volatility asymmetry motivated several nonlinear GARCH type models, like the exponential GARCH (Nelson 1991), the quadratic GARCH (Engle 1990) and the JGR-GARCH (Glosten, Jagannathan and Runke 1993).

In this work we follow the approach which uses an observable proxy for the return volatility. A standard choice is daily squared returns which will be referred as volatility in the rest of the paper.

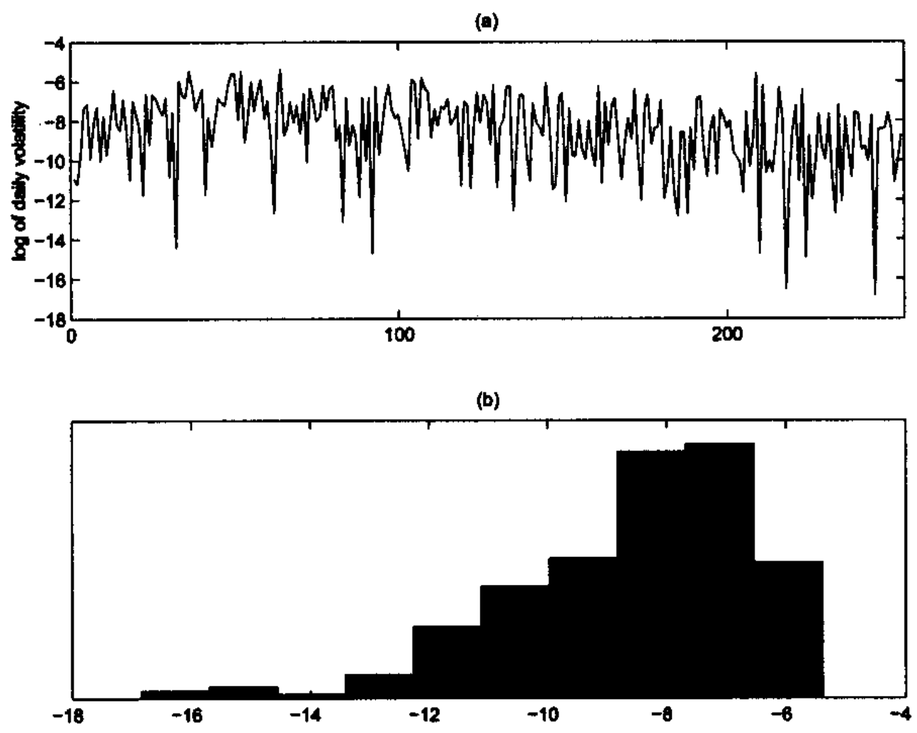

Figure 1: Plot (a) and histogram (b) of the logarithm of daily return volatility (Nasdaq index from May 15, 2002 to May 15, 2003)

We consider daily returns r_t on Nasdaq index from May 15, 2002 to May 15, 2003 (253 observations) and we take the logarithm of volatility: $y_t = ln(r_t^2)$. Time series plot (a) and histogram (b) of y_t are displayed in Figure 1.

About the fixed hyperparameters in the distributions (5), (6) and (7), we set $\delta_j = 1$, for every $j = 1, \ldots, k$, $\tau = R_y$, $\mu_0 = \min(y) + R_y/2$, $\alpha = 2$ and $\beta = 0.1 \cdot R_y^{-2}$. The hyperparameters of the prior on the model indexes (8) and (9) are $k_{max} = 5$ and $\rho_{max} = 9$: we believe these numbers are sufficiently high, since simulations with higher values gave the same results.

The number of iterations of the MCMC algorithm conditional on k was 50000, with a burn-in period of 10000. The successive reduced MCMC algorithms for the marginal likelihood were run with 15.000 iterations each one.

Remember that prior distributions for the model indexes are uniform, as reported in equa-

k	$p(k\|y)$
1	0
2	0.81
3	0.18
4	0.01
5	0

Table 1: Posterior probabilities of the number of mixture components

tions (9) and (8). The resulting joint prior follows a parsimony principle, penalizing complex models with a high number of components

$$p(\rho, k) = p(\rho|k)p(k) \propto \left(\frac{1}{\rho_{max}+1}\right)^k,$$

where $\propto$ denotes "proportional to". Anyway it is easy to convert results to those corresponding to alternative priors by the identity

$$\tilde{p}(\theta, m|y) \propto p(\theta, m|y)\frac{\tilde{p}(m)}{p(m)},$$

where $\tilde{p}(\cdot|y)$ is the posterior for a different prior $\tilde{p}$.

Table 1 shows the posterior probabilities of the number of mixture components k. The model with two components is largely preferred (0.81). On the grounds of many different simulations, we noted a sensitivity of these results with respect to some hyperparameters. An obvious result is that smaller values of the prior variance τ increase the preference for the one-component model. At any rate, in all the simulations we performed the one-component model never had the highest posterior probability. $p(k|y)$ is also sensitive with respect to the prior on the variances σ^2. The posterior probabilities of the models with a number of components greater than 2 increases if the prior mean of σ^2 decreases. This is due to the fact that smaller the variances of the components, the higher the number of components needed to fit the data.

Posterior probabilities of the orders of the autoregressive components are given in Table 2 for different values of k.

Information about the posterior distributions of some model parameters conditional on $k = 2$ are showed in Table 3 (posterior means and variances) and in Figure 2 (histograms of the posterior distributions).

Tables 2 and 3 suggest several considerations. First of all, conditional on $k = 1$, order zero has an high posterior probability (0.85). In other words, the one-component model, that is a simple autoregressive model, does not seem able to capture volatility persistence.

$k=1$		$k=2$		$k=3$	
(ρ_1)	$p(\rho_1\|k,y)$	(ρ_1,ρ_2)	$p(\rho_1,\rho_2\|k,y)$	(ρ_1,ρ_2,ρ_3)	$p(\rho_1,\rho_2,\rho_3\|k,y)$
(0)	0.85	(0,0)	0.46	(0,0,0)	0.09
(1)	0.09	(0,1)	0.03	(0,1,0)	0.04
(2)	0.03	(1,0)	0.24	(1,0,0)	0.08
(3)	0.02	(2,0)	0.06	(2,0,0)	0.05
		(3,0)	0.04	(3,0,0)	0.05
		(4,0)	0.03	(4,0,0)	0.04
		(5,0)	0.02	(5,0,0)	0.05
				(6,0,0)	0.04
				(7,0,0)	0.03
				(8,0,0)	0.03
				(9,0,0)	0.03

Table 2: Posterior probabilities (> 0.02) of autoregressive components' orders conditional on k.

	$\mathrm{E}(\cdot\|k=2,y)$	$\sqrt{Var(\cdot\|k=2,y)}$
w_1	0.31	0.10
w_2	0.69	0.10
μ_1	-11.07	2.23
μ_2	-7.69	0.21
σ_1^2	5.11	1.19
σ_2^2	1.65	0.36

Table 3: Posterior mean and standard deviation of the model parameters conditional on $k=2$

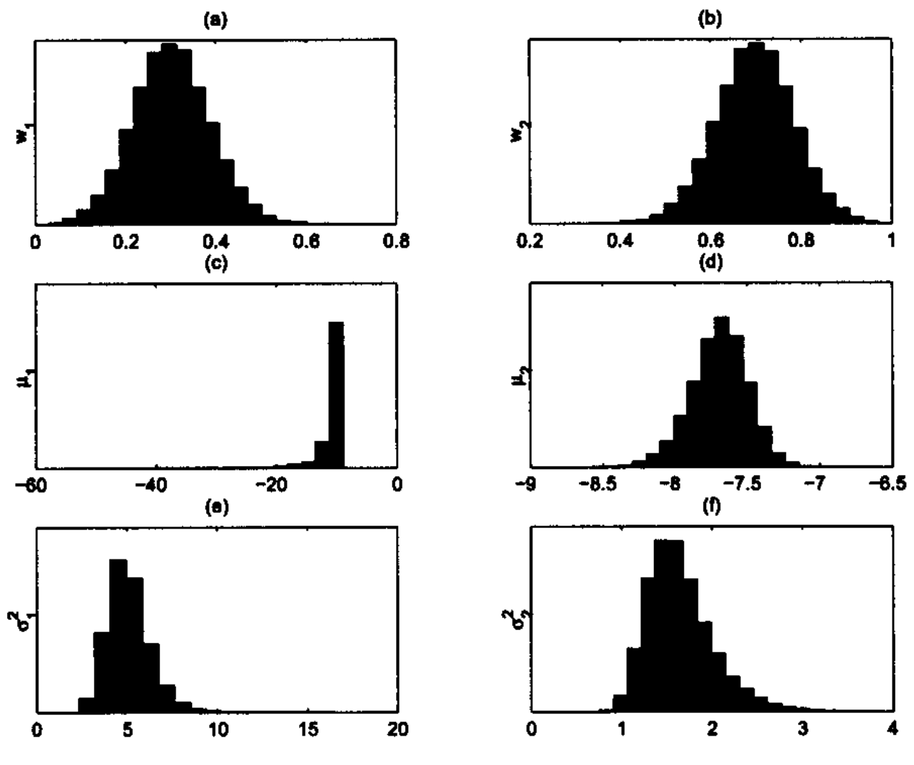

Figure 2: Posterior distributions of the weights [(a) and (b)], the means [(c) and (d)] and the variances [(e) and (f)] conditional on $k = 2$

On the other hand, with $k \geq 2$, the results show different levels of persistence according to a threshold effect. The threshold effect, which depends on the asymmetry of information among market agents, means that large shocks are less persistent than small shocks, so when volatility increases due to a big market movement, it comes back rapidly to its long-term level. Consider the model formed by two components: the first one corresponds to low volatility values ($\mu_1 = -11.07$, see Table 3) and it shows higher persistence with respect to the second component. In fact, models of the form $(\rho_1, \rho_2) = (r, 0)$, with $r = 0, \ldots, 9$, have a total posterior probability of 0.883. Simulations with $k \geq 3$ lead to a similar structure, in which a first component formalizes low and persistent volatility values while the remaining components are associated with zero autoregressive orders.

Table 4 shows posterior means and standard deviations of the autoregressive coefficients ϕ conditional on $k = 2$ and on different values of ρ. Figure 3 shows the posterior distributions of ϕ conditional on $k = 2$ and on $\rho = (1, 0)$ and $\rho = (2, 0)$.

Predictive distributions are evaluated by Monte Carlo estimators. For instance, the one-step predictive distribution for a future variable $\tilde{y}$ conditional on k

$$f(\tilde{y}|k, y) = \sum_{\rho} \int f(\tilde{y}|\theta, \rho, k, y) p(\theta, \rho|k, y) d\psi, \tag{20}$$

is estimated by

$$N^{-1} \sum_{i=1}^{N} f(\tilde{y}|\theta^{(i)}, \rho^{(i)}, k, y), \tag{21}$$

where $\theta^{(i)}$ and $\rho^{(i)}$, for $i = 1, \ldots, N$, are samples from the conditional posterior $p(\theta, \rho|k, y)$ and they are available from the MCMC output. The one-step predictive distributions conditional

(ρ_1, ρ_2)	$E(\phi_{1,\cdot}\|\rho, k, y)$	$E(\phi_{2,\cdot}\|\rho, k, y)$	$\sqrt{Var(\phi_{1,\cdot}\|\rho, k, y)}$	$\sqrt{Var(\phi_{2,\cdot}\|\rho, k, y)}$
$(0,0)$	-	-	-	-
$(0,1)$	-	0.015	-	0.054
$(1,0)$	-0.25	-	0.205	-
$(2,0)$	-0.182	-	0.256	-
	0.111		0.212	
$(3,0)$	-0.161	-	0.261	-
	0.156		0.214	
	0.211		0.217	
$(4,0)$	-0.197	-	0.19	-
	0.126		0.183	
	0.208		0.196	
	0.278		0.209	
$(5,0)$	-0.205	-	0.205	-
	0.094		0.194	
	0.124		0.236	
	0.295		0.179	
	0.177		0.181	

Table 4: Posterior means and standard deviations of the autoregressive coefficients ϕ conditional on $k = 2$ and on different values of ρ

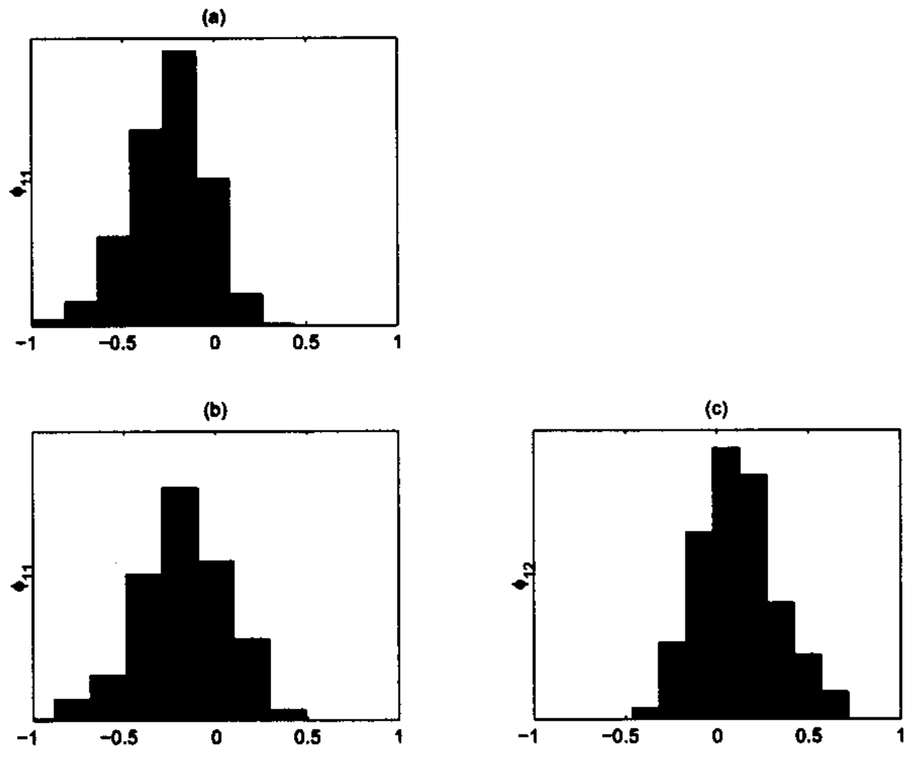

Figure 3: Posterior distributions of the autoregressive parameters conditional on $k = 2$ and $\rho = (1, 0)$ [(a)] and conditional on $\rho = (2, 0)$ [(b) and (c)]

on k are showed in figure 4. With a number of components greater than one, the predictive distribution is asymmetric and substantially unimodal.

The unconditional predictive distribution is similarly estimated and it is given in figure 5.

An out of sample analysis is summarized in figure 6, where the one-step predictive distributions conditional on $k = 2$ for some values of t are displayed. Actual values of y_t are also shown at the times t-1, t and t+1.

Discussion

We presented a fully Bayesian analysis to model nonlinear time series data. Both parameter estimation and model selection are based on MCMC techniques. In order to assess the performance of the algorithm, we tested it on several simulated data sets. We found satisfactory results in terms of model selection and our analysis does not seem to suffer from the difficulties inherent in more standard approaches.

The model can be extended in several ways. For instance, a mixture of autoregressive moving average (ARMA) components could be developed: the invertibility conditions of the moving average part of the model can be handled by a generalization of the reparametrization we used (Monaham 1984).

The normality of the component densities can also be easily relaxed and different conditional distributions can be used. Note that consideration of other distributions implies suitable changes in the moves of the MCMC algorithm.

Another possible development is related to the stationarity conditions. In this work, we

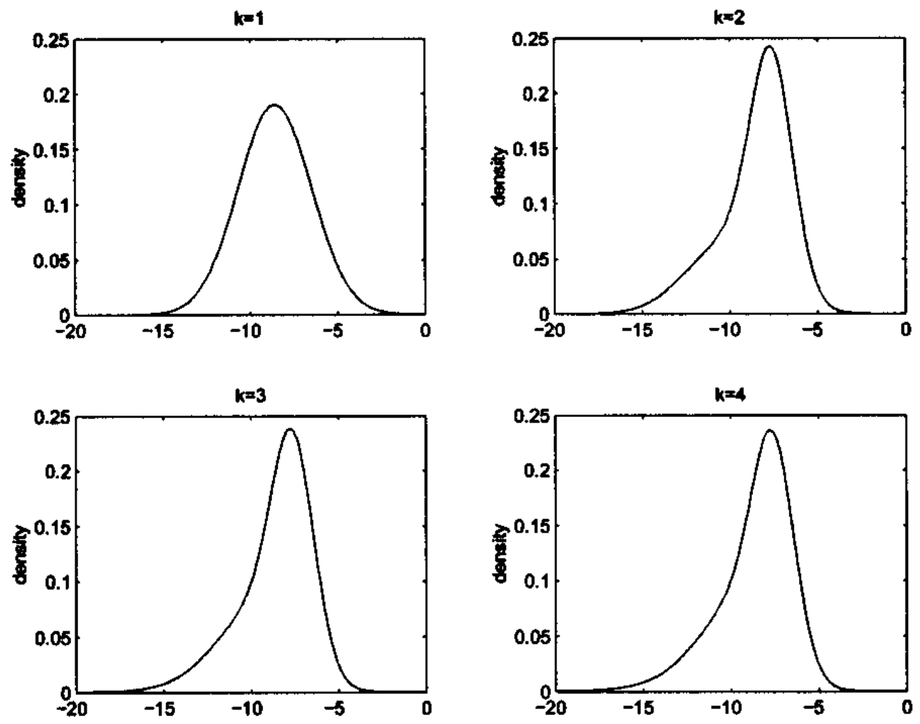

Figure 4: One-step predictive distributions conditional on the number of autoregressive components k

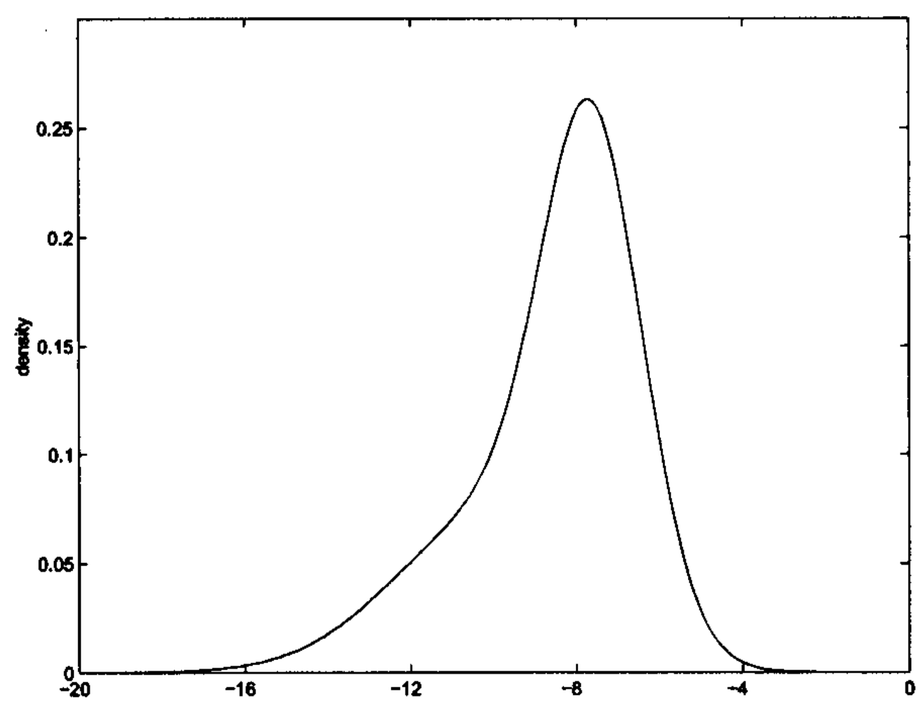

Figure 5: One-step unconditional predictive distribution

consider local conditions, that is within each autoregressive components. This local stationarity seems to be a sufficient condition for the global stationarity, but, as showed by Wong and Li (2000), a mixture of nonstationarity components can be stationary. A Bayesian analysis which takes into account global stationarity conditions should be feasible, although far from straightforward.

Acknowledgements

I would like to thank Professor Guido Consonni (University of Pavia) for his help and suggestions. I also thank Professor Petros Dellaportas (Athens University of Economics and Business) and Professor Fabio Trojani (University of the Southern Switzerland) for very useful discussions on the subject of this paper.

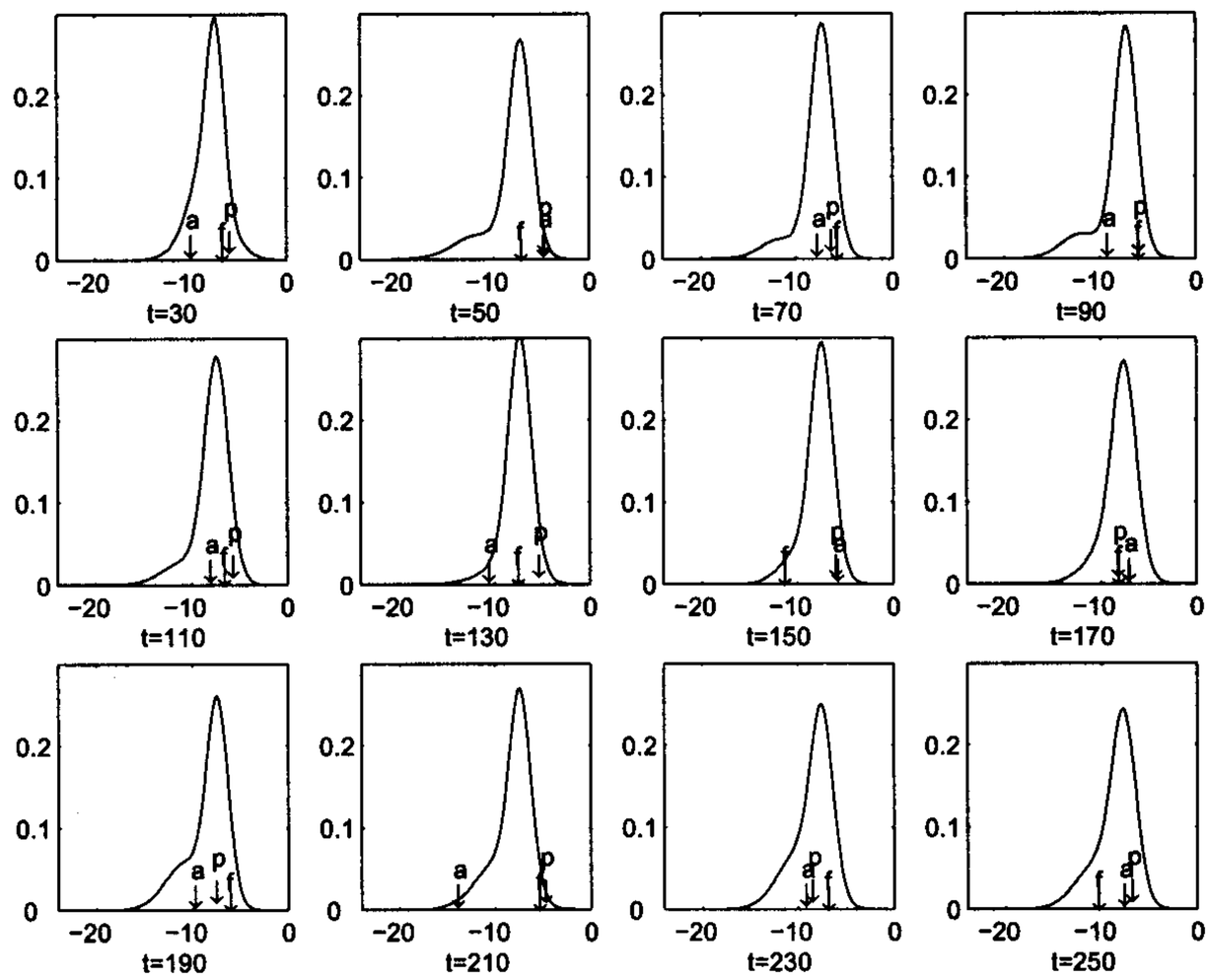

Figure 6: One-step predictive distributions conditional on $k = 2$ for different values of t. The actual values are shown (labels 'p', 'a' and 'f' stand for observations at times t-1, t and t+1 respectively)

References

Akaike, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle," in *Proceedings of the 2nd International Symposioum in Information Theory*, eds. B. N. Petrov and F. Csaki, Budapest: Akademiaia Kiado, pp. 267-281.

Barberis, N. (2000), "Investing for the Long Run when Returns are Predictable," *Journal of Finance*, 55, 225-264.

Barbieri, M. M., and O'Hagan, A. (1997), "A Reversible Jump MCMC Sampler for Bayesian Analysis of ARMA Time Series," manuscript.

Barndorff-Nielsen, O., and Schou, G. (1973), "On the Parametrization of Autoregressive Models by Partial Autocorrelation," *Journal of Multivariate Analysis*, 3, 408-419.

Bernardo, J. M., and Smith, A. F. M. (2000), *Bayesian Theory* (2nd ed.), New York: Wiley.

Bollerslev, T., Chou, R. Y., and Kroner, K. P. (1992), "ARCH Modeling in Finance: a Review of the Theory and Empirical Evidence," *Journal of Econometrics*, 52, 5-59.

Bollerslev, T., and Mikkelsen, H. O. A. (1996), "Modelling and Pricing Long Memory in Stock Market Volatility," *Journal of Econometric*, 73, 151-184.

Broemeling, L. D., and Shaaraway, S. (1988), "Time Series: a Bayesian Analysis in the Time Domain," in *Bayesian Analysis of Time Series and Dynamic Models*, ed. J.C. Spall, New York: Dekker, pp. 1-21.

Cao, C. Q., and Tsay, R. S. (1992), "Nonlinear Time-series Analysis of Stock Volatilities," *Journal of Applied Economterics*, December, Supplement, 1, 165-185.

Chib, S., and Jeliazkov, I. (2001), "Marginal Likelihood from the Metropolis-Hastings Output," *Journal of the American Statistical Association*, 96, 270-281.

Cremers, K. J. M. (2002), "Stock Return Predictability: a Bayesian Model Selection Perspective," *Review of Financial Studies*, 15, 1223-1249.

Diebolt, J., and Robert, C. P. (1994), "Estimation of Finite Mixture Distributions Through Bayesian Sampling," *Journal of the Royal Statistical Society*, Ser. B, 56, 363-375.

Ding, Z., Granger, J. W. J., and Engle, R. E. (1996), "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance*, 1, 83-106.

Engle, R. F. (1990), "Discussion: Stock Market Volatility and the Crash of 87," *The Review of Financial Studies*, 3, 103-106.

French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987), "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 19, 3-29.

Frühwirth-Schnatter, S. (2001), "Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models," *Journal of the American Statistical Association*, 96, 194-209.

Ghysels, E., Harvey, A., and Renault, E. (1996), "Stochastic Volatility," in *Handbook of Statistics: Statistical Methods in Finance*, Amsterdam: Elsevier Science, 14, 119-191.

Glosten, L. R., Jagannathan, R., and Runke, D. E. (1993), "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779-1801.

Gray, S. F. (1996), "Modeling the Conditional Distribution of Interest Rates as a Regime Switching Process," *Journal of Financial economics*, 42, 27-62.

Green, P. J. (1995), "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika* , 82, 711-732.

Hamilton, J. D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357-384.

Hastings, W. K. (1970), "Monte Carlo Sampling Methods using Markov Chain and their Applications," *Biometrika*, 57, 97-109.

Hsieh, D. A. (1991), "Chaos and Nonlinear Dynamics: Application to Financial Markets," *Journal of Finance*, 46, 1839-1877.

Jones, C., Lamont, O., and Lumsdaine, R. (1998), "Macroeconomic News and Bond Market Volatility," *Journal of Financial Economics*, 47, 315-337.

Jones, M. C. (1987), "Randomly Choosing Parameters from the Stationarity and Invertibility Region of Autoregressive-moving Average Models," *Applied Statistics*, 36, 134-138.

Longin, F. M. (1997), "The Threshold Effect in Expected Volatility: a Model based on Asymmetric Information," *The Review of Financial Studies*, 10, 837-869.

MacKinley, A. C., and Pastor, L. (2000), "Asset Pricing Models: Implications for Expected Returns and Portfolio Selection," *Review of Financial studies*, 13, 883-916.

Marin, J. M., Mengersen, K. and Robert, C. P. (in press), "Bayesian Modelling and Inference on Mixtures of Distributions," In *Handbook of Statistics*, vol. 25, Amsterdam: Elsevier-Sciences.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, W. N., and Teller, A. H. (1953), "Equations of State Calculations by Fast Computing Machines," *Journal of Chemical Physics*, 21, 1087-1091.

Monahan, J. F. (1984), "A Note on Enforcing Stationarity in Autoregressive-moving Average Models," *Biometrika*, 71, 403-404.

Nelson, D. B. (1991), "Conditional Heteroskedasticity in Asset Returns: a New Approach," *Econometrica*, 59, 347-370.

Pastor, L. (2000), "Portfolio Selection and Asset Pricing Models," *Journal of Finance*, 55, 179-224.

Pastor, L., and Stambaugh, R. F. (2000), "Comparing Asset Pricing Models: an Investment Perspective," *Journal of Financial Economics*, 56, 353-381.

Poon, S. H., and Granger, C. W. J. (2003), "Forecasting Volatility in Financial Markets," *Journal of Economic Literature*, 41, 478-539.

Poterba, J. M., and Summers, L. H. (1986), "The Persistence of Volatility and Stock Markets Fluctuations," *American Economic Review*, 76, 1142-1151.

Richardson, S., and Green, P. J.(1997), "On Bayesian Analysis of Mixtures with an Unknown Number of Components," *Journal of the Royal Statistical Society*, Ser. B, 59, 731-759.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461-464.

Schwert, G. W. (1987), "Effects of Model Specification on Tests for Unit Roots in Macroeconomic Data," *Journal of Monetary Economics*, 20, 73-103.

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

Wong, C. S., and Li, W. K. (2000), "On a Mixture Autoregressive Model," *Journal of the Royal Statistical Society*, Ser. B, 62, 95-115.

Zellner, A. (1971), *An Introduction to Bayesian Inference and Econometrics*, New York: Wiley.