On a Mixture Autoregressive Conditional Heteroscedastic Model

Chun Shan WONG and Wai Keung LI

We propose a mixture autoregressive conditional heteroscedastic (MAR-ARCH) model for modeling nonlinear time series. The models consist of a mixture of K autoregressive components with autoregressive conditional heteroscedasticity; that is, the conditional mean of the process variable follows a mixture AR (MAR) process, whereas the conditional variance of the process variable follows a mixture ARCH process. In addition to the advantage of better description of the conditional distributions from the MAR model, the MAR-ARCH model allows a more flexible squared autocorrelation structure. The stationarity conditions, autocorrelation function, and squared autocorrelation function are derived. Construction of multiple step predictive distributions is discussed. The estimation can be easily done through a simple EM algorithm, and the model selection problem is addressed. The shape-changing feature of the conditional distributions makes these models capable of modeling time series with multimodal conditional distributions and with heteroscedasticity. The models are applied to two real datasets and compared to other competing models. The MAR-ARCH models appear to capture features of the data better than the competing models.

KEY WORDS: Autocorrelation; EM algorithm; Model selection; Predictive distributions; Stationarity.

1. INTRODUCTION

In earlier work (Wong and Li 2000), we introduced the mixture autoregressive (MAR) model as a powerful tool for modeling nonlinear time series. The MAR(K; $p_1, \ldots, p_K$) model is defined as

$$F(y_t|\mathcal{F}_{t-1}) = \sum_{k=1}^{K} \alpha_k \Phi\left(\frac{y_t - \phi_{k0} - \phi_{k1}y_{t-1} - \cdots - \phi_{kp_k}y_{t-p_k}}{\sigma_k}\right)$$

where $\Phi(\cdot)$ is the (conditional) cumulative distribution function of the standard Gaussian distribution. These models can be considered as a generalization of Gaussian mixture transition distribution (GMTD) models introduced by Le, Martin, and Raftery (1996). Certain features make this a promising model: the flexibility to mix a nonstationary AR process with a stationary AR process, the ability to model multimodal conditional distributions, and the ability to capture conditional heteroscedasticity. We have also demonstrated the usefulness of the MAR model with some financial data and the classical Canadian lynx data.

Despite the aforementioned interesting properties, the MAR model suffers from a limitation in the modeling of nonlinear time series. The squared autocorrelation structure of the MAR model is quite simple and is analogous to that of the AR model (Granger and Newbold 1986, p. 309; Wong 1998). The absence of a squared autocorrelation structure is a shortfall that limits application of the MAR model to financial data especially; see Section 5.2 for an example. Gray (1996) also reported large squared residual autocorrelations when a mixture of two mean reversion models is fitted to some interest rates data.

In this article we generalize the MAR models to the mixture autoregressive conditional heteroscedastic (MAR-ARCH) models. These models consist of a mixture of K autoregressive components with autoregressive conditional heteroscedasticity; that is, the conditional mean of y_t follows an AR process, and the conditional variance of y_t follows an ARCH process (Engle 1982). Unlike Cai (1994), Hamilton and Susmel (1994), and Gray (1996), we do not assume a latent Markov structure for the regime indicator variable at each t. Rather, we assume that these indicators are independent and that each follows the discrete distribution with probabilities $\alpha_1, \ldots, \alpha_K$, which can be regarded as an exchangeable prior (Chen and Liu 1995). This results in a somewhat simpler, yet sufficiently flexible model. The predictive distribution at each t has a nice closed form, and (as in Le et al. 1996), flat stretches, bursts, and outliers can be easily incorporated into the model. Furthermore, a latent Markov structure may be absent or inappropriate in, say, a short series. Recently, Huerta, Jiang, and Tanner (1999) considered a hierarchical mixture time series model where the covariate space is partitioned into O overlapping regions and within each region M models compete with each other. Their model is clearly more general and worthy of further investigation.

The stationarity properties and the autocorrelation function of a MAR-ARCH model can be easily derived. The squared autocorrelation function of a MAR-ARCH model can also be derived. The estimation can be done via the EM algorithm as in the case of the MAR model, with some modification. The MAR-ARCH model retains all of the interesting properties of the MAR model. The flexible squared autocorrelation structure makes the MAR-ARCH model a promising model for financial time series. In fact, it is possible to mix both explosive and inexplosive ARCH components, and yet the series is still second-order stationary. On the other hand, it is not clear whether a general Markov switching model has this property. For model selection, we use a minimum BIC procedure.

In Section 2 we describe the MAR-ARCH model, the stationarity conditions, the autocorrelation function, the squared autocorrelation function, and the predictive distributions. We discuss the estimation procedures in Section 3 and also derive the observed information matrix is also derived. We discuss

Chun Shan Wong is Research Assistant Professor and Wai Keung Li is Professor, Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong (E-mail: *chun-shan.wong@graduate.hku.hk*). The authors thank the editor, the associate editor, and two referees for comments that led to improvement of the article. Li's research is supported by the Hong Kong Research Grants Council, and Wong's research is supported by the Committee on Research and Conference Grants, The University of Hong Kong.

Journal of the American Statistical Association
September 2001, Vol. 96, No. 455, Theory and Methods

the model selection problem in Section 4 and illustrate the model's usefulness by two examples in Section 5. In the Appendixes, we give the proofs of the theorems and the justification of our EM procedure.

2. THE MIXTURE AUTOREGRESSIVE CONDITIONAL HETEROSCEDASTIC MODEL

The K-component MAR-ARCH model under consideration is defined by

$$F(y_t|\mathcal{F}_{t-1}) = \sum_{k=1}^{K} \alpha_k \Phi\left(\frac{e_{k,t}}{\sqrt{h_{k,t}}}\right),$$

$$e_{k,t} = y_t - \phi_{k0} - \phi_{k1} y_{t-1} - \cdots - \phi_{kp_k} y_{t-p_k},$$

and

$$h_{k,t} = \beta_{k0} + \beta_{k1} e_{k,t-1}^2 + \cdots + \beta_{kq_k} e_{k,t-q_k}^2. \tag{1}$$

We denote this model by MAR-ARCH $(K; p_1, p_2, \ldots, p_K; q_1, q_2, \ldots, q_K)$. Here $F(y_t|\mathcal{F}_{t-1})$ is the conditional cumulative distribution function of Y_t given the past information, evaluated at y_t; $\mathcal{F}_t$ is the information set up to time t; $\Phi(\cdot)$ is the (conditional) cumulative distribution function of the standard Gaussian distribution; and $\alpha_1 + \cdots + \alpha_K = 1$, $\alpha_k > 0$ $(k = 1, \ldots, K)$. Let $p = \max(p_1, \ldots, p_K)$, $q = \max(q_1, \ldots, q_K)$, and $\phi(\cdot)$ be the probability density function of a standard normal distribution. To avoid the possibility of zero or negative conditional variance, the following conditions for β_{ki}'s must be imposed: $\beta_{k0} > 0$ $(k = 1, \ldots, K)$, and $\beta_{ki} \geq 0$ $(i = 1, \ldots, q_k; k = 1, \ldots, K)$.

The MAR-ARCH model is actually a mixture of K autoregressive models with ARCH innovations. The properties of the MAR-ARCH model are similar to those of the MAR model of Wong and Li (2000). For example, the shape of the conditional distribution of the series changes over time as the conditional means and variances of the components, which depend on past values of time series in different ways, differ. The shape of the conditional distribution can be multimodal. Note that the conditional expectation of y_t given the past,

$$\sum_{k=1}^{K} \alpha_k (\phi_{k0} + \phi_{k1} y_{t-1} + \cdots + \phi_{kp_k} y_{t-p_k}) = \sum_{k=1}^{K} \alpha_k \mu_{k,t},$$

may not be the best predictor of the future values.

An important feature of the MAR-ARCH model is its greater flexibility in modeling changing conditional variance. The conditional variance of y_t is given by

$$\operatorname{var}(y_t|\mathcal{F}_{t-1}) = \sum_{k=1}^{K} \alpha_k h_{k,t} + \sum_{k=1}^{K} \alpha_k \mu_{k,t}^2 - \left(\sum_{k=1}^{K} \alpha_k \mu_{k,t}\right)^2.$$

The first term allows modeling of the dependence of the conditional variance on the past "errors." The second and third terms model the change of the conditional variance resulting from the difference in the conditional means of the components.

Now we give the results for the stationarity of the MAR-ARCH model. As we show later, it is possible that some or all components are nonstationary but the MAR-ARCH model is stationary. The results also confirm the intuition that the MAR-ARCH model is (first- and second-order) stationary if each AR-ARCH component is stationary.

The first-order stationarity conditions for the MAR-ARCH model (1) are given in Theorem 1. Derivation of the second-order stationarity conditions for the general MAR-ARCH model is more complicated, because it involves the autocovariance terms and the unconditional expectation of the $e_{k,t}$'s. We give only the second-order stationarity condition for the MAR-ARCH $(K; 1, \ldots, 1; q_1, \ldots, q_K)$ models in Theorem 2. The second-order stationarity condition for the MAR-ARCH $(K; 0, \ldots, 0; q_1, \ldots, q_K)$ can be obtained by letting $\phi_{k1} = 0$, $k = 1, \ldots, K$, in Theorem 2. The proofs of the theorems are similar to those of Le et al. (1996) and Wong and Li (2000), who used a result from Beněs (1967), and are given in Appendix A. The derivation of the second-order stationarity conditions for the general MAR-ARCH models is similar to the derivation of the conditions for the MAR-ARCH $(K; 1, \ldots, 1; q_1, \ldots, q_K)$ model.

Theorem 1. A necessary and sufficient condition for the process Y_t to be stationary in the mean is that the roots $z_1, \ldots, z_p$ of the equation

$$1 - \sum_{i=1}^{p} \left(\sum_{k=1}^{K} \alpha_k \phi_{ki}\right) z^{-i} = 0 \tag{2}$$

all lie inside the unit circle, where $\phi_{ki} = 0$ for $i > p_k$.

Theorem 2. Suppose that the process Y_t follows a MAR-ARCH $(K; 1, \ldots, 1; q_1, \ldots, q_K)$ model is first-order stationary. The necessary and sufficient condition for the process to be second-order stationary is that the roots $z_1, \ldots, z_{q+1}$ of the equation

$$\begin{aligned} 1 &- \sum_{k=1}^{K} \alpha_k (\beta_{k1} + \phi_{k1}^2) z^{-1} \\ &- \sum_{i=2}^{q} \sum_{k=1}^{K} \alpha_k \left[\beta_{ki} + \beta_{k,i-1} \phi_{k1} \left\{\phi_{k1} - 2\left(\sum_{j=1}^{K} \alpha_j \phi_{j1}\right)\right\}\right] z^{-i} \\ &- \sum_{k=1}^{K} \alpha_k \beta_{kq} \phi_{k1} \left\{\phi_{k1} - 2\left(\sum_{j=1}^{K} \alpha_j \phi_{j1}\right)\right\} z^{-q-1} = 0 \end{aligned} \tag{3}$$

all lie inside the unit circle, where $\beta_{ki} = 0$ for $i > q_k$.

It can be observed from the theorems that the stationarity conditions depend largely on the mixing proportions of the components, that is, the values of the α_k's. For the MAR-ARCH (2;0,0;1,1) model, the condition for second-order stationarity is $|\alpha_1 \beta_{11} + \alpha_2 \beta_{21}| < 1$ or, equivalently, $\alpha_1 \beta_{11} + \alpha_2 \beta_{21} < 1$, as the term on the left side must be nonnegative. Note that the MAR-ARCH (2;0,0;1,1) model must be first-order stationary. It is possible for one of the components to assume an "explosive" ARCH process and yet for the time series to still be considered second-order stationary.

For the MAR-ARCH (2;1,1;1,1) model, the condition for first-order stationarity is $|\alpha_1 \phi_{11} + \alpha_2 \phi_{21}| < 1$. It is easily shown that the conditions for second-order stationarity are

$$\sum_{k=1}^{2}\alpha_k\beta_{k1}\phi_{k1}^2-2\Bigl(\sum_{k=1}^{2}\alpha_k\beta_{k1}\phi_{k1}\Bigr)\Bigl(\sum_{j=1}^{2}\alpha_j\phi_{j1}\Bigr)+\sum_{k=1}^{2}\alpha_k(\beta_{k1}+\phi_{k1}^2)<1,$$

$$\sum_{k=1}^{2}\alpha_k\beta_{k1}\phi_{k1}^2-2\Bigl(\sum_{k=1}^{2}\alpha_k\beta_{k1}\phi_{k1}\Bigr)\Bigl(\sum_{j=1}^{2}\alpha_j\phi_{j1}\Bigr)-\sum_{k=1}^{2}\alpha_k(\beta_{k1}+\phi_{k1}^2)<1,$$

$$\left|\sum_{k=1}^{2}\alpha_k\beta_{k1}\phi_{k1}^2-2\Bigl(\sum_{k=1}^{2}\alpha_k\beta_{k1}\phi_{k1}\Bigr)\Bigl(\sum_{j=1}^{2}\alpha_j\phi_{j1}\Bigr)\right|<1.$$

It is possible for a component to be an explosive AR process and/or a nonstationary ARCH process yet for the time series to still be considered second-order stationary. Clearly this does not imply stationarity in distribution, which is an interesting open problem. We illustrate this point with a simulated series from the following MAR-ARCH (2;1,1;1,1) model (A): $F(y_t|\mathcal{F}_{t-1}) = .75\Phi(e_{1,t}/\sqrt{h_{1,t}})+.25\Phi(e_{2,t}/\sqrt{h_{2,t}})$, $e_{1,t}=y_t-.5y_{t-1}$, $h_{1,t}=1.0+.5e_{1,t-1}^2$, $e_{2,t}=y_t-1.1y_{t-1}$, and $h_{2,t}=1.0+1.2e_{2,t-1}^2$. The time plot and the sample autocorrelation functions of the two simulated series are shown in Figures 1 and 2. The simulated time series appears to be stationary. Note that the sample autocorrelation functions in Figure 2 resemble those of an AR(1) process, because model (A) is a mixture of AR(1)-ARCH(1) components.

The autocorrelations of the time series generated by the MAR-ARCH model are similar to those of the MAR model, which in turn are similar to those of an AR model. The autocorrelations for the MAR-ARCH model satisfy a system of equations similar to the Yule–Walker equations for the ordinary AR(p) process, where the lag i coefficient of the AR(p) process is replaced by the coefficient $\sum_{k=1}^{K}\alpha_k\phi_{ki}$. As a generalization of the MAR model, the range of possible autocorrelations is as great as that of the standard AR process (see Le et al. 1996 and Wong and Li 2000 for discussion).

Modifying the MAR-ARCH model (1) to handle nonstationary time series is not difficult. This can be done by restricting one of the roots of the equation $1-\phi_{k1}z^{-1}-\cdots-\phi_{kp_k}z^{-p_k}=0$ to be 1 for each of the K components. This is equivalent to fitting a stationary MAR-ARCH model for the differenced series y_t-y_{t-1}.

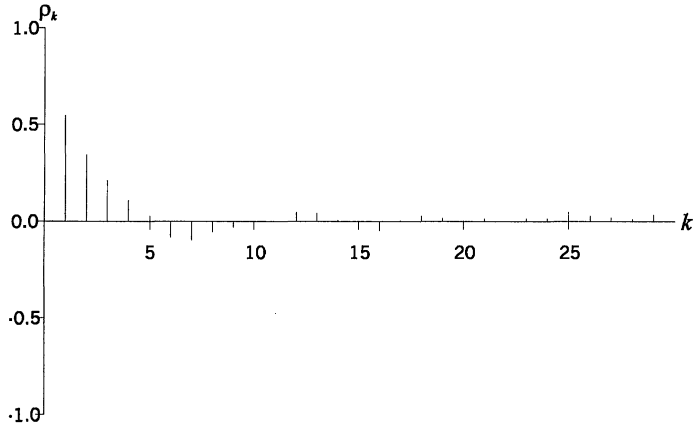

Figure 2. Sample Autocorrelations for the Simulated Time Series of Model (A).

We can also derive the condition for the existence of higher-order moments of the general MAR-ARCH model. However, the derivation is quite complicated, because it involves the expectation of the product of squared y_t's. Here we give only the condition for the existence of the fourth-order moment of a stationary MAR-ARCH $(K;0,\ldots,0;1,\ldots,1)$ model with $\phi_{k0}=0$ $(k=1,\ldots,K)$ in Theorem 3. The proof is given in Appendix A. Note that the first- and third-order moments of this model are both 0, whereas the second-order moment is given by $\sum_{k=1}^{K}\alpha_k\beta_{k0}/(1-\sum_{k=1}^{K}\alpha_k\beta_{k1})$.

Theorem 3. The fourth-order moment of a stationary MAR-ARCH $(K;0,\ldots,0;1,\ldots,1)$ model with $\phi_{k0}=0$ $(k=1,\ldots,K)$ exists if and only if

$$\sum_{k=1}^{K}\alpha_k\beta_{k1}^2<\frac{1}{3}. \tag{4}$$

Theorem 3 shows that the condition for the existence of the fourth-order moment of a MAR-ARCH $(K;0,\ldots,0;1,\ldots,1)$ model is similar to that of a first-order linear ARCH process, where the squared lag 1 coefficient of the ARCH(1) process is replaced by the coefficient $\sum\alpha_k\beta_{k1}^2$.

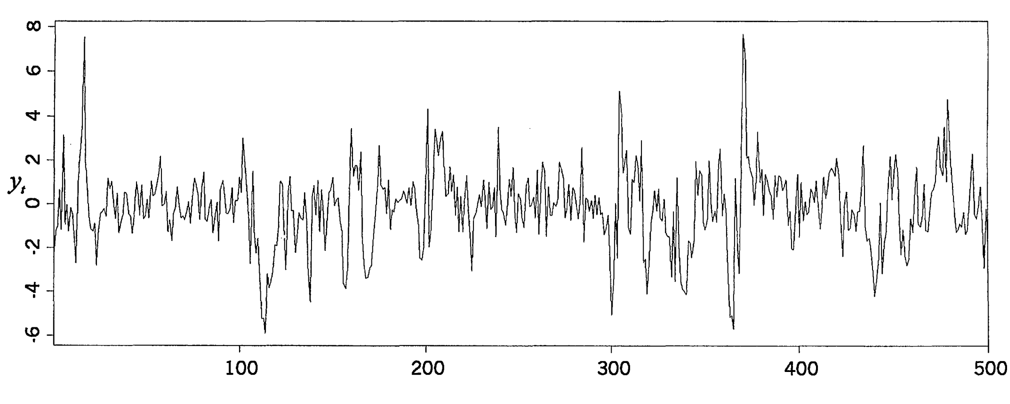

Figure 1. Simulated Time Series of Model (A).

An expression for the fourth-order moment is given in the proof. Note that the kurtosis of Y_t generally is greater than three.

The squared autocorrelations of the time series generated by a stationary MAR-ARCH model with finite fourth moment should be similar to those of an ARCH model. As an example, we consider a stationary MAR-ARCH $(K; 0, \dots, 0; 1, \dots, 1)$ model with $\phi_{k0} = 0$ $(k = 1, \dots, K)$ that satisfy the condition (4). The expectation of the product of y_t^2 and y_{t-l}^2 is given by

$$\begin{aligned}
E(Y_t^2 Y_{t-l}^2) &= E\Big\{Y_{t-l}^2 E\Big(Y_t^2 | \mathcal{F}_{t-1}\Big)\Big\} \\
&= E\Big\{Y_{t-l}^2\Big(\sum \alpha_k \beta_{k0} + \sum \alpha_k \beta_{k1} Y_{t-1}^2\Big)\Big\} \\
&= E\Big[Y_{t-l}^2\Big\{\sum \alpha_k \beta_{k0} + \sum \alpha_k \beta_{k1}\Big(\sum \alpha_k \beta_{k0} \\
&\qquad + \sum \alpha_k \beta_{k1} Y_{t-2}^2\Big)\Big\}\Big] \\
&\ \ \vdots \qquad\qquad \vdots \\
&= E\Big(Y_{t-l}^2\Big[\sum \alpha_k \beta_{k0}\Big\{1 + \sum \alpha_k \beta_{k1} + \Big(\sum \alpha_k \beta_{k1}\Big)^2 \\
&\quad + \cdots + \Big(\sum \alpha_k \beta_{k1}\Big)^{l-1}\Big\} + \Big(\sum \alpha_k \beta_{k1}\Big)^l Y_{t-l}^2\Big]\Big) \\
&= \frac{\sum \alpha_k \beta_{k0}\Big\{1 - (\sum \alpha_k \beta_{k1})^l\Big\}}{1 - \sum \alpha_k \beta_{k1}} E(Y_{t-l}^2) \\
&\quad + \Big(\sum \alpha_k \beta_{k1}\Big)^l E(Y_{t-l}^4) \\
&= \Big\{E(Y_t^2)\Big\}^2 + \Big(\sum \alpha_k \beta_{k1}\Big)^l \Big[E(Y_t^4) - \Big\{E(Y_t^2)\Big\}^2\Big].
\end{aligned}$$

Hence the autocorrelations of the squared time series are given by

$$\operatorname{corr}(Y_t^2, Y_{t-l}^2) = \Big(\sum \alpha_k \beta_{k1}\Big)^l.$$

Note that the squared autocorrelation function is similar to that of an ARCH(1) model, with the lag 1 coefficient replaced by the coefficient $\sum \alpha_k \beta_{k1}$. As a generalization of the ARCH model, the range of possible squared autocorrelations should be as great as that of the corresponding standard ARCH process. In the case of a Markov switching generalized ARCH process, Francq and Zakoian (1999) showed that the squared process behaves like a special ARMA model.

As in the case of MAR model, the m-step predictive distribution of MAR-ARCH model can be computed. It is easy to compute the one-step predictive distribution, $F(y_{t+1}|\mathcal{F}_t)$, based on (1). To compute the m-step predictive distribution, $F(y_{t+m}|\mathcal{F}_t)$, we can use three approaches, the naive, the exact, and the Monte Carlo approaches (Granger and Teräsvirta 1993). We use the two-step predictive distribution $F(y_{t+2}|\mathcal{F}_t)$ to illustrate these approaches.

In the naive approach, we simply use the one-step forecast, $\hat{y}_{t+1} = E(y_{t+1}|\mathcal{F}_t)$, as if it is the true value of y_{t+1}. So we have

$$F(y_{t+2}|\mathcal{F}_t) = F(y_{t+2}|\mathcal{F}_t, y_{t+1} = \hat{y}_{t+1}).$$

This approach is convenient, but it ignores any information provided by the shape of the one-step predictive distribution. In the case of the MAR-ARCH model, the information loss can be severe as $F(y_{t+1}|\mathcal{F}_t)$ can be multimodal. A better approach is to compute the exact two-step predictive distribution by

$$F(y_{t+2}|\mathcal{F}_t) = \int F(y_{t+2}|\mathcal{F}_t, y_{t+1}) dF(y_{t+1}|\mathcal{F}_t).$$

This integration may be intractable analytically, but we can still use numerical methods to evaluate it. Alternatively, a Monte Carlo approximation to the exact predictive distribution may be more desirable. The Monte Carlo approximation of the two-step predictive distribution is

$$F(y_{t+2}|\mathcal{F}_t) = \frac{1}{N}\sum_{j=1}^{N} F(y_{t+2}|\mathcal{F}_t, y_{t+1}^j),$$

where $\{y_{t+1}^j\}$ are sampled from $F(y_{t+1}|\mathcal{F}_t)$. In Section 5 we illustrate computation of one-step predictive distributions and the Monte Carlo approach in computing the two-step predictive distributions.

3. ESTIMATION AND THE OBSERVED INFORMATION MATRIX

Estimation of the parameters of a MAR-ARCH model is similar to that of a MAR model but with more complications. The EM algorithm (Dempster, Laird, and Rubin 1977), the most readily available procedure in estimating mixture-type models, is used. One advantage of the EM algorithm is that it ensures that the likelihood values increase monotonically. (See McLachlan and Basford 1988, and McLachlan and Krishnan 1997 for discussion of the EM algorithm and other alternatives.) Suppose that the observation $Y = (y_1, \dots, y_n)$ is generated from the MAR-ARCH model (1). Let $Z = (Z_1, \dots, Z_n)$, where Z_t is a K-dimensional unobservable random vector with its kth component equal to 1 if y_t comes from the kth component of the conditional distribution function and 0 otherwise. Denote the kth component of Z_t as $Z_{k,t}$. Note that the probability of $Z_{k,t}$ equal to 1 is α_k. Let $\alpha = (\alpha_1, \dots, \alpha_{K-1})'$, and $\theta_k = (\phi_{k0}, \phi_{k1}, \dots, \phi_{kp_k})'(k = 1, \dots, K)$, $\beta_k = (\beta_{k0}, \beta_{k1}, \dots, \beta_{kq_k})'(k = 1, \dots, K)$, and $\theta = (\alpha', \theta_1', \beta_1', \dots, \theta_K', \beta_K')'$. Here $'$ denotes the transpose of a vector or a matrix. Identifiability is ensured by imposing the condition $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_K$ (Aitkin and Rubin 1985). (See Titterington, Smith, and Makov 1985 and the works cited earlier for discussion of this problem.)

The (normalized and conditional) log-likelihood is given by

$$\ell = \frac{1}{N}\sum_{t=p+q+1}^{n} \ell_t = \frac{1}{N}\sum_{t=p+q+1}^{n} \times \left(\sum_{k=1}^{K} Z_{k,t}\log\alpha_k - \sum_{k=1}^{K}\frac{Z_{k,t}}{2}\log h_{k,t} - \sum_{k=1}^{K}\frac{Z_{k,t}e_{k,t}^2}{2h_{k,t}}\right), \quad (5)$$

where $N = n - p - q$. The first-order derivatives of the log-likelihood with respect of θ are given as follows: For $k = 1, \dots, K-1$,

$$\frac{\partial \ell}{\partial \alpha_k} = \frac{1}{N}\sum_t\left(\frac{Z_{k,t}}{\alpha_k} - \frac{Z_{K,t}}{\alpha_K}\right). \quad (6)$$

For $k=1,\ldots,K$, $i=0,\ldots,p_k$,

$$\frac{\partial \ell}{\partial \phi_{ki}} = \frac{1}{N}\left\{\sum_t \frac{Z_{k,t}}{2h_{k,t}}\frac{\partial h_{k,t}}{\partial \phi_{ki}}\left(\frac{e_{k,t}^2}{h_{k,t}}-1\right) - \sum_t \frac{Z_{k,t}e_{k,t}}{h_{k,t}}\frac{\partial e_{k,t}}{\partial \phi_{ki}}\right\}. \tag{7}$$

For $k=1,\ldots,K$, $i=0,\ldots,q_k$,

$$\frac{\partial \ell}{\partial \beta_{ki}} = \frac{1}{N}\sum_t \frac{Z_{k,t}}{2h_{k,t}}\frac{\partial h_{k,t}}{\partial \beta_{ki}}\left(\frac{e_{k,t}^2}{h_{k,t}}-1\right). \tag{8}$$

Here $\partial e_{k,t}/\partial \phi_{ki} = -1$ for $i=0$, $\partial e_{k,t}/\partial \phi_{ki} = -y_{t-i}$ for $0<i\le p_k$, $\partial h_{k,t}/\partial \phi_{ki} = 2\sum_{j=1}^{q_k}\beta_{kj}e_{k,t-j}(\partial e_{k,t-j}/\partial \phi_{ki})$, $\partial h_{k,t}/\partial \beta_{ki} = 1$ for $i=0$, and $\partial h_{k,t}/\partial \beta_{ki} = e_{k,t-i}^2$ for $0<i\le q_k$.

The iterative EM procedure estimates the parameters by maximizing the log-likelihood function (5). It comprises an E step and an M step described as follows:

E step: Suppose that θ is known. The missing data Z are then replaced by their conditional expectations, conditional on the parameters and on the observed data Y. In this case the conditional expectation of the kth component of Z_t is just the conditional probability that the observation y_t comes from the kth component of the mixture distribution, conditional on θ and Y. Let $\tau_{k,t}$ be the conditional expectation of $Z_{k,t}$. Then, similar to Le et al. (1996) and Wong and Li (2000), the E step equations are

$$\tau_{k,t} = \frac{\alpha_k}{\sqrt{h_{k,t}}}\phi\left(\frac{e_{k,t}}{\sqrt{h_{k,t}}}\right)\left\{\sum_{l=1}^{K}\frac{\alpha_l}{\sqrt{h_{l,t}}}\phi\left(\frac{e_{l,t}}{\sqrt{h_{l,t}}}\right)\right\}^{-1}$$

for $k=1,\ldots,K$. In practice, θ is set to $\hat\theta$ from the previous M step of the EM procedure.

M step: Suppose that the missing data are known. Then the estimates of the parameters θ can be obtained by maximizing the log-likelihood ℓ. This can be done by equating (6), (7), and (8) to 0. The parameter estimates of α are

$$\hat\alpha_k = \frac{1}{n-p-q}\sum_{t=p+q+1}^{n}\tau_{k,t} \quad (k=1,\ldots,K).$$

For the θ_k's and β_k's, there are no explicit solutions of the likelihood equations. Numerical methods must be used to find the parameter estimates. Here we use the Newton–Raphson method. Note that as we have for $k,l=1,\ldots,K$ and $k\ne l$, $\partial^2\ell/(\partial\theta_k\partial\theta_l) = \partial^2\ell/(\partial\beta_k\partial\beta_l) = \partial^2\ell/(\partial\theta_k\partial\beta_l) = 0$, we can consider estimation of the parameters θ_k and β_k in each component separately. We show in Appendix B that the second-order derivatives of the log-likelihood with respect to the parameters in the kth component can be approximated by the following quantities:

For $i,j=0,\ldots,p_k$,

$$-\frac{\partial^2\ell}{\partial\phi_{ki}\partial\phi_{kj}} \approx \frac{1}{N}\sum_t Z_{k,t}\left(\frac{1}{2h_{k,t}^2}\frac{\partial h_{k,t}}{\partial \phi_{ki}}\frac{\partial h_{k,t}}{\partial \phi_{kj}} + \frac{1}{h_{k,t}}\frac{\partial e_{k,t}}{\partial \phi_{ki}}\frac{\partial e_{k,t}}{\partial \phi_{kj}}\right). \tag{9}$$

For $i,j=0,\ldots,q_k$,

$$-\frac{\partial^2\ell}{\partial\beta_{ki}\partial\beta_{kj}} \approx \frac{1}{N}\sum_t \frac{Z_{k,t}}{2h_{k,t}^2}\frac{\partial h_{k,t}}{\partial \beta_{ki}}\frac{\partial h_{k,t}}{\partial \beta_{kj}}. \tag{10}$$

For $i=0,\ldots,p_k$ and $j=0,\ldots,q_k$,

$$-\frac{\partial^2\ell}{\partial\phi_{ki}\partial\beta_{kj}} \approx 0. \tag{11}$$

This suggests that starting with the initial values $\theta_k^{(0)}$ and $\beta_k^{(0)}$, the values of θ_k and β_k in the subsequent iterations can be given as

$$\theta_k^{(i+1)} = \theta_k^{(i)} + \left\{\left.\frac{\partial^2\ell}{\partial\theta_k\partial\theta_k}\right|_{\theta_k^{(i)},\beta_k^{(i)}}\right\}^{-1}\left.\frac{\partial\ell}{\partial\theta_k}\right|_{\theta_k^{(i)},\beta_k^{(i)}} \tag{12}$$

and

$$\beta_k^{(i+1)} = \beta_k^{(i)} + \left\{\left.\frac{\partial^2\ell}{\partial\beta_k\partial\beta_k}\right|_{\theta_k^{(i+1)},\beta_k^{(i)}}\right\}^{-1}\left.\frac{\partial\ell}{\partial\beta_k}\right|_{\theta_k^{(i+1)},\beta_k^{(i)}}, \tag{13}$$

where $\theta_k^{(i)}$ and $\beta_k^{(i)}$ are the values in the ith iteration. The parameter estimates $\hat\theta_k$ and $\hat\beta_k$ in a particular M-step iteration are obtained by iterating (12) and (13) until convergence. In practice, the $Z_{k,t}$'s are set to the $\tau_{k,t}$'s from the previous E step of the EM procedure.

The estimates of the parameters θ are obtained by iterating the E and M steps until convergence. The performance of the EM algorithm will be assessed by some simulation experiments. Note that nonconvergence may result if the series is actually nonstationary. But from our experience, the problem can be easily removed by differencing.

The standard errors of the parameter estimates can be computed by the missing information principle (Louis 1982). The observed information matrix, I, can be computed from the complete information matrix, I_c, and the missing information matrix, I_m, with the relation

$$I = I_c - I_m = E\left(-N\frac{\partial^2\ell}{\partial\theta^2}\middle|\theta, Y\right)_{\hat\theta} - \operatorname{var}\left(N\frac{\partial\ell}{\partial\theta}\middle|\theta, Y\right)_{\hat\theta}.$$

The formulas for computing I_c and I_m were given in earlier work (Wong 1998).

The variance matrix of the estimate $\hat\theta$ is given by the inverse of the observed information matrix, I. The large-sample variance of the estimate $\hat\alpha_K$ (which is $1-\hat\alpha_1-\cdots-\hat\alpha_{K-1}$) is given by

$$\sum_{k=1}^{K-1}\operatorname{var}(\hat\alpha_k) + \sum_{k=1}^{K-1}\sum_{\substack{l=1\\ l\ne k}}^{K-1}\operatorname{cov}(\hat\alpha_k,\hat\alpha_l).$$

We have done some simulation studies to assess the performance of the EM estimation method and the estimated standard errors based on work of Louis (1982). We report only four of these simulation studies here. (Results of the other simulation studies are given in Wong 1998.) In all simulation experiments, 1,000 independent sample paths are generated from the MAR-ARCH model (1). In the first two experiments, 500 data points are generated in each sample path, whereas in the last two experiments 1000 data points are used. The models that we used in the simulation studies are as follows:

- Model (I): a MAR-ARCH (2;1,1;1,1) model with parameter values $(\alpha_1, \alpha_2, \phi_{10}, \phi_{11}, \beta_{10}, \beta_{11}, \phi_{20}, \phi_{21}, \beta_{20}, \beta_{21}) = (.5, .5, 1.0, .7, 1.0, .5, -1.0, -.7, 1.0, .5)$
- Model (II): a MAR-ARCH (2;1,1;1,1) model (A) in Section 2
- Model (III): a MAR-ARCH (3;1,1,0;1,1,0) model with parameter values $(\alpha_1, \alpha_2, \alpha_3, \phi_{10}, \phi_{11}, \beta_{10}, \beta_{11}, \phi_{20}, \phi_{21}, \beta_{20}, \beta_{21}, \phi_{30}, \beta_{30}) = (.5, .4, .1, 0, .6, 1.0, 0.3, 0, -.7, 1.0, .4, 0, 10.0)$
- Model (IV): a MAR-ARCH (3;1,1,1;1,2,1) model with parameter values $(\alpha_1, \alpha_2, \alpha_3, \phi_{10}, \phi_{11}, \beta_{10}, \beta_{11}, \phi_{20}, \phi_{21}, \beta_{20}, \beta_{21}, \beta_{22}, \phi_{30}, \phi_{31}, \beta_{30}, \beta_{31}) = (.4, .3, .3, 1.0, .9, 1.0, .3, 0, -.7, 1.0, .2, .4, -1.0, .2, 1.0, .5)$.

The results of the simulation studies are shown in Tables 1–4.

Table 1 illustrates the performance of the EM algorithm for a MAR-ARCH model with two stationary AR-ARCH processes, and Table 2 illustrates the situation of mixing a stationary AR-ARCH process with an explosive AR and a nonstationary ARCH process. It can be observed that the EM estimation method has small bias and reasonable standard errors. The simulation also suggests a small downward bias of the theoretical standard errors. However, the average large-sample theoretical standard errors computed by Louis's method are generally close to the empirical standard errors. Note that the result of the simulation studies will be improved by using a larger sample size.

Model (III) comprises two stationary AR-ARCH components mixed with a small portion of observations generated from a normal distribution with mean 0 and a large variance. The last component can be interpreted as a model of independent pure replacement-type outliers (Le et al. 1996; Martin and Yohai 1985, 1986). Model (IV) is a MAR-ARCH model with three stationary AR-ARCH processes. In the last two simulation studies, we have used a larger sample size to ensure sufficient observations to estimate the parameters in each component. Tables 3 and 4 show that the EM estimation method performs well and that the empirical standard errors match the average large-sample theoretical standard errors very well.

4. MODEL SELECTION CRITERION

In this section we investigate the problem of model selection for MAR-ARCH models. We consider the Bayesian information criterion (BIC) (Schwarz 1978) which has be shown to be useful in the case of MAR model (Wong and Li 2000).

There are two aspects of model selection, the number of components and the order of each AR-ARCH component. Here we do not discuss the selection problem for the number of components, K, because this is difficult to handle even in the special case of the MAR model. Using BIC to choose K is somewhat nonstandard, because it corresponds to testing problems with nuisance parameters that do not exist under the null hypothesis (Davies 1977, 1987). Rather, we concentrate on the order selection of each component.

We use the following definition of BIC:

$$\mathrm{BIC} = -2N\ell^* + \log(n - p_{\max} - q_{\max})\left(3K - 1 + \sum_{k=1}^{K} p_k + \sum_{k=1}^{K} q_k\right).$$

Here ℓ^* is the maximized log-likelihood computed from the (conditional) probability density function of the MAR-ARCH model. It is defined as

$$\ell^* = \frac{1}{N} \sum_{t=p+q+1}^{n} \log f(y_t|\mathcal{F}_{t-1}) = \frac{1}{N} \sum_{t=p+q+1}^{n} \log \frac{d}{dy_t} F(y_t|\mathcal{F}_{t-1}), \tag{14}$$

and $F(y_t|\mathcal{F}_{t-1})$ is given in (1). The order of the components is chosen to be that minimizing the value of BIC.

We have performed a small simulation study to illustrate the performance of the minimum BIC procedure. We simulated model (I) with 50 replications and sample size of 500. The number of components is set to two, and hence no selection

Table 1. Results of the Simulation Study With Model (I)

k		α_k	ϕ_{k0}	ϕ_{k1}	β_{k0}	β_{k1}
1	True values	.5000	1.0000	.7000	1.0000	.5000
	Mean of estimates	.5006	1.0038	.6956	.9977	.5004
	Empirical s.e.	.0368	.1337	.0380	.2754	.0858
	Theoretical s.e.	.0352	.1250	.0318	.2567	.0795
2	True values	.5000	−1.0000	−.7000	1.0000	.5000
	Mean of estimates	.4994	−.9995	−.6961	.9966	.4991
	Empirical s.e.	.0368	.1303	.0411	.3017	.0871
	Theoretical s.e.	.0352	.1177	.0325	.2796	.0797

NOTE: Sample size = 500; number of replications = 1,000.

Table 2. Results of the Simulation Study With Model (II)

k		α_k	ϕ_{k0}	ϕ_{k1}	β_{k0}	β_{k1}
1	True values	.7500	.0000	.5000	1.0000	.5000
	Mean of estimates	.7312	.0025	.4851	.9962	.4692
	Empirical s.e.	.0880	.0852	.0519	.2004	.1254
	Theoretical s.e.	.0839	.0798	.0468	.1816	.1131
2	True values	.2500	.0000	1.1000	1.0000	1.2000
	Mean of estimates	.2688	.0001	1.1038	.8491	1.2206
	Empirical s.e.	.0880	.2050	.1197	.4875	.4613
	Theoretical s.e.	.0839	.1655	.0768	.3857	.3991

NOTE: Sample size = 500; number of replications = 1,000.

Table 3. Results of the Simulation Study With Model (III)

k		α_k	ϕ_{k0}	ϕ_{k1}	β_{k0}	β_{k1}
1	True values	.5000	.0000	.6000	1.0000	.3000
	Mean of estimates	.4974	.0007	.6002	.9996	.2994
	Empirical s.e.	.0456	.0915	.0536	.1971	.0632
	Theoretical s.e.	.0424	.0894	.0468	.1959	.0583
2	True values	.4000	.0000	−.7000	1.0000	.4000
	Mean of estimates	.3988	−.0047	−.7040	.9841	.3992
	Empirical s.e.	.0438	.0998	.0703	.2598	.0902
	Theoretical s.e.	.0408	.0969	.0570	.2507	.0779
3	True values	.1000	.0000		10.0000	
	Mean of estimates	.1039	.0199		10.1160	
	Empirical s.e.	.0405	.7528		3.0038	
	Theoretical s.e.	.0386	.5440		2.8806	

NOTE: Sample size = 1,000; number of replications = 1,000.

Table 4. Results of the Simulation Study With Model (IV)

k		α_k	ϕ_{k0}	ϕ_{k1}	β_{k0}	β_{k1}	β_{k2}
1	True values	.4000	1.0000	.9000	1.0000	.3000	
	Mean of estimates	.4020	1.0030	.8969	.9803	.2949	
	Empirical s.e.	.0400	.1206	.0291	.2249	.0571	
	Theoretical s.e.	.0348	.1158	.0241	.2058	.0554	
2	True values	.3000	.0000	−.7000	1.0000	.2000	.4000
	Mean of estimates	.3015	−.0145	−.7082	1.0132	.1888	.4011
	Empirical s.e.	.0441	.1773	.1056	.5458	.0816	.1054
	Theoretical s.e.	.0416	.1760	.0792	.5418	.0711	.1020
3	True values	.3000	−1.0000	.2000	1.0000	.5000	
	Mean of estimates	.2964	−1.0183	.1956	.9371	.4836	
	Empirical s.e.	.0582	.2194	.0818	.3961	.1596	
	Theoretical s.e.	.0511	.1942	.0722	.3363	.1519	

NOTE: Sample size = 1,000; number of replications = 1,000.

of K is needed. For each component, the maximum AR order $p_{\max}$ is five, and the maximum ARCH order $q_{\max}$ is three. The BIC correctly identifies the p_k's and q_k's at a rate of 90%.

The associate editor raised some concern of the quality of the parameter estimates and the standard errors when the order selection is done by the minimum BIC procedure. We would expect that the quality of the parameter estimates via the EM algorithm and the standard errors computed using the method of Louis (1982) are still good when BIC chooses a model with the true or higher orders in all components. But the quality of the parameter estimates and the standard errors deteriorates when BIC chooses a model with lower orders in some components. We have performed some simulation experiments using model (I), which confirm that this indeed is the case. The results of these simulation studies are available from the authors.

5. EXAMPLES

5.1 Example 1: Chemical Process Temperature Series

The first example is the temperature readings taken every minute from a chemical process (226 observations), reported by Box, Jenkins, and Reinsel (1994) as series C. This set of data has been widely analyzed (e.g., Gray, Kelly, and McIntire 1978; Tsay 1987; Tsay and Tiao 1984). The original series, y_t, the first differenced series ,$y_t - y_{t-1}$, and the second differenced series, $y_t - 2y_{t-1} + y_{t-2}$, are shown in Figure 3. A number of observations can be made from these figures. First, the original series is clearly nonstationary. Second, after performing the first differencing, it is still hard to determine whether the resulting series is stationary, and a second differencing of the series may be required. Box et al. (1994) also suggested that either first- or second-order differencing of the series is required; that is, the series is either a $I(1)$ or $I(2)$ series, where $I(d)$ denotes the integrated order d. Finally, the variability of the second differenced series is not uniform.

Box et al. (1994) suggested two autoregressive integrated moving average (ARIMA) models to the series. The first model is the ARIMA(1,1,0) model, namely

$$y_t - 1.82y_{t-1} + .82y_{t-2} = e_t,$$

where $e_t \sim N(0, .018)$, and the second model is the ARIMA(0,2,2) model, namely

$$y_t - 2y_{t-1} + y_{t-2} = e_t - .13e_{t-1} - .12e_{t-2},$$

where $e_t \sim N(0, .019)$. But these models are not adequate with large squared residual autocorrelations, possibly because of the presence of conditional heteroscedasticity (Tsay 1987).

We consider a two-component MAR-ARCH model for the first differenced series. The orders of the AR-ARCH components are chosen by the minimum BIC criterion. The best MAR-ARCH model is a MAR-ARCH (2;1,1;0,1) model with $\phi_{10} = \phi_{20} = 0$ and BIC of -700.73. The parameter estimates and standard errors for (α_1, α_2, ϕ_{11}, β_{10}, ϕ_{21}, β_{20}, β_{21}) are (.2738, .7262, .5377, .0037, .9966, .0102, .4725) and (.0865, .0865, .0487, .0016, .0499, .0018, .1557). The two-component MAR-ARCH model for the first differencing series is then transformed back to a two-component MAR-ARCH model for the original series, namely

$$F(y_t|\mathcal{F}_{t-1}) = .2738\ \Phi\left(e_{1,t}/\sqrt{h_{1,t}}\right) + .7262\ \Phi\left(e_{2,t}/\sqrt{h_{2,t}}\right),$$
$$e_{1,t} = y_t - 1.5377y_{t-1} + .5377y_{t-2},\ h_{1,t} = .0037,$$

and

$$e_{2,t} = y_t - 1.9966y_{t-1} + .9966y_{t-2},$$
$$h_{2,t} = .0102 + .4725e^2_{2,t-1}.$$

Here the first component is a $I(1)$ process, and the second component can be considered as a $I(2)$ process, because the parameter estimate of ϕ_{21} is close to 1. Moreover, the second component suggests that the process has conditional heteroscedasticity after a second differencing of the process variable. Note that if we restrict $\phi_{21} = 1$ in the estimation, then the value of BIC is -706.13.

The conditional variance of y_t, given by $\mathrm{var}(y_t|\mathcal{F}_{t-1})$, is shown in Figure 4. The pattern of conditional variance computed from the fitted MAR-ARCH model resembles the variability of the second differenced data.

The one-step predictive distributions, $F(y_{t+1}|\mathcal{F}_t)$, at times $t = 199$, $t = 220$, and $t = 221$ are shown in Figure 5, together with the actual values for the series at time $t + 1$. The shapes of the predictive distributions are either unimodal or bimodal.

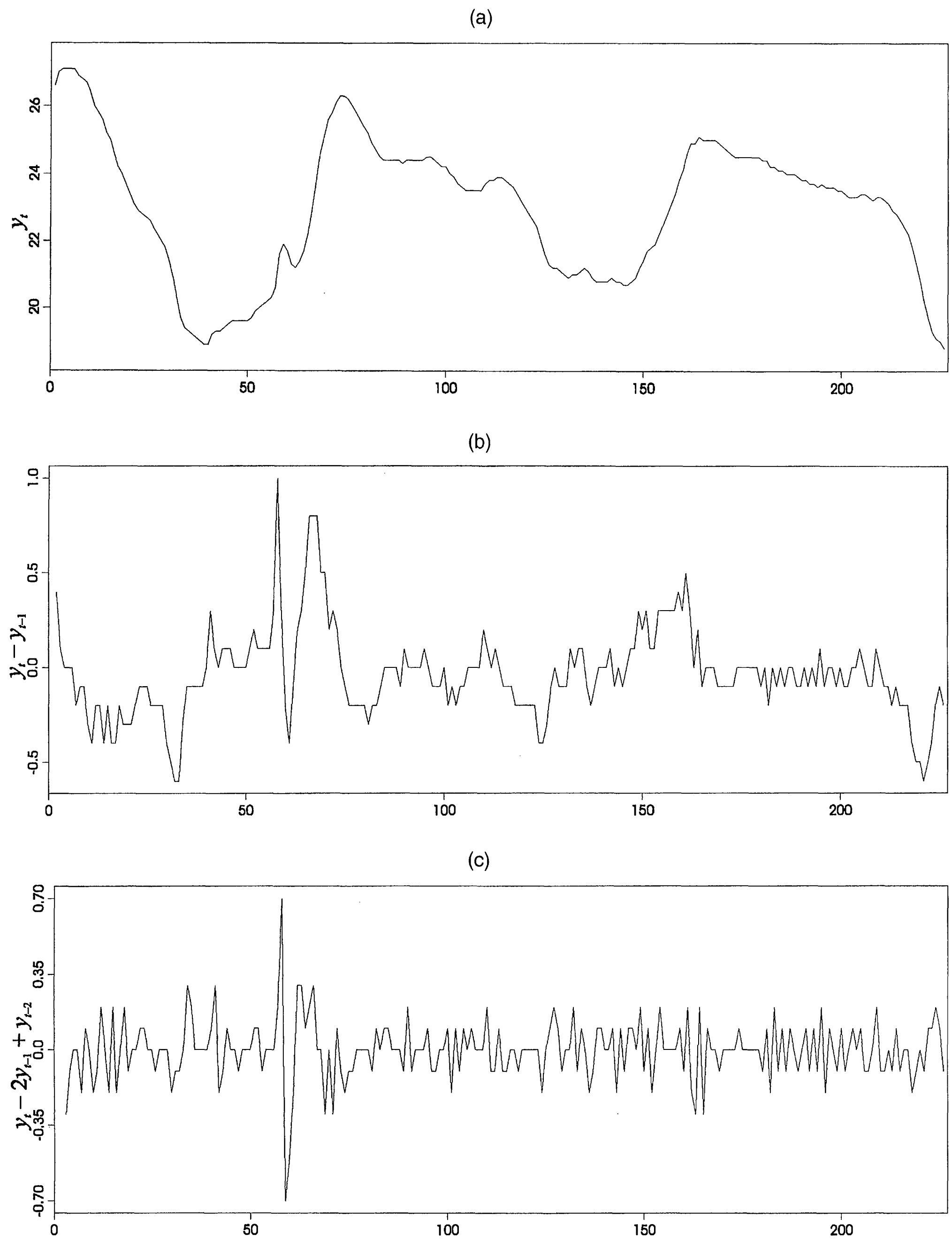

Figure 3. Chemical Process Temperature Series: (a) The Original Series; (b) The First Differenced Series; (c) The Second Differenced Series.

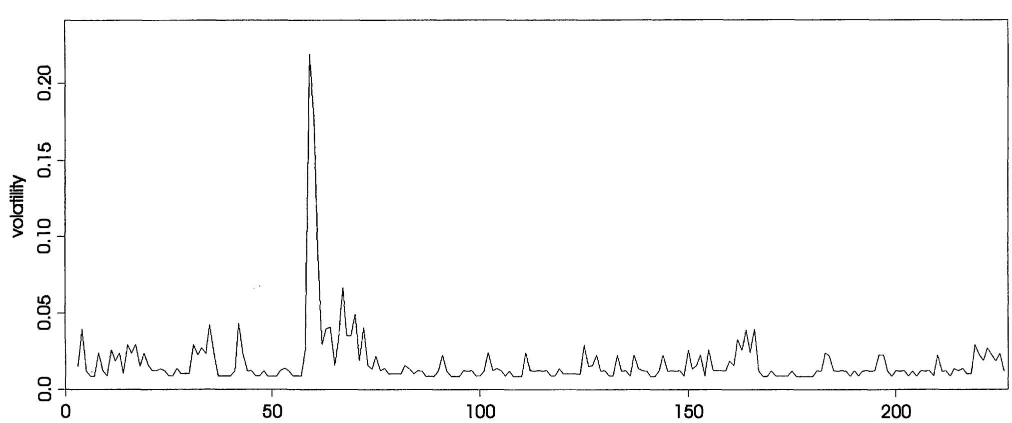

Figure 4. Volatility Computed From the Fitted MAR-ARCH Model for the Chemical Process Temperature Series.

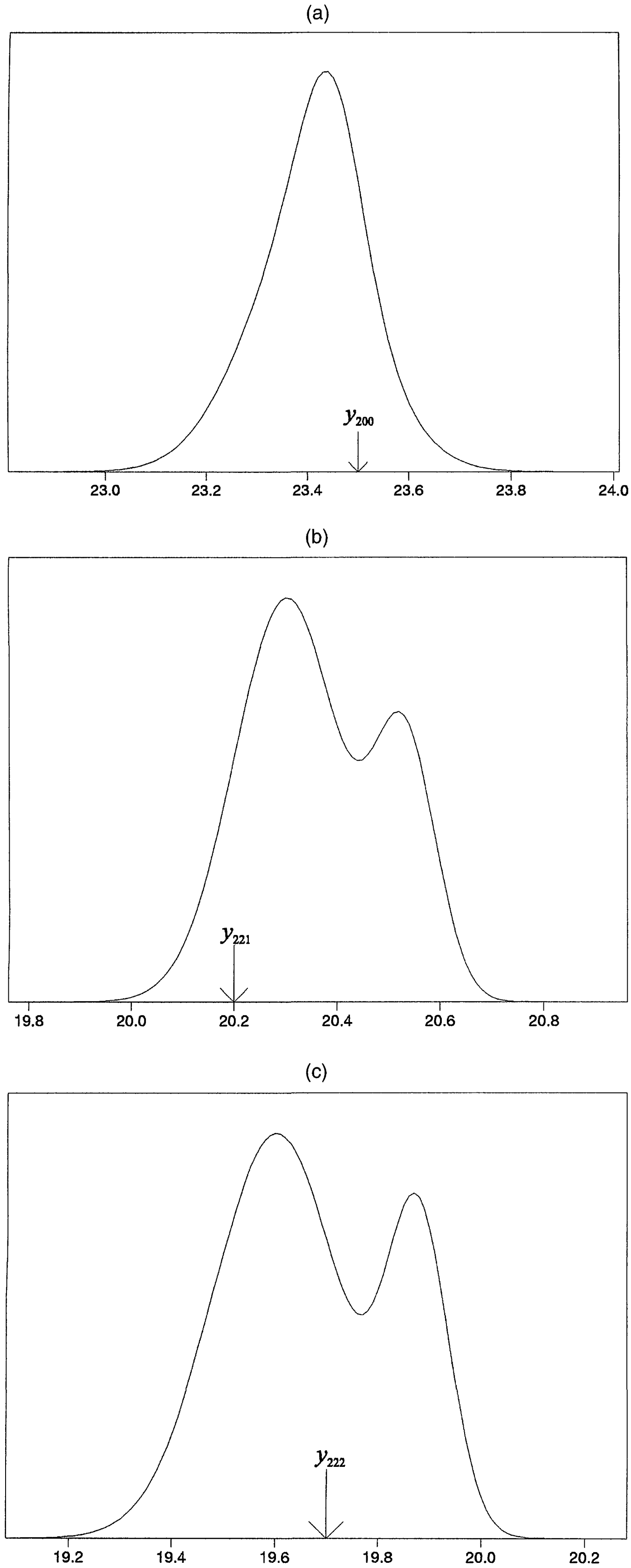

Figure 5. The One-Step Predictive Distributions of the Chemical Process Temperature Series (a) at $t = 199$, (b) at $t = 220$, and (c) at $t = 221$. The actual values of the time series at $t + 1$ are also shown.

We compare the MAR-ARCH model with the models fitted by Box et al. (1994) and an AR-ARCH model based on the forecast accuracy and the ability to describe the predictive distribution of the series. The fitted AR-ARCH model is

$$y_t = 1.8427 y_{t-1} - .8427 y_{t-2} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, h_t),$$
$$h_t = E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = .0098 + .4101 \varepsilon_{t-1}^2.$$

The value of BIC is -705.88. Although the value of BIC for the AR-ARCH model is smaller than the value of BIC for the (unrestricted) MAR-ARCH model, it is not appropriate to say that the AR-ARCH model is a better model. Using the BIC to choose the AR-ARCH model ($K = 1$) versus the MAR-ARCH model ($K = 2$) is a nonstandard problem, as discussed in Section 4.

We use the following procedure to assess the forecast accuracy of these models. We divide the entire dataset into two parts: the initialization part of the first 200 observations and the test part of the last 26 observations. The foregoing models are refitted to the initialization part of the data, and one- to five-step-ahead forecasts are generated for the test part. Table 5 gives the mean squared errors of these forecasts. The forecasts generated from the MAR-ARCH model perform best when the forecast horizon is small, whereas the forecasts generated from the ARIMA(0,2,2) model perform best when the forecast horizon is large. Although the merit of the MAR-ARCH model lies in its ability to describe conditional distribution of the time series, its point prediction is consistently better than that of the ARIMA(1,1,0) model or the AR-ARCH model.

Models' ability to describe the predictive distribution can be compared by computing the empirical coverages of the one-step-ahead and two-step-ahead prediction intervals constructed by these models. For each time t, we computed the one-step and two-step predictive distributions, [i.e., $F(y_{t+1}|\mathcal{F}_t)$ and $F(y_{t+2}|\mathcal{F}_t)$], using the models fitted to the entire dataset. Based on these predictive distributions, we constructed the one-step and two-step prediction intervals. The empirical coverage of the prediction intervals is defined as the proportion of the data lying within their corresponding prediction intervals. The empirical coverages of the one-step-ahead and two-step-ahead prediction intervals for the series generated by each model are given in Tables 6 and 7. As a whole, the empirical coverages of the MAR-ARCH-based prediction intervals are closer to the nominal coverages than are the ARIMA-based and the AR-ARCH-based prediction intervals. This is particularly true for the two-step-ahead prediction intervals. This may

Table 5. Comparison of Mean Squared Errors of One- to Five-Step-Ahead Forecasts for the Chemical Process Temperature Series

	Number of steps				
Model	*1*	*2*	*3*	*4*	*5*
MAR-ARCH	.0108	.0554	.1571	.3321	.5909
AR-ARCH	.0111	.0574	.1640	.3509	.6327
ARIMA(1,1,0)	.0117	.0616	.1775	.3844	.7002
ARIMA(0,2,2)	.0111	.0600	.1661	.3143	.4876

Table 6. Empirical Coverages of the $(1-\alpha)100\%$ One-Step-Ahead Prediction Intervals for the Chemical Process Temperature Series

	$(1-\alpha)100\%$					
Model	*95*	*90*	*80*	*70*	*60*	*50*
MAR-ARCH	93.27	87.44	83.86	73.99	60.99	50.22
AR-ARCH	92.83	88.79	82.96	76.68	59.19	50.22
ARIMA(1,1,0)	96.86	92.38	82.51	77.58	69.51	57.40
ARIMA(0,2,2)	94.62	93.72	81.17	78.03	72.65	53.36

Table 7. Empirical Coverages of the $(1-\alpha)100\%$ Two-Step-Ahead Prediction Intervals for the Chemical Process Temperature Series

	$(1-\alpha)100\%$					
Model	*95*	*90*	*80*	*70*	*60*	*50*
MAR-ARCH	95.50	92.34	84.68	72.52	60.81	50.90
AR-ARCH	94.14	91.89	85.14	77.03	66.67	56.31
ARIMA(1,1,0)	95.05	92.79	88.29	80.18	68.92	57.21
ARIMA(0,2,2)	95.05	90.09	85.14	80.18	72.52	61.71

be partly explained by the fact that ARIMA models and AR-ARCH model can describe only a unimodal predictive distribution. Note that we have used the Monte Carlo method in computing the two-step-ahead prediction intervals.

5.2 Example 2: Standard and Poor Composite 500 Stock Price Index

Figure 6(a) shows the daily Standard and Poor Composite 500 (S&P 500) stock price closing index from January 2, 1970 to January 15, 1999 (6,340 observations). The closing index is clearly nonstationary. We transform the series into the return series, $y_t = \log(\text{index}_t) - \log(\text{index}_{t-1})$ and divide the dataset into six nonoverlapping 5-year periods: 1970–1974, 1975–1979, 1980–1984, 1985–1989, 1990–1994, and 1995–1999 with 1,262, 1,262, 1,264, 1,262, 1,264, and 1,020 observations. The return series is shown in Figure 6(b). In general, the return series has clusters of volatility. In the period 1985–1989, some outliers are observed that correspond to the stock market crash in October 1987, and at the end of the period 1995–1999, clusters of high volatility can also be observed that correspond to the Asian financial crisis.

We fit two-component and three-component MAR-ARCH models to the return series in each period. Table 8 gives the best MAR-ARCH models chosen by the minimum BIC procedure. The BIC values for the best MAR-ARCH models are $-10{,}791.51$, $-11{,}092.84$, $-10{,}541.15$, $-10{,}519.62$, $-11{,}161.63$, and $-8{,}584.12$ for the six periods.

A number of interesting observations can be made from the fitted MAR-ARCH models. First, the fitted MAR-ARCH models reveal evidence of conditional heteroscedasticity in four of the six periods. In the periods 1975–1979 and 1980–1984, MAR models alone are sufficient to model the return series.

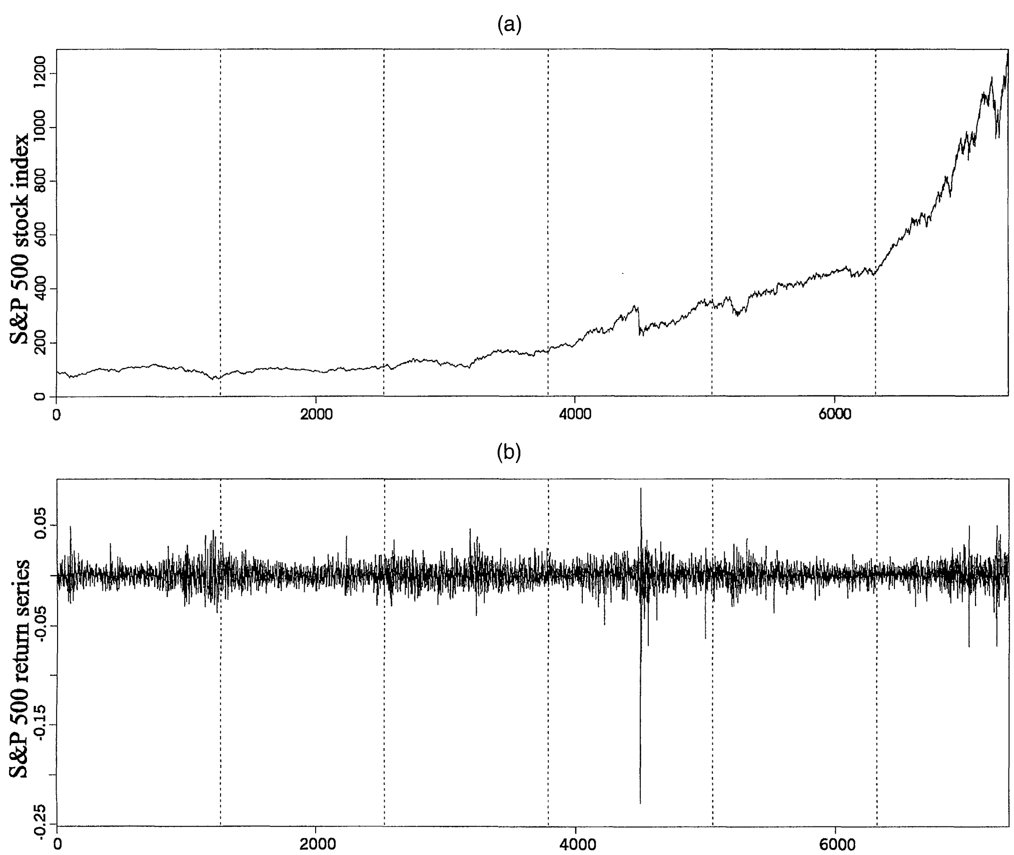

Figure 6. The S&P 500 Stock Index Series: (a) The Original Series; (b) The Return Series. The vertical dotted lines indicate the partitions of the dataset into six 5-year periods.

Table 8. MAR-ARCH Model Fitted to the Return Series of the S&P 500 Stock Price Index

Period	Orders	k		α_k	ϕ_{k1}	β_{k0}	β_{k1}	β_{k2}	β_{k3}	β_{k4}
1970–1974	(2;1,0;4,0)	1	Estimate	.9486	.3368	.000024	.0744	.1258	.1491	.2250
			s.e.	.0296	.0299	.000003	.0318	.0464	.0404	.0445
		2	Estimate	.0514		.000285				
			s.e.	.0296		.000104				
1975–1979	(2;1,0;0,0)	1	Estimate	.5033	.3448	.000035				
			s.e.	.1610	.1103	.000005				
		2	Estimate	.4967		.000072				
			s.e.	.1610		.000014				
1980–1984	(2;1,0;0,0)	1	Estimate	.4900	.2033	.000135				
			s.e.	.1613	.0861	.000020				
		2	Estimate	.5100		.000041				
			s.e.	.1613		.000011				
1985–1989	(3;1,0,0;2,0,0)	1	Estimate	.5624	.1773	.000095	.1488	.2829		
			s.e.	.0599	.0541	.000010	.0603	.0891		
		2	Estimate	.4313		.000020				
			s.e.	.0602		.000004				
		3	Estimate	.0063		.008569				
			s.e.	.0030		.004952				
1990–1994	(2;0,0;0,1)	1	Estimate	.4183		.000013				
			s.e.	.0759		.000004				
		2	Estimate	.5817		.000075	.2554			
			s.e.	.0759		.000008	.0875			
1995–1999	(2;1,0;1,1)	1	Estimate	.9071	.1137	.000042	.1824			
			s.e.	.0313	.0356	.000004	.0409			
		2	Estimate	.0929		.000237	4.0317			
			s.e.	.0313		.000067	2.2107			

Second, for the period 1990–1994, the fitted MAR-ARCH model is a mixture of a white noise process and a ARCH noise process. This fitted MAR-ARCH noise model is superior to a MAR noise model, because it can model the conditional heteroscedasticity in the return series. Third, for the period 1970–1974, the fitted MAR-ARCH model is a mixture of AR-ARCH model with a white noise process with large variance. The mixing coefficient of the white noise process is quite small, which corresponds to about 65 observations. This white noise process can be interpreted as a model for the independent pure replacement-type outliers in the return series (Le et al. 1996; Martin and Yohai 1985, 1986). Fourth, the fitted MAR-ARCH model for the period 1985–1989 comprises three components. As implied from the model, about half of the return series follows an AR-ARCH(1,2) model (the first component) and the other half of the return series is unpredictable with small variance (the second component). But, the third component is just a white noise process with very large variance. The mixing coefficient for the third component is small, which corresponds to only eight observations (mostly) coming from the period around October 1987. Hence it can be interpreted as a model for the outliers during that period. Finally, the second component of the fitted model in the period 1995–1999 has a very large estimated value for the ARCH(1) parameter (i.e., $\hat{\beta}_{21} = 4.0317$). This nonstationary ARCH component can be interpreted as a model for the cluster of extreme volatility during the Asian financial crisis.

6. CONCLUDING REMARKS

We have discussed a potentially useful class of time series models, the MAR-ARCH models, for modeling nonlinear time series. The models allow much flexibility that other models do not have. The fuller range of shape-changing predictive distribution and the more flexible squared autocorrelation structure are especially attractive. The EM algorithm allows easy estimation of the parameters in the model, and model selection can be done with the BIC criterion. We have demonstrated the potential usefulness of the MAR-ARCH models with the chemical process temperature data and a high-frequency financial data.

Many problems for the MAR-ARCH models remain open. For example, new approaches for selecting the number of components in the MAR-ARCH model should be developed, and generalization to multivariate time series is currently under investigation.

APPENDIX A: PROOFS OF THEOREMS IN SECTION 2

Theorems 1 and 2 are proven by using similar techniques to the proof of theorems 1, 2, and 3 of Wong and Li (2000); that is, using the result due to Beněs (1967) stated by Le et al. (1996). Theorem 3 is proven by direct evaluation of the expectation of Y_t^4 and searching for the condition of its existence.

Proof of Theorem 1.

See appendix A of Le et al. (1996).

Proof of Theorem 2.

Without loss of generality, let $\phi_{k0} = 0$ $(k = 1, \ldots, K)$. Let $\gamma_{i,t} = E(Y_t Y_{t-i})$, $g_k(y_t) = (1/\sqrt{h_{k,t}})\phi\{(y_t - \phi_{k1} y_{t-1})/\sqrt{h_{k,t}}\}$, and C be a constant that is independent of t. The conditional variance of y_t is

$$E(Y_t^2|\mathcal{F}_{t-1}) = \int y_t^2 f(y_t|\mathcal{F}_{t-1})dy_t$$

$$= \int y_t^2 \sum_{k=1}^{K} \alpha_k g_k(y_t)dy_t$$

$$= \sum_{k=1}^{K} \alpha_k \int (y_t - \phi_{k1} y_{t-1} + \phi_{k1} y_{t-1})^2 g_k(y_t) dy_t$$

$$= \sum_{k=1}^{K} \alpha_k \int \{(y_t - \phi_{k1} y_{t-1})^2 + 2\phi_{k1} y_{t-1}(y_t - \phi_{k1} y_{t-1}) + \phi_{k1}^2 y_{t-1}^2\} g_k(y_t) dy_t$$

$$= \sum_{k=1}^{K} \alpha_k h_{k,t} + \sum_{k=1}^{K} \alpha_k \phi_{k1}^2 y_{t-1}^2$$

as $\int \phi_{k1} y_{t-1}(y_t - \phi_{k1} y_{t-1}) g_k(y_t) dy_t = 0$ $(k = 1, \ldots, K)$. To obtain the variance of y_t, we first consider

$$E(e_{k,t}^2) = E\{(Y_t - \phi_{k1} Y_{t-1})^2\}$$
$$= E(Y_t^2) - 2\phi_{k1} E(Y_t Y_{t-1}) + \phi_{k1}^2 E(Y_{t-1}^2)$$
$$= \gamma_{0,t} + \left\{\phi_{k1}^2 - 2\phi_{k1}\left(\sum_{k=1}^{K} \alpha_k \phi_{k1}\right)\right\}\gamma_{0,t-1},$$

because

$$E(Y_t Y_{t-1}) = E\{Y_{t-1} E(Y_t | \mathcal{F}_{t-1})\} = E\left\{Y_{t-1}\left(\sum_{k=1}^{K} \alpha_k \phi_{k1} Y_{t-1}\right)\right\}$$
$$= \left(\sum_{k=1}^{K} \alpha_k \phi_{k1}\right) E(Y_{t-1}^2).$$

The variance of y_t is

$$\gamma_{0,t} = E\left\{E(Y_t^2 | \mathcal{F}_{t-1})\right\}$$
$$= E\left\{\sum_{k=1}^{K} \alpha_k \left(\beta_{k0} + \beta_{k1} e_{k,t-1}^2 + \cdots + \beta_{kq} e_{k,t-q}^2\right) + \sum_{k=1}^{K} \alpha_k \phi_{k1}^2 Y_{t-1}^2\right\}$$
$$= \sum_{k=1}^{K} \alpha_k \beta_{k0} + \sum_{i=1}^{q}\sum_{k=1}^{K} \alpha_k \beta_{ki} E(e_{k,t-i}^2) + \sum_{k=1}^{K} \alpha_k \phi_{k1}^2 E(Y_{t-1}^2)$$
$$= \sum_{k=1}^{K} \alpha_k \beta_{k0} + \sum_{i=1}^{q}\sum_{k=1}^{K} \alpha_k \beta_{ki} \times \left[\gamma_{0,t-i} + \left\{\phi_{k1}^2 - 2\phi_{k1}\left(\sum_{j=1}^{K} \alpha_j \phi_{j1}\right)\right\}\gamma_{0,t-i-1}\right] + \sum_{k=1}^{K} \alpha_k \phi_{k1}^2 \gamma_{0,t-1}$$
$$= C + \sum_{k=1}^{K} \alpha_k (\beta_{k1} + \phi_{k1}^2)\gamma_{0,t-1} + \sum_{i=2}^{q}\sum_{k=1}^{K} \alpha_k \left[\beta_{ki} + \beta_{k,i-1} \times \left\{\phi_{k1}^2 - 2\phi_{k1}\left(\sum_{j=1}^{K} \alpha_j \phi_{j1}\right)\right\}\right]\gamma_{0,t-i}$$
$$+ \sum_{k=1}^{K} \alpha_k \beta_{kq}\left\{\phi_{k1}^2 - 2\phi_{k1}\left(\sum_{j=1}^{K} \alpha_j \phi_{j1}\right)\right\}\gamma_{0,t-q-1}.$$

The necessary and sufficient condition for this nonhomogeneous difference equation to have a stable solution that is finite and independent of t is that the roots $z_1, \ldots, z_{q+1}$ of the difference equation (3) all lie inside the unit circle (Goldberg 1958). Let μ be the measure corresponding to the distribution of the process at time 0; that is, a finite mixture of Gaussian distribution, and hence $\mu \in C^+$. Here C^+ is the strictly positive cone of the Banach space as defined by Beněs (1967). Thus, by the result of Beněs, there exists a finite invariant measure for the process Y_t.

Proof of Theorem 3.

It is easy to show that $E(Y_t) = 0$ and $\mu_{k,t} = 0$ $(k = 1, \ldots, K)$. Let $g_k(y_t) = (1/\sqrt{h_{k,t}})\phi(y_t/\sqrt{h_{k,t}})$. The fourth-order moment of Y_t is given by

$$E(Y_t^4) = E\{E(Y_t^4 | \mathcal{F}_{t-1})\} = E\left\{\int y_t^4 \sum_{k=1}^{K} \alpha_k g_k(y_t) dy_t\right\}$$
$$= \sum_{k=1}^{K} \alpha_k E\left\{\int y_t^4 g_k(y_t) dy_t\right\} = \sum_{k=1}^{K} \alpha_k E\left(3h_{k,t}^2\right)$$
$$= 3\sum_{k=1}^{K} \alpha_k E\left(\beta_{k0}^2 + 2\beta_{k0}\beta_{k1} Y_{t-1}^2 + \beta_{k1}^2 Y_{t-1}^4\right)$$
$$= 3\left\{\sum_{k=1}^{K} \alpha_k \beta_{k0}^2 + 2\sum_{k=1}^{K} \alpha_k \beta_{k0}\beta_{k1} \frac{\sum_{k=1}^{K} \alpha_k \beta_{k0}}{1 - \sum_{k=1}^{K} \alpha_k \beta_{k1}} + \sum_{k=1}^{K} \alpha_k \beta_{k1}^2 E(Y_{t-1}^4)\right\}.$$

A necessary and sufficient condition for this nonhomogeneous difference equation to have a stable solution that is finite and independent of t is that condition (4) holds (Goldberg 1958). The fourth-order moment of Y_t is then given by

$$3\frac{\sum \alpha_k \beta_{k0}^2 + 2(\sum \alpha_k \beta_{k0}\beta_{k1})(\sum \alpha_k \beta_{k0})/(1 - \sum \alpha_k \beta_{k1})}{1 - 3\sum \alpha_k \beta_{k1}^2}.$$

APPENDIX B: SECOND-ORDER DERIVATIVES OF THE LOG-LIKELIHOOD

We now give the justification of the approximation to the second derivatives of the log-likelihood in (9), (10), and (11). We have, for $i, j = 0, \ldots, p_k$,

$$\frac{\partial^2 \ell_t}{\partial\phi_{ki}\partial\phi_{kj}} = -\frac{Z_{k,t}}{2h_{k,t}^2}\frac{\partial h_{k,t}}{\partial\phi_{ki}}\frac{\partial h_{k,t}}{\partial\phi_{kj}}\left(\frac{e_{k,t}^2}{h_{k,t}} - 1\right) + \frac{Z_{k,t}}{2h_{k,t}}\frac{\partial^2 h_{k,t}}{\partial\phi_{ki}\partial\phi_{kj}}\left(\frac{e_{k,t}^2}{h_{k,t}} - 1\right)$$
$$+ \frac{Z_{k,t}e_{k,t}}{h_{k,t}^2}\frac{\partial h_{k,t}}{\partial\phi_{ki}}\frac{\partial e_{k,t}}{\partial\phi_{kj}} - \frac{Z_{k,t}e_{k,t}^2}{2h_{k,t}^3}\frac{\partial h_{k,t}}{\partial\phi_{ki}}\frac{\partial h_{k,t}}{\partial\phi_{kj}} - \frac{Z_{k,t}}{h_{k,t}}\frac{\partial e_{k,t}}{\partial\phi_{ki}}\frac{\partial e_{k,t}}{\partial\phi_{kj}} + \frac{Z_{k,t}e_{k,t}}{h_{k,t}^2}\frac{\partial e_{k,t}}{\partial\phi_{ki}}\frac{\partial h_{k,t}}{\partial\phi_{kj}}$$
$$- \frac{Z_{k,t}e_{k,t}}{h_{k,t}}\frac{\partial^2 e_{k,t}}{\partial\phi_{ki}\partial\phi_{kj}};$$

for $i, j = 0, \ldots, q_k$,

$$\frac{\partial^2 \ell_t}{\partial\beta_{ki}\partial\beta_{kj}} = -\frac{Z_{k,t}}{2h_{k,t}^2}\frac{\partial h_{k,t}}{\partial\beta_{ki}}\frac{\partial h_{k,t}}{\partial\beta_{kj}}\left(\frac{e_{k,t}^2}{h_{k,t}} - 1\right) + \frac{Z_{k,t}}{2h_{k,t}}\frac{\partial^2 h_{k,t}}{\partial\beta_{ki}\partial\beta_{kj}}\left(\frac{e_{k,t}^2}{h_{k,t}} - 1\right) - \frac{Z_{k,t}e_{k,t}^2}{2h_{k,t}^3}\frac{\partial h_{k,t}}{\partial\beta_{ki}}\frac{\partial h_{k,t}}{\partial\beta_{kj}};$$

and for $i = 0, \ldots, p_k$ and $j = 0, \ldots, q_k$,

$$\frac{\partial^2 \ell_t}{\partial \phi_{ki} \partial \beta_{kj}} = -\frac{Z_{k,t}}{2h_{k,t}^2} \frac{\partial h_{k,t}}{\partial \phi_{ki}} \frac{\partial h_{k,t}}{\partial \beta_{kj}} \left(\frac{e_{k,t}^2}{h_{k,t}} - 1 \right) + \frac{Z_{k,t}}{2h_{k,t}} \frac{\partial^2 h_{k,t}}{\partial \phi_{ki} \partial \beta_{kj}} \left(\frac{e_{k,t}^2}{h_{k,t}} - 1 \right) + \frac{Z_{k,t} e_{k,t}}{h_{k,t}^2} \frac{\partial e_{k,t}}{\partial \phi_{ki}} \frac{\partial h_{k,t}}{\partial \beta_{kj}} - \frac{Z_{k,t} e_{k,t}^2}{2h_{k,t}^3} \frac{\partial h_{k,t}}{\partial \phi_{ki}} \frac{\partial h_{k,t}}{\partial \beta_{kj}}.$$

Recall that $Z_{k,t}$ is an unobserved random variable that indicates whether y_t comes from the kth component of the conditional distribution function. In the M step of the EM algorithm, the realization of this random variable $Z_{k,t}$ is assumed known. If $Z_{k,t}$ is 0, then all of the foregoing second-order derivatives of ℓ_t are 0. On the other hand, if $Z_{k,t}$ is 1, then the conditional density of y_t is given by $f(y_t | \mathcal{F}_{t-1}, Z_{k,t} = 1) = (1/\sqrt{h_{k,t}}) \phi(e_{k,t}/\sqrt{h_{k,t}})$. Hence the conditional expectations of the second-order derivatives of ℓ_t are

$$E\left(-\frac{\partial^2 \ell_t}{\partial \phi_{ki} \partial \phi_{kj}} \middle| \mathcal{F}_{t-1}, Z_{k,t} \right) = \frac{Z_{k,t}}{2h_{k,t}^2} \frac{\partial h_{k,t}}{\partial \phi_{ki}} \frac{\partial h_{k,t}}{\partial \phi_{kj}} + \frac{Z_{k,t}}{h_{k,t}} \frac{\partial e_{k,t}}{\partial \phi_{ki}} \frac{\partial e_{k,t}}{\partial \phi_{kj}},$$

$$E\left(-\frac{\partial^2 \ell_t}{\partial \beta_{ki} \partial \beta_{kj}} \middle| \mathcal{F}_{t-1}, Z_{k,t} \right) = \frac{Z_{k,t}}{2h_{k,t}^2} \frac{\partial h_{k,t}}{\partial \beta_{ki}} \frac{\partial h_{k,t}}{\partial \beta_{kj}},$$

and

$$E\left(-\frac{\partial^2 \ell_t}{\partial \phi_{ki} \partial \beta_{kj}} \middle| \mathcal{F}_{t-1}, Z_{k,t} \right) = \frac{Z_{k,t}}{2h_{k,t}^2} \frac{\partial h_{k,t}}{\partial \phi_{ki}} \frac{\partial h_{k,t}}{\partial \beta_{kj}},$$

and the expectations of the second-order derivatives of l_t given $Z_{k,t}$ are

$$E\left(-\frac{\partial^2 \ell_t}{\partial \phi_{ki} \partial \phi_{kj}} \middle| Z_{k,t} \right) = \frac{Z_{k,t}}{2} E\left(\frac{1}{h_{k,t}^2} \frac{\partial h_{k,t}}{\partial \phi_{ki}} \frac{\partial h_{k,t}}{\partial \phi_{kj}} \right) + Z_{k,t} E\left(\frac{1}{h_{k,t}} \frac{\partial e_{k,t}}{\partial \phi_{ki}} \frac{\partial e_{k,t}}{\partial \phi_{kj}} \right), \tag{B.1}$$

$$E\left(-\frac{\partial^2 \ell_t}{\partial \beta_{ki} \partial \beta_{kj}} \middle| Z_{k,t} \right) = \frac{Z_{k,t}}{2} E\left(\frac{1}{h_{k,t}^2} \frac{\partial h_{k,t}}{\partial \beta_{ki}} \frac{\partial h_{k,t}}{\partial \beta_{kj}} \right), \tag{B.2}$$

and

$$E\left(-\frac{\partial^2 \ell_t}{\partial \phi_{ki} \partial \beta_{kj}} \middle| Z_{k,t} \right) = \frac{Z_{k,t}}{2} E\left(\frac{1}{h_{k,t}^2} \frac{\partial h_{k,t}}{\partial \phi_{ki}} \frac{\partial h_{k,t}}{\partial \beta_{kj}} \right). \tag{B.3}$$

Equations (B.1) and (B.2) justify the approximations to the second-order derivatives of the log-likelihood given by (9) and (10). Note that closed-form expressions in terms of model parameters could be possible for some special cases.

To justify (11), we make use of a lemma of Engle (1982): Let u and v be any two random variables. $E\{g(u, v)|v\}$ will be an antisymmetric function of v if g is antisymmetric in v, the conditional density of $u|v$ is symmetric in v, and the expectation exists.

We can write (B.3) as

$$E\left(-\frac{\partial^2 l_t}{\partial \phi_{ki} \partial \beta_{kj}} \middle| Z_{k,t} \right) = Z_{k,t} \sum_{m=1}^{q_k} \beta_{km} E\left(\frac{1}{h_{k,t}} e_{k,t-j}^2 e_{k,t-m} y_{t-m-i} \right)$$

by substituting the first derivatives of $h_{k,t}$. If on the right side the expectation of the term inside brackets conditional on $\mathcal{F}_{t-m-1}$ is 0 for all i, j, t, and m, then (B.3) is equal to 0, and hence this justifies the approximation given in (11). Note that the foregoing expectation is finite, because the time series is stationary and ergodic. Let

$$\psi(e_{k,t-m}) = E\left(\frac{1}{h_{k,t}} e_{k,t-j}^2 e_{k,t-m} \middle| \mathcal{F}_{t-m}, Z_{k,t} = 1 \right).$$

Because $h_{k,t}^{-1}$ is symmetric in $e_{k,t-m}$ and $e_{k,t-m}$ is antisymmetric, the entire expression is antisymmetric in $e_{k,t-m}$, which is part of the information set $\mathcal{F}_{t-m}$. Because $h_{k,t}$ is symmetric, the conditional density must be symmetric in $e_{k,t-m}$, and the lemma can be invoked to show that $\psi(e_{k,t-m})$ is antisymmetric. Finally, taking expectation of ψ conditional on $\mathcal{F}_{t-m-1}$ gives 0, because $e_{k,t-m}$ conditional on the past has a symmetric (normal) density. This shows that (B.3) is equal to 0.

[Received April 1999. Revised March 2000.]

REFERENCES

Aitkin, M., and Rubin, D. B. (1985), "Estimation and Hypothesis Testing in Finite Mixture Models," *Journal of the Royal Statistical Society*, Ser. B, 47, 67–75.

Beneš, V. E. (1967), "Existence of Finite Invariant Measures for Markov Processes," *Proceedings of the American Mathematical Society*, 18, 1058–1061.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994), *Time Series Analysis Forecasting and Control* (3rd ed.), Englewood Cliffs, NJ: Prentice-Hall.

Cai, J. (1994), "A Markov Model of Switching-Regime ARCH," *Journal of Business and Economic Statistics*, 12, 309–316.

Chen, R., and Liu, J. S. (1996), "Predictive Updating Methods With Application to Bayesian Classification," *Journal of the Royal Statistical Society*, Ser. B, 58, 397–415.

Davies, R. B. (1977), "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative," *Biometrika*, 64, 247–254.

——— (1987), "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative," *Biometrika*, 74, 33–43.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 39, 1–38.

Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

Francq, C., and Zakoian, J. M. (1999), "Linear-Representations Based Estimation of Switching-Regime GARCH Models," in *Proceedings of the 1999 NBER/NSF Time Series Conference*, Taipei.

Goldberg, S. (1958), *Difference Equations*, New York: Wiley.

Granger, C. W. J., and Newbold, P. (1986), *Forecasting Economic Time Series*, San Diego: Academic Press.

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

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

Gray, H. L., Kelly, G. D., and McIntire, D. D. (1978), "A New Approach to ARMA Modeling," *Communications in Statistics, Part B*, 7, 1–77.

Hamilton, J. D., and Susmel, R. (1994), "Autoregressive Conditional Heteroskedasticity and Changes in Regime," *Journal of Econometrics*, 64, 307–333.

Huerta, G., Jiang, W., and Tanner, M. A. (1999), "Time Series Modeling via Hierarchical Mixtures," preprint, Northwestern University, Dept. of Statistics.

Le, N. D., Martin, R. D., and Raftery, A. E. (1996), "Modeling Flat Stretches, Bursts, and Outliers in Time Series Using Mixture Transition Distribution Models," *Journal of the American Statistical Association*, 91, 1504–1514.

Louis, T. A. (1982), "Finding the Observed Information Matrix When Using the EM Algorithm," *Journal of the Royal Statistical Society*, Ser. B, 44, 226–233.

Martin, R. D., and Yohai, V. J. (1985), "Robustness in Time Series and Estimating ARMA Models," in *Handbook of Statistics 5*, eds. E. J. Hannan, P. R. Krishnaiah, and M. M. Rao, New York: Elsevier, pp. 119–155.

——— (1986), "Influence Functionals for Time Series," *The Annals of Statistics*, 14, 781–855.

McLachlan, G. J., and Basford, K. E. (1988), *Mixture Models: Inference and Applications to Clustering*, New York: Marcel Dekker.

McLachlan, G. J., and Krishnan, T. (1997), *The EM Algorithm and Extensions*, New York: Wiley.

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

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

Tsay, R. S. (1987), "Conditional Heteroscedastic Time Series Models," *Journal of the American Statistical Association*, 82, 590–604.

Tsay, R. S., and Tiao, G. C. (1984), "Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models," *Journal of the American Statistical Association*, 79, 84–96.

Wong, C. S. (1998), "Statistical Inference for Some Nonlinear Time Series Models", doctoral thesis, The University of Hong Kong.

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