

# EUI WORKING PAPERS

ECO No. 2006/3



A Mixture Multiplicative Error Model  
for Realized Volatility

MARKKU LANNE



**EUROPEAN UNIVERSITY INSTITUTE**

Department of Economics

**EUROPEAN UNIVERSITY INSTITUTE**  
**DEPARTMENT OF ECONOMICS**

*A Mixture Multiplicative Error Model for Realized Volatility*

**MARKKU LANNE**

EUI Working Paper **ECO** No. 2006/3

This text may be downloaded freely only for personal academic purposes. Any reproductions, whether in hard copies or electronically, require the consent of the author. If cited or quoted, reference should be made to the full name of the author(s), editor(s), the title, the working paper, or other series, the year and the publisher.

ISSN 1725-6704

© 2006 Markku Lanne

Printed in Italy  
European University Institute  
Badia Fiesolana  
I – 50016 San Domenico di Fiesole (FI)  
Italy

<http://www.iue.it/>  
<http://cadmus.iue.it/dspace/index.jsp>

# A Mixture Multiplicative Error Model for Realized Volatility\*

Markku Lanne<sup>†</sup>  
European University Institute,  
University of Jyväskylä, RUESG and HECER

## Abstract

A multiplicative error model with time-varying parameters and an error term following a mixture of gamma distributions is introduced. The model is fitted to the daily realized volatility series of Deutschemark/Dollar and Yen/Dollar returns and is shown to capture the conditional distribution of these variables better than the commonly used ARFIMA model. The forecasting performance of the new model is found to be, in general, superior to that of the set of volatility models recently considered by Andersen et al. (2003) for the same data.

JEL classification: C22, C52, C53, G15

Keywords: Mixture model, Realized volatility, Gamma distribution

---

\*I thank Helmut Lütkepohl and Pentti Saikkonen for useful comments. I am grateful for financial support from Suomen Arvopaperimarkkinoiden Edistämissäätiö and the Research Unit of Economic Structures and Growth (RUESG) in the University of Helsinki. Part of this research was done while the author was a Jean Monnet Fellow at the Economics Department of the European University Institute whose hospitality is acknowledged.

<sup>†</sup>Economics Department, European University Institute, Villa San Paolo, Via della Piazzuola 43, 50133 Florence, Italy. e-mail: markku.lanne@iue.it

## 1 Introduction

Being able to accurately forecast the volatility of asset returns is vital in many fields of finance, such as derivative pricing and risk management. For modeling daily data, different variants of the generalized autoregressive conditional variance (GARCH) model (Bollerslev, 1986) and the stochastic volatility model (Taylor, 1986) have been employed. These are effectively models for the squared returns, where the conditional variance is latent, and they produce relatively noisy forecasts. With the recent wide availability of intradaily data, it has become possible to measure the daily volatility more accurately. Andersen and Bollerslev (1998) and Bandorff-Nielsen and Shephard (2001) show that the realized variance, computed as the sum of squared intradaily returns, provides one such measure with desirable properties.

After the introduction of the realized variance into the financial econometrics literature, a burgeoning literature concerned with modeling and forecasting it has emerged. Andersen et al. (2001a, 2001b) have demonstrated that the square root of realized variance (henceforth realized volatility) of stock and exchange rate returns can be described as unconditionally lognormally distributed and exhibiting long memory. Following these papers, the most commonly employed conditional model, incorporating both long memory and lognormality, for the realized volatility is the Gaussian autoregressive fractionally integrated moving average (ARFIMA) model specified for the logarithm of realized volatility. It has also been used in the multivariate framework by Andersen et al. (2003) and, in addition to that paper, shown to produce superior forecasts by Koopman et al. (2005), *inter alia*. However, the ARFIMA model may not be optimal. Pong et al. (2004) have recently shown that a short-memory autoregressive moving average (ARMA) model can be as good in forecasting realized volatility of stock returns as a long-memory ARFIMA model. This is in line with the findings of Bos et al. (2002) that the parameters of the ARFIMA model may not be well identified empirically because the FI part and ARMA part can capture the same characteristics. Also, Morana and Beltratti (2004)

have shown structural breaks to be capable of explaining the long-memory property of the realized volatility of exchange rate returns to a large extent, which is consistent with Maheu and McCurdy's (2002) result that the realized foreign exchange volatility contains nonlinearities. Moreover, the ARFIMA model involves an infinite-order lag polynomial that in practice needs to be truncated, which produces an approximation error.

In this paper we introduce a new kind of time series model for realized volatility. The model belongs to the family of Multiplicative Error Models (MEM) of Engle (2002). As pointed out by Engle (2002), the multiplicative structure is convenient as it avoids the logarithmic transformation. Even though we explicitly consider only modeling realized volatility, the model is expected to be suitable for some other positive-valued time series, such as range-based volatility measures (Alizadeh et al., 2002) and trading volume. Time series of these variables may contain occasional zeros, precluding the logarithmic transformation. Compared to the model suggested by Engle (2002) the new feature of our model is a full mixture structure. In addition to modeling the error term as a mixture of gamma distributions (as opposed to the exponential distribution), we also allow for a time-varying conditional mean. The generalizations bring about considerable flexibility that seems to be required for adequately modelling the realized volatility.

We estimate the mixture-MEM model for the realized volatility of two spot foreign exchange rates, the Deutschemark and Japanese Yen against the U.S. Dollar, and compare the estimated models to ARFIMA models assuming lognormality. The latter are shown to provide a poor description of the conditional distribution of the realized volatilities. In out-of-sample forecasting experiments we compare the new model to the full set of models recently considered by Andersen et al. (2003) and find little evidence against the mixture-MEM model being superior.

The plan of the paper is as follows. In Section 2 the mixture-MEM model is introduced. Estimation and inference are discussed in Section 3. As the model does not produce normally distributed residuals even if the specification is correct,

standard diagnostic checks are not available, and suitable procedures are described in Section 4. Section 5 contains the empirical results, and finally, Section 6 concludes with some suggestions for potential extensions.

## 2 The Model

In this section we introduce the econometric model for the evolution of the realized volatility. Our model is a multiplicative model belonging to the MEM family of Engle (2002). In other words, the realized volatility,  $v_t$ , is assumed to evolve as

$$v_t = \mu_t \varepsilon_t, \quad t = 1, 2, \dots, T, \quad (1)$$

where the conditional mean

$$\mu_t = \omega + \sum_{i=1}^q \alpha_i v_{t-i} + \sum_{j=1}^p \beta_j \mu_{t-j} \quad (2)$$

and  $\varepsilon_t$  is a stochastic positive-valued error term with mean unity. In the sequel we will call this specification the MEM( $p, q$ ) model. The model has the same structure as the ACD model (Engle and Russell, 1998) for durations, and similar models have also been applied to the range of transaction prices (Chou, 2005) and transaction volume (Manganelli, 2005).

In the previous literature employing MEM models, various distributional assumptions on the error term  $\varepsilon_t$  have been entertained. Although Engle (2002) has shown that under regularity conditions, the consistent quasi maximum likelihood estimator is obtained by assuming  $\varepsilon_t$  to be exponentially distributed, this choice has been found inadequate in some empirical applications where, among others, the gamma (e.g. Engle and Gallo, 2003) and mixture of exponential distributions (De Luca and Gallo, 2004) have been assumed. To allow for even more flexibility, we consider a mixture of gamma-distributed random variables that has also the exponential (mixture) distribution as a special case, making comparisons between distributional assumptions straightforward. For simplicity and because it is likely to be adequate in a wide range

of applications, we restrict the number of mixture components to be two. Hence, we assume that  $\varepsilon_t$  is a mixture of  $\varepsilon_{1t}$  and  $\varepsilon_{2t}$  such that  $\varepsilon_{1t} \sim \text{Gamma}(\gamma_1, \delta_1)$  with probability  $\pi$  and  $\varepsilon_{2t} \sim \text{Gamma}(\gamma_2, \delta_2)$  with probability  $1 - \pi$  ( $0 < \pi < 1$ ). Since the error term needs to have mean unity, we impose the restrictions that  $\gamma_1 = 1/\delta_1$  and  $\gamma_2 = 1/\delta_2$ , i.e., the shape parameters are the reciprocals of the scale parameters.<sup>1</sup>

Additional flexibility into the model can easily be brought by allowing also (at least some of) the parameters of the conditional mean equation (2) to vary along with the parameters of the error distribution such that the conditional mean equals  $\mu_{1t}$  with probability  $\pi$  and  $\mu_{2t}$  with probability  $(1 - \pi)$  where

$$\mu_{1t} = \omega_1 + \sum_{i=1}^q \alpha_{1i} v_{t-i} + \sum_{j=1}^p \beta_{1j} \mu_{1,t-j} \quad (3)$$

and

$$\mu_{2t} = \omega_2 + \sum_{i=1}^q \alpha_{2i} v_{t-i} + \sum_{j=1}^p \beta_{2j} \mu_{2,t-j}. \quad (4)$$

The conditional mean can alternatively be written in matrix form as

$$\boldsymbol{\mu}_t = \boldsymbol{\omega} + \boldsymbol{\alpha}(L) v_{t-1} + \boldsymbol{\beta}(L) \boldsymbol{\mu}_{t-1}, \quad (5)$$

where  $\boldsymbol{\mu}_t = (\mu_{1t}, \mu_{2t})'$ ,  $\boldsymbol{\omega} = (\omega_1, \omega_2)'$  and  $\boldsymbol{\alpha}(L)$  and  $\boldsymbol{\beta}(L)$  are  $q$ th and  $p$ th order lag polynomials with coefficients

$$\boldsymbol{\alpha}_i = \begin{bmatrix} \alpha_{1i} \\ \alpha_{2i} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\beta}_i = \begin{bmatrix} \beta_{1i} & 0 \\ 0 & \beta_{2i} \end{bmatrix},$$

respectively

Provided  $E(\varepsilon_{1t}) = E(\varepsilon_{2t}) = 1$ , which is guaranteed by the restrictions  $\gamma_1 = 1/\delta_1$  and  $\gamma_2 = 1/\delta_2$ , conditional on past information  $\mathfrak{F}_{t-1} = \{v_{t-j}, j \geq 0\}$ ,

$$\begin{aligned} E_{t-1}(v_t) &= \pi E_{t-1}(\mu_{1t} \varepsilon_{1t}) + (1 - \pi) E_{t-1}(\mu_{2t} \varepsilon_{2t}) \\ &= \pi \mu_{1t} + (1 - \pi) \mu_{2t} \\ &= \boldsymbol{\pi}' \boldsymbol{\mu}_t, \end{aligned}$$

---

<sup>1</sup>The error term  $\varepsilon_t$  would have mean unity even under the more general restriction  $\pi \gamma_1 \delta_1 + (1 - \pi) \gamma_2 \delta_2 = 1$ . However, as shown below, the stricter constraint is required if, in addition, the conditional mean parameters are not constant across the mixture components.

where  $\boldsymbol{\pi} = (\pi, 1 - \pi)'$ . By the law of iterated expectations, the unconditional mean of  $v_t$  equals

$$E(v_t) = E[E_{t-1}(v_t)] = \boldsymbol{\pi}' E(\boldsymbol{\mu}_t).$$

The unconditional mean of vector  $\boldsymbol{\mu}_t$ ,  $E(\boldsymbol{\mu}_t)$ , can be obtained by straightforward calculation from (5). Another way is to make use of the following vector first-order representation of the conditional expectation of  $\boldsymbol{\mu}_t$  which is also convenient for studying the dynamic properties of the model,

$$E_{t-2}(\mathbf{M}_t) = \boldsymbol{\varpi} + \mathbf{C}\mathbf{M}_{t-1}, \quad (6)$$

where  $\mathbf{M}_t = (\boldsymbol{\mu}_t, \boldsymbol{\mu}_{t-1}, \dots, \boldsymbol{\mu}_{t-s+1}, \boldsymbol{\mu}_{t-s})'$ ,  $\boldsymbol{\varpi} = (\boldsymbol{\omega}, \mathbf{0}, \dots, \mathbf{0})'$  and

$$\mathbf{C} = \begin{bmatrix} \boldsymbol{\alpha}_1 \boldsymbol{\pi}' + \boldsymbol{\beta}_1 & \boldsymbol{\alpha}_2 \boldsymbol{\pi}' + \boldsymbol{\beta}_2 & \cdots & \boldsymbol{\alpha}_{s-1} \boldsymbol{\pi}' + \boldsymbol{\beta}_{s-1} & \boldsymbol{\alpha}_s \boldsymbol{\pi}' + \boldsymbol{\beta}_s \\ \mathbf{I}_2 & \mathbf{0} & & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 & & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & & \mathbf{I}_2 & \mathbf{0} \end{bmatrix}.$$

Here  $s = \max(p, q)$  and  $\alpha_{ki} = \beta_{kj} = 0$  for  $i > q$ ,  $j > p$  and  $k = 1, 2$ . By recursive substitution of (6) we obtain

$$E_{t-\tau-1}(\mathbf{M}_t) = \sum_{j=1}^{\tau-1} \mathbf{C}^j \boldsymbol{\varpi} + \mathbf{C}^\tau \mathbf{M}_{t-\tau}$$

for arbitrary  $\tau$ . This shows that the persistence of  $\boldsymbol{\mu}_t$  (and hence of  $v_t$ ) is governed by the largest eigenvalue of  $\mathbf{C}$ . Moreover, assuming covariance-stationarity,

$$E(\mathbf{M}_t) = \lim_{\tau \rightarrow \infty} E[E_{t-\tau-1}(\mathbf{M}_t)] = (\mathbf{I} - \mathbf{C})^{-1} \boldsymbol{\varpi}$$

so that the unconditional mean of  $\boldsymbol{\mu}_t$ ,

$$E(\boldsymbol{\mu}_t) = [\mathbf{I}_2 - \boldsymbol{\alpha}(1) - \boldsymbol{\beta}(1)]^{-1} \boldsymbol{\omega}.$$

Stationarity and the existence of moments in this kind of models have recently been studied by Meitz and Saikkonen (2004), Bougerol and Picard (1992) and Carrasco and Chen (2002).

The autocorrelation function of  $v_t$  implied by the model can be computed analytically in a straightforward manner in at least two special cases. First, when the parameters of the conditional expectation  $\mu_t$  are constant across the mixture components, standard methods derived for GARCH models (see, e.g. He and Teräsvirta, 1999) can be directly employed. Second, in the case of the first-order model the approach of Haas et al. (2004) can be used. Otherwise, the autocorrelation function can be obtained by means of simulation as exemplified in Section 5.2.

To guarantee positivity of  $v_t$ , a number of restrictions must be placed on the parameters. The constraints typically imposed in the closely related GARCH models are, however, far too severe, as pointed out by Nelson and Cao (1992) who derive less restrictive conditions. As the MEM model has a similar structure, their results should be directly applicable here. When parameters are allowed to vary across the mixture components, positivity is guaranteed if both sets of parameter values separately satisfy their conditions, i.e., if  $\mu_{1t}$  and  $\mu_{2t}$  each remain positive. For instance, in the case of the mixture-MEM(1,2) model found adequate for the DM/\$ and Yen/\$ volatilities in Section 5, these conditions require that  $\omega_i \geq 0$ ,  $\alpha_{i1} \geq 0$ ,  $0 \leq \beta_{i1} < 1$  and  $\beta_{i1}\alpha_{i1} + \alpha_{i2} > 0$  for  $i = 1, 2$ .

### 3 Estimation and Inference

The mixture-MEM model can be estimated in a straightforward way by the method of maximum likelihood (ML). Under the mixture of gamma distributions assumption, the realized volatility  $v_t$  has the following conditional distribution,

$$f_{t-1}(v_t; \theta) = \pi \frac{1}{\mu_{1t} \Gamma(\gamma_1) \delta_1^{\gamma_1}} \left( \frac{v_t}{\mu_{1t}} \right)^{\delta_1 - 1} \exp\left(-\frac{v_t}{\delta_1 \mu_{1t}}\right) + (1 - \pi) \frac{1}{\mu_{2t} \Gamma(\gamma_2) \delta_2^{\gamma_2}} \left( \frac{v_t}{\mu_{2t}} \right)^{\delta_2 - 1} \exp\left(-\frac{v_t}{\delta_2 \mu_{2t}}\right), \quad (7)$$

where  $\Gamma(\cdot)$  is the gamma function and  $\theta$  is a vector consisting of all the parameters in the model. Thus, the logarithmic likelihood function can be written as

$$l_T(\theta) = \sum_{t=1}^T \ln [f_{t-1}(v_t)],$$

and maximizing this function subject to the constraints  $\gamma_1 = 1/\delta_1$  and  $\gamma_2 = 1/\delta_2$  guaranteeing that the error term has mean unity, yields the maximum likelihood estimator.

The likelihood function is clearly twice continuously differentiable so that, assuming that the  $v_t$  series is generated by a stationary and ergodic process, it is reasonable to apply standard large sample results in testing. In particular, asymptotic standard errors can be obtained from the diagonal elements of the matrix  $-\left[\partial^2 l_T(\hat{\theta})/\partial\theta\partial\theta'\right]^{-1}$  where  $\hat{\theta}$  denotes the ML estimate of  $\theta$ . Likewise, Wald and likelihood ratio tests for more general hypotheses with conventional asymptotic  $\chi^2$  null distributions can be obtained.

As mentioned above, two mixture components should be sufficient in most applications, but sometimes more components might be needed to reach adequate fit while in some other cases a single component suffices. Because of the well-known problem of unidentified parameters under the null hypothesis (see, e.g., Davies, 1977) likelihood ratio (LR) test statistics of hypotheses restricting the number of mixture components do not have the usual asymptotic  $\chi^2$  null distributions. Hence, in the two-component model, for instance, the hypothesis  $\pi = 1$  cannot be tested in the usual way. One possibility would be to consider some computationally intensive testing procedures (see, e.g. Hansen, 1996). However, in this paper we opt for the use of diagnostic checks (see Section 4) to deem the adequacy of the estimated model. Furthermore, assuming that a subset of the parameters vary in time allows for testing for the constancy of others using standard asymptotics. So, assuming that the parameters of the error distribution vary, say, the constancy of (some of) the parameters of the conditional mean equation can be tested by the LR test in the usual way.

## 4 Diagnostic Checks

Checking for the adequacy of the conditional distribution of the estimated MEM model is complicated by the fact that its error term is not normal, while most existing diagnostic tests are based on normality. Furthermore, in the case of more than one mixture component, it is not even obvious how residuals can be obtained because each mixture component has its own error term and switching between the components is random. Therefore, instead of regular standardized residuals we consider diagnostic checks based on the transformed residuals suggested by Palm and Vlaar (1997). Analogous density forecast evaluation methods have recently been popularized in financial econometrics by Diebold et al. (1998) and Berkowitz (2001), *inter alia*, and applied to MEM-type duration models by Bauwens et al. (2004). However, as we are not primarily interested in forecasting the density of realized volatility, we concentrate on in-sample diagnostic checks and evaluate out-of-sample forecast performance in terms of point forecasts only.

The diagnostic checks are based on the probability integral transform of the data,

$$z_t = \int_{-\infty}^{v_t} f_{t-1}(u) du, \quad t = 1, 2, \dots, T,$$

where  $f_{t-1}(\cdot)$  is the conditional density of the realized volatility,  $v_t$ , implied by the model (given by (7) for the mixture-MEM model). If the conditional distribution is correct, the distribution of the sequence  $\{z_t\}_{t=1}^T$  is iid  $U[0, 1]$ . While a test for this hypothesis could be devised, Diebold et al. (1998) recommend considering the uniformity and independence properties separately and, in particular, emphasize the visual inspection of the histograms and autocorrelation functions of demeaned probability integral transform series and its squares. These procedures may help in finding out which part of the hypothesis is potentially violated better than mere test results.

To check for uniformity, Pearson's goodness-of-fit test can be used. The test statistic is based on a histogram of  $\{z_t\}_{t=1}^T$  consisting of  $m$  bins,

$$\sum_{t=1}^T \frac{(T_i - T/m)}{T/m},$$

where  $T_i$  is the number of observations in the  $i$ th bin. If  $z_t$  really is uniformly distributed,  $T_i$  should equal  $T/m$  for all  $i$ , i.e., there should be equal number of observations in each bin. Under the null hypothesis this test statistic follows chi-square distribution with  $m - 1$  degrees of freedom. Note that this test does not take into account parameter estimation uncertainty, i.e., the fact that  $z_t$  is not directly observable but is based on an estimated model.

## 5 Empirical Results

### 5.1 Data

The mixture-MEM model is applied to the time series of the realized volatility of two spot foreign exchange rates, Deutschemark and Japanese Yen against the U.S. Dollar. Following Andersen et al. (2003), we compute a measure of the daily realized variance by summing squared thirty-minute returns over each trading day. The use of thirty-minute returns is a compromise between the theoretical considerations recommending sampling at very high frequencies and the desire to avoid contamination by microstructure effects. The returns are based on interbank bid and ask quotes displayed on Reuters FAFX screen. These quotes are only *indicative* rather than *firm* in that they are not binding commitments to trade. Hence, as recently pointed out by Daniélsson and Payne (2002), at very high frequencies they may not accurately measure tradeable exchange rates. Daniélsson and Payne (2002), however, show that at levels of aggregation of five minutes and above, returns computed from these data are a fairly good proxy for firm returns which is a further argument against using very disaggregated data.

For comparability, we use exactly the same data as Andersen et al. (2003), covering the period from the beginning of December 1986 until the end of June 1999. All the returns between Friday 21:00 GMT and Sunday 21:00 GMT are excluded as well as a number of inactive days such as holiday periods.<sup>2</sup> This leaves us 3,045 ob-

---

<sup>2</sup>The raw data comprise millions of quotes compiled by Olsen & Associates. For details of

servations in total, of which, following Andersen et al. (2003), 2,449 (from December 1, 1986 through December 1, 1996) form the estimation period, while the remaining 596 observations (from December 2, 1996 through June 30, 1999) are left for forecast evaluation. The time series of the square root of realized volatility are plotted in Figure 1.

## 5.2 Estimation Results

In this section we present the estimation results for the proposed mixture-MEM models. Following standard practice in the previous literature, the dependent variable in all the models is  $v_t$ , the realized standard deviation. Furthermore, statistical fit turned out to be inferior for the realized variance that was also considered as the dependent variable. For comparison, we also estimated ARFIMA models that have previously been proposed as an to model the time series behavior realized volatility (see Andersen et al. (2003) and the references therein).

Estimation results for the mixture-MEM models are presented in Table 1.<sup>3</sup> Using diagnostic checks (see below), the MEM(1,2) specification with two mixture components turned out to be adequate for both the DM/\$ and Yen/\$ realized volatility series.<sup>4</sup> The parameters are, in general, very accurately estimated, and the estimates satisfy the conditions for positivity discussed in Section 2. The estimated densities of the error terms  $\varepsilon_{1t}$  and  $\varepsilon_{2t}$  in the model for the DM/\$ volatility depicted in Figure 2 attest to considerable differences across the mixture components; they are virtually indiscernible from those in the model for the Yen/\$ volatility. Approximately 87% 

---

constructing the realized variance series, see Andersen et al. (2003). The data set was downloaded from <http://www.ssc.upenn.edu/~fdiebold>.

<sup>3</sup>The MEM models were estimated using the BHHH algorithm implemented in the CML package of GAUSS.

<sup>4</sup>Among other things, we estimated MEM(1,1) models, which suffer from unmodeled autocorrelation, and attempted to estimate MEM models with three mixture components, which reduced to the two-component models presented in Table 1, i.e., the estimated probability of third component was estimated zero.

(75%) of the time the errors to the DM/\$ (Yen/\$) volatility come from the almost symmetric distribution with relatively small variance located around unity. The rest of the time the errors are generated from the right-skewed distribution with higher variance that is probably the source of the occasional spikes in Figure 1. In each model the two mixture components have quite different dynamics. The prevalent components are highly persistent for both currencies, while for the DM/\$ volatility the other component is considerably less persistent and for the Yen/\$ volatility it almost exhibits unit root type behavior. The largest eigenvalues of the  $\mathbf{C}$  matrix (see Section 2), 0.961 and 0.972 for the DM/\$ and Yen/\$ volatility, respectively, indicate also relatively high overall persistence. Moreover, the estimate of the  $\beta$  parameter in the latter component in the model for the DM/\$ volatility ran into the zero constraint required by positivity, indicating that the conditional expectation of volatility here only depends on past observed volatilities which makes this component potentially very erratic. Simulations of long realizations from either model indicate no deviations from stationarity.

As mentioned above, in the previous literature the long-memory ARFIMA model has been found to fit realized volatility series rather well. Therefore, to get a benchmark for comparisons, we also estimated ARFIMA( $r, d, 0$ ) models for both series.<sup>5</sup> In particular, these models were estimated for the logarithm of the realized standard deviation, assuming a normally distributed error term. With a maximum of 5 lags, the Akaike information criterion (AIC) selected 3 and 0 lags for the DM/\$ and Yen/\$ realized volatility series, respectively.<sup>6</sup> The estimated orders of fractional integration,  $d$ , were 0.476 and 0.413. These figures are in line with those obtained in the previous literature and suggest the presence of long memory in the realized volatility series.

---

<sup>5</sup>All the computations of the ARFIMA models are implemented using the ARFIMA package of Doornik and Ooms (2001) within the programming environment of Ox, see Doornik (2001).

<sup>6</sup>Following Andersen et al. (2003), we also estimated a higher-order model with five lags. However, the extra lags had virtually no effect on the results. Most results on the ARFIMA models are not reported in detail, but they are available upon request.

The long-memory property is also consistent with the estimated autocorrelation functions in Figure 3. Although the mixture-MEM model is a short-memory model, the estimated models produce rather slowly decaying autocorrelation functions. To verify that the observed autocorrelation functions could have been generated by the estimated mixture-MEM models, we simulated 10,000 realizations and computed their autocorrelation functions. Then, bands including 95% of the middle autocorrelation coefficients at each lag were formed. As Figure 3 shows, the observed autocorrelation functions are included in the 95% bands, indicating that the data from which they have been estimated could plausibly have been generated by the mixture-MEM models.

While the mixture-MEM model is rather complicated, no obvious simplifications are supported by the data. Tests of the same dynamics in the two mixture components, i.e., the hypothesis  $\omega_1 = \omega_2$ ,  $\alpha_{11} = \alpha_{21}$ ,  $\alpha_{12} = \alpha_{22}$  and  $\beta_1 = \beta_2$ , were rejected with p-values very close to zero for both currencies. Likewise, the hypothesis for  $\varepsilon_{1t}$  and  $\varepsilon_{2t}$  having the same probability distribution was clearly rejected in both cases. This outcome also lends support to the insufficiency of the exponential distribution suggested by Engle (2002). The mixture-MEM models are capable of generating a wide variety of predictive densities, as exemplified by the upper panel of Figure 4 for the DM/\$ volatility for two dates (January 20, 1988 and May 7, 1990) the first of which represents a relatively volatile day, while the latter one is close to the average. It is interesting to note that the mixture model may also produce a bimodal predictive density as happened to be the case for January 20, 1988. In the lower panel of Figure 4 are depicted the 5th and 95th percentiles of the predictive distributions of the two models along with the realized values for the 50 days following that day. The probability mass of the conditional distribution of the mixture-MEM model is highly concentrated compared to the corresponding densities implied by the ARFIMA models that give a relatively high probability for a wide range of values. Therefore, the observed values are always included in this range for the ARFIMA models, when volatility is high. On the other hand, small observed values often fall outside the range

which rarely is the case with the range implied by the mixture-MEM model. Similar figures for other time periods and also the Yen/\$ realized volatility (not shown) reconfirm these findings that suggest better forecast accuracy of the mixture-MEM model, particularly on days when the true volatility is not very high.

The diagnostic checks for the models for the DM/\$ and Yen/\$ volatility are presented in Figures 5 and 6, respectively. We use these checks also to compare the mixture-MEM and ARFIMA models. A formal test does not seem feasible as the models not only are not nested but are also specified for different variables. For both currencies the general result is the same. The histograms of the mixture-MEM models lend support to the specified distribution, whereas the ARFIMA models seem to be clearly misspecified. This impression is confirmed by the p-values of Pearson's goodness-of-fit test based on these histograms. For the mixture-MEM models for the DM/\$ and Yen/\$ volatility they are 0.16 and 0.25, respectively, whereas the corresponding figures for the ARFIMA models are  $3.30 \cdot 10^{-12}$  and  $1.83 \cdot 10^{-14}$ . While virtually all bins in the histograms based on the MEM models lie within the 95% confidence interval, the ARFIMA models exhibit several violations. In particular, for both currencies the histogram takes a hump shape with the exception of the rightmost bin being clearly above the confidence interval, indicating that the tails are not adequately accounted for, with the exception of the ultimate upper tail which is overemphasized. In other words, the predictive distributions implied by the ARFIMA models are too narrow, but have too much probability mass on the extreme right tail. This is in accordance with the examples of predictive distributions depicted in Figure 4. There seems to be no evidence of autocorrelation of the demeaned probability integral transform series, but for both currencies and models their squares exhibit some autocorrelation. Thus, although Andersen et al. (2001) have shown that these realized exchange rate volatility series can unconditionally be characterized as log-normally distributed, the ARFIMA model coupled with lognormality does not seem to provide a good description of their conditional distribution. The mixture-MEM model, on the other hand, seems to fit to both series adequately.

### 5.3 Forecasting Performance

To study the performance of the mixture-MEM model in forecasting out-of-sample exchange rate volatility, we compare it to the estimated ARFIMA models. In the previous literature the ARFIMA model specified for the logarithmic realized volatility has been found to have superior forecasting performance (see e.g., Andersen et al. (2003) and Koopman et al. (2005) and their references), and, hence, we take that as the natural contender to compare the mixture-MEM model to. In addition, comparisons to the models considered by Andersen et al. (2003) for the same data are conducted. Thus, the set of alternative models comprises univariate (mixture-MEM and ARFIMA) and multivariate (VAR-RV) models for realized volatility as well as different models estimated using daily return data (GARCH(1,1), RiskMetrics, FIEGARCH(1,  $d$ , 0) and VAR-ABS) and a FIEGARCH model for intradaily data. VAR-RV is a fifth-order long-memory Gaussian vector autoregressive model for the logarithmic DM/\$, Yen/\$ and Yen/DM realized volatilities with the coefficient of fractional integration fixed at 0.401, while VAR-ABS is the corresponding model for daily logarithmic absolute returns. Two ARFIMA specifications for the logarithmic realized volatility are considered: those presented in Section 5.2 and ARFIMA(5, 0.401, 0) models of Andersen et al. (2003). The RiskMetrics model corresponds to an IGARCH(1,1) with intercept fixed at zero and the moving average coefficient in the ARIMA(0, 1, 1) representation for the squared returns equal to  $-0.94$ . The intradaily FIEGARCH model is based on deseasonalized and filtered half-hour returns. For details of the alternative specifications, see Andersen et al. (2003).<sup>7</sup> In all comparisons the out-of-sample forecast period covers the days from December 2, 1996 through June 30, 1999 (the 596 last observations) and the models have been estimated using data from the estimation period only.

In evaluating the forecasting performance we concentrate on the mean square

---

<sup>7</sup>The forecasts of the models estimated by Andersen et al. (2003) were downloaded from <http://www.ssc.upenn.edu/~fdiebold>.

error,

$$MSE = \frac{1}{T^*} \sum_{t=1}^{T^*} (v_t - \hat{v}_t)^2,$$

where  $T^*$  is the length of the forecast period and  $\hat{v}_t$  is the volatility forecast implied by the model. It is easy to show that this loss function satisfies Hansen and Lunde's (in press) sufficient conditions for correct ranking of volatility forecasts when they are measured against an imperfect proxy such as the realized volatility  $v_t$ . Some other commonly employed loss functions, including the mean absolute error (MAE), on the other hand, do not satisfy these conditions and their use can lead to the incorrect model being selected. Following Andersen et al. (2003), one- and ten-day-ahead forecasts are compared.

We start the forecast comparisons by reporting the results of the pairwise test due to Diebold and Mariano (1995) for forecast accuracy. As the null model we take the mixture-MEM(1,2) model so that negative values of the Diebold-Mariano test statistic indicate that this model has smaller MSE than the contender. For the DM/\$ volatility (Table 2) the mixture-MEM model is significantly (at the 5% level) more accurate than models based on daily returns, while differences against other models are not statistically significant. For the ten-day-ahead volatility the mixture-MEM model is more accurate than any other model, and the differences are significant at the 5% level against all but the ARFIMA(5, 0.401, 0), VAR-RV and GARCH(1,1) models. As far as the Yen/\$ volatility (Table 3) is concerned, for one-day-ahead volatility the mixture-MEM model has a smaller MSE than the VAR-RV or the univariate ARFIMA models, but the differences are not significant. For the ten-day-ahead volatility, on the other hand, the mixture-MEM model is significantly more accurate than the ARFIMA(5, 0.401,0), VAR-RV and VAR-ABS models at the 5% level, while the differences against other models are not significant. Thus, the results suggest that the mixture-MEM model performs at least as well as the VAR-RV model that Andersen et al. (2003) selected the best. Moreover, the VAR-RV model does not seem to perform any better against the mixture-MEM model

than the ARFIMA(5, 0.401,0) model which is its univariate counterpart. Hence, accounting for the multivariate interaction across the realized volatilities does not seem to improve forecasts.

The Diebold-Mariano tests provide only pairwise comparisons between the mixture-MEM and competing models that does not take into account the entire universe of models being compared. Essentially, we are interested in finding out whether any of the previously presented models provide more accurate forecasts than the mixture-MEM model. Therefore, we next report results of Hansen's (2005) test for superior predictive ability (SPA) that allows for controlling for the full set of models and their interdependence when evaluating the significance of relative forecasting performance. The null hypothesis is that the benchmark is not inferior to any alternative forecast. As the VAR-RV model was selected the best by Andersen et al. (2003) and the Diebold-Mariano test results indicate that the mixture-MEM model is at least as accurate as that model, it is natural to consider these two models, in turn, as benchmark models. The results are presented in Table 4.<sup>8</sup> There is no evidence against the mixture-MEM model being superior to the competing models for either currency or forecast horizon. However, there is strong evidence against the VAR-RV model being the superior model for the Yen/\$ volatility, while its superiority cannot be rejected for the DM/\$ volatility. Hence, the SPA test leads to the same general conclusion as the Diebold-Mariano test that the mixture-MEM model is at least as accurate as the other models and in some cases even more accurate. Note also, that the results are robust in that both the liberal and conservative p-values lead to the same conclusion as the consistent p-value is each case.<sup>9</sup>

---

<sup>8</sup>I am grateful to Peter Hansen for the Ox code for the SPA test.

<sup>9</sup>We also ran both the Diebold-Mariano and SPA tests in the universe including, in addition, the ten models considered but not reported by Andersen et al. (2003), and the conclusions remained intact.

## 6 Conclusion

In this paper we have introduced a new kind of multiplicative mixture model for realized volatility. An application to foreign exchange rate volatility illustrates the model's good fit and forecast performance, even compared to the commonly employed long-memory ARFIMA model. Moreover, there is little evidence of it producing forecasts inferior to those produced by any of the volatility models considered by Andersen et al. (2003) for the same data. Although the mixture model is a short-memory model, it is demonstrated that it can plausibly generate the slowly decaying autocorrelation pattern characteristic of realized volatility series. In addition to the apparent long memory, another feature of realized volatility series is the fact that the lognormal distribution provides a good description of their unconditional distribution (see, Andersen et al., 2001a, 2001b). This observation has been used to motivate the use of the ARFIMA model with a Gaussian error term for the logarithmic realized volatility in the previous literature. However, our results suggest that conditionally lognormality is not an adequate description.

We have considered univariate models with fixed mixing probabilities, but the model could be extended in several directions. A corresponding multivariate model would be a natural extension, while just including exogenous explanatory variables in the univariate model could provide interesting empirical applications. The mixing probabilities could also be made time-varying by allowing them to depend on lagged values of realized volatility or some exogeneous variables, such as the lagged return. For risk measurement, including Value-at-Risk (VaR) analysis, the mixture model should ideally be augmented with a model for the return. Andersen et al. (2003), on the basis of the observation that returns standardized by realized volatility seem to be normally distributed, suggested computing VaR predictions assuming normality. However, Giot and Laurent (2004) recently showed that for the exchange rate and stock returns they modeled, standardized by realized volatility forecasts from ARFIMAX models are not Gaussian, and our results reconfirmed this finding for the

mixture-MEM and ARFIMA models for exchange rate returns. Finally, it would be interesting to see how the model fits to realized stock return volatility as well as certain other positive-valued financial time series, such as range-based volatility and trading volume.

## References

- Alizadeh, S., M. Brandt, and F.X. Diebold (2002), Range-based estimation of stochastic volatility models. *Journal of Finance* 57, 1047–1092.
- Andersen, T.G., and T. Bollerslev (1988), Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39, 885–905.
- Andersen, T.G., T. Bollerslev, F.X. Diebold, and H. Ebens (2001a), The distribution of realized stock return volatility. *Journal of Financial Economics* 61, 43–76.
- Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys (2001b), The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* 96, 42–55.
- Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys (2003), Modeling and forecasting realized volatility. *Econometrica* 71, 579–625.
- Barndorff-Nielsen, O.E., and N. Shephard (2001), Non-Gaussian OU based models and some of their uses in financial economics (with discussion). *Journal of the Royal Statistical Society. Series B* 63, 167–241.
- Bauwens, L., P. Giot, J. Grammig, and D. Veredas (2004), A comparison of financial duration models via density forecasts. *International Journal of Forecasting* 20, 589–609.

- Berkowitz, J. (2001), Testing density forecast, with applications to risk management. *Journal of Business and Economic Statistics* 19, 465–474.
- Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.
- Bougerol, P., and N. Picard (1992), Stationarity of GARCH processes and some nonnegative time series. *Journal of Econometrics* 52, 115–127.
- Bos, C.S., P.H. Franses, and M. Ooms (2002), Inflation forecast intervals and long memory regression models. *International Journal of Forecasting* 18, 243–264.
- Carrasco, M., and X. Chen (2002), Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17–39.
- Chou, R.Y. (2005), Forecasting financial volatilities with extreme values: The conditional autoregressive range (CARR) model. *Journal of Money, Credit and Banking* 37, 561–582.
- Daniélsson, J., and R. Payne (2002), Real trading patterns and prices in spot foreign exchange markets. *Journal of International Money and Finance* 21, 203–222.
- Davies, R.B. (1977), Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64, 247–254.
- De Luca, G., and G.M. Gallo (2004), Mixture processes for financial intradaily durations. *Studies in Nonlinear Dynamics & Econometrics* 8, 1–18.
- Diebold, F.X., T. Gunther, and A.S. Tay, Evaluating density forecasts, with applications to financial risk management. *International Economic Review* 39, 863–883

- Diebold, F.X., and R.S. Mariano (1995), Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13, 253–263.
- Doornik, J.A. (2001), *Ox: Object oriented matrix programming 3.0*. Timberlake Consultants Press, London.
- Doornik, J.A., and M. Ooms (2001), A package for estimating forecasting and simulating ARFIMA models: ARFIMA package 1.01 for Ox, Working Paper. Nuffield College, Oxford.
- Engle, R. (2002), New frontiers for ARCH models. *Journal of Applied Econometrics* 17, 425–446.
- Engle, R.F., and G.M. Gallo (2003), A multiple indicators model for volatility using intra-daily data. Department of Statistics Working Paper 2003/7, Università degli Studi di Firenze.
- Engle, R.F., and J.E. Russell (1998), Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* 66, 1127–1162.
- Giot, P., and S. Laurent (2004), Modeling daily Value-at-Risk using realized volatility and ARCH type models. *Journal of Empirical Finance* 11, 379–398.
- Haas, M., S. Mittnik, and M.S. Paolella (2004), Mixed normal conditional heteroskedasticity. *Journal of Financial Econometrics* 2, 211–250.
- Hansen, B.E. (1996), Inference when a nuisance parameter is present only under the alternative. *Econometrica* 64, 413–430.
- Hansen, P.R. (2005), A test for superior predictive ability. *Journal of Business and Economic Statistics* 23, 365–380.
- Hansen, P.R., and A. Lunde (in press), Consistent ranking of volatility models. *Journal of Econometrics*.

He, C., and T. Teräsvirta (1999), Fourth moment structure of the GARCH(p,q) process. *Econometric Theory* 15, 824–846.

Koopman, S.J., B. Jungbacker, and E. Hol (2005), Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance* 12, 445–475.

Manganelli, S. (2005), Duration, volume and volatility impact of trades. *Journal of Financial Markets* 8, 377–399.

Maheu, J.M., and T.H. McCurdy (2002), Nonlinear features of realized FX volatility. *Review of Economics and Statistics*, 84, 668–681.

Meitz, M., and P. Saikkonen (2004), Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models. Unpublished manuscript. University of Helsinki.

Morana, C., and A. Beltratti (2004), Structural change and long-range dependence in volatility of exchange rates: either, neither or both? *Journal of Empirical Finance* 11, 629–658.

Nelson, D.B., and C.Q. Cao (1992), Inequality constraints in the univariate GARCH model. *Journal of Business and Economic Statistics* 10, 229–235.

Palm, F.C., and P.J.G. Vlaar (1997), Simple diagnostic procedures for modeling financial time series. *Allgemeines Statistisches Archiv* 81, 85–101.

Pong, S., M.B. Shackleton, S.J. Taylor, and X. Xu (2004), Forecasting currency volatility: A comparison of implied volatilities and AR(FI)MA models. *Journal of Banking and Finance* 28, 2541–2563.

Taylor, S.J. (1986), *Modeling Financial Time Series*. John Wiley and Sons, Chichester.

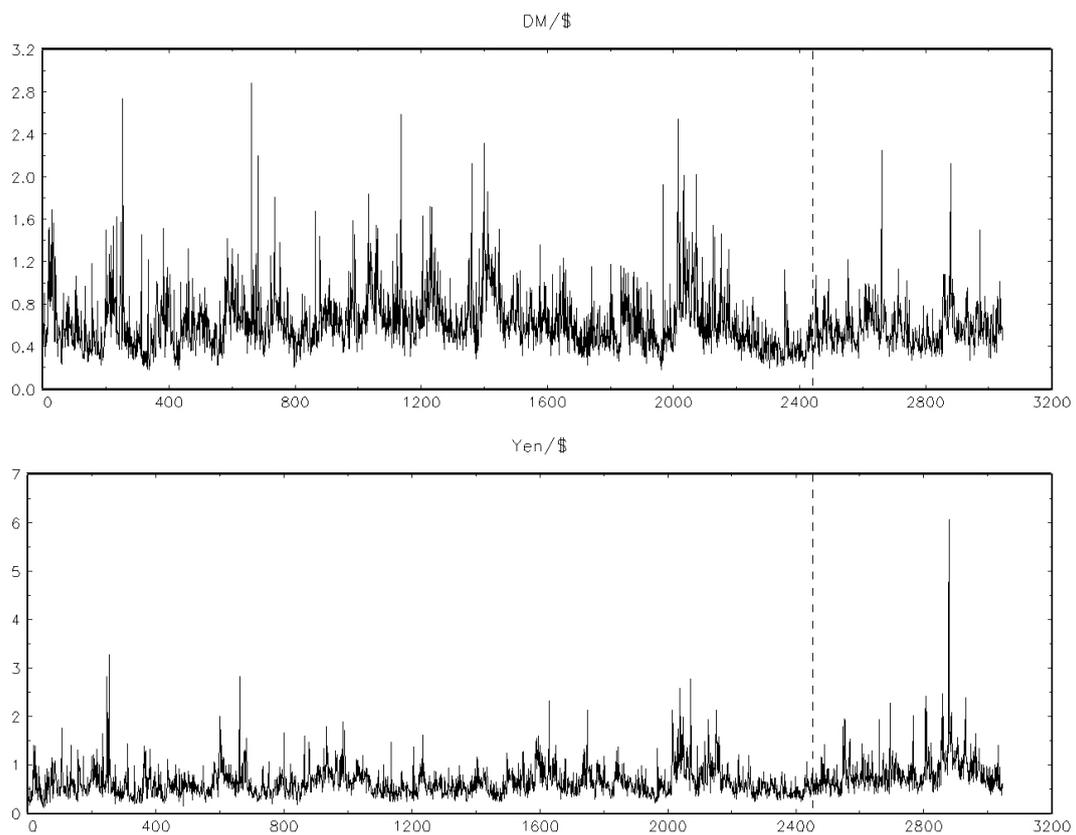


Figure 1: Time series of daily realized volatilities from January 1, 1986 through June 30, 1999. The dashed vertical line shows the last observation of the estimation period (December 1, 1996).

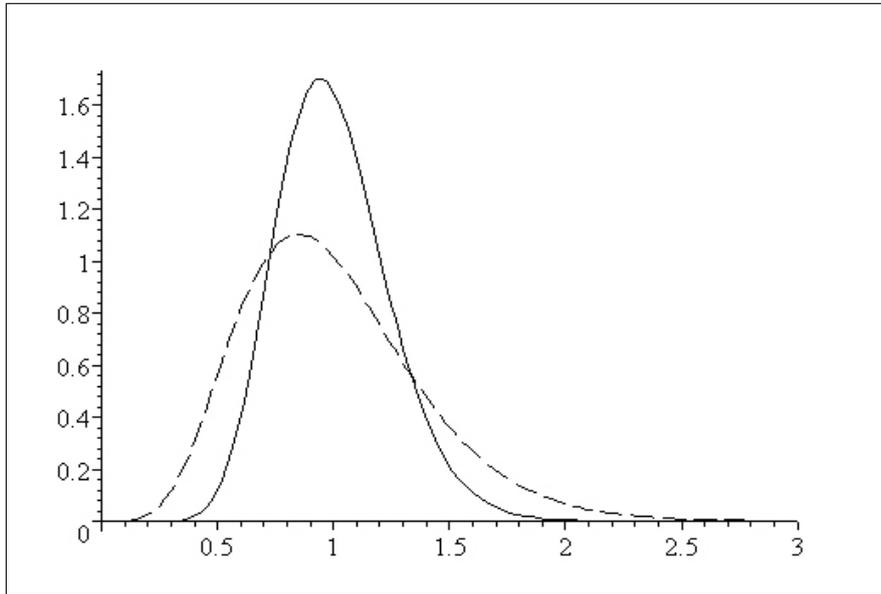


Figure 2: The estimated densities of the error terms of the mixture-MEM(1,2) model for the DM/\$ realized volatility. The solid and dashed lines are the densities of  $\varepsilon_{1t}$  and  $\varepsilon_{2t}$ , respectively.

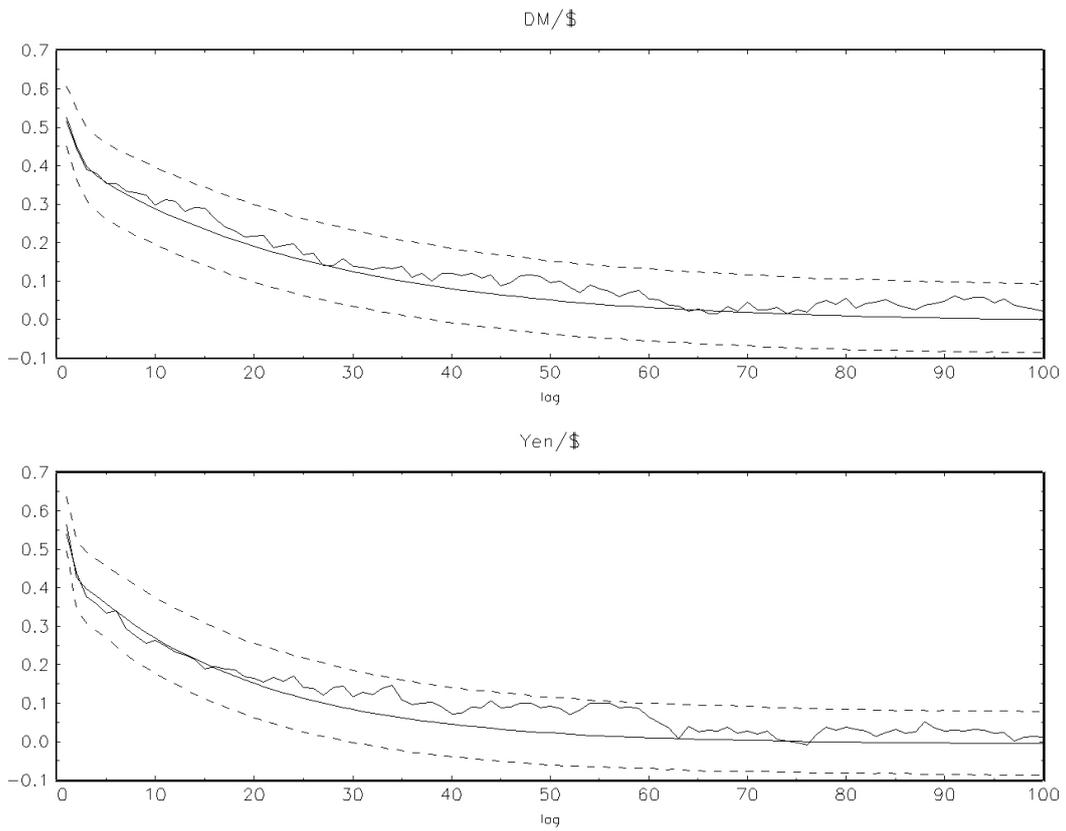


Figure 3: Autocorrelation functions of the realized volatility series estimated from the data (solid line) and implied by the mixture-MEM-models (smooth solid line) and the bands including 95% of the autocorrelations implied by the estimated mixture-MEM models (dashes).

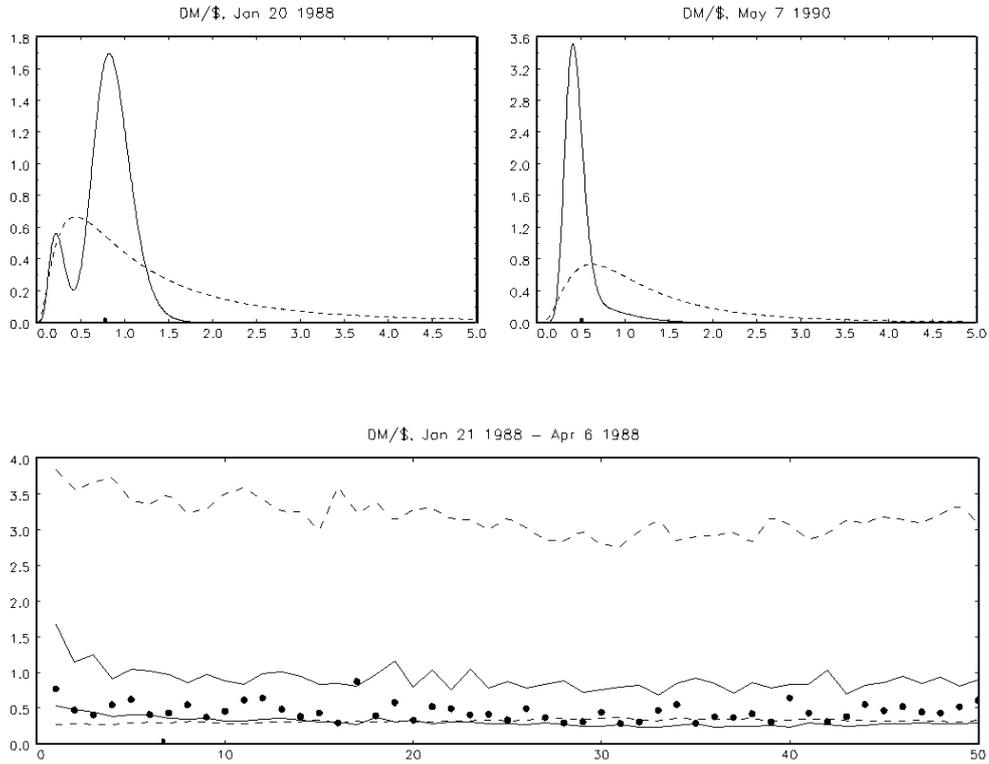


Figure 4: Examples of predictive densities for the DM/\$ realized volatility. In the upper panel the solid and dashed lines denote the densities implied by the mixture-MEM and ARFIMA models, respectively. In the lower panel the solid and dashed lines give the 5th and 95th percentiles of the predictive distributions of the mixture-MEM and ARFIMA models, respectively. The dots denote the observed values.

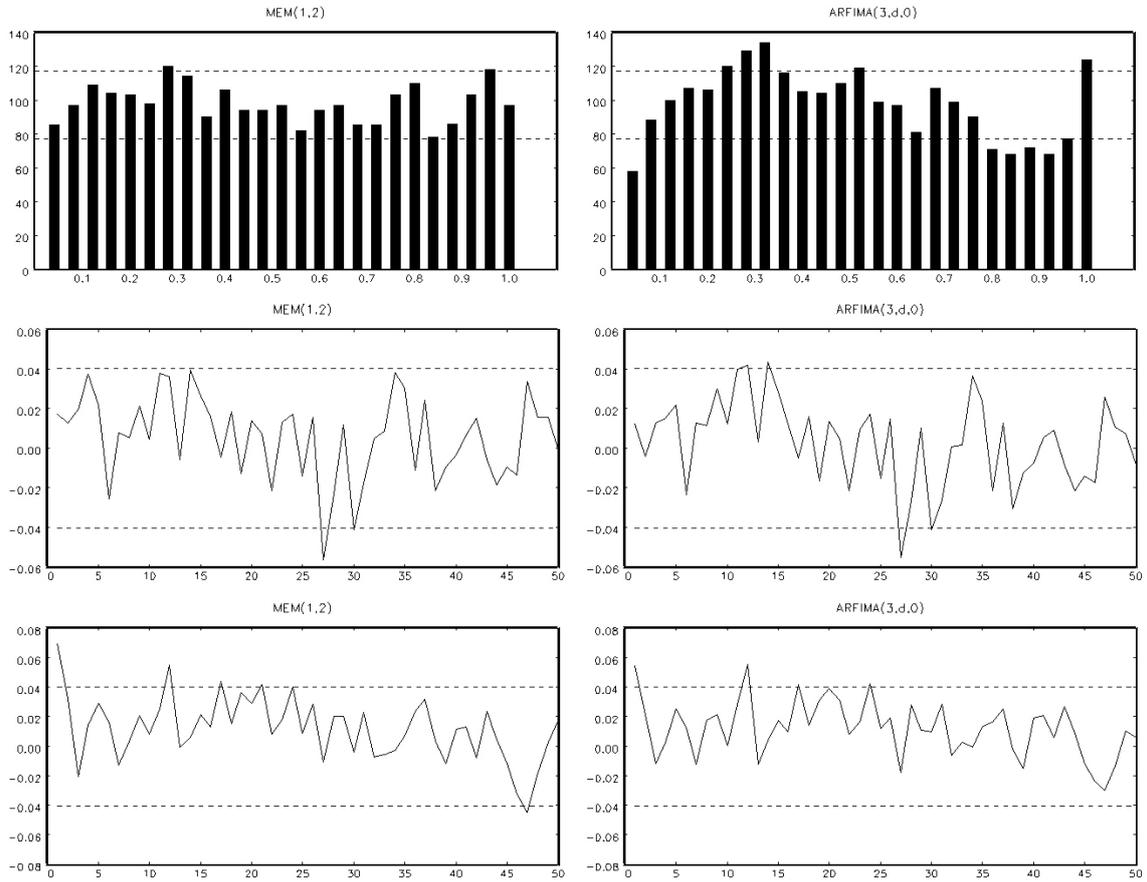


Figure 5: Diagnostics for the probability integral transforms of the DM/\$ data using the estimated mixture-MEM and ARFIMA models. The upper panel depicts their frequency distributions, and the middle and lower panels the autocorrelation functions of the demeaned transforms and their squares, respectively. The dashed lines are the 95% confidence intervals.

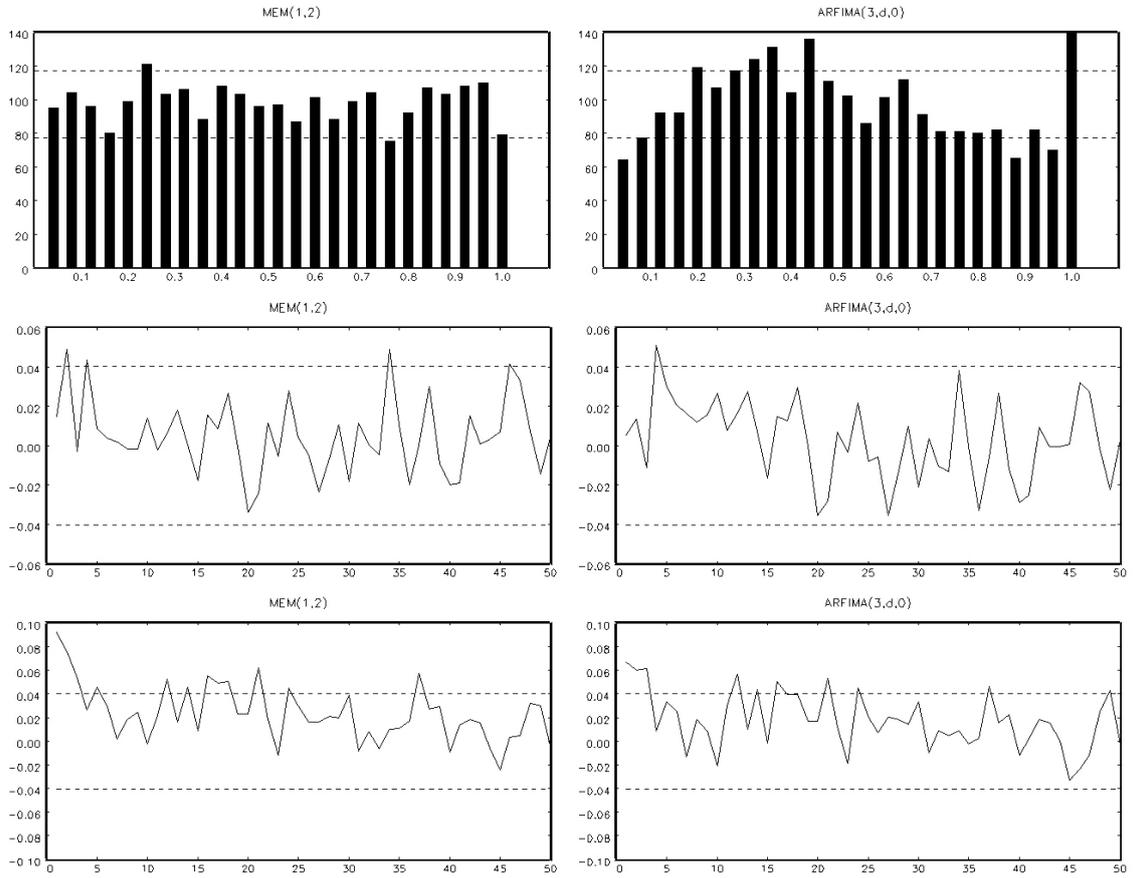


Figure 6: Diagnostics for the probability integral transforms of the Yen/\$ data using the estimated mixture-MEM and ARFIMA models. The upper panel depicts their frequency distributions, and the middle and lower panels the autocorrelation functions of the demeaned transforms and their squares, respectively. The dashed lines are the 95% confidence intervals.

Table 1: Estimation results for the mixture-MEM(1,2) models for the realized volatilities.

	DM/\$		Yen/\$	
$\pi$	0.870	(0.023)	0.738	(0.042)
$\gamma_1$	17.326	(0.832)	18.379	(1.429)
$\omega_1$	0.010	(0.003)	0.013	(0.004)
$\alpha_{11}$	0.325	(0.020)	0.372	(0.024)
$\alpha_{12}$	-0.179	(0.025)	-0.183	(0.034)
$\beta_1$	0.826	(0.018)	0.767	(0.028)
$\gamma_2$	6.664	(0.816)	6.549	(0.529)
$\omega_2$	0.446	(0.073)	0.014	(0.013)
$\alpha_{21}$	0.294	(0.143)	0.498	(0.081)
$\alpha_{22}$	0.484	(0.138)	-0.430	(0.104)
$\beta_2$			0.929	(0.047)
log likelihood	854.49		759.84	

The figures in parentheses are standard errors computed from the inverse of the final Hessian matrix.

Table 2: Out-of-sample forecast evaluation: the Diebold-Mariano test for the models for the DM/\$ volatility.

Model	One-Day-Ahead Forecast			Ten-Day-Ahead Forecast		
	MSE	D-M Statistic	p-value	MSE	D-M Statistic	p-value
Mixture-MEM(1,2)	0.0291			0.1522		
ARFIMA(3, $d$ ,0)	0.0290	0.289	0.773	0.2056	-2.543	0.011
ARFIMA(5,0.401,0)	0.0291	-0.092	0.926	0.1638	-0.844	0.399
VAR-RV	0.0288	0.664	0.507	0.1618	-0.795	0.427
Intraday FIEGARCH	0.0317	-1.936	0.053	0.2343	-2.474	0.013
Daily GARCH(1,1)	0.0360	-3.153	0.002	0.185	-1.530	0.126
Daily RiskMetrics	0.0358	-2.781	0.005	0.2193	-2.474	0.013
Daily FIEGARCH(1, $d$ ,0)	0.0411	-4.600	< 0.001	0.2428	-3.154	0.002
VAR-ABS	0.0293	-10.939	< 0.001	1.021	-6.524	< 0.001

The null model in the Diebold-Mariano test is the mixture-MEM(1,2) model.

Table 3: Out-of-sample forecast evaluation: the Diebold-Mariano test for the models for the Yen/\$ volatility.

Model	One-Day-Ahead Forecast			Ten-Day-Ahead Forecast		
	MSE	D-M Statistic	p-value	MSE	D-M Statistic	p-value
Mixture-MEM(1,2)	0.1135			0.9002		
ARFIMA(0, $d$ ,0)	0.1151	-0.520	0.603	0.9837	-1.363	0.173
ARFIMA(5,0.401,0)	0.1200	-1.141	0.254	0.9768	-2.388	0.017
VAR-RV	0.1199	-1.340	0.180	0.9640	-2.465	0.014
Intraday FIEGARCH	0.1254	-1.966	0.049	1.2604	-1.897	0.058
Daily GARCH(1,1)	0.1218	-1.999	0.046	0.8504	0.621	0.535
Daily RiskMetrics	0.1327	-3.060	0.002	1.0252	-0.942	0.346
Daily FIEGARCH(1, $d$ ,0)	0.1125	0.172	0.864	0.8678	0.524	0.600
VAR-ABS	0.3553	-6.494	< 0.001	3.316	-5.997	< 0.001

The null model in the Diebold-Mariano test is the mixture-MEM(1,2) model.

Table 4: Out-of-sample forecast evaluation: the test for superior predictive ability.

	Benchmark	Naive	$SPA_l$	$SPA_c$	$SPA_u$
DM/\$, One-Day Ahead Forecast	Mixture-MEM(1,2)	0.254	0.414	0.427	0.724
	VAR-RV	0.723	0.684	0.974	0.997
DM/\$, Ten-Day Ahead Forecast	Mixture-MEM(1,2)	0.921	0.559	0.943	1.000
	VAR-RV	0.079	0.082	0.185	0.300
Yen/\$, One-Day Ahead Forecast	Mixture-MEM(1,2)	0.403	0.685	0.832	0.933
	VAR-RV	0.224	0.030	0.033	0.039
Yen/\$, Ten-Day Ahead Forecast	Mixture-MEM(1,2)	0.204	0.247	0.403	0.665
	VAR-RV	0.077	< 0.001	< 0.001	< 0.001

The null hypothesis is that the benchmark model is the best model, i.e., produces the smallest MSE. The figures are p-values. The naive p-value compares the model with the smallest MSE to the benchmark, ignoring the other models. The columns  $SPA_l$  and  $SPA_u$  contain the p-values of a liberal and conservative test, respectively, whereas the column  $SPA_c$  gives the consistent p-value (see Hansen, 2005, for details). The p-values are based on 1,000 bootstrap resamples.