

EUI WORKING PAPERS

ECO No. 2006/20



Forecasting Realized Volatility by Decomposition

MARKKU LANNE



EUROPEAN UNIVERSITY INSTITUTE

Department of Economics

EUROPEAN UNIVERSITY INSTITUTE
DEPARTMENT OF ECONOMICS

Forecasting Realized Volatility by Decomposition

MARKKU LANGE

This text may be downloaded for personal research purposes only. Any additional reproduction for such purposes, whether in hard copy or electronically, requires the consent of the author(s), editor(s). 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.

The author(s)/editor(s) should inform the Economics Department of the EUI if the paper is to be published elsewhere, and should also assume responsibility for any consequent obligation(s).

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>

Forecasting Realized Volatility by Decomposition*

Markku Lanne[†]

European University Institute,
University of Jyväskylä, RUESG and HECER

Abstract

Forecasts of the realized volatility of the exchange rate returns of the Euro against the U.S. Dollar obtained directly and through decomposition are compared. Decomposing the realized volatility into its continuous sample path and jump components and modeling and forecasting them separately instead of directly forecasting the realized volatility is shown to lead to improved out-of-sample forecasts. Moreover, gains in forecast accuracy are robust with respect to the details of the decomposition.

*I thank Helmut Lütkepohl for useful comments. Financial support from the Yrjö Jahnsson Foundation and Suomen Arvopaperimarkkinoiden Edistämissäätiö is gratefully acknowledged. Part of this research was done while the author was a Jean Monnet Fellow at the Economics Department of the European University Institute.

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

1 Introduction

Recently, Andersen et al. (2005) suggested modeling and forecasting the realized volatility of exchange rate, stock and bond returns by extracting the component due to jumps and including it as an explanatory variable in a HAR-RV regression model of Müller et al. (1997) and Corsi (2003). In some cases, the jump component turned out to be highly significant and considerable increases in the coefficient of determination were observed. This suggests that gains in forecasting the realized volatility could be made by separately modeling and forecasting the jump and continuous sample path components and obtaining forecasts of the realized volatility as their sum instead of considering the aggregate realized volatility, as conjectured by Andersen et al. (2005). The purpose of this paper is to study whether such an approach really would be beneficial and whether the potential gains in forecast accuracy depend on the way the decomposition is made. To this end, we examine the returns of the Euro against the U.S. Dollar. To model the realized volatility and the continuous components, the mixture-MEM model previously shown to fit well to comparable exchange rate data by Lanne (2006) is employed. The jump components are modeled by means of standard Markov-switching models.

The potential improvement in forecast accuracy due to decomposition can be seen as resulting from two factors. First, once the variation due to jumps has been eliminated from the realized volatility series, the process of the remaining continuous sample path component may be more easily captured, i.e., its process may be more easily estimable. Second, the jump component itself may contain predictable variation that contributes toward the forecast of the realized volatility. We show that at least with these data, statistically significant gains in out-of-sample forecast accuracy can be

made by decomposing, and this finding is fairly robust with respect to the details of the decomposition. However, if the jump component is very tightly defined, i.e., it takes nonzero values on only the days with the very greatest jumps, it has very little predictable variation so that virtually all the gains in forecast accuracy come from the improvements in estimating the process of the continuous component. Although the results are clear in showing the benefits of the decomposition, the diagnostic tests suggest that as far as the jump component is concerned, even further improvements might be attainable by more sophisticated models. While the results are in a sense specific to the chosen econometric models, they should be rather general in that the mixture-MEM model has previously been shown to fit comparable exchange rate data at least as well as relevant alternatives in the previous literature, and also here diagnostic checks indicate its adequacy.

The plan of the paper is as follows. In Section 2 the decomposition methods put forth by Andersen et al. (2005) are reviewed. Section 3 introduces the mixture-MEM and Markov-switching models and reports the estimation results, while Section 4 presents the forecast comparisons. Finally, Section 5 concludes.

2 Decomposition of Realized Volatility

In this section we discuss different decompositions of the daily return variance and introduce the data set. As a starting point for the analysis we have the realized variance of discretely sampled Δ -period returns $r_{t,\Delta} \equiv p(t) - p(t - \Delta)$,

$$RV_{t+1}(\Delta) \equiv \sum_{j=1}^{1/\Delta} r_{t+j\Delta,\Delta}^2, \quad (1)$$

where $p(t)$ is the price of the asset at time point t . As Barndorff-Nielsen and Shephard (2004) have shown, the difference between this measure and the stardized bi-power variation,

$$BV_{t+1}(\Delta) \equiv \mu_1^{-2} \sum_{j=2}^{1/\Delta} |r_{t+j\Delta, \Delta}| |r_{t+(j-1)\Delta, \Delta}|$$

where $\mu_1 \equiv \sqrt{2/\pi}$, consistently estimates the component of total return variation due to discrete jumps. Hence, it is natural to base the decomposition of $RV_{t+1}(\Delta)$ on $RV_{t+1}(\Delta) - BV_{t+1}(\Delta)$. As this difference can also take negative values, the measure

$$J_{t+1}(\Delta) \equiv \max[RV_{t+1}(\Delta) - BV_{t+1}(\Delta), 0] \quad (2)$$

suggested by Bandorff-Nielsen and Shephard (2004) can be used instead to ensure nonnegativity of the jump component. The continuous sample path component $C_{t+1}(\Delta)$ simply equals $RV_{t+1}(\Delta) - J_{t+1}(\Delta)$.

One problem with $J_{t+1}(\Delta)$ is that it typically takes positive values too frequently to be characterized as a component due to jumps. Instead, it would be desirable to identify only the significant jumps. To this end, Andersen et al. (2005) suggest employing the following test statistic

$$Z_{t+1}(\Delta) \equiv \Delta^{-1/2} \frac{[RV_{t+1}(\Delta) - BV_{t+1}(\Delta)] RV_{t+1}(\Delta)^{-1}}{\{(\mu_1^{-4} + 2\mu_1^{-2} - 5) \max[1, TQ_{t+1}(\Delta) BV_{t+1}(\Delta)^{-2}]\}^{1/2}} \quad (3)$$

whose distribution Huang and Tauchen (2005) find well approximated by the standard normal distribution. Here $TQ_{t+1}(\Delta)$ is the standized realized tri-power quarticity,

$$TQ_{t+1}(\Delta) \equiv \Delta^{-1} \mu_{4/3}^{-3} \sum_{j=3}^{1/\Delta} |r_{t+j\Delta, \Delta}|^{4/3} |r_{t+(j-1)\Delta, \Delta}|^{4/3} |r_{t+(j-2)\Delta, \Delta}|^{4/3},$$

$\mu_{4/3} \equiv 2^{2/3} \Gamma(7/6) \Gamma(1/2)^{-1}$ and $\Gamma(\cdot)$ is the gamma function. The idea is that the jump component defined by significance level α , $J_{t+1, \alpha}(\Delta)$, takes

positive values only on days on which the above test statistic is significant and equals zero otherwise, i.e.,

$$J_{t+1,\alpha}(\Delta) \equiv I[Z_{t+1}(\Delta) > \Phi_\alpha] [RV_{t+1}(\Delta) - BV_{t+1}(\Delta)], \quad (4)$$

where $I(\cdot)$ denotes the indicator function and Φ_α is the critical value at the significance level α . This means that the definition of $J_{t+1,\alpha}(\Delta)$ depends on the chosen significance level, or in other words, on how big jumps are considered significant. In order to make sure that the components sum to $RV_{t+1}(\Delta)$, the continuous sample path component has to be redefined accordingly,

$$C_{t+1,\alpha}(\Delta) \equiv I[Z_{t+1}(\Delta) \leq \Phi_\alpha] RV_{t+1}(\Delta) + I[Z_{t+1}(\Delta) > \Phi_\alpha] BV_{t+1}(\Delta). \quad (5)$$

Note that $J_{t+1,0.5}(\Delta)$ and $C_{t+1,0.5}(\Delta)$ equal $J_{t+1}(\Delta)$ and $C_{t+1}(\Delta)$, respectively. Following Andersen et al. (2005), instead of $BV_{t+1}(\Delta)$ and $TQ_{t+1}(\Delta)$ defined above we use the corresponding measures based on staggered returns,

$$BV_{1,t+1}(\Delta) \equiv \mu_1^{-2} (1 - 2\Delta)^{-1} \sum_{j=3}^{1/\Delta} |r_{t+j\Delta,\Delta}| |r_{t+(j-2)\Delta,\Delta}|$$

and

$$TQ_{1,t+1}(\Delta) \equiv \Delta^{-1} \mu_{4/3}^{-3} (1 - 4\Delta)^{-1} \sum_{j=5}^{1/\Delta} |r_{t+j\Delta,\Delta}|^{4/3} |r_{t+(j-2)\Delta,\Delta}|^{4/3} |r_{t+(j-4)\Delta,\Delta}|^{4/3}$$

in the empirical analysis to mitigate the effects of microstructure noise. Otherwise the statistic (3) tends to find too few jumps, as pointed out by Huang and Tauchen (2005).

The data set consists of thirty-minute returns ($\Delta = 1/48$ in the above formulas) of the Euro against the U.S. Dollar covering the period October 1, 1994 to September 30, 2004.¹ Thirty-minute returns are used, following

¹For the period until the end of 1998, the returns are computed from the Deutschmark/Dollar rate.

Andersen et al. (2003) and Lanne (2006), as a compromise between the theoretical considerations recommending sampling at very high frequencies and the desire to avoid contamination by microstructure effects. The returns are computed from a five-minute return data set compiled by Olsen and Associates. These returns are based on interbank bid and ask quotes displayed on Reuters FAFX screen. The quotes are thus only indicative rather than firm in that they are not binding commitments to trade. Hence, as recently pointed out by Danielsson and Payne (2002), at very high frequencies they may not accurately measure tradeable exchange rates. Danielsson 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. Following the common practice in the literature, certain inactive periods have been discarded. First, all the returns between Friday 21:00 GMT and Sunday 21:00 GMT are excluded. Second, we eliminated the following slow trading days associated with holidays: Christmas (December 24–26), New Year (December 31 and January 1–2), Good Friday, Easter Monday, Memorial Day, July Fourth, Labor Day, and Thanksgiving and the following day. This leaves us 2,496 observations in total, of which 1,998 (from October 1, 1994 through September 30, 2002) form the estimation period, while the remaining 498 observations (from October 1, 2002 through September 30, 2004) are left for forecast evaluation.

The realized variance, bi-power variation and the jump component $J_{t+1}(\Delta)$ are depicted in Figure 1. The maximum of all the series occurred on September 22, 2000. On that day, the European Central Bank, the Federal Reserve, the Bank of Japan, the Bank of England and the Bank of Canada bought euros in a coordinated intervention, presumably causing an abrupt

increase in volatility. As the bottom panel shows, the jump component defined in (2) takes a positive value on almost every day, which is very different from the conventional idea of an infrequently occurring jump. Therefore, in the empirical analysis we concentrate on the continuous sample path and jump components defined in (5) and (4), respectively. Following Andersen et al. (2005), three significance levels, α , are considered, 0.95, 0.99 and 0.999. Moreover, because of better fit we will consider the realized volatility, $RV_{t+1}^{1/2}(\Delta)$ decomposed into the sum of (with slight abuse of notation)

$$C_{t+1,\alpha}^{1/2}(\Delta) \equiv I[Z_{t+1}(\Delta) \leq \Phi_\alpha] RV_{t+1}^{1/2}(\Delta) + I[Z_{t+1}(\Delta) > \Phi_\alpha] BV_{t+1}^{1/2}(\Delta) \quad (6)$$

and

$$J_{t+1,\alpha}^{1/2}(\Delta) \equiv I[Z_{t+1}(\Delta) > \Phi_\alpha] \left[RV_{t+1}^{1/2}(\Delta) - BV_{t+1}^{1/2}(\Delta) \right] \quad (7)$$

instead of modeling the realized variance and $J_{t+1,\alpha}(\Delta)$ and $C_{t+1,\alpha}(\Delta)$ directly. These components for the different significance levels are plotted in Figure 2. The continuous components, in general, resemble the bi-power variation series, while the appearance of the jump components greatly depends on the significance level α . As the top panel shows, quite a few significant jumps (586) are still found when $\alpha = 0.95$, whereas the number declines to 328 when α equals 0.99, and at the 99.9% level only 80 significant jumps are detected. Still, visual inspection of the series suggests that all the jump component series exhibit some clustering and are thus potentially predictable.

3 Modeling Realized Volatility

In this section we estimate models for the realized volatility, $RV_{t+1}^{1/2}(\Delta)$, and its continuous sample path and jump components $C_{t+1,\alpha}^{1/2}(\Delta)$ and $J_{t+1,\alpha}^{1/2}(\Delta)$, defined as the square root of (1), (6) and (7), respectively. For the realized

volatility and continuous components, we estimate mixture multiplicative error models that were shown to fit realized exchange rate volatility series quite well by Lanne (2006). Compared to the HAR-RV model employed by Andersen et al. (2005), this model has the additional advantage that the positivity of the volatility forecasts can easily be guaranteed. Standard Markov-switching models (e.g. Hamilton, 1989), on the other hand, are shown to give a reasonable fit to the jump component.

3.1 Mixture Multiplicative Error Model

Denoting by v_t the variable to be modeled (realized volatility or the continuous component), the mixture multiplicative error model can be written as

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

where the conditional mean is parametrized as

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

and the stochastic error term ε_t is a mixture of ε_{1t} and ε_{2t} such that $\varepsilon_{1t} \sim \text{Gamma}(\gamma_1, \delta_1)$ with probability π and $\varepsilon_{2t} \sim \text{Gamma}(\gamma_2, \delta_2)$ with probability $1 - \pi$ ($0 < \pi < 1$). In order for ε_t 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. Furthermore, we allow the conditional mean to switch accordingly, i.e., the conditional mean equals μ_{1t} with probability π and μ_{2t} with probability $(1 - \pi)$ where

$$\mu_{1t} = \omega_1 + \sum_{i=1}^{q_1} \alpha_{1i} v_{t-i} + \sum_{j=1}^{p_1} \beta_{1j} \mu_{1,t-j}$$

and

$$\mu_{2t} = \omega_2 + \sum_{i=1}^{q_2} \alpha_{2i} v_{t-i} + \sum_{j=1}^{p_2} \beta_{2j} \mu_{2,t-j}.$$

This specification will be called the mixture-MEM($p_1, q_1; p_2, q_2$) model. It can be estimated in a straightforward manner by the method of maximum likelihood (see Lanne, 2006). In estimation, the parameters of both mixture components must be restricted such that they satisfy the conditions of Nelson and Cao (1992) to ensure positivity of v_t .

The estimation results are presented in Table 1. Based on diagnostic checking, the mixture-MEM(1, 6; 1, 2) models were selected for all the series. Presumably the sixth lag is required for modeling some kind of seasonality in the series. The coefficients of the lags between 2 and 6 of the first mixture component turned out to be insignificant, so they are restricted to zero. The error distributions of the mixture components are distinctly different, but the differences are similar across the series. Plots of the error distributions (not shown) indicate that most of the time the errors come from a distribution relatively tightly concentrated around unity, whereas somewhat less than 20% of the time the errors are generated from a right-skewed distribution with clearly fatter tails. Moreover, as could be expected, the latter mixture component of each continuous sample path component has less probability mass on the tails than that of the realized volatility.

As pointed out by Lanne (2006), the persistence of the series is measured by the largest eigenvalue of the first-order vector representation of the model, and these values equal 0.987, 0.982, 0.983 and 0.981 for the realized volatility and the continuous components with significance levels 0.95, 0.99 and 0.999, respectively. This confirms the expectation that all the series are highly persistent and suggests that the persistence of the realized volatility is partly brought about by jumps (cf. Vlaar and Palm (1993) who attribute a part of the high persistence implied by GARCH models for exchange rate returns to ignoring jumps). There are also differences in persistence between the

mixture components for all the series. While the first component is very persistent with estimates of β_{11} between 0.922 and 0.935, the corresponding figures for β_{21} range from 0.526 to 0.874, and the differences are considerably greater for the continuous components than the realized volatility.

According to the diagnostic checks depicted in Figure 3 all the models can, in general, be deemed adequate. As the mixture model does not yield conventional standardized residuals, diagnostics are based on the so-called probability integral transform (for details, see e.g. Lanne, 2006). For the adequacy of the specification, the transformed data should be independently uniformly distributed. Diebold et al. (1998) recommend checking this by plotting a histogram of the transformation and computing the autocorrelation function of the demeaned residuals and their squares. Virtually all the bins of the histograms lie within in the 95% confidence intervals, indicating no violation of the uniformity requirement. This can also be tested using Pearson's goodness-of-fit test. The p-values for the models for the realized volatility and the continuous components with significance levels 0.95, 0.99 and 0.999 equal 0.68, 0.90, 0.28 and 0.52, respectively, reinforcing the impression given by visual inspection. Likewise, the autocorrelation in the transform series is minor, attesting to the adequacy of the mixture-MEM specifications. There is, however, some evidence of autocorrelation in the squared series which was also detected by Lanne (2006) in realized volatility of other exchange rate series.

3.2 Modeling the Jump Component

As pointed out above, the jump component series in Figure 2 seem to exhibit some serial dependence that could be exploited in forecasting. To this end, the standard two-regime Markov-switching model (e.g. Hamilton,

1989) is employed. In other words, the process is assumed to switch between two regimes characterized by $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ distributions, respectively.² The switching is assumed to be governed by a Markov chain such that the probability of staying in regime 1 in the next period if regime 1 prevails in this period equals p_{11} and the corresponding probability of staying in regime 2 equals p_{22} .

The estimation results³ are presented in the upper panel of Table 2. For all values of α , the process switches between regime 1 with mean (very close to) zero and small variance and regime 2 with greater mean and higher variance. The mean and variance parameters in regime 2 increase with α as the average size and variability of the jump components increase. This regime is also not persistent, especially with $\alpha = 0.999$, whereas regime 1 is highly persistent with the estimates of p_{11} exceeding 0.8, reinforcing the interpretation of the two regimes as the “normal” and “jump” regimes, respectively.

The diagnostic tests proposed by Hamilton (1996) are employed to check the adequacy of the Markov-switching models, and the results are reported in the lower panel of Table 2. When α equals 95% or 99%, the simple model seems, in general, to sufficiently capture the dynamics in the jump component. There is only some evidence of unmodeled conditional heteroskedasticity, indicating that further refinements might be possible. This feature may, however, be difficult model adequately, and because even with this deficiency, the forecasting performance is good, we proceed with the simple model spec-

²Because the normality assumption is not realistic in that it allows the process to take negative values, the maximum likelihood (ML) estimation of the model can only be given a quasi ML interpretation. The distributional assumption is not likely to have a big effect on the results in practice.

³I am grateful to James Hamilton for making available on his homepage the GAUSS software for estimating the Markov-switching models and computing the diagnostics.

ification. When $\alpha = 0.999$, on the other hand, also the hypotheses of no autocorrelation in regime 2 and across the regimes are clearly rejected. It is likely that these findings are indications of the general lack of predictable variation in the jump component process in this case. As a matter of fact, the forecasts of the jump component generated by this model do not contribute at all toward the accuracy of the realized volatility forecasts, as will be seen in Section 4. Therefore, also in this case we settle for this specification.

4 Forecasts

In order to answer the question posed in the Introduction of whether more accurate forecasts can be obtained by modeling the components of the realized volatility separately, we conduct some forecast experiments. In evaluating the forecasting performance we concentrate on the mean square error,

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

where T^* is the length of the forecast period, v_t the realized volatility and \hat{v}_t is the volatility forecast either implied by the model for the realized volatility or computed as the sum of the forecasts of the models for the continuous and jump components. It is easy to show that this loss function satisfies Hansen and Lunde's (2006) sufficient conditions for correct ranking of volatility forecasts when they are measured against an imperfect proxy such as the realized volatility. 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 forecast we take the combination forecast so that negative values of the Diebold-Mariano test statistic indicate that by summing the component forecasts, a smaller MSE is obtained than by forecasting the realized volatility directly. The results in Table 3 show that out of sample, considerable gains in predictability can be made by separately modeling the continuous and jump components, whereas the differences in the in-sample figures are minor and not statistically significant. This is consistent with the finding in Andersen et al. (2005) that the lags of the jump component are not significant in their regression model for the realized volatility with $\alpha = 0.999$. Out of sample at the one-day horizon, the MSE is significantly smaller at any sensible significance level irrespective of the exact definition of the jump component. For $\alpha = 0.999$, the reduction in the MSE compared to directly modeling the realized volatility is approximately 5.5%, whereas in the other two cases the corresponding figure is around 8%. As far as the 10-day volatility is concerned, the reductions in out-of-sample forecast accuracy are even greater, ranging from 9.8% ($\alpha = 0.99$) to 10.6% ($\alpha = 0.999$), but according to the Diebold-Mariano test the differences are not statistically significant.

While we are mainly interested in finding out, whether significant gains in forecast accuracy can be obtained by decomposing the conditional volatility, the Diebold-Mariano tests only provide pairwise comparisons between the mixture-MEM model for the realized volatility directly and forecasts obtained through the different decompositions. Hence, that test may not be optimal, and to answer the question of interest more directly and to avoid potential data snooping biases, we also computed Hansen's (2005) test statistics 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. In our case, the natural benchmark is the mixture-MEM model for the realized volatility. This test rejects for both the one and ten-day out-of-sample volatilities with p-values 0.003 and 0.021, respectively, indicating that significant gains can be made by decomposing, also for the ten-day volatility. The corresponding in-sample p-values equal 0.411 and 0.303, respectively, confirming that the benchmark model is not surpassed by the decomposition forecasts.

Although there are small differences between the MSE's produced by different decompositions, these seem to be minor, and as a general conclusion it could be said that at least when the combination of MEM and Markov-switching models is used, it is of lesser importance how the decomposition is done. This is probably due to the flexibility of the models to fit series with somewhat different properties which was also indicated by the favorable diagnostic test results in all cases considered. The forecasting benefits of decomposing can be seen as coming from two sources, the better fit of the MEM model due to the purging of the series of extreme observations and the predictability of the jump component. However, as mentioned in Section 3, there does not seem to be much predictable variation in the jump component when only the greatest jumps are included ($\alpha = 0.999$), so that in this case the benefits almost exclusively come from the first factor. This was reconfirmed by computing the MSE's of forecasts of the realized volatility based on the continuous component only. In the $\alpha = 0.999$ case the figures were virtually unchanged, while in the other cases dismissing the jump component led to considerable loss in forecast accuracy.

5 Conclusion

With EUR/USD exchange rate data it has been shown that by decomposing the realized volatility into its jump and continuous sample path components, considerable gains in out-of-sample forecast accuracy can be reached. Moreover, this finding seems to be relatively independent of the details of the decomposition in the range typically considered. Hence, we have been able to answer in the affirmative the question posed by Andersen et al. (2005) of whether separately modeling and forecasting the two components is beneficial. However, further work along these lines is called for as our results may be specific to the particular data set and models. Although diagnostic tests suggest the adequacy of the chosen specification, comparable gains might not be possible when other commonly used models are employed.

References

- [1] Andersen, T.G., T. Bollerslev, and F.X. Diebold (2005), Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility. NBER Working Paper 11775.
- [2] Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys (2003), Modeling and forecasting realized volatility. *Econometrica* 71, 579–625.
- [3] Barndorff-Nielsen, O.E., and N. Shephard (2004), Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 2, 1–37.
- [4] Corsi, F. (2003), A simple long memory model of realized volatility. Manuscript, University of Southern Switzerland.

- [5] 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.
- [6] Diebold, F.X., T. Gunther, and A.S. Tay, Evaluating density forecasts, with applications to financial risk management. *International Economic Review* 39, 863–883.
- [7] Diebold, F.X., and R.S. Mariano (1995), Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13, 253–263.
- [8] Hamilton, J.D. (1989), A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357–384.
- [9] Hamilton, J.D. (1996), Specification testing in Markov-switching time-series models. *Journal of Econometrics* 70, 127–157.
- [10] Hansen, P.R. (2005), A test for superior predictive ability. *Journal of Business and Economic Statistics* 23, 365–380.
- [11] Hansen, P.R., and A. Lunde (2006), Consistent ranking of volatility models. *Journal of Econometrics* 131, 97–121.
- [12] Huang, X., and G. Tauchen (2005), The relative contribution of jumps to total price variance. *Journal of Financial Econometrics* 3, 456–499.
- [13] Lanne, M. (2006), A Mixture multiplicative error model for realized volatility. *European University Institute Working Paper ECO 2006/3*.
- [14] Müller, U.A., M.M. Dacorogna, R.D. Davé, R.B. Olsen, O.V. Pictet, and J. von Weizsäcker (1997), Volatilities of different time resolutions - Analyzing the dynamics of market components. *Journal of Empirical Finance* 4, 213–239.

- [15] Nelson, D.B., and C.Q. Cao (1992), Inequality constraints in the univariate GARCH model. *Journal of Business and Economic Statistics* 10, 229–235.
- [16] Vlaar, P.J.G, and F.C. Palm, The message in weekly exchange rates in the European Monetary System: Mean reversion, conditional heteroskedasticity and jumps. *Journal of Business and Economic Statistics* 11, 351–360.

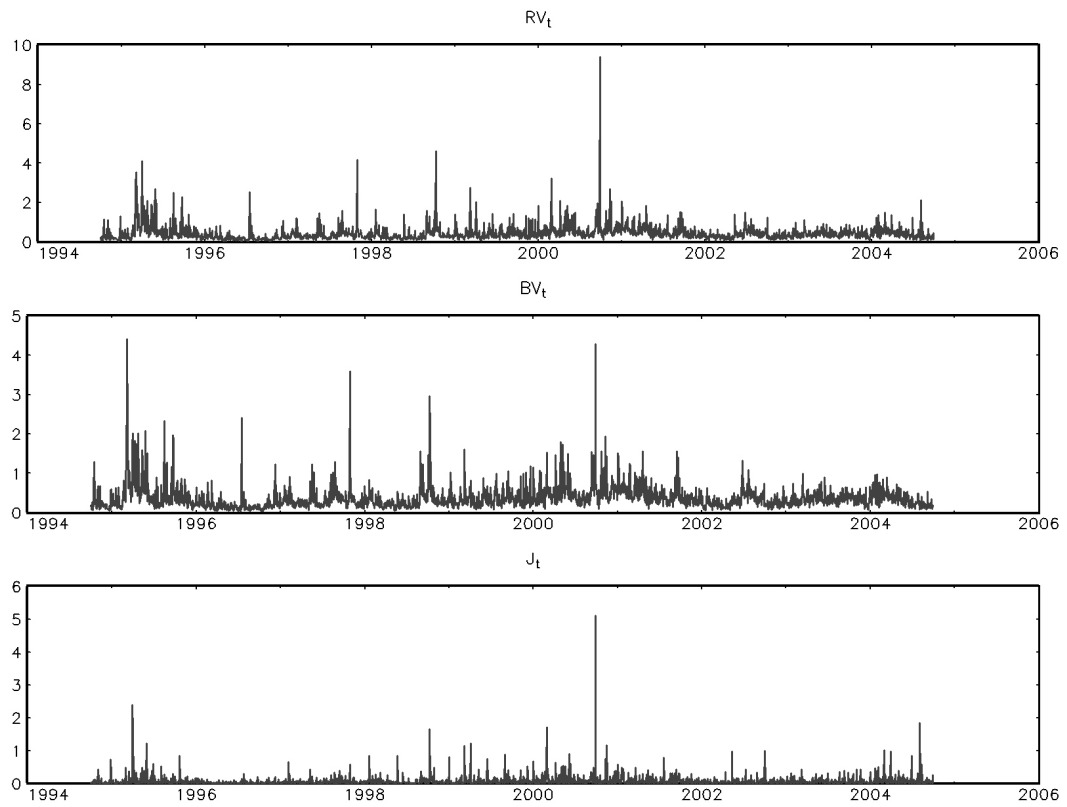


Figure 1: The realized variance, bi-power variation and jump component of EUR/USD returns.

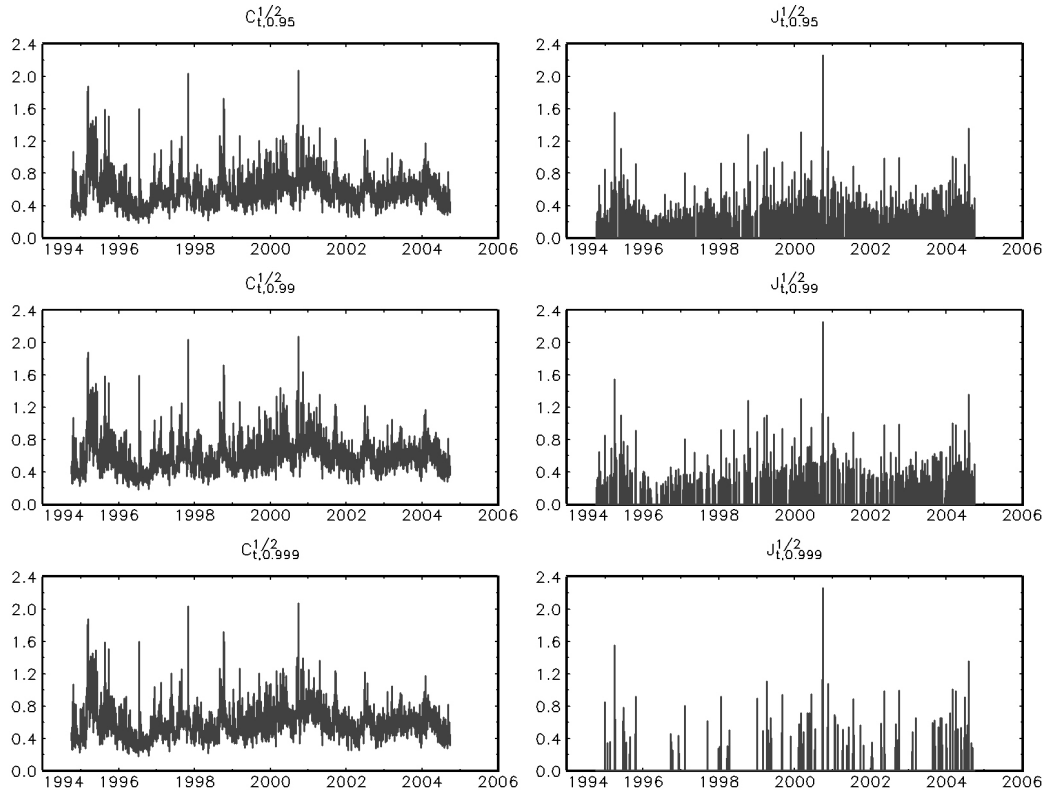


Figure 2: The decomposition of the EUR/USD realized volatility into continuous and jump components based on the Z_t statistic at the 95, 99 and 99.9 % significance levels.

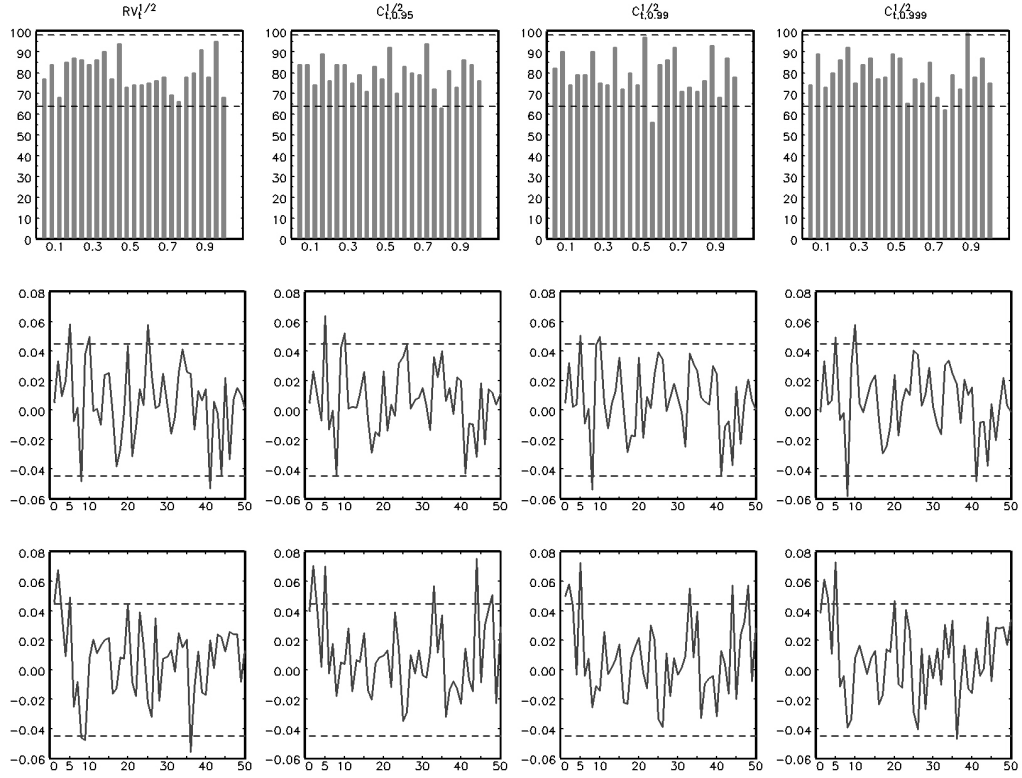


Figure 3: Diagnostics for the probability integral transforms of mixture-MEM models for the realized volatility and continuous components. The upper panel depicts their frequency distributions, and the middle and lower panels the autocorrelation functions of the demeaned transforms and their squares, respectively. The dashed lines are the 95% confidence intervals.

Table 1: Estimation results of the mixture-MEM models for the realized volatility and continuous sample path components.

	$RV_t^{1/2}$	$C_{t,0.95}^{1/2}$	$C_{t,0.99}^{1/2}$	$C_{t,0.999}^{1/2}$
π	0.843 (0.036)	0.844 (0.043)	0.807 (0.052)	0.838 (0.041)
γ_1	22.390 (1.475)	22.777 (1.510)	22.903 (1.711)	22.035 (1.450)
ω_1	0.002 (0.001)	0.003 (0.001)	0.003 (0.001)	0.003 (0.001)
α_{11}	0.272 (0.023)	0.294 (0.024)	0.282 (0.025)	0.282 (0.023)
α_{12}	-0.170 (0.026)	-0.190 (0.027)	-0.177 (0.028)	-0.173 (0.026)
α_{16}	-0.045 (0.010)	-0.043 (0.010)	-0.043 (0.010)	-0.042 (0.010)
β_{11}	0.935 (0.011)	0.929 (0.013)	0.928 (0.014)	0.922 (0.015)
γ_2	8.883 (1.163)	10.212 (1.442)	10.753 (1.418)	9.828 (1.365)
ω_2	0.031 (0.037)	0.106 (0.084)	0.076 (0.067)	0.089 (0.121)
α_{21}	0.586 (0.114)	0.623 (0.113)	0.596 (0.097)	0.593 (0.117)
α_{22}	-0.465 (0.185)	-0.185 (0.309)	-0.232 (0.270)	-0.273 (0.411)
β_{21}	0.874 (0.135)	0.526 (0.319)	0.614 (0.279)	0.650 (0.430)

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

Table 2: Estimation and diagnostic test results of the Markov-switching models for the jump components of the realized volatility.

	$J_{t,0.95}^{1/2}$	$J_{t,0.99}^{1/2}$	$J_{t,0.999}^{1/2}$
Estimation Results ^a			
μ_1	0.006 (0.001)	0.001 (0.001)	6.52e-7 (0.001)
μ_2	0.145 (0.008)	0.164 (0.010)	0.269 (0.028)
σ_1^2	0.001 (6.07e-5)	0.001 (2.36e-5)	0.001 (1.66e-5)
σ_2^2	0.016 (0.001)	0.020 (0.002)	0.046 (0.009)
p_{11}	0.826 (0.014)	0.877 (0.009)	0.968 (0.004)
p_{22}	0.214 (0.030)	0.165 (0.028)	0.018 (0.012)
Diagnostic Tests ^b			
Autocorrelation in regime 1	0.477	0.979	0.055
Autocorrelation in regime 2	0.242	0.236	5.63e-15
Autocorrelation across regimes	0.525	0.221	5.68e-15
ARCH	0.006	0.008	0.004

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

^bThe figures are marginal significance levels.

Table 3: Out-of-sample forecast evaluation: the Diebold-Mariano test for the direct and combined forecasts.

	One-Day-Ahead Forecast			Ten-Day-Ahead Forecast		
Model	MSE	D-M	p-value	MSE	D-M	p-value
	Statistic			Statistic		
	In-Sample					
RV_t	0.0380			1.250		
$J_{t,0.95} + C_{t,0.95}$	0.0378	-0.508	0.612	1.261	0.330	0.741
$J_{t,0.99} + C_{t,0.99}$	0.0379	-0.301	0.763	1.242	-0.284	0.777
$J_{t,0.999} + C_{t,0.999}$	0.0381	0.230	0.818	1.239	-0.476	0.634
	Out-of-Sample					
RV_t	0.0242			0.4413		
$J_{t,0.95} + C_{t,0.95}$	0.0224	-3.120	0.002	0.3986	-1.085	0.278
$J_{t,0.99} + C_{t,0.99}$	0.0225	-3.029	0.002	0.3999	-1.210	0.226
$J_{t,0.999} + C_{t,0.999}$	0.0229	-2.291	0.002	0.3968	-1.406	0.160

The null forecast in the Diebold-Mariano test is the combination forecast.