MEM or/and LogARMA: Which Model for Realized Volatility?

Stanislav Anatolyev

Abstract There coexist two popular autoregressive conditional density model classes for series of positive financial variables such as realized volatility. One is a class of multiplicative error models (MEM), where the conditional mean is modelled autoregressively, while the specified shape of conditional distribution imposes evolution on higher order moments. The other class contains LogARMA models—ARMA models for logarithms of the original series, with a possibly time varying conditional distribution imposed on top of it. For MEM models, generating forecasts is straightforward, while for LogARMA models, additional numerical integration may be required. We compare small and big models from the two classes, along with their combinations, in terms of in-sample fit and out-of-sample predictability, using real data on realized volatility. The forecast combination weights show that both model classes are able to generate competitive forecasts, but the class of LogARMA models appears more reliable in forecasting than the class of MEM models.

1 Introduction

When the notion and analysis of realized volatility came to play in Andersen et al. (2003), the leading idea of how to model it was the use of *LogARMA models*, i.e., the class of ARMA models applied to the logarithm of the original series. Andersen and coauthors (Andersen et al. 2003) employed a LogARMA-spirited VAR-RV model for logarithmic volatility, somewhat motivated by the fact that the realized volatility is approximately log-normally distributed (see also Andersen et al. (2001)). Later, the LogARMA model took various fancy forms, such as the heterogeneous autoregressive (HAR) model (Corsi 2009) and its extensions.

However, realized volatility is a positive series. There is another option for modeling the dynamics of positive series. The story started developing from the work of Engle (Engle and Russell 1998) who proposed a class of autoregressive con-

S. Anatolyev (✉)
CERGE-EI, Politickych vězňů 7, 11121 Prague 1, Czech Republic
e-mail: stanislav.anatolyev@cerge-ei.cz

N. Ngoc Thach et al. (eds.), *Financial Econometrics: Bayesian Analysis, Quantum Uncertainty, and Related Topics*, Studies in Systems, Decision and Control 427,
https://doi.org/10.1007/978-3-030-98689-6_16

ditional duration (ACD) models for intradaily trade durations, a positive variable. Later, Engle (2002) renamed the class as *multiplicative error models (MEM)* noting that these models are convenient to use for any serially correlated stationary positive series. Various more generalized MEM-based models for realized volatility in levels were introduced into empirical work (e.g., Engle and Gallo (2006), Hautsch (2011)).

Nowadays, both approaches to modeling realized volatility coexist. Engle (Engle 2002) made a quick comparison of attractive and unpleasant features of MEM with LogARMA models, while Allen and coauthors (Allen et al. 2008) showed how the two model classes overlap (see also Sect. 2.3). Below we list in comparison four separate aspects related to both classes, the first two important for modeling decisions, and the other two less critical.

1. The MEM class is targeted primarily to model the conditional mean of the series, and hence is natural and sufficient for modeling the dynamics for the purposes of forecasting. At the same time, a model for the conditional mean of a log-transformed variable such as LogARMA needs to model the whole conditional distribution if one wants to eventually forecast the volatility in levels. This may be cumbersome to do, except under conditional homoskedasticity and/or conditional normality—the situations, when the conditional mean can be easily translated to the conditional mean of an exponent.
2. The simplest MEM model (so called exponential MEM) describing the dynamics of the conditional mean of the series automatically describes its higher order conditional moments, at least when the multiplicative innovation's conditional distribution is time-independent. The class of LogARMA models, in contrast, allows independent modeling higher order moments, at least its conditional variance by, for instance, an ARCH-type evolution, even when the standardized innovation's conditional distribution is time-independent. Thus, all else equal, the dynamics of conditional distribution is more flexible within the LogARMA framework.
3. If the support of the variable modelled contains values close to zero, construction of an LogARMA model may be problematic because of taking a logarithm of very small values, and zero values are totally prohibitive. At the same time, the MEM model successfully adapts for zeros in the support of the conditional distribution even in a logarithmic version that we use here. Moreover, Hautsch and coauthors (Hautsch et al. 2014) show how to take care of a probability mass at zero if zero values have a non-zero probability of occurrence.
4. Extensions of a scalar heteroskedastic LogARMA to a multivariate framework are familiar: the mean equation extends from ARMA to VARMA, and the variance equation from GARCH to multivariate GARCH. Extensions of a scalar MEM to a multivariate MEM are trickier and require the copula machinery (see, for example, Engle and Gallo (2006)).

The first two critical aspects suggest a trade-off between flexibility of modeling and complexity of two types: the usual one—degree of parameterization, and another one—producing forecasts. In this paper, we play one model class off against the other, using real data on realized volatility, to try to determine the 'optimal' modeling strategy. We specify two members of each class of models—one is a ‘small’ model and the other a ‘big’ model. Both use plausible specifications of volatility dynamics, but differ in specifications of the conditional distribution. The ‘small model’ uses a baseline conditional distribution that, in particular, allows quasi-maximum likelihood estimation, while the ‘big model’ uses a sophisticated conditional distribution with additional shape parameters—not the fanciest that one can find in the literature but one that is likely to be exploited by a practitioner. For the MEM class, these are the exponential and Burr distributions, respectively, while for the LogARMA class, these are the normal and skewed Student distributions. Both sophisticated distributions—Burr and skewed Student—possess two additional shape parameters.

We compare the in-sample and out-of-sample performance of the resulting four models, paying special attention to forecasting quality. Towards this end, we compare forecasts produced by the four models together with two types of model combinations, and construct the model confidence sets (MCS, Hansen et al. (2011)) of the best performing models. To do the numerical evaluation, we use a popular data-set of realized stock market volatility on ten stocks. We find that in terms of in-sample fit, the ‘small’ MEM model does not fit well compared to the other three, while among these three, usually two of three models tend to stand out, depending on the stock. In terms of forecasting quality, the four models perform quite similarly, and usually multiple models can be deemed the best in terms of MCS. Overall, the class of LogARMA models seems to be more reliable in forecasting than the class of MEM models, and small models tend to dominate big models from the same class. While model averaging using in-sample quality of fit does not tend to improve forecasting performance above the performance of best individual best models, model averaging using out-of-sample quality of fit is able, sometimes, to slightly improve forecasts.

The article is organized as follows. Section 2 describes the models, together with estimation and forecasting methods. Section 3 contains empirical results. Section 4 concludes.

2 Models

Denote the realized volatility by rv_t. We will compare four individual (‘pure’) models and two combinations of those. The individual models are: two MEM models based on the conditional exponential (‘small’) and Burr (‘big’) distributions and a linear dynamics in logs for the conditional mean of rv_t, and two LogARMA models based on the conditional normal (‘small’) and skewed Student (‘big’) distributions and an ARMA-EGARCH dynamics for logs of rv_t. The two model combinations are based on individual model performance: one on the in-sample performance as judged by the smoothed Takeuchi information criterion, and the other on the out-of-sample performance as judged by the forecasting quality in a validation subsample. For simplicity, all dynamic models have orders (1,1).

2.1 MEM

In both MEM models, $rv_t = \mu_t \varepsilon_t$, where μ_t is the conditional mean of rv_t, and ε_t has a positive distribution with conditional mean unity. The dynamics of μ_t is logarithmic:

$$\log \mu_t = \omega + \alpha \log rv_{t-1} + \beta \log \mu_{t-1}.$$

A small MEM is represented by the *ExpMEM* model

$$\varepsilon_t | I_{t-1} \sim \mathscr{E},$$

where $\mathscr{E}$ denotes standard exponential distribution having the density

$$f_{\mathscr{E}}(\varepsilon) = \exp(-\varepsilon).$$

A big MEM is represented by the *BurrMEM* model[1]

$$\varepsilon_t | I_{t-1} \sim \mathscr{B}(\zeta, \varrho),$$

where $\mathscr{B}(\zeta, \varrho)$ denotes Burr distribution with mean unity and shape parameters ζ and ϱ, thus having the density

$$f_{\mathscr{B}}(\varepsilon) = \frac{\zeta}{\chi^{\zeta}} \varepsilon^{\zeta-1} \left(1 + \varrho \left(\frac{\varepsilon}{\chi}\right)^{\zeta}\right)^{-1-\varrho^{-1}}, \tag{1}$$

where

$$\chi = \frac{\Gamma(1+\varrho^{-1})\varrho^{1+\zeta^{-1}}}{\Gamma(1+\zeta^{-1})\Gamma(\varrho^{-1}-\zeta^{-1})}, \quad \zeta, \varrho > 0.$$

For both ExpMEM and BurrMEM, the one-step forecast of realized volatility then has the following form:

$$\widehat{rv}_{t+1} = \hat{\mu}_{t+1} = \exp\left(\hat{\omega} + \hat{\alpha} \log rv_t + \hat{\beta} \log \hat{\mu}_t\right).$$

2.2 LogARMA

In both LogARMA models, $\log rv_t$ follows an ARMA dynamics

[1] This distribution was proposed in Grammig and Maurer (2000) for ACD models for trade durations in order to account for non-monotonicity of the conditional hazard function. An alternative flexible distribution is normalized generalized gamma (see Hautsch (2011)). We utilize Burr because it exhibits much higher stability than the generalized gamma in experiments with real data.

$$\log rv_t = \mu + \phi \log rv_{t-1} + e_t + \theta e_{t-1},$$

with semi-strong white noise innovations $e_t = \sigma_t \eta_t$, whose conditional variance ('volatility of volatility') follows EGARCH dynamics:

$$\log \sigma_t^2 = \omega + \alpha|\eta_{t-1}| + \gamma\eta_{t-1} + \beta \log \sigma_{t-1}^2,$$

where η_t is standardized innovation with conditional mean zero and conditional variance unity.

A small LogARMA is represented by the *NormLogARMA* model

$$\eta_t | I_{t-1} \sim \mathscr{N},$$

where $\mathscr{N}$ denotes standard normal distribution having the density

$$f_{\mathscr{N}}(\eta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\eta^2}{2}\right).$$

A big LogARMA is represented by the *SkStLogARMA* model[2]

$$\eta_t | I_{t-1} \sim \mathscr{S}\mathscr{S}(\lambda, \nu).$$

where $\mathscr{S}\mathscr{S}(\lambda, \nu)$ denotes standardized (to have zero mean and unit variance) skewed Student distribution (Hansen 1994) with shape parameters λ (responsible for asymmetry) and ν (degrees of freedom, responsible for tail thickness), thus having the density

$$f_{\mathscr{S}\mathscr{S}}(\eta, \lambda, \nu) = c_0 c_1 \left(1 + \frac{\xi^2}{\nu - 2}\right)^{-(\nu+1)/2}, \tag{2}$$

where $\xi = (c_1\eta + c_2)/(1-\lambda)$ if $\eta < -c_2/c_1$ and $\xi = (c_1\eta + c_2)/(1+\lambda)$ otherwise, $c_0 = \Gamma((\nu+1)/2)/\Gamma(\nu/2)/\sqrt{\pi(\nu-2)}$, $c_2 = 4c_0\lambda(\nu-2)/(\nu-1)$, and $c_1 = \sqrt{1 + 3\lambda^2 - c_2^2}$.

For both LogARMA models, the one-step forecast of logarithmic realized volatility is given by

$$\widehat{\log rv}_{t+1} = \hat{\mu} + \hat{\phi} \log rv_t + \hat{\theta} e_t,$$

while the volatility prediction is

$$\hat{\sigma}_{t+1}^2 = \exp\left(\hat{\omega} + \hat{\alpha}|\hat{\eta}_t| + \hat{\gamma}\hat{\eta}_t + \hat{\beta} \log \hat{\sigma}_t^2\right).$$

[2] An alternative flexible distribution is normalized skewed generalized error distribution (SGED) (see, e.g., Anatolyev and Petukhov (2016)). We utilize skewed Student because it exhibits much higher stability than the SGED in experiments with real data.

For the NormLogARMA model, because of conditional normality, these forecasts are translated into the forecast for realized volatility via

$$\widehat{rv}_{t+1} = \hat{E}[rv_{t+1}|I_t] = \exp\left(\widehat{\log rv}_{t+1} + \frac{1}{2}\hat{\sigma}^2_{t+1}\right).$$

For the SkStLogARMA model, the conditional expectation $E[rv_{t+1}|I_t]$ does not have a closed form. Therefore, we form the forecasts using

$$\widehat{rv}_{t+1} = \exp\left(\widehat{\log rv}_{t+1}\right) E\left[\exp\widehat{(\sigma_{t+1}\eta_{t+1})}\,|I_t\right],$$

where the integral $E\left[\exp\widehat{(\sigma_{t+1}\eta_{t+1})}\,|I_t\right] = \int_{-\infty}^{+\infty} \exp\left(\hat{\sigma}_{t+1}\eta\right) f_{\mathscr{S}\mathscr{S}}(\eta, \hat{\lambda}, \hat{\nu})d\eta$ is computed using the Gauss-Chebychev quadrature (see, e.g., Judd (1998)):

$$\int_a^b g(x)dx \approx \frac{\pi(b-a)}{2n}\sum_{i=1}^{n}(1-x_i^2)^{\frac{1}{2}} g\left(a + \frac{(x_i+1)(b-a)}{2}\right),$$

where $x_i = \cos\left((2i-1)\pi/(2n)\right)$, $i = 1, \ldots, n$. We set $n = 100$, $a = -8$, and $b = 8$; these values deliver sufficient computational precision.

2.3 Reconciliation of MEM and LogARMA

Let us reconcile the dynamics of realized volatility in the two model classes. From the MEM multiplicative structure, it follows that

$$\log rv_t = \log\mu_t + \log\varepsilon_t = (1-\beta L)^{-1}(\omega + \alpha\log rv_{t-1}) + \log\varepsilon_t,$$

or

$$\log rv_t = \omega + (\alpha+\beta)\log rv_{t-1} + \log\varepsilon_t - \beta\log\varepsilon_{t-1},$$

which has a homoskedastic ARMA(1,1) for $\log rv_t$. On the other hand, in the LogARMA model class, $\log rv_t$ follows a heteroskedastic ARMA dynamics

$$\log rv_t = \mu + \phi\log rv_{t-1} + e_t + \theta e_{t-1},$$

with innovations e_t, whose conditional variance follows EGARCH dynamics. Hence, as viewed from the perspective of the mean logarithmic volatility dynamics, the MEM and LogARMA models are equally flexible, but are different in the flexibility of the volatility-of-volatility dynamics; LogARMA is more flexible in this respect.

However, even if the heteroskedasticity was shut down in LogARMA, the models would still not be equivalent in conditional distributional features, as the exponential/Burr distribution of ε_t does not correspond to normal/skewed Student distribution of e_t. Thus, neither model is a special case of the other.

2.4 Model Averaging

Along with the four pure models, we use their combinations and corresponding combined forecasts with weights based on in-sample and out-of-sample performance. In both cases, predictions are formed as a linear combination of predictions from individual models:

$$\widehat{rv}_{t+1} = \sum_{i=1}^{M} w_i \widehat{rv}_{t+1,i},$$

where M is number of pure models (4 in our case), and w_i and $\widehat{rv}_{t+1,i}$, $i = 1, \ldots, M$, are the model weights and individual forecasts, respectively.

The first model averaging combination based on an in-sample smoothed information criterion (Buckland et al. 1997), which is a convenient tool to track relative in-sample fit of several models. Denote by $f_{t-1}(rv_t|\theta)$ the conditional density of realized volatility at period t, where θ is a vector of all parameters in a given model, and let $\ell_n(\hat{\theta}) = \sum_t \log f_{t-1}(rv_t|\hat{\theta})$ be the loglikelihood function. The 5th set of predictions is produced by model averaging using smoothed Takeuchi information criterion (STICMA). The Takeuchi information criterion (TIC, Takeuchi (1976)) is a more general version of the familiar Akaike information criterion (AIC) that acknowledges misspecification of the conditional density, which is important in our setup:

$$TIC = -2\ell_n\big(\hat{\theta}\big) + 2\mathrm{tr}\big(\hat{J}^{-1}\hat{I}\big),$$

where $\hat{J}$ and $\hat{I}$ are empirical analogs of $J = -E\left[\partial^2 \log f_{t-1}(rv_t|\theta)/\partial\theta\partial\theta'\right]$ and $I = E\left[\partial \log f_{t-1}(rv_t|\theta)/\partial\theta\, \partial \log f_{t-1}(rv_t|\theta)/\partial\theta'\right]$, the ingredients of the asymptotic variance of the quasi-ML estimator. When the given model is correctly specified, $J = I$ and $\mathrm{tr}\left(J^{-1}I\right) = K$, and TIC reduces to AIC, $AIC = -2\ell_n(\hat{\theta}) + 2K$, where $K = \dim(\theta)$ is a total number of parameters in the model under consideration. Implementationwise, we compute $\hat{J}$ and $\hat{I}$ by using numerical derivatives (see, e.g., Judd (1998)). For the MEM models, formulas for the TICs are straightforward, but for LogARMA models formulated for log-transformed variables, the TIC that contains densities of observables should be adjusted, according to the relation between densities of original and transformed variables: if $z = \exp(x)$, then $f_Z(z) = f(\log z)/z$, and so $E[\log f_Z(z)] = E[\log f(x) - x]$. The STICMA weights w_i are given by

$$w_i = \frac{\exp(-TIC_i/2)}{\sum_{j=1}^{M} \exp(-TIC_j/2)}, \quad i = 1, \ldots, M.$$

The second model averaging combination, which produces the 6th set of predictions, is based on the out-of-sample performance of individual models. To this end, we adapt the jackknife model averaging (JMA) machinery (Hansen and Racine 2012) to the nonlinear modeling setup. Here, the weights w_i, $i = 1, \ldots, M$, are determined by minimizing the cross-validation (CV) criterion, which is computed from all M models' forecast errors on the validation subsample; for exact formulation, see equations (5) and (6) in Hansen and Racine (2012). The quadratic optimization problem subject to the constraint that all the M weights are non-negative and sum to unity, is a nice easily implementable problem even when numerically solved repeatedly in a rolling window.

3 Empirical Evaluation

3.1 Data

We perform the estimation and forecasting exercises using the popular elsewhere data-set of realized stock market volatility from (Noureldin et al. 2012). This dataset contains daily realized volatilities on 10 stocks: BAC, JPM, IBM, MSFT, XOM, AA, AXP, DD, GE, KO, and covers the period from February 1, 2001 to December 31, 2009. Because towards the end of the sample volatilities exhibit turbulence, we cut it at May 31, 2007. This leaves 1589 observations, of which the last 589 we use for forecast evaluation, and estimate the models in a 1000-observations window. The length of the validation subsample is set at 100 observations. The ten volatility series are depicted in Fig. 1.

3.2 Model Estimation

Box 1 shows estimation results for BAC, as a typical example, in the first estimation window. For the two MEM models, there is a stark difference in the degree of in-sample fit between the small and big models. The large gap between the corresponding loglikelihood values and information criteria is delivered by the two shape parameters of the Burr density sharply different from those implied by the standard exponential density. The parameters between the two conditional mean equations are very similar and well identified.

In contrast, the differences in the degree of fit between the two LogARMA models are modest at most, even though the density shape parameters of the skewed Student distribution statistically significantly differ from those implied by the standard normal density. The two conditional mean equations are very similar and pretty well identified, while some estimates in the variance equation differ quite a lot and have big standard errors. Evidently, the dynamics of 'volatility of volatility' is quite hard to identify.

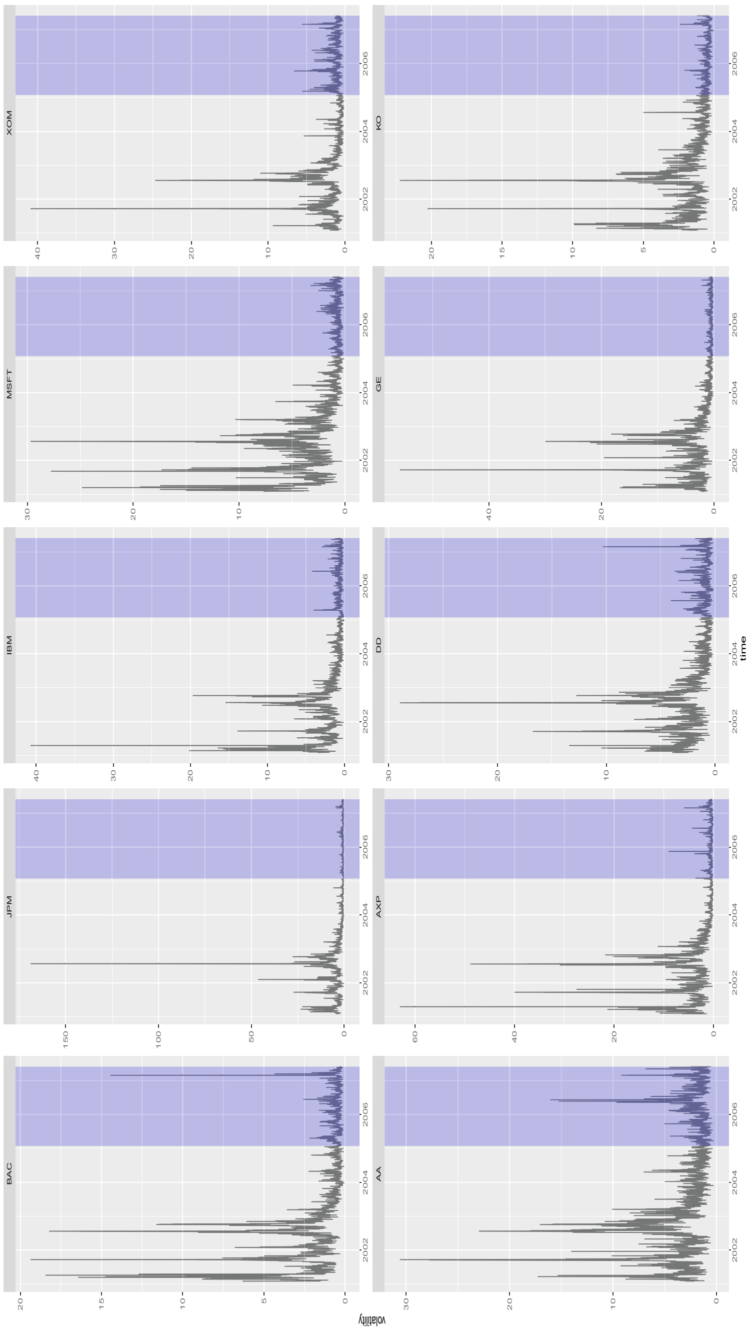

Fig. 1 Data on realized volatilities (shaded is out-of-sample period)

3.3 Model Average Weights

Figures 2 and 3 depict, correspondingly, the values of STICMA and JMA weights in a rolling window for all the stocks. When the weights are formed from the in-sample performance (Fig. 2), the small MEM model (ExpMEM) always has zero weights as a result of its having a much smaller in-sample quality than the big MEM model (BurrMEM).

ExpMEM model

$$\varepsilon_t | I_{t-1} \sim \mathscr{E}$$

$$\log \mu_t = \underset{(0.0096)}{0.0420} + \underset{(0.193)}{0.363} \log rv_{t-1} + \underset{(0.215)}{0.615} \log \mu_{t-1}$$

$$LL = -1165.7, \ TIC = 2333.8$$

BurrMEM model

$$\varepsilon_t | I_{t-1} \sim \mathscr{B}\left(\underset{(0.159)}{3.081}, \underset{(0.198)}{1.277}\right)$$

$$\log \mu_t = \underset{(0.0117)}{0.0438} + \underset{(0.140)}{0.371} \log rv_{t-1} + \underset{(0.150)}{0.611} \log \mu_{t-1}$$

$$LL = -675.0, \ TIC = 1361.2$$

NormLogARMA model

$$\log rv_t = \underset{(0.00569)}{-0.00267} + \underset{(0.009)}{0.977} \log rv_{t-i} + e_t - \underset{(0.040)}{0.611} e_{t-1}$$

$$\log \sigma_t^2 = \underset{(2.375)}{-0.155} + \underset{(0.275)}{0.071} |\eta_{t-1}| - \underset{(0.021)}{0.013} \eta_{t-1} + \underset{(0.979)}{0.938} \log \sigma_{t-1}^2$$

$$\eta_t | I_{t-1} \sim \mathscr{N}$$

$$LL = -625.6, \ TIC = 1383.4$$

SkStLogARMA model

$$\log rv_t = \underset{(0.00591)}{-0.00077} + \underset{(0.009)}{0.981} \log rv_{t-1} + e_t - \underset{(0.033)}{0.611} e_{t-1}$$

$$\log \sigma_t^2 = -\underset{(10.231)}{0.809} + \underset{(0.179)}{0.112} |\eta_{t-1}| - \underset{(0.046)}{0.000} \eta_{t-1} + \underset{(5.810)}{0.545} \log \sigma_{t-1}^2$$

$$\eta_t | I_{t-1} \sim \mathscr{S}\mathscr{S}\left(\underset{(0.0370)}{0.0927}, \underset{(0.102)}{1.546}\right)$$

$$LL = -615.3, \ TIC = 1361.4$$

Box 1: Results of estimation of the four models for BAC in the first window.

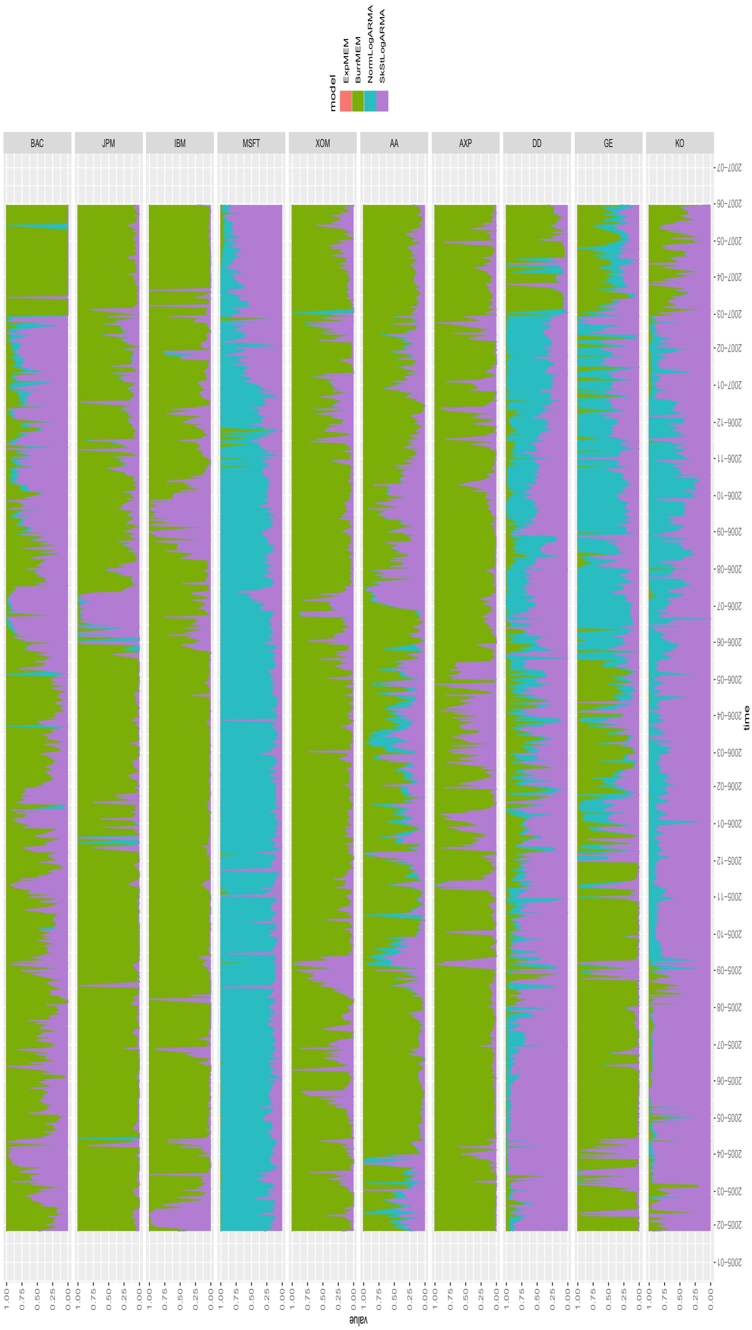

Fig. 2 Values of STICMA weights in rolling window

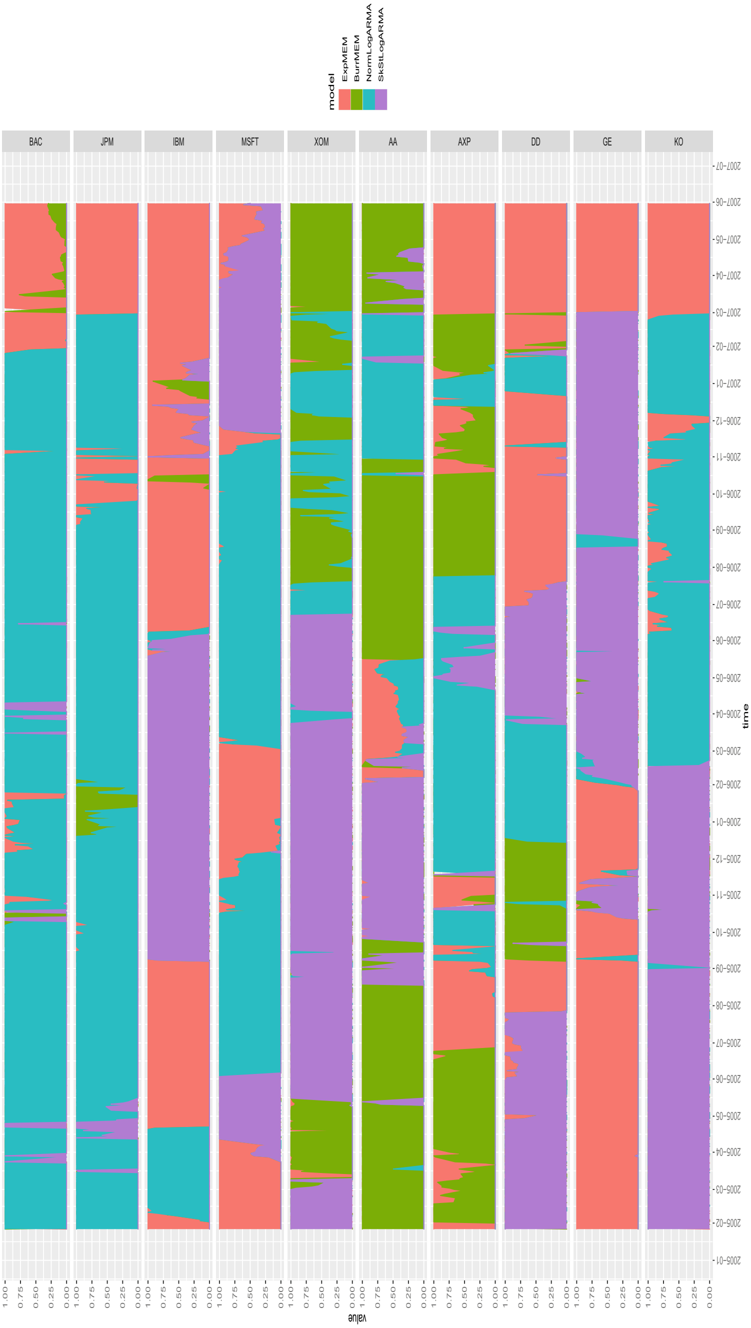

Fig. 3 Values of JMA weights in rolling window

The big MEM and both LogARMA models perform in-sample on par with each other, though the parity depends on the asset under consideration. For most stocks, two of the three models stand out—for example, for MSFT, only the two LogARMAs have nontrivial weights, while, for example, for IBM, it is the two big models that perform better. For other stocks, for example, for GE, all the three models are balanced. The balance between the non-trivially weighted models usually varies in time in clusters, with weights for one model being able to take values close to zero at times and values near unity at other times. Table 1 presents the average, together with standard deviations, STICMA weights from different models. Any of the models other than ExpMEM is able to dominate on average, although BurrMEM seems to be doing it most often.

When the weights are formed from the out-of-sample performance (Fig. 3), the role of the small MEM model (ExpMEM) ceases to be trivial, and for some stocks in some periods this models dominates. The other three models are keeping up, so for each of the four models there are combinations of stocks and periods, albeit short, when this model dominates the other three. There are much fewer instances when two models are on par with each other, than when the weights are driven by in-sample performance. Table 2 presents the average, together with standard deviations, JMA weights from different models. Each of the four models is able to exhibit average weights of 50% or higher for some of the stocks, and at the same time average weights close to zero for others. The standard deviations also point at the instability of weights as functions of a position of the validation subsample.

Table 1 Average (and standard deviations of) STICMA weights from different models

	BAC	JPM	IBM	MSFT	XOM	AA	AXP	DD	GE	KO
ExpMEM	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
BurrMEM	0.56 (0.33)	0.87 (0.24)	0.86 (0.26)	0.02 (0.10)	0.86 (0.19)	0.70 (0.26)	0.85 (0.23)	0.22 (0.29)	0.51 (0.43)	0.10 (0.20)
NormLogARMA	0.03 (0.11)	0.02 (0.11)	0.00 (0.02)	0.72 (0.26)	0.00 (0.04)	0.06 (0.12)	0.00 (0.00)	0.25 (0.26)	0.26 (0.31)	0.21 (0.24)
SkStLogARMA	0.41 (0.31)	0.12 (0.22)	0.14 (0.26)	0.26 (0.24)	0.13 (0.19)	0.24 (0.21)	0.15 (0.23)	0.53 (0.29)	0.23 (0.25)	0.69 (0.28)

Table 2 Average (and standard deviations of) JMA weights from different models

	BAC	JPM	IBM	MSFT	XOM	AA	AXP	DD	GE	KO
ExpMEM	0.15 (0.33)	0.16 (0.36)	0.56 (0.48)	0.24 (0.39)	0.01 (0.08)	0.07 (0.20)	0.29 (0.42)	0.40 (0.48)	0.50 (0.49)	0.15 (0.33)
BurrMEM	0.02 (0.09)	0.03 (0.14)	0.01 (0.09)	0.00 (0.00)	0.33 (0.45)	0.54 (0.49)	0.35 (0.45)	0.12 (0.32)	0.00 (0.03)	0.00 (0.02)
NormLogARMA	0.80 (0.39)	0.79 (0.39)	0.10 (0.29)	0.50 (0.49)	0.15 (0.33)	0.17 (0.35)	0.32 (0.46)	0.15 (0.36)	0.02 (0.13)	0.40 (0.47)
SkStLogARMA	0.03 (0.17)	0.02 (0.13)	0.33 (0.46)	0.26 (0.42)	0.51 (0.50)	0.22 (0.39)	0.04 (0.17)	0.33 (0.46)	0.47 (0.49)	0.45 (0.50)

3.4 Forecasting Performance

Now we turn to comparing the forecasting performance of individual models and their model averages. Table 3 contains out-of-sample average squared errors from different models for all the stocks, and Table 4 reports model confidence sets (MCS) for the 25% confidence level, which is a conventional level in volatility analysis (see Laurent et al. (2012)). The MCS machinery (Hansen et al. 2011) allows statistically correct multiple hypotheses testing, and the MCS is a subset of models from the pool of all predictive models under consideration that are statistically insignificanly different by their forecasting performance. The null hypothesis states that all the models inside the MCS perform equally well, while any model outside of the MCS performs worse; for more details, see Hansen et al. (2011).

One can immediately see that the four 'pure' models are very similar in forecasting performance, and for most stocks multiple models can be deemed the best. Often, it is two or three models out of the four, but it may be also all four. The model that always belongs to the MCS at the 25% level is the NormLogARMA model, i.e. small LogARMA. This happens despite the estimation noise in barely identifiable volatility-of-volatility dynamics. Also, recall that an additional advantage of this particular LogARMA model is that forecasts can be computed without numerical integration. The SkStLogARMA model, i.e. big LogARMA, is a bit less likely to be among the best, so the MCS contains this model for 7 out of 10 stocks. The ExpMEM model, i.e. small MEM, is much less likely to be among the best, and only for half of the stocks is this model contained in the MCS. Finally, the BurrMEM model, i.e. big MEM, enters the MCS only for 3 out of 10 stocks. Overall, the class of LogARMA

Table 3 Average out-of-sample losses from different models

	BAC	JPM	IBM	MSFT	XOM	AA	AXP	DD	GE	KO
ExpMEM	0.4364	0.2097	0.1538	0.1537	0.4946	1.4307	0.3422	0.4118	0.0748	0.0570
BurrMEM	0.4376	0.2090	0.1541	0.1555	0.4926	1.4329	0.3423	0.4136	0.0758	0.0571
NormLogARMA	0.4372	0.2070	0.1536	0.1525	0.4922	1.4318	0.3423	0.4136	0.0752	0.0565
SkStLogARMA	0.4378	0.2076	0.1532	0.1527	0.4907	1.4349	0.3424	0.4133	0.0754	0.0564
STICMA	0.4377	0.2084	0.1540	0.1525	0.4928	1.4325	0.3421	0.4136	0.0756	0.0564
JMA	0.4376	0.2069	0.1534	0.1527	0.4932	1.4332	0.3426	0.4121	0.0750	0.0563

Table 4 Predictive model confidence sets for 25% confidence level

	BAC	JPM	IBM	MSFT	XOM	AA	AXP	DD	GE	KO
ExpMEM	*					*	*	*	*	
BurrMEM					*	*	*	*		
NormLogARMA	*	*	*	*	*	*	*	*	*	*
SkStLogARMA		*	*		*	*	*	*	*	*
STICMA				*	*	*	*	*		*
JMA	*	*	*		*	*	*	*	*	*

models seems to be more reliable in forecasting than the class of MEM models. There is a slight tendency of small models to dominate big models from the same classes.

The two bottom lines in Tables 3 and 4 show the figures for the model averaging forecasts that are based on STICMA and JMA. The model average combinations do sometimes, though not always, improve the forecasting performance relative to individual models, confirming the common wisdom. It is also intuitive that model averaging based on out-of-sample performance fairs better than model averaging based on in-sample criteria. In fact, the JMA forecasts entered the 25% MCS for almost all stocks—9 out of 10, in contrast to STICMA, which is contained in the MCS only for 6 stocks.

4 Concluding Remarks

We have run a mini-competition among several models from two popular model classes—MEM and LogARMA—for realized volatility of ten liquid stocks, paying main attention to the forecasting quality. Overall, the class of LogARMA models seems to be more reliable in forecasting than the class of MEM models, while small models tend to dominate big models from the same class. The small LogARMA model and the model average based on forecasting performance have more chances to yield best predictions than the other individual models or the information criterion based model average, although this tendency is unstable through time and across stocks. For some stocks, the difference across all the models seems to be immaterial. The decision about which model class to select and how complex a model within the class to use does not seem empirically that big of a deal.

Acknowledgements The author thanks organizers and audiences of the 41st International Symposium on Forecasting (virtual, June 2021), 7th International Conference on Time Series and Forecasting (Gran Canaria, Spain, July 2021) and 5th International Conference on Financial Econometrics (hybrid, Ho Chi Minh City, Vietnam, January 2022). This research was supported by grant 20-28055S from the Czech Science Foundation. I also thank Filip Staněk for excellent research assistance.

References

Allen, D., Chan, F., McAleer, M., Peiris, S.: Finite sample properties of the QMLE- for the LogACD model: application to Australian stocks. J. Econometr. **147**, 163–185 (2008)

Anatolyev, S., Petukhov, A.: Uncovering the skewness news impact curve. J. Financ. Econometr. **14**(4), 746–771 (2016)

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P.: The distribution of realized exchange rate volatility. J. Amer. Stat. Assoc. **96**, 42–55 (2001)

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P.: Modeling and forecasting realized volatility. Econometrica **71**(2), 579–625 (2003)

Buckland, S.T., Burnham, K.P., Augustin, N.H.: Model selection: an integral part of inference. Biometrics **53**, 603–618 (1997)
Corsi, F.: A simple approximate long-memory model of realized volatility. J. Financ. Econometr. **7**(2), 174–196 (2009)
Engle, R.F.: New frontiers for ARCH models. J. Appl. Econometr. **17**(5), 425–446 (2002)
Engle, R.F., Russell, J.R.: Autoregressive conditional duration: a new model for irregularly spaced transaction data. Econometrica **66**(5), 1127–1162 (1998)
Engle, R.F., Gallo, G.M.: A multiple indicators model for volatility using intra-daily data. J. Econometr. **131**(1–2), 3–27 (2006)
Grammig, J., Maurer, K.-O.: Non-monotonic hazard functions and the autoregressive conditional duration model. Econometri. J. **3**, 16–38 (2000)
Hansen, B.E.: Autoregressive conditional density estimation. Int. Econ. Rev. **35**(3), 705–730 (1994)
Hansen, B.E., Racine, J.: Jackknife model averaging. J. Econ. **167**, 38–46 (2012)
Hansen, P.R., Lunde, A., Nason, J.M.: The model confidence set. Econometrica **79**(2), 453–497 (2011)
Hautsch, N.: Econometrics of Financial High-Frequency Data. Springer Science & Business Media, Berlin/Heidelberg, Germany (2011)
Hautsch, N., Malec, P., Schienle, M.: Capturing the zero: a new class of zero-augmented distributions and multiplicative error processes. J. Financ. Econometr. **12**(1), 89–121 (2014)
Judd, K.: Numerical Methods in Economics. MIT Press, Cambridge, MA (1998)
Laurent, S., Rombouts, J.V.K., Violante, F.: On the forecasting accuracy of multivariate GARCH models. J. Appl. Econometr. **27**(6), 934–955 (2012)
Noureldin, D., Shephard, N., Sheppard, K.: Multivariate high-frequency-based volatility (HEAVY) models. J. Appl. Econometr. **27**(6), 907–933 (2012)
Takeuchi, K.: Distributions of information statistics and criteria for adequacy of models (in Japanese). Suri-Kagaku (Math. Sci.) **153**, 12–18 (1976)