



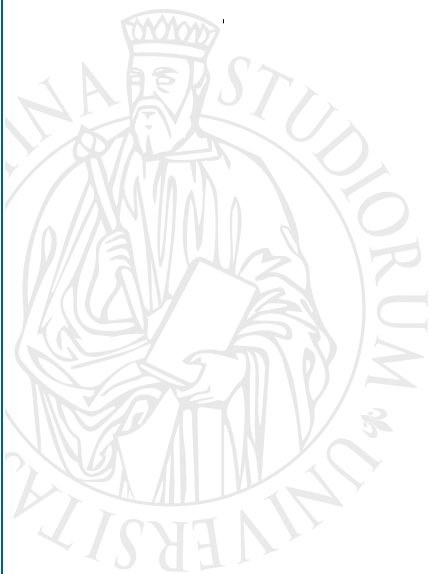
UNIVERSITÀ
DEGLI STUDI
FIRENZE

DISIA

DIPARTIMENTO DI STATISTICA,
INFORMATICA, APPLICAZIONI
"GIUSEPPE PARENTI"

Realized variance modeling: decoupling forecasting from estimation

Fabrizio Cipollini, Giampiero M. Gallo,
Alessandro Palandri



**DISIA WORKING PAPER
2019/05**

© Copyright is held by the author(s).

Realized variance modeling: decoupling forecasting from estimation

Fabrizio Cipollini¹, Giampiero M. Gallo^{2,*} and Alessandro Palandri^{1,3}

¹Dipartimento di Statistica, Informatica, Applicazioni (DiSIA)
G. Parenti, Università di Firenze

²Italian Court of Audits, and New York University in Florence

³DCU Business School, Dublin City University

This Version: June 25, 2019

Abstract

In this paper we evaluate the in-sample fit and out-of-sample forecasts of various combinations of realized variance models and *estimation criteria*. Our empirical findings highlight that: independently of the econometrician's *forecasting loss* function, certain *estimation criteria* perform significantly better than others; the simple ARMA modeling of the log realized variance generates superior forecasts than the HAR family, for any of the *forecasting loss* functions considered; the (2,1) parameterizations with negative lag-2 coefficient emerge as the benchmark specifications generating the best forecasts and approximating long-run dependence as well as the HAR family.

Keywords: Variance modeling; Variance forecasting; Heterogeneous Autoregressive (HAR) model; Multiplicative Error Model (MEM); Realized variance

JEL classification: C32, C53, C58, G17

*Address correspondence to: Giampiero M. Gallo, NYU in Florence, Italy, Email: giampiero.gallo@nyu.edu. The views expressed in the article are those of the authors and do not involve the responsibility of the Corte dei conti.

1 Introduction

The technical capability of recording and storing intradaily data has given a tremendous boost to the literature both on measurement issues (various forms of aggregation of ultra-high frequency data into a daily value) and on modeling their dynamics (extending the autoregressive flavor behind the GARCH class of models). There exist recent refinements within a somewhat consolidated menu of realized variance choices (plain vanilla, robustness to market microstructure noise, jumps, etc.; cf. Park and Linton, 2012, for a survey); in this work we focus on the realized kernel variance (Barndorff-Nielsen *et al.*, 2008) to investigate a series of aspects related to modeling and forecasting. We will take it either as such, or as its square root, or yet as its logarithm, but for the sake of simplicity we generically talk here about realized variance modeling.

In this paper we consider that evaluating forecasts *out-of-sample* (OOS) is a matter of subjective taste about how to judge the distance between a predicted outcome and the actual value. In this respect, several examples of *forecasting loss* functions are considered, with the idea that a particular *a priori* choice is a consequence of individual preferences, hence it is not subject to qualitative assessment. In order to produce forecasts, two other elements are important: the *model specification* (the equation reproducing conditional variance dynamics), and the in-sample (IS) *estimation criterion*, i.e. the distance between observed and fitted values delivering parameter estimates. As for the adoption of a model specification and of an *estimation criterion*, we take the view that there are no natural *a priori* choices, but they must be geared toward obtaining the best results in terms of the given *forecasting loss*. In this respect, we argue that the choice of the latter should not force the same function to be repeated as the *estimation criterion*.

A first contribution of the paper is to explore whether, for a given *forecasting loss*, the IS *estimation criterion* with the same functional form produces the best OOS results by specification or, rather, other *estimation criteria* are to be preferred. A second contribution is to investigate the capability of a popular model selection tool such as the BIC to identify the best OOS specification. Third, we compare the benchmark HAR specifications of Corsi (2009) and Andersen *et al.* (2007) to several realized variance models inspired by the functionals of successful GARCH parameterizations which, for the most part, correspond to the simple ARMA modeling of either realized variance, its square root or its logarithm. Across classes, here

we focus on the *core specifications* that could be extended to accommodate several refinements (asymmetry, see Engle and Gallo (2006), jumps, see Andersen *et al.* (2007), measurement errors, see Bollerslev *et al.* (2016), etc.).

We explore these issues on a panel of 28 constituents of the Dow Jones 30 Index using daily realized kernel variance observations from January 2005 to December 2015. The sample is split into six 5-year IS periods; we estimate all combinations of specifications and *estimation criteria* on the first IS period, generating OOS static forecasts for the ensuing 1-year period, then we move the IS window by one year and repeat the procedure. Our results may be summarized as follows: there are *forecasting loss* functions not particularly apt to be repeated as *estimation criteria*; we find encouraging results for the BIC ability to identify *ex-ante* the best OOS specification; the ARMA modeling of the log realized variance provides the best IS and OOS results. In general, we identify the (2,1) structure with a negative lag-2 coefficient to be a good parameterization. As a reading key to these results, we find it informative to relate the goodness of the OOS forecasts to the structure's capability to mimic long memory features: as a matter of fact, these specifications deliver a long memory approximation which is equivalent to that of the HAR family, but overall superior OOS forecasts across assets.

The paper is structured as follows. Section 2 discusses the *forecasting losses* employed to compare model forecasts and the *estimation criteria* used to estimate parameters. Section 3 presents the specifications for the variance dynamics and Section 4 the information criteria used for model selection. Data are introduced in Section 5 and results are presented in Section 6. In Section 7 we provide a general discussion on the *estimation criteria* and the (2,1) parameterizations and their ability to approximate long memory. Section 8 concludes.

2 Forecasting Loss and Estimation Criterion

A loss function maps events onto real numbers according to an individual's preference orderings. When estimating, values of the unknown coefficients of a parametric model are obtained from the minimization of a loss function IS. When forecasting, the quality of a specification is assessed through the calculation of a loss function based on the distance between the predicted outcomes and the actual values, OOS. In what follows, we argue that the priority given to forecasting requires to define separately the *forecasting loss* and the *estimation criterion*.

With the expression *forecasting loss* (henceforth FL) we refer to the loss function used to evaluate the model performance OOS. Due to its subjective nature, the *forecasting loss* does not satisfy any criteria other than representing an individual's preferences. With the term *estimation criterion* (henceforth EC) we refer to the objective function used to obtain the parameter estimates. In general, the choice of this function responds to some features: theoretical properties of the resulting estimator, tractability, etc. Distinguishing between forecast evaluation and model estimation allows us to investigate, empirically and in the context of variance modeling, whether the EC that coincides with the FL produces the best OOS results. For FL and EC we consider some of the most common functionals adopted in the literature to measure the distance between observed and predicted values. Prevailing measures are quadratic, which corresponds to least-squares estimation, and Kullback-Leibler, which corresponds to quasi-maximum likelihood estimation when the chosen density function is linear-exponential.

2.1 Quadratic Distance

We characterize the general quadratic *forecasting loss* by:

$$\sum_{t=T_1+1}^{T_2} [f(RV_t) - f(\sigma_t^2)]^2$$

where $t = T_1 + 1, \dots, T_2$ is the OOS period and $f(\sigma_t^2) \equiv \mathbb{E}_{t-1}[f(RV_t)]$. Common choices for the monotonic function $f(\cdot)$ are the identity, the square-root and the logarithm, which lead to what we define as the LS, SDLS and LNLS *forecasting losses*, respectively. As an example, for the square-root function we have that the *forecasting loss* synthesizes the distance between $RV_t^{1/2}$ and $\sigma_t \equiv E_{t-1}[RV_t^{1/2}]$, $t = T_1 + 1, \dots, T_2$.

We define the quadratic *estimation criterion* in the same way but without constraining it to have the same functional $f(\cdot)$ of the FL:

$$\sum_{t=1}^{T_1} [g(RV_t) - g(\sigma_t^2)]^2$$

where $t = 1, \dots, T_1$ is the IS period and $g(\cdot)$ is a monotonic function. Choosing $g(\cdot)$ from the f -menu of identity, square-root and logarithmic functions, leads to LS, SDLS and LNLS *estimation criteria*, respectively. Since they allow for the simple OLS estimation of the models' parameters, LS is the *estimation criterion* of choice for the HAR of Corsi (2009) and LNLS for the LOG – HAR of Andersen *et al.* (2007). Do notice

that σ_t^2 in the EC is that defined by the *forecasting loss*: $\sigma_t^2 = f^{-1}(\mathbb{E}_{t-1}[f(RV_t)])$. Therefore, unless $g = f$, there is a misalignment between what the model parameters are trained to predict, $\mathbb{E}_{t-1}[g(RV_t)]$, and the object of interest $\mathbb{E}_{t-1}[f(RV_t)]$.

2.2 Kullback-Leibler Distance

Of the many density functions that may characterize a Kullback-Leibler distance we focus on the Gaussian density to generate the QML *forecasting loss*:

$$\sum_{t=T_1+1}^{T_2} \left[\ln \sigma_t^2 + \frac{RV_t}{\sigma_t^2} \right]$$

where $t = T_1 + 1, \dots, T_2$ is the OOS period and σ_t^2 is the conditional expectation $\mathbb{E}_{t-1}[RV_t]$ of realized variance. Similarly, the QML *estimation criterion* is given by:

$$\sum_{t=1}^{T_1} \left[\ln \sigma_t^2 + \frac{RV_t}{\sigma_t^2} \right] \quad (1)$$

where $t = 1, \dots, T_1$ is the IS period and $\sigma_t^2 = f^{-1}(\mathbb{E}_{t-1}[f(RV_t)])$, as defined by the FL. Since the Gaussian belongs to the family of *linear-exponential* distributions, minimization of the QML *estimation criterion* is in fact a quasi-maximum-likelihood estimation with associated properties: Gouriéroux *et al.* (1984). If intraday observations are normally distributed, Gaussian QML is maximum likelihood. Although differently motivated, the EC in equation (1) is the estimator of choice for the univariate MEM of Brownlees *et al.* (2012).

3 Variance Modeling

In this section we present the *cores* of the HAR and several other models inspired by successful GARCH parameterizations, which nest various specifications that have been put forward in the literature on realized variance modeling. Motivated by the stylized facts of GARCH modeling, we focus on the (1, 1) and (2, 1) parameterizations where the former is found to be well suited to generate good forecasts, as highlighted by Hansen and Lunde (2005), while the latter occasionally provided better fit and forecasts.

We present all models as parameterizing σ_t^2 . Although this is non-standard, it is consistent with our setup in which the model parameters are chosen IS to generate the best predictions $\mathbb{E}_{t-1}[g(\sigma_t^2)]$ but are ultimately used to produce OOS forecasts

$\mathbb{E}_{t-1}[f(\sigma_t^2)]$. While our empirical analysis focuses on one-step ahead forecasts, k -step-ahead forecasts $\mathbb{E}_{t-1}[f(\sigma_{t+k-1}^2)]$ may be generated for any of the following parameterizations, granting that in general their calculation requires numerical integration for $k > 1$.

3.1 HAR Specifications

The HAR, introduced by Corsi (2009), has rapidly achieved the status of benchmark specification for the modeling of realized variances. Features that contribute to its role as a point of reference are the simplicity with which its parameters may be estimated and its ability to reproduce long memory features: “[...] *the mixing of relatively few volatility components is capable of reproducing a remarkably slow volatility autocorrelation decay that is almost indistinguishable from that of a hyperbolic pattern over most empirically relevant forecast horizons*” (Andersen *et al.*, 2007). The HAR models σ_t^2 as a function of past realizations over daily, weekly and monthly time intervals:

$$\sigma_t^2 = \omega + \alpha_1 \cdot RV_{t-1} + \alpha_2 \cdot \frac{1}{5} \sum_{i=1}^5 RV_{t-i} + \alpha_3 \cdot \frac{1}{22} \sum_{i=1}^{22} RV_{t-i} \quad (2)$$

corresponding to an $AR(22)$ process for RV_t with parameter constraints. Its parameters may be estimated by ordinary least squares when the criterion used in estimation is LS, whereas for different choices of the EC the estimates are not available in closed form. Necessary and sufficient conditions for the positivity of σ_t^2 are $\omega > 0$, $\alpha_3 \geq 0$, $\alpha_2/5 + \alpha_3/22 \geq 0$ and $\alpha_1 + \alpha_2/5 + \alpha_3/22 \geq 0$.

The LOG – HAR, introduced by Andersen *et al.* (2007), is an alternative specification linear in the logarithms:

$$\sigma_t^2 = \exp \left\{ \omega + \alpha_1 \cdot \ln RV_{t-1} + \alpha_2 \cdot \ln \left(\frac{1}{5} \sum_{i=1}^5 RV_{t-i} \right) + \alpha_3 \cdot \ln \left(\frac{1}{22} \sum_{i=1}^{22} RV_{t-i} \right) \right\} \quad (3)$$

Notice that the presence of the logarithms of averages places the LOG – HAR outside the class of AR processes. Its parameters may be estimated by ordinary least squares for LNLS *estimation criterion*, while for different choices of the EC the estimates are not available in closed form.

3.2 MVAR Specification

With MVAR we indicate the parameterization of σ_t^2 in terms of its lags and lags of RV_t or, equivalently, the ARMA modeling of realized variance RV_t . It is inspired

by the GARCH(p, q) of Bollerslev (1986) from which it may be obtained by substituting the more precise realized variances for the squared open-to-close returns. This GARCH would arise as the limiting case of MVAR when the number of intradaily observations is one and realized variance collapses to the squared open-to-close return. The MVAR(2, 1) specification is given by:

$$\sigma_t^2 = \omega + \alpha_1 \cdot RV_{t-1} + \alpha_2 \cdot RV_{t-2} + \beta_1 \cdot \sigma_{t-1}^2. \quad (4)$$

Necessary and sufficient conditions for the positivity of σ_t^2 are $\omega > 0$, $\alpha_1, \beta_1 \geq 0$ and $\alpha_2 \geq -\alpha_1\beta_1$. MVAR estimated by QML coincides with the MEM of Cipollini *et al.* (2013) while associated to the LS criterion it reduces to standard ARMA modeling and estimation.

3.3 MVOL Specification

With MVOL we denote the the parameterization of σ_t in terms of its lags and lags of $RV_t^{1/2}$ or, equivalently, the ARMA modeling of realized volatility $RV_t^{1/2}$. It is related to the TGARCH(p, q) of Zakoian (1994), but without the asymmetric term, from which it may be obtained by substituting realized volatilities for the absolute value of the returns. Alternatively, the TGARCH may be seen as the limiting case of MVOL when there is one intradaily observation and realized volatility collapses to the absolute open-to-close return. The symmetric MVOL(2, 1) specification is given by:

$$\sigma_t^2 = \left\{ \omega + \alpha_1 \cdot RV_{t-1}^{1/2} + \alpha_2 \cdot RV_{t-2}^{1/2} + \beta_1 \cdot \sigma_{t-1} \right\}^2 \quad (5)$$

Although σ_t^2 is positive by construction, the marginal effects of its determinants do not exhibit abrupt sign changes if and only if σ_t is also positive. Necessary and sufficient conditions for the positivity of σ_t are $\omega > 0$, $\alpha_1, \beta_1 \geq 0$ and $\alpha_2 \geq -\alpha_1\beta_1$. QML estimation of MVOL coincides with the MEM in Brownlees *et al.* (2012) while adopting the SDLS criterion reduces to standard ARMA modeling and estimation.

3.4 MLOG Specification

MLOG is the ARMA modeling of the log realized variances $\ln RV_t$ or, equivalently, the parameterization of $\ln \sigma_t^2$ in terms of its lags and lags of $\ln RV_t$. It may be seen as the analogue to the log-GARCH(p, q) of Geweke (1986) from which it may be obtained by substituting log realized variances for log squared residuals. Similarly, log-GARCH is the limiting case of MLOG when the log realized variance reduces to

log squared residual for a single intradaily observation. The MLOG(2, 1) specification is given by:

$$\sigma_t^2 = \exp \left\{ \omega + \alpha_1 \cdot \ln RV_{t-1} + \alpha_2 \cdot \ln RV_{t-2} + \beta_1 \cdot \ln \sigma_{t-1}^2 \right\} \quad (6)$$

When associated to the LNLS criterion it reduces to standard ARMA modeling and estimation. QML, on the other hand, would be the natural estimator within the MEM framework. Dynamic specifications analogous to the MLOG are not uncommon in the context of Autoregressive Conditional Durations, among which Bauwens *et al.* (2008) and Taylor and Xu (2017) are examples of QML¹ and LNLS estimates, respectively.

3.5 MEXP Specification

Inspired by the EGARCH(p, q) parameterization of Nelson (1991), but without the asymmetric term, the MEXP(2, 1) specification is obtained by substituting realized volatilities for the absolute value of the return. Like for the other GARCH-inspired parameterizations, in the limiting case of a single intraday observation, MEXP reproduces the symmetric EGARCH:

$$\sigma_t^2 = \exp \left\{ \omega + \alpha_1 \cdot \frac{RV_{t-1}^{1/2}}{\sigma_{t-1}} + \alpha_2 \cdot \frac{RV_{t-2}^{1/2}}{\sigma_{t-2}} + \beta_1 \cdot \ln \sigma_{t-1}^2 \right\} \quad (7)$$

Notice how the presence of lags of the dependent variable at the denominator places MEXP outside the class of ARMA processes.

4 Model Evaluation

We perform IS model evaluation and selection by means of the Bayes (Schwarz) Information Criterion *BIC*. For the quadratic *estimation criteria* of Section 2.1 we construct the *BIC* by treating the generalized residuals $g(RV_t) - g(\sigma_t^2)$ as Gaussian:

$$BIC = T_1 \ln \left(\frac{1}{T_1} \sum_{t=1}^{T_1} [g(RV_t) - g(\hat{\sigma}_t^2)]^2 \right) + k \ln T_1$$

where k is the number of parameters, T_1 the sample size and $\hat{\sigma}_t^2$ the model's prediction. For the Kullback-Leibler *estimation criterion* of Section 2.2, the *BIC* is immediately

¹To be precise: Bauwens *et al.* (2008) estimate an ACD analogous to MLOG by minimizing the Kullback-Leibler distance based on the exponential distribution. Since the exponential belongs to the family of *linear-exponential* distributions, the resulting estimator is also quasi-maximum-likelihood.

obtained from the log-likelihood function:

$$BIC = \sum_{t=1}^{T_1} \ln \hat{\sigma}_t^2 + k \ln T_1$$

where $\hat{\sigma}_t^2$ is the prediction from the QML-estimated model. Notice how the *BIC* is calculated on the condensed *estimation criterion* in (1) from which the average of RV_t/σ_t^2 is dropped on the grounds that deviations from its limiting value of one are neither data nor model driven but only reflect initial value choices².

Since, for equally parameterized specifications, every information criterion produces identical model rankings, Akaike's and Hannan-Quinn's may produce results that differ from those we present only in the comparisons of differently parameterized specifications. Furthermore, given that *BIC* is the most conservative of the three, when it selects a richer parameterization so do Akaike's and Hannan-Quinn's.

We evaluate OOS forecasts directly from the *forecasting loss* functions of Sections 2.1 and 2.2. The reason why no penalties are needed is that, contrary to IS, OOS measures of fit are not *non-decreasing* functions of the number of parameters k .

5 Data

The data used in this study pertains to 28 of the 30 constituents of the Dow Jones 30 Index. The sample has 11 years of high-frequency daily observations from 01/03/2005 to 12/31/2015 for a total of 2768 days. The two series, with tickers TRV and V, are not included in the study because they are not available for the full sample period³. Tickers of the 28 included stocks are: AAPL, AXP, BA, CAT, CSCO, CVX, DD, DIS, GE, GS, HD, IBM, INTC, JNJ, JPM, KO, MCD, MMM, MRK, MSFT, NKE, PFE, PG, UNH, UTX, VZ, WMT, XOM. The raw tick-by-tick TAQ data is cleaned using the procedure of Brownlees and Gallo (2006) and the series of realized variances calculated following Barndorff-Nielsen *et al.* (2011) with Parzen kernel. For every stock the time series of realized variances have been Winsorized at the 99.9-th percentile (3 observations) to eliminate the possibly disturbing effects of spurious measures such as those recorded on the day of the Flash Crash of May 6, 2010.

The sample is split into six 5-year IS periods: 2005-2009 (1259 obs.), 2006-2010 (1259 obs.), 2007-2011 (1260 obs.), 2008-2012 (1259 obs.), 2009-2013 (1258 obs.) and

²In fact, for all the specifications considered, when the initial value σ_0^2 is treated as an unknown parameter and estimated, the average ratio RV_t/σ_t^2 is equal to one.

³TRV data are available only from 02/26/2007 while V data are missing from 08/04/2006 to 02/26/2007.

2010-2014 (1258 obs.). All model combinations are estimated on each of the six IS periods, and for each of them, OOS forecasts are generated for the following 1-year period: 2010 (252 obs.), 2011 (252 obs.), 2012 (250 obs.), 2013 (252 obs.), 2014 (252 obs.) and 2015 (251 obs.).

6 Results

The first result we document is the poor performance of the MEXP specification inspired by the successful EGARCH: Tables 9-12. It never provides the best IS description of the data⁴ and it always generates the largest OOS *forecasting losses*: 300% and 50% increments with respect to LS and SDLS, respectively. Perhaps this is due⁵ to the different nature of the shocks driving the processes: return shocks in the case of the EGARCH against variance shocks in the MEXP. Due to these findings we do not discuss the the MEXP specification any further in what follows.

6.1 Is the forecasting loss the best estimation criterion?

In Tables 1-8 we evaluate the conditional variance specifications (rows) for a given FL when estimated by LS, SDLS, LNLS and QML *estimation criteria* (columns). From Tables 1 and 2 it emerges that when the FL is LS or SDLS, the *estimation criterion* that produces the best OOS results in most instances is LNLS. When the FL is LNLS, Table 3 highlights that it should also be used as EC as it produces the best OOS results in 86.90% to 94.64% of the instances. On the other hand, when the FL is QML, Table 4 shows that overall QML is the preferred EC.

These results are confirmed in Tables 5-8 which, for every conditional variance specification (rows) and *estimation criterion* (columns), report the average value of the given FL over the OOS period and across assets. The lowest average OOS *forecasting loss* measured by LS, SDLS and LNLS are obtained, for every variance specification, when estimated using the LNLS *estimation criterion*. Similarly, the QML *estimation criterion* produces the lowest OOS QML *forecasting losses*. Hence, while the hypothesis that the functional form of FL provides the best EC has to be rejected as a general statement, it cannot be rejected for the QML *forecasting loss*.

⁴The only exception is observed for the LS *forecasting loss* for which MEXP(2,1) is the best IS specification 1.79% of the instances, Table 9.

⁵While another possible explanation may be related to invertibility issues (see Wintenberger (2013)), in our case it does not seem to be the driving force: focusing on one step-ahead forecasts strongly mitigates any ill behavior (even explosive processes do not explode) and imposing invertibility constraints will further reduce the already poor IS performance of MEXP.

6.2 Does BIC identify the best OOS specification?

From Tables 9-12: BIC selects the best OOS specification 26.00%, 30.00%, 38.00% and 3.00% of the instances for LS, SDLS, LNLS and QML *estimation criteria*, respectively. Hence, it clearly emerges that the ability of BIC to identify what will be the best OOS specification is rather hopeless when associated to the QML *estimation criterion*. This finding is a bit surprising since the BIC associated to the QML *estimation criterion* follows directly from the EC itself as opposed to the other cases where BIC is *constructed* by treating the generalized residuals, from the various measures of distance, as normally distributed. In Tables 9 and 12 the average OOS *forecasting loss* of the specifications selected by BIC is no smaller than that of any of the specifications considered. On the other hand, the average *forecasting loss* of the BIC selection for the SDLS and LNLS *forecasting loss* functions is the second smallest, preceded only by the MLOG(2,1): Tables 10 and 11.

Therefore, for big caps, the best OOS results are obtained by adopting the MLOG specification for every asset. Clearly, it would be a hasty generalization to extend such recommendation to medium and small caps or other asset classes. A judicious approach would suggest to rely on model selection procedures unless the superiority of a specific model for a given asset class is substantiated by empirical evidence. Similarly, the performance of a maintained model should be monitored over time and asset classes to detect significant changes and intervene when necessary.

6.3 Is there a best OOS specification?

Tables 9-12 provide further IS and OOS performance measures for when the FL functional is used as EC. To begin, HAR does not emerge as the best OOS specification for any of the FL considered. Specifically, HAR is the best OOS specification 13.69%, 8.93%, 10.71% and 21.43% of the instances for LS, SDLS, LNLS and QML, respectively. Focusing on the average OOS *forecasting losses* generated by HAR it emerges that its performance is in-between that of MVAR(1,1) and MVAR(2,1), with all other specifications exhibiting lower OOS *forecasting losses*.

LOG – HAR produces the best OOS fit 16.07%, 8.33%, 10.71% and 6.55% of the instances for LS, SDLS, LNLS and QML *forecasting losses*, respectively. Similarly to HAR, it too does not emerge as the preferred OOS specification. In terms of average OOS *forecasting losses*, LOG – HAR performs better than HAR and MVAR but does not do as well as MVOL(2,1) and MLOG.

Considering HAR and LOG – HAR jointly doesn't alter the picture as they provide the best OOS fit 29.76%, 17.26%, 21.42% and 27.98% of the instances for LS, SDLS, LNLS and QML *forecasting losses*, respectively. In contrast, MLOG produces the best OOS specifications 47.02%, 64.88%, 64.88% and 54.16% of the times. In fact, what emerges is that the specification producing the smallest average OOS values of the FL considered is MLOG(2, 1) followed nearly *ex aequo* by MLOG(1, 1) and MVOL(2, 1).

6.4 Is there a best IS specification?

Even though the main focus of this study is on OOS forecasts, we recognize that there are instances where IS fit is of primary importance, e.g., a reduced form model capturing most relevant data features to be used as *auxiliary model* in the indirect estimation of a *maintained model*. In Tables 9-12, HAR does not emerge as the best IS specification: it provides the best IS fit only 4.76%, 4.17%, 2.38% and 12.50% of the instances for LS, SDLS, LNLS and QML *estimation criteria*, respectively.

LOG – HAR provides the best IS fit 13.10%, 30.95%, 25.00% and 24.40% of the instances for LS, SDLS, LNLS and QML *estimation criteria*, respectively. While there are sizeable improvements with respect to HAR, LOG – HAR is the preferred specification (30.95% of the instances) only for the QML *estimation criterion* with MLOG(1, 1) a very close second (30.36% of the times). On the other hand, with respect to the other EC considered, MLOG is selected from two to five more times than LOG – HAR as best IS specification.

Jointly, HAR and LOG – HAR are the preferred specifications 29.76%, 28.57%, 15.48% and 43.45% of the instances for LS, SDLS, LNLS and QML *estimation criteria*, respectively. However, the corresponding performance of MLOG of 65.48%, 58.93%, 70.24% and 30.36% is substantially better (except for QML). Furthermore, if the HAR and LOG – HAR pair is compared to the MLOG and MVOL pair, the latter is found to provide the best IS for any EC considered.

7 General Discussion

7.1 On the Estimation Criteria

A possible reading key to the empirical findings promoting LNLS and QML as preferred EC is that of their sensitivity to what we call *relatively-extreme observations* to distinguish them from outliers. Outliers are those data points that are distant

from the rest of the observations, according to some measure. In our context, outliers correspond to extreme measures of RV_t which, in general, are relatively easy to detect and to deal with⁶. Instead, by *relatively-extreme observations* we mean those data points that, while not being intrinsically anomalous, are extreme with respect to certain data features, e.g., with RV_{t-1} and RV_{t+1} in the 5-th percentile of magnitudes, for an otherwise persistent process, an RV_t in the 90-th percentile constitutes a *relatively-extreme observation* but not an outlier in the usual sense.

Focusing on the fact that parameters' estimates respond to *relatively-extreme observations* by re-setting the estimators' first order conditions to zero, we elicit the following two insights⁷. First, the impact of a *relatively-extreme observation* on the generalized residual, $\hat{\varepsilon}_t = g(RV_t) - g(\hat{\sigma}_t^2)$ for the criteria of Section 2.1 and $\hat{\varepsilon}_t = RV_t - \hat{\sigma}_t^2$ for the criterion of Section 2.2, is smallest for LNLS followed by SDLS and the pair of equals LS and QML. Secondly, the contribution of a *relatively-extreme observation* to the first order conditions is proportional to $\hat{\sigma}_t^{-n}$ where $\hat{\sigma}_t^2$ is the model's prediction, and $n = \{0, 1, 2, 4\}$ for LS, SDLS, LNLS and QML, respectively. It follows that inverse powers of $\hat{\sigma}_t^2$ have a dampening effect on the generalized residuals which is strongest for QML, followed by LNLS, SDLS and LS. Hence, SDLS and LS emerge as more sensitive to *relatively-extreme observations* in line with the empirical findings which don't promote any of the two as preferred EC.

7.2 The (2, 1) Parameterizations

The positive performance of MLOG and MVOL may be ascribed, among others, to the inclusion of the components associated with the coefficients β_1 and α_2 . The relevance of the moving-average coefficient β_1 should not come as a surprise given its ubiquitous presence in GARCH models. The (2, 1) parameterizations of MVAR, MVOL and MLOG are found to produce better forecasts than their (1, 1) counterparts: Tables 5-8. Figure 1 reports the distributions of the (2, 1) specifications' parameters estimated by LNLS and QML. The striking empirical regularity is the negative sign of the $\hat{\alpha}_2$ coefficients for every specification, asset and sub-sample period. As shown by Figure 2, the effect of $\alpha_2 < 0$ is to reduce (increase) persistence at short (long) horizons with respect to the geometrically decaying correlations of the (1, 1) parameterizations. Analogously, the negative relation summarized by $\alpha_2 < 0$ is not captured by the *core*

⁶In our study, the 99.9-th percentile Winsorization of the data effectively removes the few outliers present such as that associated to the Flash Crash of May 6, 2010.

⁷For a detailed treatment of the robustness of M-estimators, see Hampel *et al.* (2005).

HAR specifications.

7.3 Long Memory Approximation

Among the features that have made the HAR specifications popular is their ability to approximate *long memory* processes and thus replicate the hyperbolic decay observed in the autocorrelation of realized variances. The cross-sectional average (over 28 assets) of the empirical autocorrelations of realized variances (DATA), for the most recent subsample (2010-2014), is plotted in Figure 2. Using the cross-sectional average of the parameters' estimates for the same subsample, 10^6 Monte Carlo simulations, from the MVAR(2, 1), MVOL(2, 1), MLOG(2, 1), HAR and LOG – HAR, are used to plot the models' autocorrelation functions in Figure 2. The decay of DATA autocorrelations from lag-12 is well approximated by all specifications with the exception of the LOG – HAR, which reproduces the correct rate of decay but under-estimates the magnitude of the autocorrelations. Most interestingly, all (2, 1) specifications have autocorrelation structures that are indistinguishable from the HAR. Hence, replication of long-range dependence is not an exclusive feature of the HAR family but one attainable by equally parsimonious specifications. Equally interesting are the results for the autocorrelations up to lag-11. At lag-1, we find autocorrelations that are approximately 0.6 for the DATA, 0.67 and 0.5 for the HAR and LOG – HAR, respectively, but only 0.3-0.4 for the (2, 1) specifications.

These findings may be reconciled with the observed superior forecasts of MLOG by realizing that replicating the autocorrelation function of the observables is not a sufficient condition for a model to produce quality forecasts. Consider the following counter-example in which RV_t and two competing forecasts $\hat{\sigma}_{1,t}^2$ and $\hat{\sigma}_{2,t}^2$ are given by:

$$\begin{aligned} RV_t &= \sigma_t^2 + \nu_t \\ \hat{\sigma}_{1,t}^2 &= \sigma_t^2 + S_1\eta_{1,t} \\ \hat{\sigma}_{2,t}^2 &= \sigma_t^2 + S_2\eta_{2,t} \end{aligned}$$

with $1 < S_2 < S_1$; ν_t and $\eta_{2,t}$ i.i.d. standardized random variables; σ_t^2 a unit variance process with lag- j autocovariance γ_j ; $\eta_{1,t}$ a standardized random variable with lag- j autocovariance $\psi_j = 0.5S_1^{-2}(S_1^2 - 1)\gamma_j$. Then, the autocorrelations of the three processes are $\rho_j(RV) = 0.5\gamma_j$, $\rho_j(\hat{\sigma}_1^2) = 0.5\gamma_j$ and $\rho_j(\hat{\sigma}_2^2) = (1 + S_2^2)^{-1}\gamma_j$, respectively. The forecasting mean square errors of the competing specifications are: $\mathbb{E}[(RV_t - \hat{\sigma}_{1,t}^2)^2] = \mathbb{E}[(RV_t - \sigma_t^2)^2] + S_1^2$ and $\mathbb{E}[(RV_t - \hat{\sigma}_{2,t}^2)^2] = \mathbb{E}[(RV_t - \sigma_t^2)^2] + S_2^2$. Therefore, for $1 < S_2 < S_1$, the autocorrelation structure of $\hat{\sigma}_{1,t}^2$ is an exact match of

that of RV_t while $\hat{\sigma}_{2,t}^2$ exhibits lower autocorrelations. Nevertheless, the forecasting mean square error of $\hat{\sigma}_{2,t}^2$ is lower than that of $\hat{\sigma}_{1,t}^2$. It is therefore possible to have alternative specifications producing more precise forecasts than HAR even though the lag-1 autocorrelation of HAR is substantially higher as a result of larger misspecification components.

This analysis is supported by the contemporaneous correlations between realized variances and the models' forecasts: 0.6153 for the MVAR(2,1), 0.6219 for the MVOL(2,1), 0.6231 for the MLOG(2,1), 0.5026 for the HAR and 0.6174 for the LOG – HAR. Hence, the similarity of the HAR correlation structure with that of realized variance is not necessarily indicative of a good specification. With respect to the latter, consider the limiting case of two independent processes y_t and x_t with identical model specifications and parameter values. By construction, the autocorrelation function of x_t is a perfect match for that of y_t . However, because the two processes are independent x_t , $\mathbb{E}_{t-1}[x_t]$ and x_{t-1} are worst forecasters of y_t than \bar{y} , which, with i.i.d. innovations, has zero autocorrelations. Therefore, while the autocorrelation structure of the *true* model matches that of the data, matching that structure is not a sufficient condition for the identification of a good model.

8 Conclusions

In this paper we advance the idea that out-of-sample forecast evaluation relies on the choice of a distance between predicted outcomes and actual values of realized variance (as well as volatility and log-variance) that reflects subjective preferences. Conditional on a specific OOS *forecasting loss*, we show that the same functional form may not be the most appropriate choice as an IS *estimation criterion* in which distance between actual and fitted values is used to deliver parameter estimates. To this end, we have examined several models inspired by well known GARCH parameterizations alongside HAR specifications and we have handled several combinations of IS *estimation criteria* and OOS *forecasting losses*.

We find the (2,1) parameterizations to be best suited to model the dynamics and forecast realized variances, volatilities and log-variances. In particular, in terms of OOS forecasts: MVAR(2,1) does generally better than HAR; MVOL(2,1) does generally better than LOG – HAR; MLOG does always better than LOG – HAR. Interestingly, our empirical findings point to the presence of a negative lag-2 coefficient in all estimated (2,1) specifications. We interpret it as a dampening agent which, by

limiting to one day *most* of the impact of a shock to variance, induces a mimicking effect of long memory properties that are indistinguishable from those of the HAR and more sustained than those of the LOG – HAR.

With respect to the estimators considered, we find that for all quadratic *forecasting losses*, models estimated using the LNLS *estimation criterion* provide the best OOS forecasts. On the other hand, for the QML *forecasting loss* the QML *estimation criterion* does better than LNLS, although marginally so. Our findings further suggest a judicious approach to model selection which should rely on information criteria unless the superiority of a specific model is substantiated by empirical evidence.

We leave the evaluation of models combining the elements of the HAR family with those of a (p, q) parameterization to future research.

References

- Andersen, T. G., Bollerslev, T. and Diebold, F. X. (2007) Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility, *Review of Economics and Statistics*, **89**, 701–720.
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2008) Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise, *Econometrica*, **76**, 1481–1536.
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2011) Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading, *Journal of Econometrics*, **162**, 149–169.
- Bauwens, L., Galli, F. and Giot, P. (2008) The moments of log-ACD models, *Quantitative and Qualitative Analysis in Social Sciences*, **2**, 1–28.
- Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31**, 307–327.
- Bollerslev, T., Patton, A. J. and Quaedvlieg, R. (2016) Exploiting the errors: A simple approach for improved volatility forecasting, *Journal of Econometrics*, **192**, 1–18.

- Brownlees, C. T., Cipollini, F. and Gallo, G. M. (2012) Multiplicative error models, in *Volatility Models and Their Applications* (Eds.) L. Bauwens, C. Hafner and S. Laurent, Wiley, pp. 223–247.
- Cipollini, F., Engle, R. F. and Gallo, G. M. (2013) Semiparametric vector MEM, *Journal of Applied Econometrics*, **28**, 1067–1086.
- Corsi, F. (2009) A simple approximate long-memory model of realized volatility, *Journal of Financial Econometrics*, **7**, 174–196.
- Engle, R. F. and Gallo, G. M. (2006) A multiple indicators model for volatility using intra-daily data, *Journal of Econometrics*, **131**, 3–27.
- Geweke, J. (1986) Modeling the persistence in conditional variances: A comment, *Econometric Reviews*, **5**, 57–61.
- Gourièroux, C., Monfort, A. and Trognon, A. (1984) Pseudo maximum likelihood methods: Theory, *Econometrica*, **52**, 681–700.
- Hampel, F., Ronchetti, E., Rousseeuw, P. and Stahel, W. (2005) *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York, 2-nd edn.
- Nelson, D. B. (1991) Conditional heteroskedasticity in asset returns: A new approach, *Econometrica*, **59**, 347–370.
- Park, S. and Linton, O. (2012) Realized volatility: Theory and applications, in *Volatility Models and Their Applications* (Eds.) L. Bauwens, C. Hafner and S. Laurent, Wiley, pp. 319–345.
- Taylor, N. and Xu, Y. (2017) The logarithmic vector multiplicative error model: an application to high frequency NYSE stock data, *Quantitative Finance*, **17**, 1021–1035.
- Wintenberger, O. (2013) Continuous invertibility and stable QML estimation of the EGARCH(1,1), *Scandinavian Journal of Statistics*, **40**, 846–867.
- Zakoian, J.-M. (1994) Threshold heteroskedastic models, *Journal of Economic Dynamics and Control*, **18**, 931 – 955.

Table 1:

LS forecasting loss for out-of-sample evaluations: for each of the four *estimation criteria* (LS, SDLS, LNLS and QML), in each row we report the number of instances (and the percentages) when that criterion provides the best performance (by row 28 tickers times 6 periods=168 possible instances). The χ_3^2 column reports p -values under the null hypothesis of a uniform best performance across the four estimation criteria.

Model		LS	SDLS	LNLS	QML	χ_3^2
MVAR	(1,1)	18 10.71%	23 13.69%	115 68.45%	12 7.14%	<0.0001
	(2,1)	22 13.10%	32 19.05%	98 58.33%	16 9.52%	<0.0001
MEXP	(1,1)	20 11.90%	20 11.90%	99 58.93%	29 17.26%	<0.0001
	(2,1)	35 20.83%	29 17.26%	45 26.79%	59 35.12%	0.0071
MVOL	(1,1)	13 7.74%	30 17.86%	113 67.26%	12 7.14%	<0.0001
	(2,1)	18 10.71%	36 21.43%	92 54.76%	22 13.10%	<0.0001
MLOG	(1,1)	14 8.33%	43 25.60%	89 52.98%	22 13.10%	<0.0001
	(2,1)	21 12.50%	46 27.38%	66 39.29%	35 20.83%	<0.0001
HAR		19 11.31%	23 13.69%	103 61.31%	23 13.69%	<0.0001
LOG-HAR		11 6.55%	40 23.81%	93 55.36%	24 14.29%	<0.0001

Table 2:

SDLS forecasting loss function for out-of-sample evaluations: for each of the four *estimation criteria* (LS, SDLS, LNLS and QML), in each row we report the number of instances (and the percentages) when that criterion provides the best performance (by row 28 tickers times 6 periods=168 possible instances). The χ_3^2 column reports p -values under the null hypothesis of a uniform best performance across the four estimation criteria.

Model		LS	SDLS	LNLS	QML	χ_3^2
MVAR	(1,1)	14 8.33%	19 11.31%	134 79.76%	1 0.60%	<0.0001
	(2,1)	6 3.57%	40 23.81%	122 72.62%	0 0.00%	<0.0001
MEXP	(1,1)	12 7.14%	12 7.14%	131 77.98%	13 7.74%	<0.0001
	(2,1)	18 10.71%	30 17.86%	75 44.64%	45 26.79%	<0.0001
MVOL	(1,1)	4 2.38%	31 18.45%	130 77.38%	3 1.79%	<0.0001
	(2,1)	2 1.19%	46 27.38%	118 70.24%	2 1.19%	<0.0001
MLOG	(1,1)	2 1.19%	49 29.17%	116 69.05%	1 0.60%	<0.0001
	(2,1)	5 2.98%	61 36.31%	99 58.93%	3 1.79%	<0.0001
HAR		5 2.98%	30 17.86%	132 78.57%	1 0.60%	<0.0001
LOG-HAR		3 1.79%	45 26.79%	117 69.64%	3 1.79%	<0.0001

Table 3:

LNLS forecasting loss for out-of-sample evaluations: for each of the four *estimation criteria* (LS, SDLS, LNLS and QML), in each row we report the number of instances (and the percentages) when that criterion provides the best performance (by row 28 tickers times 6 periods=168 possible instances). The χ_3^2 column reports p -values under the null hypothesis of a uniform best performance across the four estimation criteria.

Model		LS		SDLS		LNLS		QML		χ_3^2
MVAR	(1,1)	3	1.79%	19	11.31%	146	86.90%	0	0.00%	<0.0001
	(2,1)	1	0.60%	13	7.74%	154	91.67%	0	0.00%	<0.0001
MEXP	(1,1)	0	0.00%	10	5.95%	157	93.45%	1	0.60%	<0.0001
	(2,1)	3	1.79%	24	14.29%	122	72.62%	19	11.31%	<0.0001
MVOL	(1,1)	1	0.60%	18	10.71%	149	88.69%	0	0.00%	<0.0001
	(2,1)	0	0.00%	9	5.36%	159	94.64%	0	0.00%	<0.0001
MLOG	(1,1)	0	0.00%	17	10.12%	151	89.88%	0	0.00%	<0.0001
	(2,1)	0	0.00%	18	10.71%	150	89.29%	0	0.00%	<0.0001
HAR		0	0.00%	15	8.93%	153	91.07%	0	0.00%	<0.0001
LOG-HAR		0	0.00%	15	8.93%	153	91.07%	0	0.00%	<0.0001

Table 4:

QML forecasting loss for out-of-sample evaluations: for each of the four *estimation criteria* (LS, SDLS, LNLS and QML), in each row we report the number of instances (and the percentages) when that criterion provides the best performance (by row 28 tickers times 6 periods=168 possible instances). The χ_3^2 column reports p -values under the null hypothesis of a uniform best performance across the four estimation criteria.

Model		LS		SDLS		LNLS		QML		χ_3^2
MVAR	(1,1)	23	13.69%	45	26.79%	48	28.57%	52	30.95%	0.0036
	(2,1)	27	16.07%	38	22.62%	32	19.05%	71	42.26%	<0.0001
MEXP	(1,1)	18	10.71%	32	19.05%	32	19.05%	86	51.19%	<0.0001
	(2,1)	16	9.52%	26	15.48%	17	10.12%	109	64.88%	<0.0001
MVOL	(1,1)	33	19.64%	49	29.17%	35	20.83%	51	30.36%	0.1008
	(2,1)	39	23.21%	36	21.43%	23	13.69%	70	41.67%	<0.0001
MLOG	(1,1)	39	23.21%	38	22.62%	34	20.24%	57	33.93%	0.0700
	(2,1)	49	29.17%	27	16.07%	26	15.48%	66	39.29%	<0.0001
HAR		25	14.88%	31	18.45%	34	20.24%	78	46.43%	<0.0001
LOG-HAR		49	29.17%	24	14.29%	23	13.69%	72	42.86%	<0.0001

Table 5:

LS forecasting loss for out-of-sample evaluations: for each of the four *estimation criteria* (LS, SDLS, LNLS and QML), in each row we report the average loss (by row 28 tickers times 6 periods=168 possible instances). For every specification, the lowest loss (across *estimation criteria*) is reported in **bold**. The $F_{3,\infty}$ column reports p -values under the null hypothesis of a uniform average performance across the four estimation criteria.

Model		LS	SDLS	LNLS	QML	$F_{3,\infty}$
MVAR	(1,1)	3.2112	3.1890	3.1537	3.3789	<0.0001
	(2,1)	3.1254	3.1060	3.1008	3.2990	<0.0001
MEXP	(1,1)	13.0124	58.6417	35.2893	70.0017	0.3240
	(2,1)	12.7777	81.0910	78.6324	159.0144	0.2989
MVOL	(1,1)	3.0772	2.9607	2.9150	3.0102	<0.0001
	(2,1)	3.0108	2.9170	2.8883	2.9610	<0.0001
MLOG	(1,1)	2.9830	2.8480	2.8407	2.8517	<0.0001
	(2,1)	2.9613	2.8201	2.8241	2.8120	<0.0001
HAR		3.1432	3.1256	3.1085	3.3150	<0.0001
LOG-HAR		3.0755	2.9114	2.8828	2.9347	<0.0001

Table 6:

SDLS forecasting loss for out-of-sample evaluations: for each of the four *estimation criteria* (LS, SDLS, LNLS and QML), in each row we report the average loss (by row 28 tickers times 6 periods=168 possible instances). For every specification, the lowest loss (across *estimation criteria*) is reported in **bold**. The $F_{3,\infty}$ column reports p -values under the null hypothesis of a uniform average performance across the four estimation criteria.

Model		LS	SDLS	LNLS	QML	$F_{3,\infty}$
MVAR	(1,1)	0.1164	0.1045	0.1029	0.1096	<0.0001
	(2,1)	0.1082	0.1003	0.0993	0.1055	<0.0001
MEXP	(1,1)	0.1503	0.1629	0.1552	0.2012	0.1121
	(2,1)	0.1556	0.1710	0.1727	0.2170	0.1778
MVOL	(1,1)	0.1128	0.0992	0.0975	0.1027	<0.0001
	(2,1)	0.1056	0.0961	0.0950	0.0994	<0.0001
MLOG	(1,1)	0.1084	0.0961	0.0950	0.0986	<0.0001
	(2,1)	0.1055	0.0939	0.0933	0.0960	<0.0001
HAR		0.1082	0.1009	0.0997	0.1057	<0.0001
LOG-HAR		0.1077	0.0968	0.0956	0.0997	<0.0001

Table 7:

LNLS forecasting loss for out-of-sample evaluations: for each of the four *estimation criteria* (LS, SDLS, LNLS and QML), in each row we report the average loss (by row 28 tickers times 6 periods=168 possible instances). For every specification, the lowest loss (across *estimation criteria*) is reported in **bold**. The $F_{3,\infty}$ column reports p -values under the null hypothesis of a uniform average performance across the four estimation criteria.

Model		LS	SDLS	LNLS	QML	$F_{3,\infty}$
MVAR	(1,1)	0.2471	0.1955	0.1881	0.2074	<0.0001
	(2,1)	0.2215	0.1880	0.1824	0.2012	<0.0001
MEXP	(1,1)	0.2521	0.2042	0.1961	0.2170	<0.0001
	(2,1)	0.2613	0.2154	0.2075	0.2204	<0.0001
MVOL	(1,1)	0.2351	0.1885	0.1827	0.2004	<0.0001
	(2,1)	0.2145	0.1820	0.1776	0.1943	<0.0001
MLOG	(1,1)	0.2205	0.1847	0.1799	0.1954	<0.0001
	(2,1)	0.2114	0.1796	0.1756	0.1903	<0.0001
HAR		0.2194	0.1885	0.1825	0.2009	<0.0001
LOG-HAR		0.2113	0.1826	0.1786	0.1944	<0.0001

Table 8:

QML forecasting loss for out-of-sample evaluations: for each of the four *estimation criteria* (LS, SDLS, LNLS and QML), in each row we report the average loss (by row 28 tickers times 6 periods=168 possible instances). For every specification, the lowest loss (across *estimation criteria*) is reported in **bold**. The $F_{3,\infty}$ column reports p -values under the null hypothesis of a uniform average performance across the four estimation criteria.

Model		LS	SDLS	LNLS	QML	$F_{3,\infty}$
MVAR	(1,1)	1.1297	1.1184	1.1206	1.1181	<0.0001
	(2,1)	1.1236	1.1177	1.1199	1.1166	<0.0001
MEXP	(1,1)	1.1323	1.1244	1.1257	1.1218	<0.0001
	(2,1)	1.1362	1.1283	1.1290	1.1233	<0.0001
MVOL	(1,1)	1.1246	1.1167	1.1199	1.1165	<0.0001
	(2,1)	1.1198	1.1163	1.1193	1.1152	<0.0001
MLOG	(1,1)	1.1200	1.1164	1.1201	1.1157	<0.0001
	(2,1)	1.1184	1.1160	1.1196	1.1147	<0.0001
HAR		1.1230	1.1179	1.1200	1.1166	<0.0001
LOG-HAR		1.1191	1.1173	1.1207	1.1159	<0.0001

Table 9:

LS forecasting loss and estimation criterion. The first three columns report the percentage a specification is best: IS, in-sample (based on *BIC*); OOS, out-of-sample (LS loss); OOS|IS, best out-of-sample, limiting to the best in-sample. A-Loss and M-Loss are the average and median losses. The BIC row contains out-of-sample results for the best specifications selected in-sample by the information criterion. Individual A-Losses are tested against the corresponding value on the bottom row (a *, **, *** signifies a better performance - 10%, 5%, 1% significance - the reverse for the corresponding *, **, ***).

Model		IS	OOS	OOS IS	A-Loss	M-Loss
MVAR	(1,1)	0.00%	0.00%	0.00%	3.2112	1.1267
	(2,1)	0.00%	16.67%	0.00%	3.1254	1.0458
MEXP	(1,1)	0.00%	0.60%	0.00%	13.0124	1.1750
	(2,1)	1.79%	1.79%	0.00%	12.7777	1.2316
MVOL	(1,1)	0.00%	0.00%	0.00%	3.0772	1.1160
	(2,1)	2.98%	4.17%	0.00%	3.0108	1.0674
MLOG	(1,1)	31.55%	10.12%	18.87%	2.9830	1.0741
	(2,1)	33.93%	36.90%	38.60%	2.9613	1.0544
HAR		4.76%	13.69%	0.00%	3.1432	1.0221
LOG-HAR		25.00%	16.07%	30.95%	3.0755	1.0705
BIC			26.00%		3.2400	1.0731

Table 10:

SDLS forecasting loss and estimation criterion. The first three columns report the percentage a specification is best: IS, in-sample (based on *BIC*); OOS, out-of-sample (SDLS loss); OOS|IS, best out-of-sample, limiting to the best in-sample. A-Loss and M-Loss are the average and median losses. The BIC row contains out-of-sample results for the best specifications selected in-sample by the information criterion. Individual A-Losses are tested against the corresponding value on the bottom row (a *, **, *** signifies a better performance - 10%, 5%, 1% significance - the reverse for the corresponding *, **, ***).

Model		IS	OOS	OOS IS	A-Loss	M-Loss
MVAR	(1,1)	0.00%	0.00%	0.00%	0.1045***	0.0770
	(2,1)	0.60%	7.74%	0.00%	0.1003***	0.0756
MEXP	(1,1)	0.00%	2.38%	0.00%	0.1629*	0.0818
	(2,1)	0.00%	0.60%	0.00%	0.1710*	0.0880
MVOL	(1,1)	0.00%	0.00%	0.00%	0.0992***	0.0753
	(2,1)	11.90%	7.14%	20.00%	0.0961	0.0743
MLOG	(1,1)	7.74%	10.12%	15.38%	0.0961	0.0748
	(2,1)	51.19%	54.76%	47.67%	0.0939***	0.0736
HAR		4.17%	8.93%	0.00%	0.1009***	0.0751
LOG-HAR		24.40%	8.33%	9.76%	0.0968*	0.0744
BIC			30.00%		0.0959	0.0736

Table 11:

LNLS forecasting loss and estimation criterion. The first three columns report the percentage a specification is best: IS, in-sample (based on *BIC*); OOS, out-of-sample (LNLS loss); OOS|IS, best out-of-sample, limiting to the best in-sample. A-Loss and M-Loss are the average and median losses. The BIC row contains out-of-sample results for the best specifications selected in-sample by the information criterion. Individual A-Losses are tested against the corresponding value on the bottom row (a *, **, *** signifies a better performance - 10%, 5%, 1% significance - the reverse for the corresponding *, **, ***).

Model		IS	OOS	OOS IS	A-Loss	M-Loss
MVAR	(1,1)	0.00%	0.00%	0.00%	0.1881***	0.1809
	(2,1)	0.00%	5.36%	0.00%	0.1824***	0.1759
MEXP	(1,1)	0.00%	0.00%	0.00%	0.1961***	0.1808
	(2,1)	0.00%	1.19%	0.00%	0.2075***	0.1966
MVOL	(1,1)	0.00%	0.60%	0.00%	0.1827***	0.1770
	(2,1)	14.29%	6.55%	8.33%	0.1776***	0.1724
MLOG	(1,1)	1.79%	10.12%	0.00%	0.1799***	0.1741
	(2,1)	68.45%	54.76%	51.30%	0.1756***	0.1699
HAR		2.38%	10.71%	0.00%	0.1825***	0.1754
LOG-HAR		13.10%	10.71%	18.18%	0.1786***	0.1739
BIC			38.00%		0.1765	0.1705

Table 12:

QML forecasting loss and estimation criterion. The first three columns report the percentage a specification is best: IS, in-sample (based on *BIC*); OOS, out-of-sample (QML loss); OOS|IS, best out-of-sample, limiting to the best in-sample. A-Loss and M-Loss are the average and median losses. The BIC row contains out-of-sample results for the best specifications selected in-sample by the information criterion. Individual A-Losses are tested against the corresponding value on the bottom row (a *, **, *** signifies a better performance - 10%, 5%, 1% significance - the reverse for the corresponding *, **, ***).

Model		IS	OOS	OOS IS	A-Loss	M-Loss
MVAR	(1,1)	8.33%	2.38%	0.00%	1.1181***	1.0965
	(2,1)	0.00%	6.55%	0.00%	1.1166	1.0945
MEXP	(1,1)	0.00%	1.19%	0.00%	1.1218***	1.0991
	(2,1)	0.00%	1.19%	0.00%	1.1233***	1.0995
MVOL	(1,1)	17.86%	1.19%	0.00%	1.1165	1.0939
	(2,1)	0.00%	5.36%	0.00%	1.1152***	1.0917
MLOG	(1,1)	30.36%	17.26%	3.92%	1.1157***	1.0933
	(2,1)	0.00%	36.90%	0.00%	1.1147***	1.0908
HAR		12.50%	21.43%	9.52%	1.1166	1.0946
LOG-HAR		30.95%	6.55%	3.85%	1.1159**	1.0917
BIC			3.00%		1.1166	1.0951

Figure 1: Parameters' density estimates over the six IS periods and all the assets considered. From left to right, the columns contain the densities of the parameters α_1 , α_2 and β_1 for the three specifications reported.

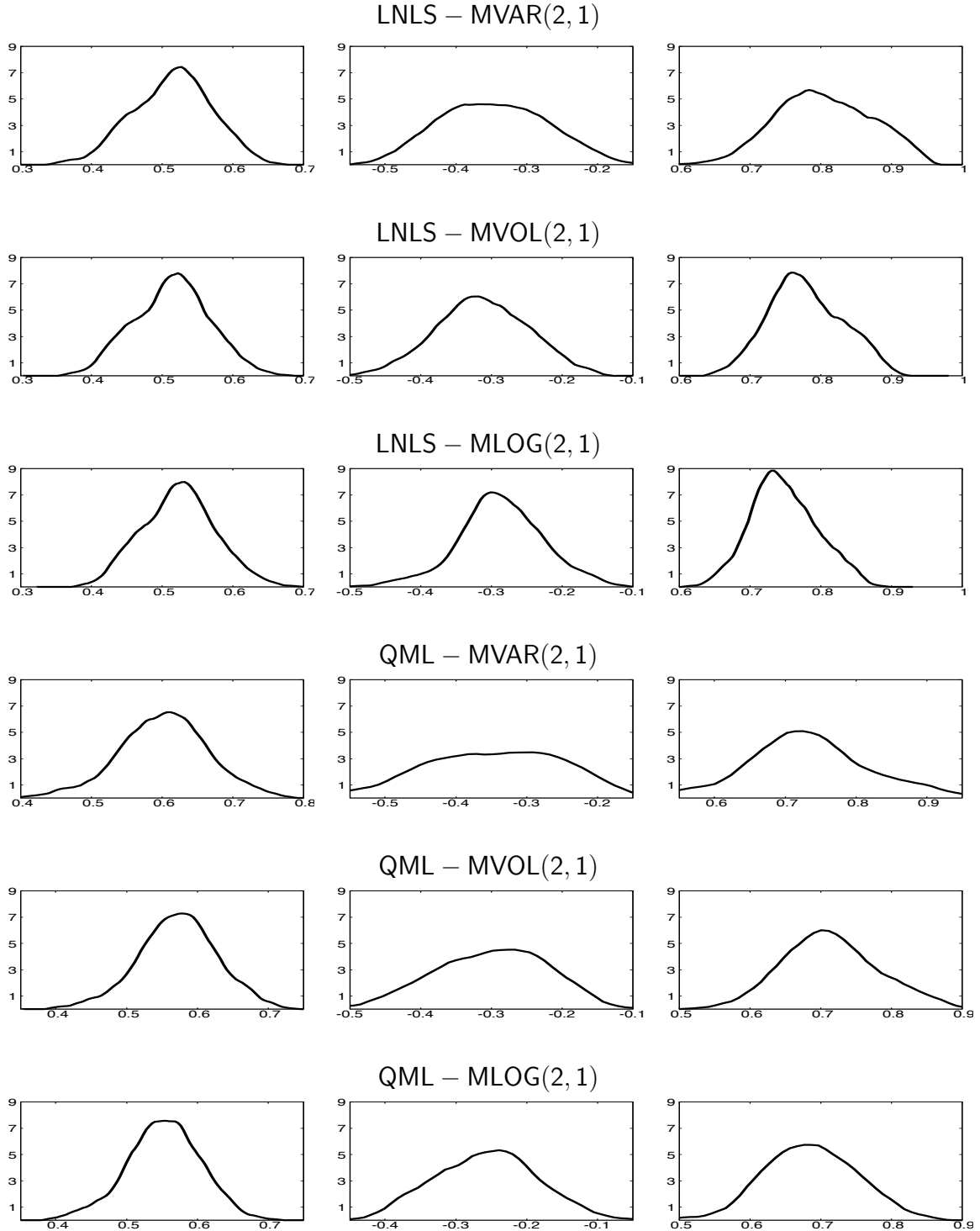


Figure 2: Autocorrelation functions for the variance processes with parameters equal to the cross-sectional averages of the LNLS estimates in the most recent subsample (2010-2014). MVAR, MVOL and MLOG specifications refer to the (2,1) parameterizations. Autocorrelations are computed over 10^6 simulations with additive standardized Gaussian innovations.

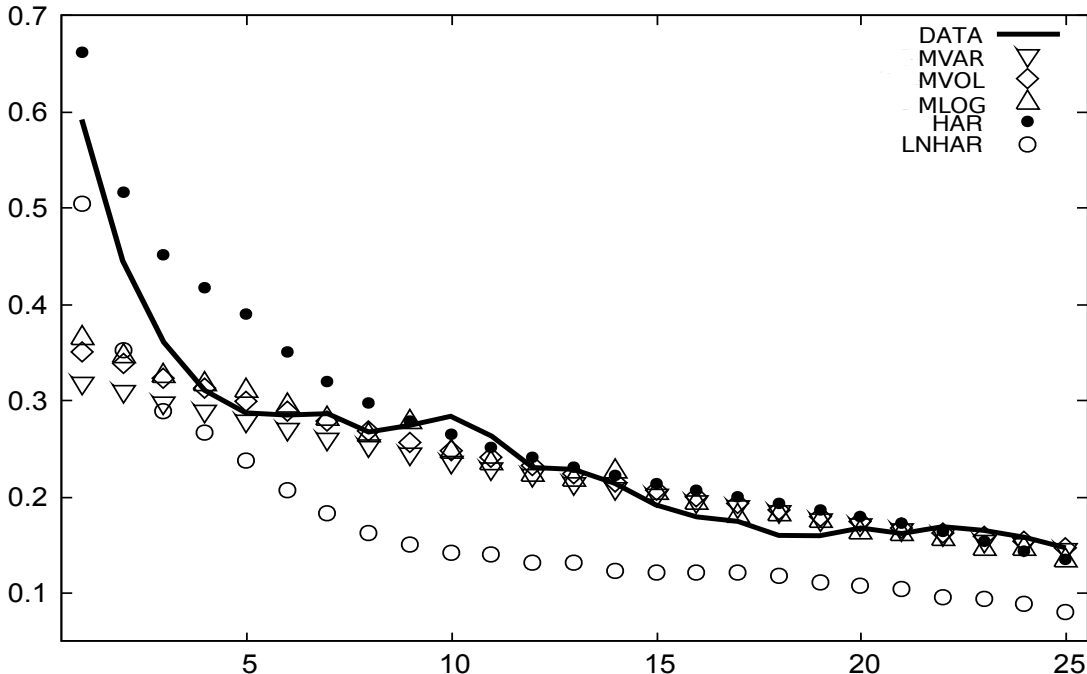


Table 13:

Tantamount to Table 1 but with no Winsorization of IS data. *LS forecasting loss* for out-of-sample evaluations: for each of the four *estimation criteria* (LS, SDLS, LNLS and QML), in each row we report the number of instances (and the percentages) when that criterion provides the best performance (by row 28 tickers times 6 periods=168 possible instances). The χ_3^2 column reports *p*-values under the null hypothesis of a uniform best performance across the four estimation criteria.

Model		LS		SDLS		LNLS		QML		χ_3^2
MVAR	(1,1)	26	15.48%	34	20.24%	95	56.55%	13	7.74%	<0.0001
	(2,1)	21	12.50%	43	25.60%	81	48.21%	23	13.69%	<0.0001
MEXP	(1,1)	31	18.45%	26	15.48%	81	48.21%	30	17.86%	<0.0001
	(2,1)	33	19.64%	43	25.60%	32	19.05%	60	35.71%	0.0092
MVOL	(1,1)	13	7.74%	29	17.26%	105	62.50%	21	12.50%	<0.0001
	(2,1)	14	8.33%	39	23.21%	85	50.60%	30	17.86%	<0.0001
MLOG	(1,1)	10	5.95%	37	22.02%	94	55.95%	27	16.07%	<0.0001
	(2,1)	15	8.93%	35	20.83%	77	45.83%	41	24.40%	<0.0001
HAR		24	14.29%	39	23.21%	78	46.43%	27	16.07%	<0.0001
LOG-HAR		10	5.95%	33	19.64%	91	54.17%	34	20.24%	<0.0001

Table 14:

Tantamount to Table 2 but with no Winsorization of IS data. *SDLS forecasting loss* for out-of-sample evaluations: for each of the four *estimation criteria* (LS, SDLS, LNLS and QML), in each row we report the number of instances (and the percentages) when that criterion provides the best performance (by row 28 tickers times 6 periods=168 possible instances). The χ_3^2 column reports *p*-values under the null hypothesis of a uniform best performance across the four estimation criteria.

Model		LS		SDLS		LNLS		QML		χ_3^2
MVAR	(1,1)	13	7.74%	23	13.69%	131	77.98%	1	0.60%	<0.0001
	(2,1)	5	2.98%	39	23.21%	124	73.81%	0	0.00%	<0.0001
MEXP	(1,1)	16	9.52%	24	14.29%	114	67.86%	14	8.33%	<0.0001
	(2,1)	20	11.90%	47	27.98%	56	33.33%	45	26.79%	0.0002
MVOL	(1,1)	3	1.79%	18	10.71%	143	85.12%	4	2.38%	<0.0001
	(2,1)	1	0.60%	30	17.86%	134	79.76%	3	1.79%	<0.0001
MLOG	(1,1)	2	1.19%	30	17.86%	132	78.57%	4	2.38%	<0.0001
	(2,1)	4	2.38%	45	26.79%	111	66.07%	8	4.76%	<0.0001
HAR		4	2.38%	37	22.02%	126	75.00%	1	0.60%	<0.0001
LOG-HAR		2	1.19%	26	15.48%	131	77.98%	9	5.36%	<0.0001

Table 15:

Tantamount to Table 3 but with no Winsorization of IS data. *LNLS forecasting loss* for out-of-sample evaluations: for each of the four *estimation criteria* (LS, SDLS, LNLS and QML), in each row we report the number of instances (and the percentages) when that criterion provides the best performance (by row 28 tickers times 6 periods=168 possible instances). The χ_3^2 column reports *p*-values under the null hypothesis of a uniform best performance across the four estimation criteria.

Model		LS		SDLS		LNLS		QML		χ_3^2
MVAR	(1,1)	3	1.79%	17	10.12%	148	88.10%	0	0.00%	<0.0001
	(2,1)	1	0.60%	12	7.14%	155	92.26%	0	0.00%	<0.0001
MEXP	(1,1)	0	0.00%	15	8.93%	152	90.48%	1	0.60%	<0.0001
	(2,1)	3	1.79%	28	16.67%	118	70.24%	19	11.31%	<0.0001
MVOL	(1,1)	1	0.60%	16	9.52%	151	89.88%	0	0.00%	<0.0001
	(2,1)	0	0.00%	8	4.76%	160	95.24%	0	0.00%	<0.0001
MLOG	(1,1)	0	0.00%	14	8.33%	154	91.67%	0	0.00%	<0.0001
	(2,1)	0	0.00%	16	9.52%	152	90.48%	0	0.00%	<0.0001
HAR		0	0.00%	15	8.93%	153	91.07%	0	0.00%	<0.0001
LOG-HAR		0	0.00%	12	7.14%	156	92.86%	0	0.00%	<0.0001

Table 16:

Tantamount to Table 4 but with no Winsorization of IS data. *QML forecasting loss* for out-of-sample evaluations: for each of the four *estimation criteria* (LS, SDLS, LNLS and QML), in each row we report the number of instances (and the percentages) when that criterion provides the best performance (by row 28 tickers times 6 periods=168 possible instances). The χ_3^2 column reports *p*-values under the null hypothesis of a uniform best performance across the four estimation criteria.

Model		LS		SDLS		LNLS		QML		χ_3^2
MVAR	(1,1)	13	7.74%	33	19.64%	49	29.17%	73	43.45%	<0.0001
	(2,1)	16	9.52%	28	16.67%	39	23.21%	85	50.60%	<0.0001
MEXP	(1,1)	14	8.33%	36	21.43%	28	16.67%	90	53.57%	<0.0001
	(2,1)	11	6.55%	29	17.26%	19	11.31%	109	64.88%	<0.0001
MVOL	(1,1)	21	12.50%	42	25.00%	39	23.21%	66	39.29%	<0.0001
	(2,1)	25	14.88%	34	20.24%	30	17.86%	79	47.02%	<0.0001
MLOG	(1,1)	25	14.88%	52	30.95%	24	14.29%	67	39.88%	<0.0001
	(2,1)	30	17.86%	44	26.19%	22	13.10%	72	42.86%	<0.0001
HAR		13	7.74%	28	16.67%	39	23.21%	88	52.38%	<0.0001
LOG-HAR		26	15.48%	50	29.76%	18	10.71%	74	44.05%	<0.0001

Table 17:

Tantamount to Table 5 but with no Winsorization of IS data. *LS forecasting loss* for out-of-sample evaluations: for each of the four *estimation criteria* (LS, SDLS, LNLS and QML), in each row we report the average loss (by row 28 tickers times 6 periods=168 possible instances). For every specification, the lowest loss (across *estimation criteria*) is reported in **bold**. The $F_{3,\infty}$ column reports p -values under the null hypothesis of a uniform average performance across the four estimation criteria.

Model		LS	SDLS	LNLS	QML	$F_{3,\infty}$
MVAR	(1,1)	3.2963	3.1822	3.1609	3.4345	<0.0001
	(2,1)	3.2135	3.1046	3.1072	3.3609	<0.0001
MEXP	(1,1)	7.2577	11.0253	32.8010	116.4256	0.0764
	(2,1)	9.1616	17.8734	70.8608	15825.0020	0.3728
MVOL	(1,1)	3.2137	2.9635	2.9166	3.0667	<0.0001
	(2,1)	3.1330	2.9307	2.8909	3.0208	<0.0001
MLOG	(1,1)	3.1247	2.8597	2.8423	2.9809	0.0020
	(2,1)	3.1610	2.8436	2.8259	2.9072	<0.0001
HAR		3.2463	3.1192	3.1141	3.3735	<0.0001
LOG-HAR		3.4370	2.9193	2.8830	3.3062	0.0321

Table 18:

Tantamount to Table 6 but with no Winsorization of IS data. *SDLS forecasting loss* for out-of-sample evaluations: for each of the four *estimation criteria* (LS, SDLS, LNLS and QML), in each row we report the average loss (by row 28 tickers times 6 periods=168 possible instances). For every specification, the lowest loss (across *estimation criteria*) is reported in **bold**. The $F_{3,\infty}$ column reports p -values under the null hypothesis of a uniform average performance across the four estimation criteria.

Model		LS	SDLS	LNLS	QML	$F_{3,\infty}$
MVAR	(1,1)	0.1342	0.1056	0.1030	0.1110	<0.0001
	(2,1)	0.1246	0.1016	0.0993	0.1073	<0.0001
MEXP	(1,1)	0.1547	0.1316	0.1527	0.2303	0.0026
	(2,1)	0.1686	0.1408	0.1701	9.0428	0.3789
MVOL	(1,1)	0.1293	0.0996	0.0975	0.1042	<0.0001
	(2,1)	0.1211	0.0972	0.0951	0.1013	<0.0001
MLOG	(1,1)	0.1209	0.0965	0.0950	0.1009	<0.0001
	(2,1)	0.1222	0.0949	0.0933	0.0984	<0.0001
HAR		0.1256	0.1022	0.0998	0.1073	<0.0001
LOG-HAR		0.1344	0.0974	0.0955	0.1036	<0.0001

Table 19:

Tantamount to Table 7 but with no Winsorization of IS data. LNLs forecasting loss for out-of-sample evaluations: for each of the four estimation criteria (LS, SDLS, LNLs and QML), in each row we report the average loss (by row 28 tickers times 6 periods=168 possible instances). For every specification, the lowest loss (across estimation criteria) is reported in **bold**. The $F_{3,\infty}$ column reports p -values under the null hypothesis of a uniform average performance across the four estimation criteria.

Model		LS	SDLS	LNLs	QML	$F_{3,\infty}$
MVAR	(1,1)	0.3113	0.2031	0.1883	0.2105	<0.0001
	(2,1)	0.2891	0.1948	0.1827	0.2052	<0.0001
MEXP	(1,1)	0.3067	0.2079	0.1976	0.2228	<0.0001
	(2,1)	0.3247	0.2240	0.2140	0.6067	0.3283
MVOL	(1,1)	0.2842	0.1905	0.1827	0.2036	<0.0001
	(2,1)	0.2663	0.1850	0.1776	0.1985	<0.0001
MLOG	(1,1)	0.2559	0.1850	0.1798	0.1990	<0.0001
	(2,1)	0.2559	0.1812	0.1756	0.1945	<0.0001
HAR		0.2914	0.1961	0.1829	0.2048	<0.0001
LOG-HAR		0.2889	0.1857	0.1785	0.1997	<0.0001

Table 20:

Tantamount to Table 8 but with no Winsorization of IS data. QML forecasting loss for out-of-sample evaluations: for each of the four estimation criteria (LS, SDLS, LNLs and QML), in each row we report the average loss (by row 28 tickers times 6 periods=168 possible instances). For every specification, the lowest loss (across estimation criteria) is reported in **bold**. The $F_{3,\infty}$ column reports p -values under the null hypothesis of a uniform average performance across the four estimation criteria.

Model		LS	SDLS	LNLs	QML	$F_{3,\infty}$
MVAR	(1,1)	1.1520	1.1226	1.1206	1.1187	<0.0001
	(2,1)	1.1461	1.1211	1.1199	1.1175	<0.0001
MEXP	(1,1)	1.1502	1.1270	1.1275	1.1236	<0.0001
	(2,1)	1.1565	1.1321	1.1319	1.1659	0.3698
MVOL	(1,1)	1.1408	1.1181	1.1199	1.1172	<0.0001
	(2,1)	1.1363	1.1176	1.1193	1.1161	<0.0001
MLOG	(1,1)	1.1310	1.1164	1.1201	1.1165	<0.0001
	(2,1)	1.1319	1.1162	1.1197	1.1156	<0.0001
HAR		1.1468	1.1213	1.1200	1.1173	<0.0001
LOG-HAR		1.1452	1.1174	1.1205	1.1172	<0.0001

Table 21:

Tantamount to Table 9 but with no Winsorization of IS data. *LS forecasting loss and estimation criterion*. The first three columns report the percentage a specification is best: IS, in-sample (based on *BIC*); OOS, out-of-sample (*LS* loss); OOS|IS, best out-of-sample, limiting to the best in-sample. A-Loss and M-Loss are the average and median losses. The BIC row contains out-of-sample results for the best specifications selected in-sample by the information criterion. Individual A-Losses are tested against the corresponding value on the bottom row (a *, **, ***, *** signifies a better performance - 10%, 5%, 1% significance - the reverse for the corresponding *, **, ***).

Model		IS	OOS	OOS IS	A-Loss	M-Loss
MVAR	(1,1)	0.00%	1.79%	0.00%	3.2963	1.2859
	(2,1)	0.00%	7.74%	0.00%	3.2135	1.1525
MEXP	(1,1)	0.00%	0.60%	0.00%	7.2577	1.2632
	(2,1)	0.60%	0.00%	0.00%	9.1616	1.3764
MVOL	(1,1)	0.00%	1.19%	0.00%	3.2137	1.1717
	(2,1)	0.00%	9.52%	0.00%	3.1330*	1.1243
MLOG	(1,1)	44.05%	32.74%	31.08%	3.1247*	1.1647
	(2,1)	39.29%	27.38%	16.67%	3.1610	1.1457
HAR		4.76%	7.74%	0.00%	3.2463	1.1433
LOG-HAR		11.31%	11.31%	42.11%	3.4370	1.1624
BIC			25.00%		3.4519	1.1554

Table 22:

Tantamount to Table 10 but with no Winsorization of IS data. *SDLS forecasting loss and estimation criterion*. The first three columns report the percentage a specification is best: IS, in-sample (based on *BIC*); OOS, out-of-sample (*SDLS* loss); OOS|IS, best out-of-sample, limiting to the best in-sample. A-Loss and M-Loss are the average and median losses. The BIC row contains out-of-sample results for the best specifications selected in-sample by the information criterion. Individual A-Losses are tested against the corresponding value on the bottom row (a *, **, ***, *** signifies a better performance - 10%, 5%, 1% significance - the reverse for the corresponding *, **, ***).

Model		IS	OOS	OOS IS	A-Loss	M-Loss
MVAR	(1,1)	0.00%	0.00%	0.00%	0.1056***	0.0817
	(2,1)	0.00%	5.36%	0.00%	0.1016***	0.0776
MEXP	(1,1)	0.00%	1.19%	0.00%	0.1316***	0.0878
	(2,1)	0.00%	0.00%	0.00%	0.1408***	0.0953
MVOL	(1,1)	0.00%	0.00%	0.00%	0.0996***	0.0775
	(2,1)	5.95%	5.95%	0.00%	0.0972	0.0759
MLOG	(1,1)	42.26%	11.90%	9.86%	0.0965	0.0757
	(2,1)	30.95%	50.60%	55.77%	0.0949***	0.0752
HAR		4.17%	10.71%	0.00%	0.1022***	0.0775
LOG-HAR		16.67%	14.29%	7.14%	0.0974	0.0752
BIC			22.00%		0.0966	0.0757

Table 23:

Tantamount to Table 11 but with no Winsorization of IS data. *LNLS forecasting loss and estimation criterion*. The first three columns report the percentage a specification is best: IS, in-sample (based on *BIC*); OOS, out-of-sample (LNLS loss); OOS|IS, best out-of-sample, limiting to the best in-sample. A-Loss and M-Loss are the average and median losses. The BIC row contains out-of-sample results for the best specifications selected in-sample by the information criterion. Individual A-Losses are tested against the corresponding value on the bottom row (a *, **, ***, *** signifies a better performance - 10%, 5%, 1% significance - the reverse for the corresponding *, **, ***).

Model		IS	OOS	OOS IS	A-Loss	M-Loss
MVAR	(1,1)	0.00%	0.00%	0.00%	0.1883***	0.1815
	(2,1)	0.00%	6.55%	0.00%	0.1827***	0.1761
MEXP	(1,1)	0.00%	0.00%	0.00%	0.1976***	0.1845
	(2,1)	0.00%	1.79%	0.00%	0.2140***	0.1974
MVOL	(1,1)	0.00%	0.60%	0.00%	0.1827***	0.1770
	(2,1)	7.74%	5.36%	0.00%	0.1776***	0.1724
MLOG	(1,1)	1.79%	10.12%	0.00%	0.1798***	0.1742
	(2,1)	75.00%	54.76%	51.59%	0.1756***	0.1699
HAR		1.19%	8.33%	0.00%	0.1829***	0.1752
LOG-HAR		14.29%	12.50%	8.33%	0.1785***	0.1739
BIC			39.00%		0.1763	0.1709

Table 24:

Tantamount to Table 12 but with no Winsorization of IS data. *QML forecasting loss and estimation criterion*. The first three columns report the percentage a specification is best: IS, in-sample (based on *BIC*); OOS, out-of-sample (QML loss); OOS|IS, best out-of-sample, limiting to the best in-sample. A-Loss and M-Loss are the average and median losses. The BIC row contains out-of-sample results for the best specifications selected in-sample by the information criterion. Individual A-Losses are tested against the corresponding value on the bottom row (a *, **, ***, *** signifies a better performance - 10%, 5%, 1% significance - the reverse for the corresponding *, **, ***).

Model		IS	OOS	OOS IS	A-Loss	M-Loss
MVAR	(1,1)	6.55%	2.38%	0.00%	1.1187***	1.0961
	(2,1)	0.00%	6.55%	0.00%	1.1175	1.0942
MEXP	(1,1)	0.00%	1.19%	0.00%	1.1236***	1.0993
	(2,1)	0.00%	1.19%	0.00%	1.1659	1.1048
MVOL	(1,1)	16.07%	1.79%	0.00%	1.1172	1.0939
	(2,1)	0.00%	4.76%	0.00%	1.1161***	1.0915
MLOG	(1,1)	32.74%	17.86%	7.27%	1.1165***	1.0933
	(2,1)	2.38%	34.52%	0.00%	1.1156***	1.0911
HAR		13.10%	22.62%	9.09%	1.1173	1.0942
LOG-HAR		29.17%	7.14%	4.08%	1.1172	1.0916
BIC			4.00%		1.1176	1.0933

