2013s-19

Forecasting financial volatility with combined QML and LAD-ARCH estimators of the GARCH model

Liam Cheung, John W. Galbraith

Série Scientifique
Scientific Series

Montréal
Juillet 2013

CIRANO
Allier savoir et décision

Centre interuniversitaire de recherche en analyse des organisations

ISSN 1198-8177

Partenaire financier

Enseignement supérieur, Recherche, Science et Technologie
Québec

Forecasting financial volatility with combined QML and LAD-ARCH estimators of the GARCH model *

Liam Cheung[†], *John W. Galbraith*[‡]

Résumé / *Abstract*

GARCH models and their variants are usually estimated using quasi-Maximum Likelihood (QML). Recent work has shown that by using estimates of quadratic variation, for example from the daily realized volatility, it is possible to estimate these models in a different way which incorporates the additional information. Theory suggests that as the precision of estimates of daily quadratic variation improves, such estimates (via LAD- ARCH approximation) should come to equal and eventually dominate the QML estimators. The present paper investigates this using a five-year sample of data on returns from all 466 S&P 500 stocks which were present in the index continuously throughout the period. The results suggest that LAD-ARCH estimates, using realized volatility on five-minute returns over the trading day, yield measures of 1-step forecast accuracy comparable or slightly superior to those obtained from QML estimates. Combining the two estimators, either by equal weighting or weighting based on cross-validation, appears to produce a clear improvement in forecast accuracy relative to either of the two different forecasting methods alone.

Mots clés/Keywords : QML and LAD-ARCH estimators, GARCH models.

* The support of the *Fonds québécois de la recherche sur la société et la culture* (FQRSC) is gratefully acknowledged. We also thank CIRANO (*Centre Interuniversitaire de recherche en analyse des organisations*) for providing research facilities. Data were provided by Tick Data (www.tickdata.com).

[†] Department of Economics, McGill University, Montreal, Quebec, Canada.

[‡] Department of Economics, McGill University, Montreal, Quebec, Canada, john.galbraith@mcgill.ca

1 Introduction

The GARCH model (Bollerslev 1986) and its numerous variants remain the most commonly used models for forecasting financial volatility. The model is generally estimated by quasi-Maximum Likelihood (QML), with volatility treated as a latent variable. With the wide availability of intraday return data, however, alternative forecasting methods have become available. Using intraday return data to estimate the daily variance (quadratic variation), for example via the daily realized variance or other estimate such as one of those proposed by Zhang et al. (2005), it is possible for example to forecast daily volatility using a simple autoregression on the daily volatility estimates. Another alternative, suggested by Galbraith, Zinde-Walsh and Zhu (2012), hereafter 'GZZ', is to use daily quadratic variation estimates to estimate a GARCH model via ARCH approximation; the resulting LAD-ARCH estimator is to some degree robust to estimation error in the estimate of daily quadratic variation (e.g., realized variance) which it uses. This method has a number of antecedents in the standard conditional mean time series literature, but has not previously been applied to conditional second moment models.

The present paper investigates the relative forecasting performance of each of these methods on a large sample of financial return data. The sample comprises the entire set of S&P 500 index stocks over a period of approximately five years; we retain for analysis the 466 stocks that remain in the index continuously throughout the period. As well, because it is well known that the combination of forecasts from different methods can often yield better results than any of the individual methods alone, we investigate combinations of forecasts based on both GARCH model estimators.

The results suggest that the new class of estimates based on intraday data does clearly have empirical value. Previous theory and simulation imply that as the quality of intraday volatility information improves, the relative performance of those estimators which exploit it should improve also, eventually coming to dominate QML. In fact the number of five-minute returns available per day in a U.S. financial return data set (i.e., 78 per day) appears, given the simulations in GZZ, to be close to the borderline at which forecasts using the LAD-ARCH estimator should come to approximately equal (and later surpass) the performance of QML, for parameter values typical in models of equity returns (i.e. a sum of GARCH parameters near unity, with the coefficient on lagged volatility $\simeq 0.9$).[1]

This result is borne out on this large set of empirical examples. The expectation that combining these two estimators should also yield better measures of forecast accuracy is also borne out in these 466 empirical examples.

The next section gives a brief description of the forecasting methods that we will use and the data available. Section 3 reports the results of the pseudo-out-of-sample forecasting exercise on the set of individual stocks. A final section concludes.

[1] GZZ find that the number of intra-day observations at which the LAD-ARCH estimator comes to dominate is lower for GARCH parameters farther from unity, so that in such cases we should see clearly better performance than for QML at 78 per day; however in US equity price data, with GARCH parameter in the neighbourhood of 0.0, we expect to observe similar performance of the two methods.

2 Estimators, forecasts and data

2.1 GARCH model estimators and forecasts

We compare forecasts of daily conditional variance. The primary object of interest is the daily integrated variance on day t, that is, $\sigma_t^2 = \int_{t-1}^{t} \sigma_s^2 ds$. An estimate of this quantity is is given by $\hat{\sigma}_t^2 = \sum_{j=(t-1)h+1}^{th} r_j^2$, with $r_j^2 = (p_j - p_{j-1})^2$, p_j indicating discretely-sampled intra-day observations on a diffusion process $\{p_t\}$ such that $p_t = p_0 + \int_0^t \sigma_s dW_s$, where $\{W_s\}$ is a Brownian motion process and σ_s^2 is the instantaneous conditional variance. The estimated quantity $\hat{\sigma}_t^2$ is known as the realized variance, or realized volatility (we will use the former term), and its probability limit as the time interval between observations decreases is, in the absence of any measurement noise, the quadratic variation.

The standard ARCH and GARCH models (Engle 1982; Bollerslev 1986) specify the daily return as $\varepsilon_t = z_t \sigma_t$, where the specification of σ_t^2 for the ARCH model is

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2, \tag{1}$$

and for the GARCH model:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2, \tag{2}$$

where ε_t is defined as $r_t - \mu_t$ for a process with conditional mean μ_t, or $\varepsilon_t = r_t$ if there is no drift. The standardized process $z_t = \varepsilon_t / \sigma_t$ is taken to follow a distribution such as the Normal or $t-$. Estimation is generally by quasi- Maximum Likelihood.

It is well known that the GARCH model can be re-written by recursive substitution to yield $\sigma_t^2 = \kappa + \sum_{\ell=1}^{\infty} \nu_\ell \varepsilon_{t-\ell}^2$. Expressions for κ and ν_ℓ are given in GZZ, where this representation is exploited for an alternative form of estimator in which the realized variance, or other estimator of the daily quadratic variation, as equal to the true daily integrated variance plus an error. In truncated finite-sample form the model becomes

$$\hat{\sigma}_t^2 = \kappa + \Sigma_{l=1}^{k} \nu_l \varepsilon_{t-l}^2 \; + u_t, \tag{3}$$

where the error u_t embodies both measurement error and truncation error. Using general results on quantile estimation of truncations of infinite-order processes from Zernov et al. 2009, GZZ show that consistent and asymptotically normal quantile (including LAD) estimation of the parameters of this truncated ARCH representation is possible. Correspondingly, consistent and asymptotically normal estimates of the GARCH parameters may be obtained by minimum distance using the parameters of the estimated ARCH representation, and the known relationship between the parameters of the GARCH model and the infinite-order ARCH approximation. Closed form expressions for these GARCH parameter estimates are given by GZZ; for the GARCH(1,1) case considered here these reduce to

$$\hat{\alpha} = \hat{\nu}_1, \quad \hat{\beta} = \left(\sum_{i=1}^{k'-1} \hat{\nu}_i \hat{\nu}_{i+1} \right) \left(\sum_{i=1}^{k'-1} \hat{\nu}_i^2 \right)^{-1}, \quad \hat{\omega} = \hat{\kappa}(1 - \hat{\beta}), \tag{4}$$

where $k' \leq k$ is the number of ARCH-approximation terms included in the computation.[2]

Combinations of forecasts generated by different procedures often outperform those obtained from the individual procedures. We include in the comparisons below two classes of combination: an equal weighting of the two forecasts, and a sample-based weighting obtained by cross-validation, updated at each sample point, for each one of the 466 equity issues included in the comparison.[3] Both the QML and LAD-ARCH estimators require selection of some choice parameters: the distribution of standardized errors to be used as a likelihood function in the first case, the truncation order for the ARCH approximation in the case of LAD-ARCH.[4] We examine sensitivity to these choices.

Finally, we emphasize that the LAD-ARCH estimator is an estimator of a GARCH model; '-ARCH' refers to estimation of that model via ARCH approximation. The results below therefore are largely devoted to exploring whether combining high-frequency information with low frequency, by this method, can produce gains in short-term volatility forecasting.

2.2 Data

The study uses a sample of intra-day data on S&P 500 securities beginning on May 1, 2006 and ending on September 16, 2011, a total of 1,357 trading days covering periods of both high volatility and relative calm. The sample contains only securities that were components of the S&P 500 for the entire period, or 466 companies. The data are obtained from intraday one-minute equity data (OMED) provided by Tick Data; OMED trade data for each one-minute interval include date, time, open, high, low, close, and volume. Tick Data provides cleaned observations that are fully adjusted for all corporate actions (such as splits and consolidations), changes in symbols, and erroneous trades that were subsequently reversed by the market regulators. [5] Compustat is used to identify the starting date and ending date for each constituent of the S&P 500

[2]We also compared with with the pure autoregressive forecasting model on the realized volatilities. In the results of Andersen et al. (2003), and in our own empirical comparisons, the AR model performs best on the logarithm of (the square root of) the realized variances, that is, as a model of $\ln(\hat{\sigma})$. However we found that the AR-based estimates were outperformed on these data by the GARCH-based forecasts, and we concentrate on the latter in the results reported below.

[3]Because of the autocorrelations in this time-series context, we use only sample data preceding the date to be forecast: that is, we obtain loss function measures for each of a set of different weights on the two models chosen from a grid of values on the [0, 1] interval, evaluating loss function measures up to time $t-1$. The weighting which provides the best forecasts on data up to $t-1$ is used to weight the two input forecasts at time t. The procedure is then repeated to re-evaluate the weights at t and at each new sample point thereafter. The cross-validated results therefore come from a sequence of weights which typically change over the interval, rather than from a fixed pair of weights.

[4]QML estimates are computed using the algorithm contained in Matlab; the LAD estimates use the quantile regression code for Matlab written by Roger Koenker.

[5]To verify data quality, we manually examined all periods where returns were over 10% and compared with historical bar quotes from Yahoo. Using this method we identified two cases where the absence of recorded trades in the first 15 minutes of trading caused the previous day's price to be used when there was a stock split. These data points were manually corrected.

from 1/1/2000 through 9/16/2011. From the list of all companies in the S&P index, we identify those for which Tick Data has continuous intraday bars from to 5/1/2006 through 9/16/2011, and this filtering leads to our subsample of 466 S&P 500 stocks that were traded continuously during the study period. Using Tick Data's *TickWrite* software, the 5-minute and daily price-fluctuation bars are extracted for each of the 466 stocks from 9:30 AM to 4:00 PM, retaining the previous value in any case where there is no trade in a given period.

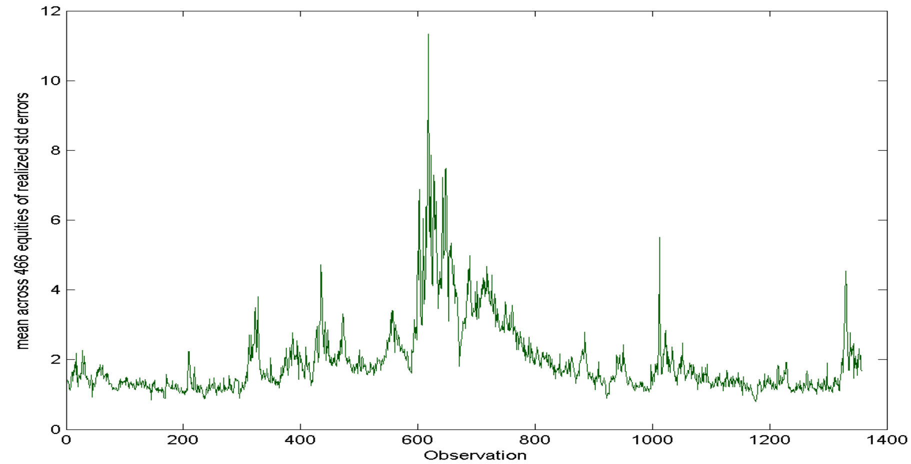
Five-minute returns are defined as the return from the first trade to the last trade of the five-minute period; the sample size at the 5-minute frequency is 105,846 for each company's stock and $466 \times 105,846$, or over 49 million, in total. The resulting data set has the form of a panel consisting of bars (daily and 5 minute; 1,357 or 105,846 rows) with 466 columns, one for each stock. Our base initial sample for forecast comparison allows 500 trading days for initial model estimation; with an initial daily observation lost for estimation and one for computation of one-period-ahead forecasts, we have 855 forecast days, and a total of 855×466 or 398,430 daily forecasts computed by each method. Results are reported for various other values of the initial sample for estimation.

Figure 1a shows the square roots of the realized variances, averaged over each of the 466 log-return series, for each of the 1,357 sample days. We see the highest values occurring between approximately sample points 600 and 700 (mid-2008), so that when we later take initial samples for estimation large enough to exclude these values from the set of pseudo-out-of-sample forecasts, we will expect to see changes in measures of forecast loss relative to samples which include these dates.

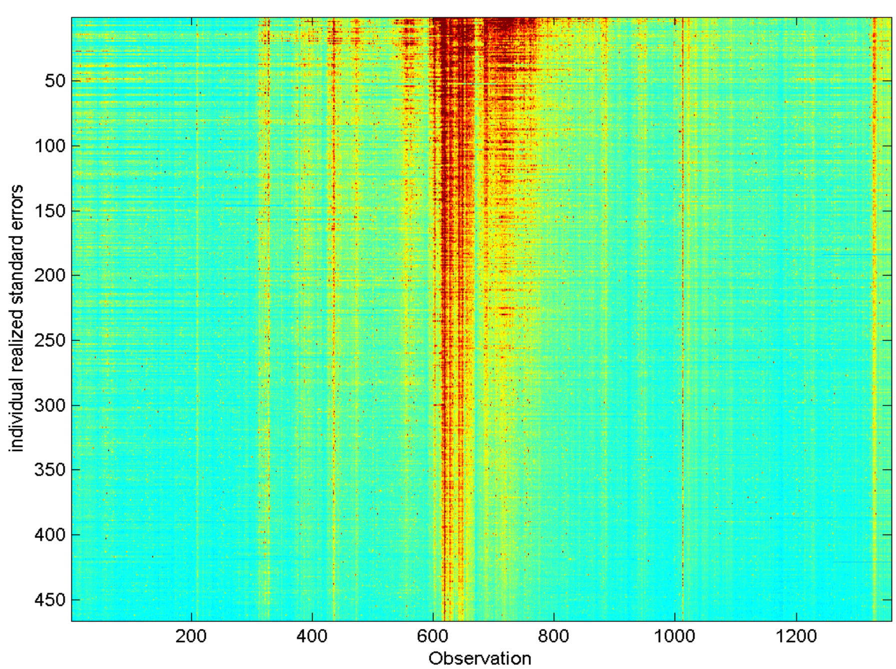
Figure 1b plots, rather than the average across all equities, the entire set of 466 $\times$ 1,357 daily (logarithmic) realized variances, so that instead of a single point at each date the figure comprises 1,357 vertical bars, each of which contains 466 coloured sections representing one of the S&P 500 securities in our sample. Colours toward the red end of the spectrum indicate higher values. The 466 equities are ordered by average realized variance across the sample days, with those showing the highest variances at the top of the figure. We see that as well as substantial variability across time, there is substantial variability in the cross-sectional dimension. Forecasting volatility on each of the 466 log-return series therefore represents a diverse set of forecasting exercises.

Figure 1
Volatility data

1a: realized volatility (square root of realized variance)
mean over 466 log-return series, 1,357 days of sample



1b: square root of realized variance
ordered (largest at top) set of 466, 1,357 days of sample



2.3 Treatment of outliers

The 'flash crash' of May 6, 2010 a number of S&P index securities in our sample experienced extraordinary prices declines, from 10% to over 95%, over time periods of around five to thirty minutes. Trades in some securities were subsequently cancelled or re-priced by securities regulators. This date corresponds with sample point 1012 in our daily data. Data points 1012 and 1013 are excluded from all samples here, on the grounds that the extremely high volatility is in some equities during the first of these days is not a result of the usual volatility process, and estimators using sample point 1012 as a lagged value would also provide anomalous forecasts for the second day. All estimation methods produced very poor estimates of volatility for the stocks having extreme movements on the first of these days.

Very high volatilities are also present in the period roughly from September 2008 through the summer of 2009, or approximately from observations 600 to 800 in the daily data. These are included in the base sample. As we move to later beginning dates for out-of-sample forecasting, or the lower forecast sample sizes, these points are progressively eliminated from the sample, and for sets of forecasts beginning around daily observation 800 this period is essentially absent. Relative performance of the estimators is little affected.

3 Empirical forecast evaluation

Each of the techniques is evaluated using both one-step-ahead mean absolute error and root mean squared error as measures of loss, relative to the daily realized volatility (the square root of the daily realized variance) as target value; that is,

$$\ell_1 = n_f^{-1} \sum_{t=t_0+1}^{T-1} \left(|\hat{\sigma}_{t+1|t} - \sigma_{rv,t+1}| \right) \tag{5}$$

$$\text{and } \ell_2 = \left[n_f^{-1} \sum_{t=t_0+1}^{T-1} \left(\hat{\sigma}_{t+1|t} - \sigma_{rv,t+1} \right)^2 \right]^{\frac{1}{2}} \tag{6}$$

where t_0 is the size of the sample used for initialization of the estimators, $T = 1357$ is the full daily sample size, $\hat{\sigma}_{t+1|t}$ is the one-step-ahead conditional standard deviation forecast made by any one of the methods, $\sigma_{rv,t+1}$ is the square root of the realized variance at time $t+1$, and $n_f = (T-1) - (t_0+1)$ is the number of pseudo-out-of-sample forecasts available for evaluation.

Before presenting the results in detail, Table 1 gives the overall average of results across all 466 stocks and the largest set (855) of pseudo-out-of-sample values; the losses are presented relative to those of QML. The base and alternative cases represent different parameter choices for the estimation methods.[6] We see in this simple aggregated comparison that the combined QML- and LAD-ARCH estimators produce the best overall results.

[6] Base case: QML: conditionally Normal; LAD-ARCH: $k = k' = 12$. Alternative case: QML: conditionally $t-$; LAD-ARCH: $k = 16, k' = 8$. Note that the number of forecasts evaluated is equal to (1357-2)- t_0, where t_o is the number of observations used for model initialization.

Table 1

Overall summary of performance by two loss measures, relative to QML: base and alternative cases, 466 $\times$ 855 forecasts

	RMSE (base)	MAE (base)	RMSE (alt)	MAE (alt)
QML	1.000	1.000	1.000	1.000
LAD-ARCH	0.940	0.886	0.944	0.891
equal weight	0.874	0.845	0.873	0.839
c.v. weight	0.888	0.837	0.888	0.834

Tables 2 and 3 present more detailed results. In Table 2, the results recorded are averages across all 466 series, for different numbers of sample points used in initial model estimation. The sample sizes recorded in the table for forecast evaluation range from 355 (1000 sample points for initial model estimation) to 855 (500 sample points for initial model estimation). We consider various initial sample sizes in order to check robustness of the results to this choice. For both base and alternative cases and at all of the sample sizes considered, the LAD-ARCH estimator shows the lower loss of the two main forecast techniques, by both RMSE and MAE. If we consider the cross-validated combined and equally weighted QML- and LAD-ARCH estimators of the GARCH model as well, then one or the other of the combined estimators shows lowest overall loss across all of these cases.

Table 3 gives an indication of the variability of these results across the 466 securities. In this table we record the proportion of cases in which a given technique shows the lowest loss, by each loss function, in a two-way comparison of the QML and LAD-ARCH estimators, also in the four-way comparison in which the combined estimators are added to the set. In the two-way comparisons, the LAD-ARCH estimator is best in a majority of cases, but there is a substantial proportion of the 466 series for which the QML-estimator is best. In the broader comparison which also includes the combined estimators, one of the two combined estimator produces the best results in the great majority of cases.

Figures 2 and 3 present information similar to that in the tables, but with results are recorded for every sample size in the interval from earliest to latest starting date, rather than at every 50th value, as in the tables. The sample sizes in the tables are therefore a subset of those recorded in the figures. As well, for the loss-function cases the figures present ratios of statistics relative to the QML values, rather than the absolute values of the loss functions. Figure 2 is analogous to Table 2, and shows that averaging over the 466 return series, the LAD-ARCH and combined estimators produce lower loss than QML at all sample sizes in the range considered. Figure 3 is analogous to Table 3, but provides a more detailed set of results across the full range of sample sizes rather than the overall average across the maximum sample size. We see again that the combined estimators reliably produce the best results for a large proportion of the 466 return series. In the two-way comparison of the techniques, we again see that the LAD-ARCH estimator tends to produce the lowest loss in the highest proportion of cases, but there are some sample sizes at which, by the RMSE criterion, the QML estimator has a slightly higher proportion of cases in which it produces the lowest loss.

It is of course a commonly observed result that an equal weighting of forecasts can outperform more sophisticated data-based weighting methods, because the efficiency cost of estimation of weighting parameters may exceed the benefit of deviating from equality. This is more likely to arise if, as is the case here, the component forecasts have approximately equal forecast error variance.

3.1 Statistical inference on differences in forecast loss

We now consider statistical inference on the differences in forecast performance among the methods considered. First we provide Diebold-Mariano (1995) pairwise tests on each of the 466 individual stocks, for various one-sided hypotheses. Although the Diebold-Mariano (DM) test is asymptotic, the overall pattern observable in these cases indicates some clear distinctions in test performance.

Each panel of Table 4 shows the results of the DM test on the null hypothesis that the mean squared errors for two given methods are equal, vs the alternative that the test named in the left-hand column shows lower loss; significance levels of 1% and 5% are reported. Number given are a count of the number of stocks, out of 466, for which we reject the null hypothesis. Large values are evidence of the superiority of the test named on the left, within the given pairing.

The results show that the loss reductions from combination versus the component methods are significant in a large number of stocks. The difference between the two combination methods, by contrast, does not show an unambiguous pattern favouring one rather than the other, although the preponderance of significant results does favour equal weighting. LAD-ARCH forecasts show a substantial degree of significant improvement over QML, but again these tend clearly to be dominated by forecast combinations.

Table 5 presents overall results, aggregating the 466 stocks, using a White (2000) reality-check test. This test takes a particular method as the basis for comparison and tests the hypothesis that all other methods considered have equal or greater forecast loss; a rejection is therefore indicative of the existence of a lower-loss method among the alternatives in the set. Failure to reject indicates lack of evidence against the hypothesis that the method of interest has the lowest forecast loss.

The results are compatible with those in Table 4; we see that we tend to reject the hypothesis that QML or LAD-ARCH forecast loss is equal to or better than that of the alternative methods, but we do not reject such a hypothesis for either forecast combination.

4 Concluding remarks

This study has attempted to report a very broad set of results in order to give a clear and reliable picture of the relative forecasting performance of GARCH-based forecasting methods based on substantially different principles. All of the methods use the parametric GARCH form.

Several results emerge. First, performance of the LAD-ARCH estimator is consistent with the performance predicted in GZZ; at the parameter values typical in daily equity return data and with 78 five-minute returns available per day, QML and LAD-ARCH should be similar in performance, with perhaps a small advantage to LAD-ARCH. This prediction is borne out by the empirical results.

Perhaps the most useful result concerns forecast combination for this problem. Although many methods are available, a simple average is often competitive with more sophisticated forecast combination devices, and we use here the sample mean of two forecasts to illustrate the potential for forecast combination. The QML and LAD-ARCH estimators are based on different information sets, and so a priori we might expect that a combined estimator could perform better than either alone, at least in regions of the parameter space where neither is dominated. This is what we observe: the combined estimator produces the lowest loss in a wide variety of cases, a result which is robust to sample period, to choice of loss function and to parameter choices for estimation methods. Cross-validated combination weights also perform well, but tend overall to be outperformed by an equally weighted combination. Statistical inference tends to confirm that the advantage of forecast combination over either constituent method, whether by equal weighting or cross-validated weights, is genuine.

We conclude that the LAD-ARCH estimator of the GARCH model provides a useful practical addition to the set of forecasting tools for conditional volatility in financial markets. Simple equally weighted average forecasts, obtained by averaging the standard QML estimator with the LAD-ARCH estimator, substantially outperform either component forecast.

It remains to be seen whether the principle of quantile-based estimators of ARCH models can provide useful gains in multivariate contexts.

References

[1] Andersen, T.G., T. Bollerslev, F.X.Diebold and P.Labys (2003) Modeling and Forecasting Realized Volatility. *Econometrica* 71, 529-626.

[2] Bollen, B. and B. Inder (2002) Estimating Daily Volatility in Financial Markets Utilizing Intraday Data. *Journal of Empirical Finance* 9, 552-562.

[3] Bollerslev, T. (1986) Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31, 307-327.

[4] Diebold, F.X. and R. Mariano (1995) Comparing Predictive Accuracy. *Journal of Business and Economic Statistics* 13, 253-265.

[5] Drost, F.C. and T.E. Nijman (1993) Temporal Aggregation of GARCH Processes. *Econometrica* 61, 909-927.

[6] Galbraith, J.W., V. Zinde-Walsh and J. Zhu (2012) GARCH Model Estimation Using Estimated Quadratic Variation. Forthcoming, *Econometric Reviews.* Presently available at `http://johngalbraith-economics.mcgill.ca/adgarch.pdf`.

[7] Koenker, R. and Q. Zhao (1996) Conditional Quantile Estimation and Inference for ARCH Models. *Econometric Theory* 12, 793–813.

[8] Nelson, D. (1990) ARCH Models as Diffusion Approximations. *Journal of Econometrics* 45, 7-38.

[9] White, H. (2000) A Reality Check for Data Snooping. *Econometrica* 68, 1097-1127.

[10] Xiao, Z. and R. Koenker (2009) Conditional Quantile Estimation for GARCH models. Working paper.

[11] Zernov, S., V. Zinde-Walsh and J.W. Galbraith (2009) Asymptotics for Estimation of Quantile Regressions with Truncated Infinite-Dimensional Processes. *Journal of Multivariate Analysis* 100, 497-508.

[12] Zhang, L., P.A. Mykland and Y. Aït-Sahalia (2005) A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data. *Journal of the American Statistical Association* 100, 1394-1411.

Table 2

Pseudo-out-of-sample forecast performance, average over 466 series
Empirical forecast losses, RMSE and MAE [7]
QML, LAD-ARCH, equal- and cv-weight combinations

Base case

Panel A: RMSE — Number of forecasts evaluated

	355	405	455	505	555	605	655	705	755	805	855
QML	0.549	0.540	0.537	0.541	0.548	0.582	0.637	0.730	0.937	0.943	0.926
LAD-ARCH	0.509	0.500	0.500	0.506	0.518	0.547	0.612	0.680	0.886	0.891	0.878
equal weight	0.479	0.471	0.470	0.474	0.483	0.508	0.561	0.627	0.820	0.826	0.814
c.v. weight	0.484	0.475	0.475	0.480	0.489	0.514	0.571	0.636	0.839	0.846	0.833

Panel B: MAE

	355	405	455	505	555	605	655	705	755	805	855
QML	0.395	0.391	0.390	0.394	0.398	0.418	0.448	0.495	0.574	0.581	0.570
LAD-ARCH	0.347	0.340	0.341	0.345	0.353	0.370	0.403	0.438	0.514	0.520	0.514
equal weight	0.335	0.330	0.330	0.334	0.339	0.353	0.380	0.414	0.482	0.488	0.481
cv weight	0.331	0.325	0.325	0.329	0.335	0.349	0.378	0.410	0.482	0.488	0.482

Alternative case

Panel A: RMSE — Number of forecasts evaluated

	355	405	455	505	555	605	655	705	755	805	855
QML	0.549	0.540	0.537	0.541	0.548	0.582	0.637	0.730	0.937	0.943	0.926
LAD-ARCH	0.508	0.499	0.500	0.506	0.518	0.549	0.617	0.688	0.896	0.900	0.887
equal weight	0.477	0.469	0.468	0.473	0.481	0.506	0.560	0.626	0.820	0.827	0.815
cv weight	0.482	0.473	0.473	0.478	0.488	0.513	0.572	0.637	0.843	0.849	0.837

Panel B: MAE

	355	405	455	505	555	605	655	705	755	805	855
QML	0.395	0.391	0.390	0.394	0.398	0.418	0.448	0.495	0.574	0.581	0.570
LAD-ARCH	0.347	0.339	0.340	0.346	0.354	0.372	0.407	0.444	0.521	0.527	0.521
equal weight	0.333	0.328	0.328	0.331	0.336	0.350	0.378	0.411	0.480	0.486	0.480
cv weight	0.328	0.322	0.322	0.326	0.332	0.347	0.377	0.410	0.483	0.489	0.483

[7] Base case: QML: conditionally Normal; LAD-ARCH: $k = k' = 12$. Alternative case: QML: conditionally $t-$; LAD-ARCH: $k = 16, k' = 8$. Note that the number of forecasts evaluated is equal to (1357-2)- t_0, where t_0 is the number of observations used for model initialization.

Table 3

Proportion of cases in which a technique shows best performance,
by the criterion of average RMSE or MAE
466 series and 855 pseudo-out-of-sample observations[8]
Compared: QML and LAD-ARCH (2),
or QML, LAD-ARCH, equal- and cv-weight combinations (4)

Base case

	by RMSE (2)	by RMSE (4)	by MAE (2)	by MAE (4)
QML	0.251	0.004	0.118	0.001
LAD-ARCH	0.749	0.006	0.882	0.028
equal weight		0.742		0.508
cv weight		0.247		0.464

Alternative case

	by RMSE (2)	by RMSE (4)	by MAE (2)	by MAE (4)
QML	0.268	0.005	0.132	0.001
LAD-ARCH	0.732	0.006	0.868	0.021
equal weight		0.753		0.520
c.v. weight		0.236		0.457

[8]Entries are the proportions of cases that a given technique has the lowest value of the given loss measure in the indicated comparison. Base case and alternative case as defined earlier. The numbers (2) or (4) in parentheses in column headings indicate the number of techniques being compared in the given column; all columns add to one.

Table 4[9]

Number of rejections out of 466 one-sided DM tests

H_0 : equal forecast loss vs. H_1 : test in left column has lower loss

	355	405	455	505	555	605	655	705	755	805	855
Panel A: QML vs LAD-ARCH											
QML 1%	1	2	1	2	8	8	13	13	1	1	1
QML 5%	4	5	7	9	14	18	31	24	7	5	6
LAD-ARCH 5%	157	163	177	176	153	153	109	173	144	152	151
LAD-ARCH 1%	98	114	119	118	101	105	71	111	76	80	78
Panel B: QML vs equal weight											
QML 1%	0	0	0	0	0	0	1	0	0	0	0
QML 5%	0	0	0	0	0	0	1	0	0	0	0
EQ 5%	436	444	444	446	435	431	431	443	433	437	435
EQ 1%	388	403	409	421	416	414	403	418	402	410	410
Panel C: LAD-ARCH vs equal weight											
LAD-ARCH 1%	3	3	3	2	3	5	1	0	0	0	0
LAD-ARCH 5%	5	9	8	3	3	7	8	4	0	0	0
EQ 5%	271	273	287	315	342	354	389	354	305	310	323
EQ 1%	161	175	197	222	263	279	332	275	192	200	212
Panel D: QML vs c.v. weight											
QML 1%	0	0	0	0	1	2	1	0	0	0	0
QML 5%	0	0	0	2	3	3	5	0	0	0	0
CV 5%	346	348	351	354	344	338	326	380	365	370	368
CV 1%	272	288	298	305	295	291	270	336	301	313	314
Panel E: LAD-ARCH vs c.v. weight											
LAD-ARCH 1%	1	1	1	1	1	1	0	1	1	1	1
LAD-ARCH 5%	2	3	1	1	1	1	0	2	6	6	4
CV 5%	377	376	387	390	400	403	412	366	280	283	285
CV 1%	323	326	349	358	368	370	385	319	243	239	242
Panel F: equal weight vs c.v. weight											
EQ 1%	41	45	58	67	88	106	129	123	120	126	129
EQ 5%	112	118	128	138	167	178	200	189	190	202	206
CV 5%	47	49	49	48	45	52	32	46	24	24	21
CV 1%	29	28	28	32	28	31	22	30	11	10	11

[9] Results are reported for the base case only; alternative case results are qualitatively similar.

Table 5[10]

White reality-check test, all 466 series combined

H_0 : all other methods have equal or greater forecast loss

	355	405	455	505	555	605	655	705	755	805	855
QML	0.570	0.562	0.559	0.565	0.577	0.617	0.688	0.792	1.017	1.029	1.010
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
LAD-ARCH	0.532	0.522	0.522	0.529	0.543	0.577	0.658	0.737	0.967	0.976	0.959
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
equal weight	0.532	0.522	0.522	0.529	0.543	0.577	0.658	0.737	0.967	0.976	0.959
	(0.91)	(0.91)	(0.94)	(0.96)	(0.94)	(0.90)	(0.94)	(0.90)	(0.94)	(0.90)	(0.91)
c.v. weight	0.532	0.522	0.522	0.529	0.543	0.577	0.658	0.737	0.967	0.976	0.959
	(0.20)	(0.22)	(0.18)	(0.12)	(0.21)	(0.25)	(0.13)	(0.23)	(0.14)	(0.18)	(0.20)

[10]Results are reported for the base case only; alternative case results are qualitatively similar. $p-$values are indicated in brackets; an indicated value of 0.00 indicates a computed $p-$value less than 0.005.

Figure 2
Ratios of RMSE's and MAE's to those of QMLE
QML, LAD-ARCH, equal-weight and cross-validation weight combinations
Varying initial estimation sample and parameter specifications

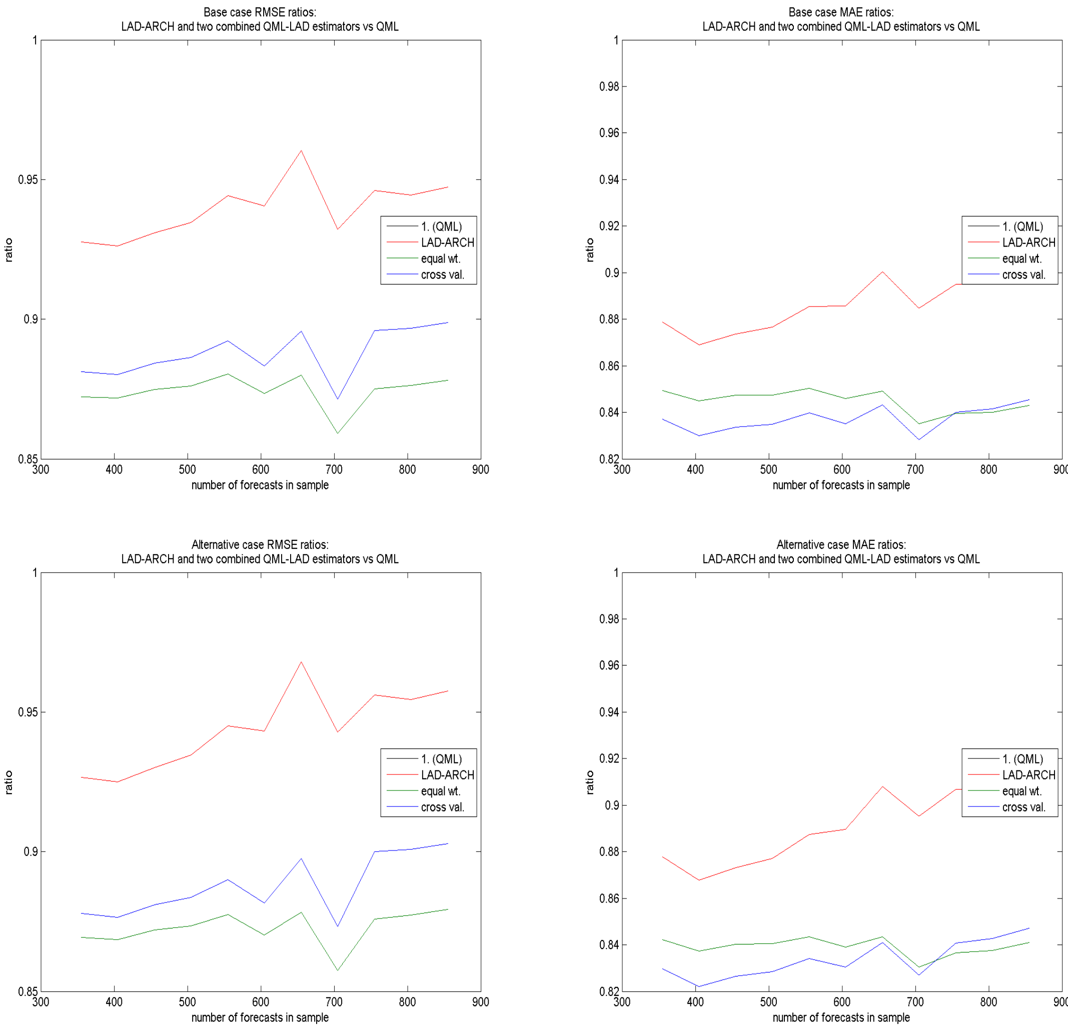


Figure 3

Proportion of cases in which a technique yields lowest loss

QML, LAD-ARCH, equal-weight and cross-validation weight combinations