Online Appendix: A Practical Guide to Volatility Forecasting through Calm and Storm

Christian Brownlees Rob Engle Bryan Kelly

1 Summary of Empirical Design

Several forecasting strategy questions naturally arise in implementing a real-time volatility forecasting exercise. These questions are the thesis of this article, and may be stated as follows: When constructing volatility forecasts for a range of assets over multiple horizons, what are the best choices for i) volatility model, ii) estimation window length, iii) frequency of parameter re-estimation and iv) innovation distribution? In addition, we ask if our conclusions regarding forecasting models or estimation strategies break down during periods of turmoil such as the volatility surge of late 2008. The set of volatility models considered GARCH, TARCH, EGARCH, NGARCH and APARCH. Estimation window lengths range from the full post-1990 sample to windows as small as four years. If parameters are unstable, it may be practical to re-estimate using a shorter, rolling window. We also consider whether daily re-estimation improves over weekly re-estimation, and similarly whether there is any significant loss to re-estimating only monthly. This question trades off computing cost against forecast accuracy, and is relevant when the collection of assets and models becomes large. To address the prevalence of fat-tailed returns, we consider taking non-Gaussian aspects of the data into account by using a Student t likelihood, which can potentially improve efficiency by correcting the specification, thus improving forecasting performance. Using a heavy-tailed likelihood, however, introduces a new parameter. Considering both Gaussian and Student t innovations allows us to assess the role played by heavy tailed innovations in the forecast optimization effort. Our test assets are eighteen domestic and international equity indices, nine sectoral equity indices, and ten exchange rates. We forecast from one day to one month ahead. Our data begins in 1990 and we use an out-of-sample interval spanning 2001 to 2008. This period contains both high volatility episodes (corresponding roughly to the NBER recessions of March 2001 to November 2001 and December 2007 through the present), as well as the protracted interval of low volatility from approximately 2003 until 2007. We devote special attention to the period of financial crisis from September 2008 to December 2008 and characterize the extremity of the crisis compared to historical standards.

Our overall assessment proceeds in two stages. We first perform exhaustive comparisons of models and estimation strategies using the S&P 500 index as the test asset. This provides the skeleton for our evaluation procedure that is then applied to all other assets. Due to the sheer volume of output, we report condensed results for the full set of test assets and more detailed results for the S&P 500. To evaluate volatility forecast accuracy we rely on ex post proxies for the true, latent volatility process. The two standard proxies are squared returns and the more precisely estimated realized variance, calculated from ultra high frequency data. Our forecast accuracy comparisons are based on squared returns for our full set of test assets. For the S&P 500 index we also use realized volatility to make our assessments - this allows us to evaluate if and how conclusions from our experiments might change when different proxies are used. Forecast accuracy is measured with robust forecast loss functions advocated by Patton (2009).

2 Related Literature

The literature on volatility forecasting and forecast evaluation is surveyed in Poon & Granger (2003), Poon & Granger (2005), Patton & Sheppard (2009). Volatility forecasting assessments are commonly structured to hold the test asset and estimation strategy fixed, focusing on model choice. We take a more pragmatic approach and consider how much data should be used for estimation, how frequently a model should be re-estimated, and what innovation distributions should be used. This is done for each model we consider. Furthermore, we do not rely on a single asset or asset class to draw our conclusions. We look beyond volatility forecasting meta-studies, in particular the Poon and Granger papers, which focus almost exclusively on one day ahead forecasts. Our work draws attention to the relevance of multi-step ahead forecast performance for model evaluation, especially in crisis periods when volatility levels can escalate dramatically in a matter of days. The issues of multiple step ahead forecasting has also been addressed by Christoffersen & Diebold (2000) and Ghysels, Rubia & Valkanov (2009). The forecast evaluation methodology employed builds upon the recent contributions on robust forecasting assessment developed in Hansen & Lunde (2005) and Patton (2009), which consistently rank volatility forecasts despite the fact volatility is not observable. An important implication of these results is that the conflicting evidence reported by some previously published empirical studies on volatility forecasting is due to the use of non robust losses, and this calls for a reassessment of previous findings in this field. The improvements in consistent volatility forecast ranking also hinge on the recent availability of high frequency based volatility measures (*inter alia* Andersen, Bollerslev, Diebold & Labys (2003), Bandi & Russel (2006), Aït-Sahalia, Mykland & Zhang (2005), Barndorff-Nielsen, Hansen, Lunde & Shephard (2009)) which provide efficient ex-post estimators of daily volatility. Other approaches to volatility forecast evaluation are based on assessing the value of predictions from economic or risk management perspectives (Fleming, Kirby & Ostdiek (2003), Kuester, Mittnik & Paolella (2006), Brownlees & Gallo (2009)). Recent developments in volatility forecasting also include a number of models based on high frequency based volatility measures. The contributions in this area include among others Andersen et al. (2003), Deo, Hurvich & Lu (2006), Engle & Gallo (2006), Ait-Sähalia & Mancini (2008), Hansen, Huang & Shek (2010), Corsi (2010), and Shephard & Sheppard (2010). While promising in terms of improved forecast accuracy, our study omits these approaches as well as approaches based on stochastic volatility models (cf. Ghysels, Harvey & Renault (1995)). ARCH models require daily frequency data which are typically easier to obtain than intra-daily data and ARCH estimation software is typical widely available and easier to implement. Moreover, using a set of models which differs in the choice of the dependent variable and the conditioning information sets makes the results of a forecasting comparison harder to interpret.

3 Additional Discussion of Results

3.1 S&P 500 Volatility

Adoption of a Student t likelihood does not significantly improve performance at short horizons, although at longer horizons significantly positive improvements are possible. The Student t down-weights extremes with respect to the Gaussian, thus it can provide a more robust estimate of the long run variance. As we see in these tables and discuss in more detail later, short horizon volatility and return realizations do not appear to violate a Gaussian assumption, though at longer horizons we see substantially more tail events than a normal curve would predict. The possibility that a Student t provides a better description of volatility at long horizons is consistent with these observations.

Using a shorter, rolling estimation window tends to weaken forecasting accuracy. In some cases, the performance decreases by as much as 20%. The window length results are non-monotonic. While the full sample dominates, we often see that the medium estimation window does worse than the short window. To understand how this may arise, Figure 1 displays the time series plot of parameter estimates for one of our models (TARCH) and the implied persistence ($\alpha + \gamma/2 + \beta$) using the base procedure. Estimates are expressed as absolute differences with respect to the estimates obtained at the beginning of the out-of-sample period. While there is no clear evidence of breaks in the parameters, the series do exhibit a slow drift: α systematically declines during this period, ending up insignificantly different from zero by the end of 2008. Movements in β and γ appear quite large and negatively correlated, transferring weight in the evolution equation away from past conditional variance toward past squared negative returns. The graph also suggests that periods of more severe financial distress are associated with higher weight on past squared negative returns. This suggests that the short window has some ability to capture variation in parameters, though at the cost of less precise estimates. Ultimately, the noisiness overwhelms the gains from parameter variation, and the net effect is slightly worse performance by the short window. The medium window, on the other hand, uses a long enough history to miss much of the time variation in parameters and at the same time loses accuracy compared to the full sample window, resulting in the worst forecasting performance of the three windows. We also see more deterioration in short and medium window performance at longer forecast horizons. Lastly, the update frequency results show that more frequent updating tends to modestly improve performance, but the difference is insignificant in many cases.

Table 6 shows losses for each GARCH model over the full out-of-sample period 2001 to 2008. We report results for both QL and MSE loss functions with r^2 and rv proxies. To compare accuracy across models, we use the Diebold-Mariano test to detect if a given model provides significantly lower average losses compared to the GARCH model. For comparison purposes the table also includes the naïve 60 days historical variance forecast. Significant outperformance at the 10%, 5% and 1% significance level are denoted by (*), (**), and (***), respectively.

Asymmetric specifications provide lower out-of-sample losses, especially over one day and one week. At longer horizons, recent negative returns are less useful for predicting future volatility. At the one month horizon, the mean reversion effect begins to dominate as the difference between asymmetric and symmetric GARCH becomes insignificant, and historical variance becomes competitive.

The choice of loss function does not change rankings, but MSE loss seems to provide more mixed evidence than QL. There is some discrepancy in the rankings however when using different proxies. Squared returns favor TARCH while realized volatility selects EGARCH. The discrepancy should not be overstated, however, as the methods do not significantly outperform each other. Results do suggest model rankings are stable over various forecasting horizons.

Table 7 repeats the direct GARCH comparison from Table 6, but focuses on the extreme volatility interval from September 2008 through December 2008. During this time, forecasting losses at all horizons are systematically larger than in the overall sample. Recall that QL is unaffected by changes in the level of volatility, so that changes in average losses purely represent differences in forecasting accuracy. We find that one step ahead losses during fall 2008 are only modestly higher than those registered in the full sample. At one month, however, QL losses are twice as big as the full sample using squared returns and four times as big using realized volatility. An important feature of this table is that our conclusions about model ranking remain largely unchanged during the crisis. TARCH tends to be most accurate, though differences with symmetric GARCH and historical variance start to lose significance at long horizons. MSE gives a much more confused picture of volatility in Fall 2008. Most noticeably, the MSE level during the crisis is an order of magnitude higher than the in full sample. GARCH specifications systematically outperform historical variance only at very short horizons, though at such horizons we again find that the asymmetric versions are superior.

Tables 8 and 9 contain the summary forecasting results for a broader collection of assets (see Table 1 in main article). Volatility forecast losses for exchange rates, US equity sector indices and international equity indices from each model averaged not only across time, but also averaged over all assets in the same class. Asterisks denote that a model significantly outperformed GARCH based on the Diebold-Mariano test. We also report the relative winning frequency for each model, defined as the number of assets in a class for which a given model provided the best out-of-sample forecasts.

Of all asset classes, exchange rates appear to be most forecastable as they give the smallest losses according to QL. For several cases, the average loss point estimate is lower for asymmetric models, despite the fact that leverage effects for exchange rates are not well defined. While we find that no specification obtains significantly lower average losses than GARCH according to the QL loss, though there is some significant evidence in favor of asymmetric models based on MSE. Also, asymmetric models demonstrate success in terms of winning frequency.

Tables 8 and 9 are strong evidence in favor of using asymmetric models for sectoral equity indices. TARCH emerges as the best performer, closely followed by APARCH. This is clearest from the QL results, which show that all asymmetric specifications (other than EGARCH) outperform GARCH over all horizons. MSE results are similar, but less statistically significant.

International equities deliver similar results. Asymmetric specifications perform better than

GARCH, with TARCH the most frequent top model according to both QL and MSE losses. Most evidence of outperformance, however, is limited to shorter horizons - at long horizons the winning frequency becomes more uniform across models.

Table 10 and 11 report average losses during the fall 2008 crisis. Results confirm our findings in the S&P500 case. One-day ahead losses are virtually unchanged from those during the full sample, while one month losses are magnified by a factor of about two. TARCH appears to be the best performer at all horizons for all asset classes, although the margin appears to remain small for exchange rates.

References

Ait-Sähalia, Y. & Mancini, L. (2008), 'Out of sample forecasts of Quadratic Variation', *Journal of Econometrics* **147**, 17–33.

Aït-Sahalia, Y., Mykland, P. A. & Zhang, L. (2005), 'How often to sample a continuous–time process in the presence of market microstructure noise', *The Review of Financial Studies* **28**, 351–416.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003), 'Modelling and Forecasting Realized Volatility', *Econometrica* **71**(2), 579–625.

Bandi, F. M. & Russel, J. (2006), 'Separating microstructure noise from volatility', *Journal of Financial Economics* **79**, 655–692.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2009), 'Realised kernels in practice: trades and quotes', *Econometrics Journal* **12**.

Brownlees, C. T. & Gallo, G. M. (2009), 'Comparison of Volatility Measures: A Risk Management Perspective', *Journal of Financial Econometrics* (), forthcoming.

Christoffersen, P. F. & Diebold, F. X. (2000), 'How Relevant is Volatility Forecasting for Financial Risk Management?', *Review of Economics and Statistics* **82**(1), 12–22.

Corsi, F. (2010), 'A simple approximate long-memory model of realized volatility', *Journal of Financial Econometrics* **7**, 174–196.

Deo, R., Hurvich, C. & Lu, Y. (2006), 'Forecasting Realized Volatility Using a Long-Memory Stochastic Volatility Model: Estimation, Prediction and Seasonal Adjustment', *Journal of Econometrics* **131**(1-2), 29–58.

Engle, R. F. & Gallo, G. M. (2006), 'A Multiple Indicators Model for Volatility Using Intra-Daily Data', *Journal of Econometrics* **131**(1-2), 3–27.

Fleming, J., Kirby, C. & Ostdiek, B. (2003), 'The Economic Value of Volatility Timing', *The Journal of Finance* **56**(1), 329–352.

Ghysels, E., Harvey, A. & Renault, E. (1995), Stochastic Volatility, *in* G. Maddala & C. Rao, eds, 'Handbook of Statistics 14, Statistical Methods in Finance,', North Holland, Amsterdam,, pp. 119–191.

Ghysels, E., Rubia, A. & Valkanov, R. (2009), Multi-period forecasts of volatility: Direct, iterated, and mixed-data approaches, Technical report, .

Hansen, P. R., Huang, Z. & Shek, H. H. (2010), Realized garch: A complete model of returns and realized measures of volatility, Technical report.

Hansen, P. R. & Lunde, A. (2005), 'Consistent ranking of volatility models', *Journal of Econometrics* **131**(1–2), 97–121.

Kuester, K., Mittnik, S. & Paolella, M. S. (2006), 'Value-at-Risk Prediction: A Comparison of Alternative Strategies', *Journal of Financial Econometrics* **4**(1), 53–89.

Patton, A. (2009), Volatility Forecast Comparison using Imperfect Volatility Proxies, Technical report, University of Oxford.

Patton, A. J. & Sheppard, K. (2009), Evaluating Volatility and Correlation Forecasts , *in* T. Mikosch, J.-P. Kreis, R. A. Davis & T. G. Andersen, eds, 'Hanbook of Financial Time Series', Elsevier, pp. 801–838.

Poon, S. & Granger, C. W. J. (2003), 'Forecasting Volatility in Financial Markets: A Review', *Journal of Economic Literature* **51**(2), 478–539.

Poon, S. & Granger, C. W. J. (2005), 'Practical Issues in Forecasting Volatility', *Financial Analysts Journal* **61**(1), 45–56.

Shephard, N. & Sheppard, K. (2010), 'Realising the future: forecasting with high frequency based volatility (heavy) models', *Journal of Applied Econometrics* **25**, 197–231.

Loss	$\hat{\sigma}^2$	Est. Strategy	Horizon				
			1 d	1 w	2 w	3 w	1 m
QLL	r^2	base	1.460	1.481	1.520	1.574	1.645
		Student t news	-0.16	-0.10	0.08	0.41	0.83*
		medium window	-0.71	-1.05	-2.12	-3.07	-4.41
		long window	-0.71	-1.06	-1.63	-2.28	-3.12
		monthly update	-0.02	-0.04	-0.02	-0.03	0.02
		daily update	0.01	0.01	0.04*	0.01	0.02
QLL	rv	base	0.273	0.310	0.343	0.373	0.414
		Student t news	-3.22	-1.96	-0.99	-0.37	0.41
		medium window	0.97	-2.36	-5.30	-7.23	-9.70
		long window	-8.95	-12.55	-16.06	-18.58	-20.85
		monthly update	-0.05	-0.02	-0.07	-0.05	-0.01
		daily update	0.11	0.03	0.02	0.01	0.03
MSE	r^2	base	27.533	30.050	31.980	33.828	36.347
		Student t news	-0.28	-0.20	-0.14	-0.26	-0.93
		medium window	0.18	0.81*	-0.08	-0.05	0.28
		long window	0.37	0.66	0.044	-0.23	-0.35
		monthly update	-0.10	-0.09	-0.01	-0.08	-0.20
		daily update	-0.04	0.09	0.22	0.08	-0.02
MSE	rv	base	4.357	4.998	5.984	6.901	7.950
		Student t news	-3.71	-4.25	-5.06	-6.33	-9.35
		medium window	-0.34	-0.39	-0.67	-0.01	1.85
		long window	-4.93	-6.19	-6.87	-7.07	-6.19
		monthly update	1.04**	1.36	1.22	0.97	0.45
		daily update	0.16	0.10	0.23	-0.08	-0.12

Table 1: GARCH Estimation Strategy Assessment. For each loss function and volatility proxy, the table reports the out-of-sample losses at multiple horizons of the Vlab estimation strategy and the percentage gains derived by modifying the estimation strategy with (i) Student t innovations, (ii) medium estimation window, (iii) long estimation window, (iv) monthly estimation update and (v) daily estimation update. Asterisks beneath the percentage gains denote the significance of a Diebold-Mariano test under the null of equal or inferior predictive ability with respect to the baseline Vlab strategy (level of statistical significance denoted by *=10%, **=5%, ***=1%).

Loss	$\hat{\sigma}^2$	Est. Strategy	Horizon				
			1 d	1 w	2 w	3 w	1 m
QLL	r^2	base	1.415	1.442	1.478	1.547	1.624
		Student t news	0.01	0.02	0.30 *	0.59 *	1.21 **
		medium window	-0.86	-1.07	-1.79	-3.06	-4.49
		long window	-1.06	-1.46	-2.06	-2.65	-3.35
		monthly update	0.01	-0.01	-0.01	-0.01	0.03
		daily update	-0.01	-0.01	0.02	0.01	0.01
QLL	rv	base	0.243	0.289	0.328	0.368	0.415
		Student t news	-2.84	-2.20	-1.45	-0.913	-0.16
		medium window	5.80 ***	2.79 *	1.03	-1.052	-3.74
		long window	-10.58	-13.65	-16.33	-18.13	-19.64
		monthly update	-0.12	-0.06	-0.08	-0.05	-0.03
		daily update	0.05 ***	0.01	0.01	-0.010	0.02
MSE	r^2	base	25.583	28.874	31.151	33.197	36.043
		Student t news	0.31	0.57	0.37	0.07	-0.80
		medium window	0.16	-0.60	-1.41	-1.07	-0.17
		long window	0.31	0.05	-0.22	-0.34	-0.59
		monthly update	-0.03	-0.02	-0.03	-0.06	-0.07
		daily update	-0.05	0.05	0.14	0.05	-0.02
MSE	rv	base	3.647	4.550	5.687	6.474	7.312
		Student t news	-10.2	-10.50	-11.38	-12.78	-15.209
		medium window	5.84	4.05	4.32 *	5.87 *	8.92 *
		long window	-2.71	-4.49	-4.69	-4.85	-4.84
		monthly update	0.70 **	0.66	0.16	0.18	0.03
		daily update	-0.18	0.75	-0.11	-0.21	-0.13

Table 2: TGARCH Estimation Strategy Assessment. For each loss function and volatility proxy, the table reports the out-of-sample losses at multiple horizons of the Vlab estimation strategy and the percentage gains derived by modifying the estimation strategy with (i) Student t innovations, (ii) medium estimation window, (iii) long estimation window, (iv) monthly estimation update and (v) daily estimation update. Asterisks beneath the percentage gains denote the significance of a Diebold-Mariano test under the null of equal or inferior predictive ability with respect to the baseline Vlab strategy (level of statistical significance denoted by *=10%, **=5%, ***=1%).

Loss	$\hat{\sigma}^2$	Est. Strategy	Horizon				
			1 d	1 w	2 w	3 w	1 m
QLL	r^2	base	1.420	1.458	1.505	1.592	1.684
		Student t news	-0.87	0.30	1.45*	2.73**	4.23**
		medium window	-2.31	-3.76	-5.14	-7.64	-10.31
		long window	-1.29	-1.75	-1.93	-2.53	-2.88
		monthly update	-0.22	-0.05	0.02	-0.01	-0.06
		daily update	-0.17	0.07	0.14	0.22	0.22
QLL	rv	base	0.234	0.282	0.320	0.365	0.413
		Student t news	-26.74	-15.22	-8.31	-3.50	1.10
		medium window	-1.66	-3.47	-6.69	-10.18	-14.93
		long window	-12.98	-12.45	-12.90	-13.57	-13.89
		monthly update	0.50	0.68***	0.56**	0.42*	0.25
		daily update	-4.75	-2.76	-2.13	-1.52	-1.11
MSE	r^2	base	26.746	31.328	33.984	35.912	37.828
		Student t news	3.36*	5.63**	5.17**	4.79**	3.65*
		medium window	2.471*	-2.31	-3.70	-3.825	-3.224
		long window	1.284*	-0.31	-0.28	-0.26	-0.18
		monthly update	-0.34	-0.23	-0.14	-0.14	-0.10
		daily update	0.10	0.30**	0.02	0.18**	0.13**
MSE	rv	base	2.653	3.859	4.758	5.428	6.085
		Student t news	-37.92	-14.12	-4.37	-0.81	1.01
		medium window	13.62**	-2.19	-6.73	-8.27	-7.59
		long window	2.17**	-3.27	-2.30	-2.17	-1.68
		monthly update	0.51	0.36	0.19	0.03	-0.08
		daily update	-0.53	-0.1 4	0.13	0.02	0.12

Table 3: EGARCH Estimation Strategy Assessment. For each loss function and volatility proxy, the table reports the out-of-sample losses at multiple horizons of the Vlab estimation strategy and the percentage gains derived by modifying the estimation strategy with (i) Student t innovations, (ii) medium estimation window, (iii) long estimation window, (iv) monthly estimation update and (v) daily estimation update. Asterisks beneath the percentage gains denote the significance of a Diebold-Mariano test under the null of equal or inferior predictive ability with respect to the baseline Vlab strategy (level of statistical significance denoted by *=10%, **=5%, ***=1%).

Loss	$\hat{\sigma}^2$	Est. Strategy	Horizon				
			1 d	1 w	2 w	3 w	1 m
QLL	r^2	base	1.417	1.446	1.485	1.557	1.633
		Student t news	0.04	0.12	0.49 **	0.67 **	1.06 **
		medium window	-1.07	-0.76	-1.23	-2.25	-3.65
		long window	-1.14	-1.61	-2.24	-3.00	-3.84
		monthly update	-0.03	-0.05	-0.06	-0.07	-0.06
		daily update	0.06	0.02	0.07	0.04	0.04
QLL	rv	base	0.249	0.299	0.340	0.385	0.435
		Student t news	-2.96	-0.73	1.27 **	2.59 **	4.03 ***
		medium window	7.49 ***	5.87 **	5.83 **	4.81 **	2.98
		long window	-7.69	-9.56	-11.09	-12.66	-14.08
		monthly update	-0.22	-0.17	-0.30	-0.27	-0.27
		daily update	0.24	0.07 **	0.07 *	0.09	0.11
MSE	rv	base	25.678	29.421	32.098	33.949	36.408
		Student t news	0.63	1.02 *	0.91	0.60	-0.08
		medium window	0.75	1.09	0.69	0.50	0.36
		long window	0.27	-0.01	-0.54	-0.50	-0.64
		monthly update	-0.15	-0.08	-0.11	-0.18	-0.21
		daily update	0.01	0.18	0.35	0.15	0.02
MSE	rv	base	3.384	4.614	5.543	6.234	6.868
		Student t news	-14.62	-10.38	-9.21	-8.72	-8.86
		medium window	0.05	4.12	2.42	2.64	4.19
		long window	-3.32	-3.62	-4.81	-5.37	-5.15
		monthly update	1.00 **	0.83 *	0.46	0.24	-0.32
		daily update	0.08	0.286	0.36	0.10	0.08

Table 4: APARCH Estimation Strategy Assessment. For each loss function and volatility proxy, the table reports the out-of-sample losses at multiple horizons of the Vlab estimation strategy and the percentage gains derived by modifying the estimation strategy with (i) Student t innovations, (ii) medium estimation window, (iii) long estimation window, (iv) monthly estimation update and (v) daily estimation update. Asterisks beneath the percentage gains denote the significance of a Diebold-Mariano test under the null of equal or inferior predictive ability with respect to the baseline Vlab strategy (level of statistical significance denoted by *=10%, **=5%, ***=1%).

Loss	$\hat{\sigma}^2$	Est. Strategy	Horizon				
			1 d	1 w	2 w	3 w	1 m
QLL	r^2	base	1.422	1.459	1.498	1.574	1.659
		Student t news	-0.02	0.06	0.39**	0.71**	1.27**
		medium window	-0.80	-1.21	-1.99	-3.26	-4.65
		long window	-0.97	-1.98	-2.72	-3.44	-4.07
		monthly update	-0.00	-0.02	-0.03	-0.04	-0.02
		daily update	0.02	0.01	0.04*	0.02	0.02
QLL	rv	base	0.244	0.296	0.337	0.380	0.432
		Student t news	-2.19	-1.48	-0.73	-0.18	0.54
		medium window	3.30	0.015	-2.28	-4.02	-6.37
		long window	-12.93	-16.54	-19.14	-20.35	-20.82
		monthly update	-0.10	-0.064	-0.138	-0.139	-0.140
		daily update	0.132**	0.02	0.036*	0.034	0.068
MSE	r^2	base	27.060	30.027	32.054	33.948	36.171
		Student t news	0.26	0.58	0.66	0.60	0.25
		medium window	0.28	-0.47	-1.27	-1.37	-0.88
		long window	0.40**	-0.25	-0.53	-0.66	-0.64
		monthly update	-0.06	-0.08	-0.06	-0.16	-0.15
		daily update	-0.00	0.08	0.15	0.08	0.03
MSE	rv	base	3.667	4.337	5.146	5.856	6.525
		Student t news	-5.42	-5.32	-5.58	-6.05	-7.09
		medium window	7.56**	5.42**	5.04**	5.05	6.23
		long window	2.04**	-0.29	-0.68	-0.93	-0.89
		monthly update	0.66**	0.75**	0.49	0.25	-0.080
		daily update	0.09	0.05	0.14	-0.08	0.016

Table 5: NGARCH Estimation Strategy Assessment. For each loss function and volatility proxy, the table reports the out-of-sample losses at multiple horizons of the Vlab estimation strategy and the percentage gains derived by modifying the estimation strategy with (i) Student t innovations, (ii) medium estimation window, (iii) long estimation window, (iv) monthly estimation update and (v) daily estimation update. Asterisks beneath the percentage gains denote the significance of a Diebold-Mariano test under the null of equal or inferior predictive ability with respect to the baseline Vlab strategy (level of statistical significance denoted by *=10%, **=5%, ***=1%).

Loss	$\hat{\sigma}^2$	Model	Horizon 1 d	1 w	2 w	3 w	1 m
QLL	r^2	GARCH	1.462	1.481	1.520	1.574	1.645
		TGARCH	**1.415** ***	**1.442** ***	**1.478** ***	**1.547** ***	**1.624**
		EGARCH	1.420 ***	1.458 ***	1.505	1.592	1.684
		APARCH	1.417 ***	1.446 ***	1.485 ***	1.557 *	1.633
		NGARCH	1.422 ***	1.459 ***	1.498 *	1.574 ***	1.659
		HIS	1.518	1.541	1.577	1.626	1.692
QLL	rv	GARCH	0.273	0.310	0.343	0.373	0.414
		TGARCH	0.243 ***	0.289 ***	0.328	0.368	0.415
		EGARCH	**0.234** ***	**0.282** ***	**0.320**	**0.365**	**0.413**
		APARCH	0.249 ***	0.299 *	0.340	0.385	0.435
		NGARCH	0.244 ***	0.296 **	0.337	0.380	0.432
		HIS	0.314	0.337	0.360	0.385	0.420
MSE	r^2	GARCH	27.533	30.050	31.980	33.828	36.347
		TGARCH	**25.583** **	**28.874** *	**31.151**	**33.197**	**36.043** *
		EGARCH	26.746	31.328	33.984	35.912	37.828
		AGARCH	25.678 **	29.421 **	32.098	33.949	36.408
		NGARCH	27.060 ***	30.027	32.054	33.948	36.171
		HIS	29.862	32.817	34.189	35.646	37.649
MSE	rv	GARCH	4.357	4.998	5.984	6.901	7.950
		TGARCH	3.647 **	4.550	5.687	6.474 **	7.312 *
		EGARCH	**2.653** ***	**3.859** *	**4.758** *	**5.428** *	**6.085**
		AGARCH	3.384 ***	4.614	5.543 **	6.234 *	6.868
		NGARCH	3.667 ***	4.337 **	5.146 **	5.856 *	6.525 *
		HIS	4.632	5.002	5.687	6.390	7.415

Table 6: S&P 500 volatility prediction comparison of GARCH models from 2001 to 2008. For each loss and volatility proxy the table reports the out-of-sample loss at multiple horizons of the GARCH models as well as 60 days Historical Variance (HIS). Asterisks beneath GARCH models losses denote the significance of a Diebold-Mariano test under the null of equal or inferior predictive ability with respect to the GARCH model (level of statistical significance denoted by *=10%, **=5%, ***=1%). The best forecasting performance for each loss and proxy pair is highlighted in bold.

Loss	$\hat{\sigma}^2$	Model	Horizon				
			1 d	1 w	2 w	3 w	1 m
QLL	r^2	GARCH	1.664	1.788	2.304	2.593	3.324
		TGARCH	**1.461**	**1.560**	**1.985**	**2.311**	**2.875**
		EGARCH	1.701	1.992	2.809	3.453	4.386
		APARCH	1.565	1.695	2.233	2.631	3.306
		NGARCH	1.653	1.815	2.376	2.781	3.561
		HIS	2.267	2.522	3.091	3.426	4.001
QLL	rv	GARCH	0.344	0.417	0.676	0.798	1.630
		TGARCH	0.304	**0.353**	**0.590**	**0.672**	**1.380**
		EGARCH	**0.255**	0.405	0.798	1.058	1.954
		APARCH	0.293	0.381	0.652	0.810	1.552
		NGARCH	0.314	0.398	0.682	0.832	1.697
		HIS	0.485	0.624	0.946	1.170	1.877
MSE	r^2	GARCH	704.130	786.109	839.163	869.790	823.869
		TGARCH	**653.564**	**758.170**	**821.906**	**856.299**	**816.990**
		EGARCH	692.028	836.393	911.788	942.909	874.657
		AGARCH	658.838	776.873	852.660	880.727	830.446
		NGARCH	696.678	792.012	848.020	878.589	821.554
		HIS	771.160	869.098	907.807	928.700	867.009
MSE	rv	GARCH	102.628	116.548	144.042	166.307	193.307
		TGARCH	88.820	108.576	140.737	158.336	178.265
		EGARCH	**60.960**	**91.652**	**118.044**	**133.112**	**147.591**
		AGARCH	79.553	109.399	135.607	150.671	164.647
		NGARCH	88.126	102.324	124.960	140.879	156.213
		HIS	103.100	114.192	133.902	150.646	176.237

Table 7: S&P 500 volatility prediction comparison of GARCH models in Fall 2008. For each loss and volatility proxy the table reports the out-of-sample loss at multiple horizons of the GARCH models as well as 60 days Historical Variance (HIS). The best forecasting performance for each loss and proxy pair is highlighted in bold.

QL Loss

Model	Average Loss					Winning Frequency				
	1 d	1 w	2 w	3 w	1 m	1 d	1 w	2 w	3 w	1 m
	Exchange Rates									
GARCH	**1.970**	**2.038**	2.096	2.120	2.139	37	25	13	25	25
TGARCH	1.976	**2.038**	**2.093**	2.118	2.138	38	50	37	13	25
EGARCH	1.991	2.073	2.133	2.171	2.226	0	0	0	0	0
APARCH	1.987	2.050	2.099	**2.117**	**2.130**	25	25	25	37	50
NGARCH	1.971	2.046	2.097	2.120	2.140	0	0	25	25	0
	Equity Sectors									
GARCH	2.259	2.290	2.340	2.380	2.434	0	0	0	0	0
TGARCH	2.236 ***	**2.266** ***	**2.313** ***	**2.356** ***	**2.412** ***	44	56	67	78	56
EGARCH	2.270	2.303	2.349	2.392	2.447	0	0	0	0	11
APARCH	**2.235** ***	2.268 ***	2.314 ***	2.359 ***	2.414 ***	56	44	33	22	33
NGARCH	2.238 ***	2.277 ***	2.323 ***	2.369 ***	2.426 *	0	0	0	0	0
	International Equities									
GARCH	2.272	2.303	2.351	2.400	2.478	6	6	6	12	6
TGARCH	**2.253** ***	**2.289** ***	**2.337** ***	**2.389** **	**2.464** **	59	82	65	53	24
EGARCH	2.262	2.300	2.352	2.406	2.475	6	0	0	0	24
APARCH	2.259 ***	2.295 *	2.344	2.396	2.467	6	6	18	24	29
NGARCH	2.255 ***	2.293 **	2.342 *	2.394	2.469	23	6	12	12	18

Table 8: QL loss volatility prediction comparison of GARCH models from 2001 to 2008 across asset classes. For each asset class the table reports the out-of-sample average QL loss at multiple horizons as well as the relative frequency of cases in which a model achieved the best performance in a given asset class. Asterisks beneath losses denote the significance of a Diebold-Mariano test under the null of equal or inferior predictive ability with respect to the GARCH model (level of statistical significance denoted by *=10%, **=5%, ***=1%). The best average forecasting performance for each loss and proxy pair is highlighted in bold.

MSE Loss

Model	Average Loss					Winning Frequency				
	1 d	1 w	2 w	3 w	1 m	1 d	1 w	2 w	3 w	1 m
	Exchange Rates									
GARCH	4.018	4.193	4.282	4.655	4.870	13	25	38	38	37
TGARCH	3.964 ***	4.146 **	4.233 **	4.583 *	4.789	25	25	13	25	25
EGARCH	4.056	4.165	4.317	4.536	4.656	0	0	12	0	12
APARCH	**3.925**	**4.087** *	**4.172**	**4.469** *	**4.641**	50	50	37	37	25
NGARCH	4.010	4.188	4.273	4.640	4.848	12	0	0	0	0
	Equity Sectors									
GARCH	63.926	62.231	66.511	69.412	73.035	0	0	0	11	11
TGARCH	**61.779** ***	**61.236** *	**65.697**	**68.592** **	**72.707**	33	78	67	44	33
EGARCH	63.462	61.880	66.226	68.903	72.550	11	11	11	11	11
APARCH	61.991 **	61.522	66.067	69.032	73.040	56	11	11	22	0
NGARCH	63.385 *	62.078	66.223	69.141	72.456	0	0	11	11	44
	International Equities									
GARCH	75.983	77.004	82.756	88.570	95.092	0	0	6	12	6
TGARCH	**73.565** **	**75.907**	**81.941**	**87.358** **	93.789	65	88	59	53	12
EGARCH	75.406	79.831	85.803	90.045	94.266	0	0	0	6	18
APARCH	73.790 *	76.827	82.900	87.903	**93.665**	24	6	6	0	12
NGARCH	75.119 ***	76.793	82.469	87.866	93.467	12	6	29	29	53

Table 9: MSE loss volatility prediction comparison of GARCH models from 2001 to 2008 across asset classes. For each asset class the table reports the out-of-sample average MSE loss at multiple horizons as the relative frequency of cases in which a model achieved the best performance in a given asset class. Asterisks beneath losses denote the significance of a Diebold-Mariano test under the null of equal or inferior predictive ability with respect to the GARCH model (level of statistical significance denoted by *=10%, **=5%, ***=1%). The best average forecasting performance for each loss and proxy pair is highlighted in bold.

Model	Horizon				
	1 d	1 w	2 w	3 w	1 m
	Exchange Rates				
GARCH	1.950	2.102	2.459	2.638	3.250
TARCH	**1.925**	**2.081**	**2.454**	**2.636**	**3.244**
EGARCH	2.050	2.314	2.898	3.309	4.410
APARCH	1.928	2.100	2.489	2.690	3.366
NGARCH	1.937	2.097	2.456	2.648	3.264
	Equity Sectors				
GARCH	2.346	2.305	2.723	3.056	4.057
TARCH	**2.217**	**2.120**	**2.523**	**2.834**	**3.870**
EGARCH	2.314	2.264	2.710	3.036	4.007
APARCH	2.239	2.158	2.579	2.908	3.934
NGARCH	2.310	2.258	2.705	3.057	4.085
	Equity Sectors				
GARCH	2.234	2.146	2.749	3.409	4.447
TARCH	**2.153**	**2.054**	**2.574**	**3.349**	**4.250**
EGARCH	2.259	2.321	3.094	4.020	4.997
APARCH	2.166	2.105	2.652	3.405	4.280
NGARCH	2.179	2.098	2.659	3.407	4.372

Table 10: QL loss at multiple horizons volatility prediction comparison of GARCH models in Fall 2008 across asset classes. For each asset class the table reports the out-of-sample average. The best average forecasting performance for each loss and proxy pair is highlighted in bold.

Model	Horizon				
	1 d	1 w	2 w	3 w	1 m
	Exchange Rates				
GARCH	110.044	114.499	116.987	127.002	132.491
TARCH	108.383	113.066	115.490	124.790	129.967
EGARCH	109.409	113.732	118.584	123.752	126.294
APARCH	**107.159**	**111.600**	**114.081**	**121.754**	**125.944**
NGARCH	109.882	114.285	116.648	126.482	131.721
	Equity Sectors				
GARCH	1405.665	1426.406	1534.140	1570.666	1531.705
TARCH	**1358.083**	**1414.462**	**1530.184**	**1564.682**	**1535.560**
EGARCH	1397.247	1445.153	1558.041	1595.051	1542.011
APARCH	1365.171	1423.885	1542.536	1577.000	1542.832
NGARCH	1400.709	1434.603	1539.774	1573.340	1518.103
	International Equities				
GARCH	1746.705	1760.143	1919.864	2060.783	1970.426
TARCH	**1680.130**	**1730.626**	**1899.445**	**2027.356**	**1935.291**
EGARCH	1742.464	1861.786	2031.493	2124.680	1966.910
APARCH	1688.682	1762.531	1932.974	2048.238	1936.920
NGARCH	1730.617	1762.325	1920.797	2048.917	1932.315

Table 11: MSE loss at multiple horizons volatility prediction comparison of GARCH models in Fall 2008 across asset classes. For each asset class the table reports the out-of-sample average. The best average forecasting performance for each loss and proxy pair is highlighted in bold.

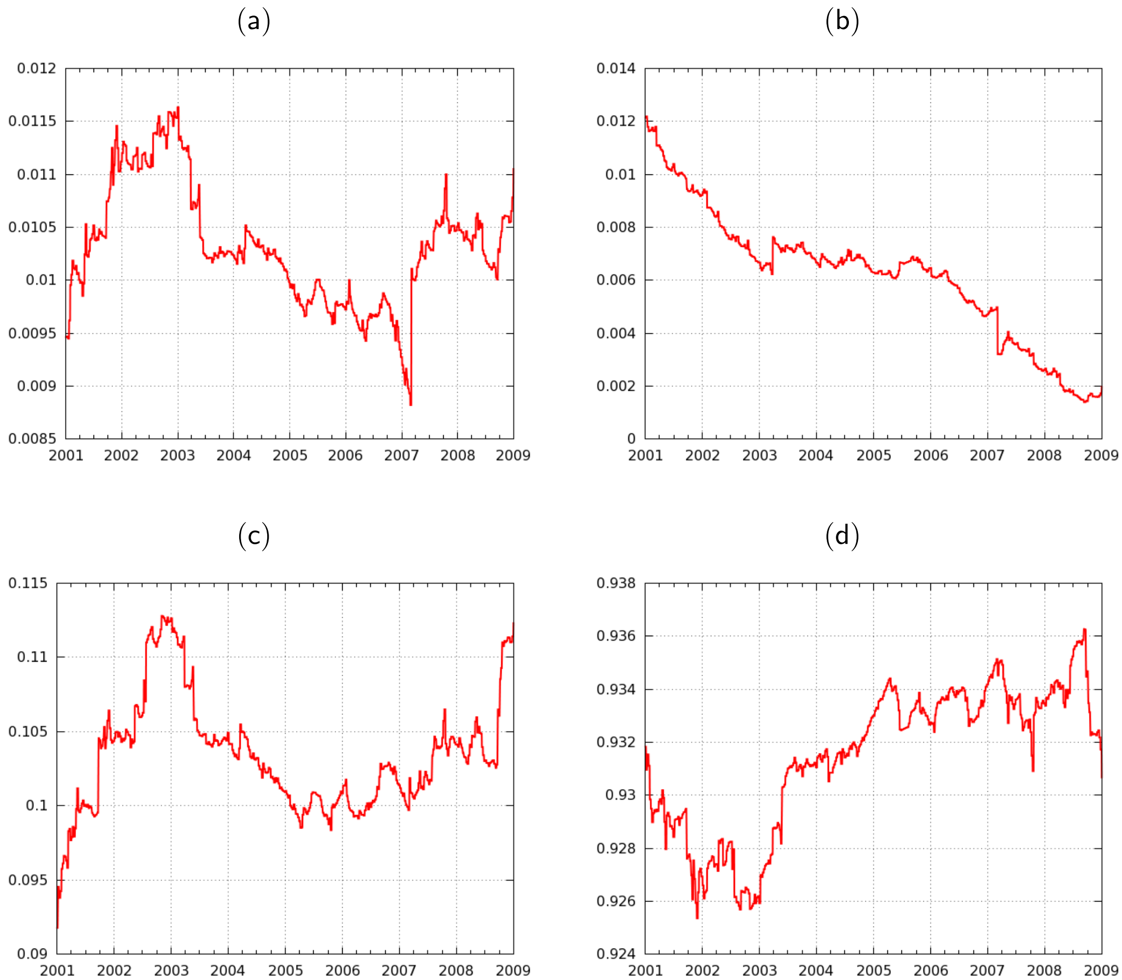

Figure 1: TARCH parameter estimates series. The graph plots the series of the ω (a), α (b), γ (c), β (d) parameters estimates obtained by the base estimation strategy using the TARCH model from 2001 to 2008.