EC4304 Final Report

Forecasting the volatility of Tesla stocks

Rooban Pannerselvam	A0218170M	e0544206@u.nus.edu
Chew Shi Zhan	A0216042W	e0538143@u.nus.edu
Lim Tze Hao	A0188738H	tzehao.lim@u.nus.edu
Mohamed Jasim	A0216261N	jasim@u.nus.edu

Abstract:

Our aim is volatility forecasting, which is the crux to a multitude of finance applications. We will compare EGARCH, HAR-RV, HAR-RAV, a simple average forecast combination, and Granger-Ramanathan forecast combination models to the GARCH model at a 1, 5 and 20 step horizon under QLIKE loss and MSE by using 5 minute intra-daily data of Tesla Inc. stock to construct Realised Volatility and Realised Absolute Value as predictors. We found that a Granger-Ramanathan combination forecast beats out both the GARCH(1,1) and a benchmark simple average combination forecast, under both the MSE and QLike loss functions when tested against them via a Diebold-Mariano test for all horizons considered. As for the single models, HAR-RAV conclusively beats out the benchmark GARCH(1, 1) model at the 5 and 20 step horizons under both MSE and QLIKE loss.

Keywords: GARCH, HAR, Granger-Ramanathan, Volatility, QLIKE, MSE, RV, RAV

1. INTRODUCTION

We decided to test the GARCH(1,1) model as it is known to be a benchmark model that is hard to beat. In doing so, we may be able to gain some insight into where certain tweaks or alternative models may beat the benchmark model and in which scenarios it may be better to rely on an alternative model instead.

This could assist greatly in future forecasts as we will be able to better discern which models to use with respect to the data presented. Tesla is famous for having its stock value skyrocketing during the COVID-19 pandemic, and it would be interesting to observe how this would affect the future stock price/volatility in the near future, especially so since we are still in the midst of a pandemic. Some notable papers which we referenced from includes Forsberg & Ghysels (2006), which we used to collate the definitions and methods of calculating RV and RAV and Corsi(2009), which we used to garnered the knowledge of executing 5-step ahead and 20-step ahead forecasts with HAR models.

Following our analysis, the results have allowed us to conclude the improved forecasting accuracy of a Granger-Ramanathan model at a 1, 5 and 20 step ahead forecast compared to a GARCH(1,1) model, along with the other alternative methods beating out the GARCH(1,1) benchmark model at various forecast horizons. We will be briefly explaining our methodology, as well as the preparation of our data, followed by delving into the results we have obtained from each of the aforementioned models. Finally, we will be comparing the forecast errors of each model and concluding with the best models for our dataset.

2. METHODOLOGY

We adopted a hybrid window for GARCH(1, 1) and EGARCH(1, 1) models to account for the spikes in volatility, rolling window for reestimating the model every 300 observations and a fixed window to produce forecasts while a recursive window method for both HAR-RV and HAR-RAV models to take advantage of the variance reduction effects by including more data. For forecast combinations, we will use the Granger-Ramanathan combination as well as simple average combination as the baseline. We will produce forecasts with prediction horizons of 1, 5 and 20 steps, representative of 1 trading day, week and month respectively, and compare the performance of the aforementioned models for each considered horizon using Diebold-Mariano Test for MSE and QLIKE loss functions. Both of these loss functions have the benefit of being robust to measurement errors as well as being independent to the unit of measurement (Patton 2011). This is especially useful in this context as variance cannot be directly observed in the stock market. Any error resulting from the estimator (RV/RAV) that we have used can be minimised by the usage of these particular loss functions. The QLIKE loss function is defined as

$$QLIKE = \frac{Y}{\hat{Y}} - log(\frac{Y}{\hat{Y}}) - 1$$

Where Y is the RV/RAV and $\hat{Y}$ is the variance forecast

While MSE is defined as

$$MSE = N^{-1} \sum_{i=1}^{N} (RV_{i,i+H} - \widehat{RV}_{i,i+H})^{2}$$

Where N = Sample size, RV=Realised Volatility

3. DATA

For our project, we will be using a 5-minute Tesla stock price dataset sourced from FirstRate Data. This dataset covers the open, high, low and close prices along with volume over a period of 12 years from 2010 to 2022. We have cleaned the dataset by omitting any values beyond 3 standard deviations. The volatility of stocks over a day, integrated volatility, is generally modelled as follows:

$$IV_T = \int_{t-1}^{t} \sigma_s^2 \, ds \tag{1}$$

However, integrated volatility is not directly observable from this dataset, it has to be estimated. To get started with this estimation, we need to first calculate the intra-daily log returns:

$$r_{t,j}^{M} = \ln P_{t-j/M} - \ln P_{t-(j-1)/M} \tag{2}$$

In equation 2, t is the end time index, j is the start time index and M is the intra-daily sampling frequency. While M = 78 is generally expected of 5-minute stock price data, in the dataset M varies from this value on any particular day. This is because the dataset excludes minutes with zero trading volume. In addition, it includes after-hours trading data. Now the realised variance (RV) can be estimated by summing the squared intra-daily returns for the day:

$$RV_{t,t+D}^{M} = \sum_{j=1}^{MD} (r_{t,j}^{M})^2 \tag{3}$$

Realised absolute volatility (RAV) can also be estimated using the dataset in the following way:

$$RAV_{t,t+D}^{M} = \mu_1^{-1} M^{-1/2} \sum_{j=1}^{M} \left| r_{t,j}^{M} \right| \tag{4}$$

where $\mu_1 = \sqrt{2/\pi}$

It is also significant that we have identified events caused spikes in RV as shown in Fig. 1

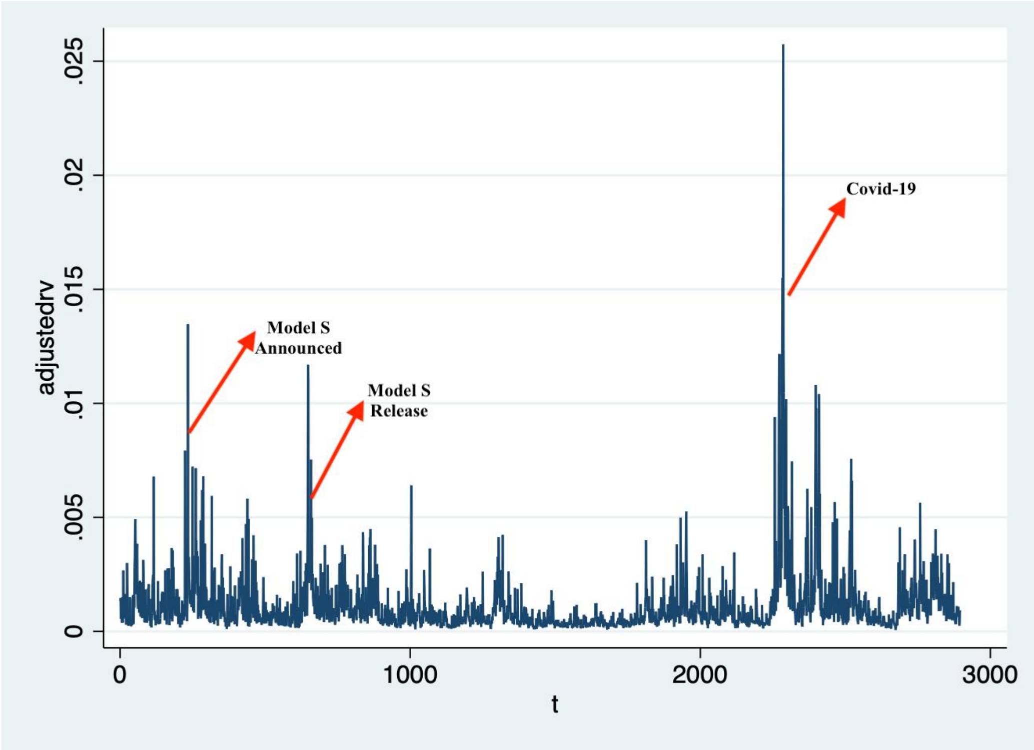


Fig. 1

4. EMPIRICAL REGRESSION MODELS

As aforementioned, GARCH(1, 1) proposed by Bollerslev (1986) is widely used in modelling economic and financial time series data, and has long been considered as a standard workhorse volatility model. GARCH is modelled by

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

Where **σ** = The conditional variance model,

ω = the variance intercept,

ε = the error term,

α and β = coefficients.

For an independent variable Y, it can be modelled using a conditional mean model. The variance of the error term from this model can then be modelled as a conditional variance model with the additional first lag of the variance giving rise to GARCH(1, 1). This is particularly useful in modelling volatility of stocks as volatility is likely to be serially correlated. For example, if the stock the market is volatile today, so it would be rational to assume that there will be a higher chance that it will be volatile tomorrow as well.

However, a stark weakness of GARCH(1, 1) is that it fails to model the asymmetric effect of positive and negative shocks in returns while empirically, volatility tends to increase significantly more post-negative shocks than positive shocks. This is also known as the “leverage effect”. As such, we have decided to consider EGARCH(1, 1) proposed by Nelson (1991). EGARCH(1, 1) is modelled by

$$ln(\sigma_t^2) = \omega + \beta ln(\sigma_{t-1}^2) + \alpha\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| + \gamma\frac{\varepsilon_{t-1}}{\sigma_{t-1}}$$

Where **σ** = The conditional variance model,

ω = the variance intercept,

ε = the error term,

α, β and ɣ = coefficients.
This allows volatility to be dependent on both absolute magnitude and sign of shocks, allowing for asymmetric response.

However, both GARCH(1, 1) and EGARCH(1, 1) are estimated using Maximum Likelihood Estimation (MLE), this poses mainly 2 problems, namely potential multiple local maxima, as well as potential flat regions and ridges thus making the estimates potentially not robust or even fail to converge entirely.

This motivates HAR-RV proposed by Corsi(2009), which summarises high persistence in volatility by incorporating daily, weekly, and monthly averages of RV. HAR-RV is modelled by

$$RV_{t,t+1} = \alpha + \beta_D RV_t + \beta_W RV_{t-5,t} + \beta_M RV_{t-22,t} + \varepsilon_{t+1}$$

Where RV = Realised Volatility.

ε = the error term,

α and β = coefficients.

Other than avoiding multiple long lags of RV to model long memory behaviour of volatility, a crucial strength of HAR-RV is that it is built on the Ordinary Least Squares (OLS) method, which avoids convergence issues present in MLE models altogether.

Adding on to the HAR-RV model, Forsberg & Ghysels (2007) proposed using Realised Absolute Values(RAV) as predictors instead of RV as RAV is more persistent than RV. The nature of stock markets means that discrete jumps are common. Forsberg and Ghysels(2007) noted that the persistence of squared returns, which is used often as a measure of volatility, tends to fade quickly. The RAV model proposed by them is at least asymptotically immune to the presence of jumps and helps to mitigate this problem. Hence, it is useful to consider RAV in our analysis as well. HAR-RAV is modelled as

$$RV_{t,t+1} = \alpha + \beta_D RAV_t + \beta_W RAV_{t-5,t} + \beta_M RAV_{t-22,t} + \varepsilon_{t+1}$$

Where RAV = Realised Absolute Value.

ε = the error term,

α and β = coefficients.

Diebold noted in his book that models are just an intentional abstraction of a much more complex reality, and he suggested that most models are likely to be misspecified. Hence, forecast combinations are often used to improve forecasts in practice. This motivates us to consider some possible forecast combinations, such as the Simple Average combined model and the Granger-Ramanathan combined model.

By combining the aforementioned models by taking the simple average, the variance of the forecast can be reduced, assuming unbiasedness and uncorrelated variances of the forecasts. Which can be shown by

$$Var(f) = \frac{\sigma_1^2+\sigma_2^2}{4}$$

Augmenting this idea, Granger and Ramanathan (1984) suggested using empirical weights based on out-of-sample forecast variances. While this was derived under the assumption of uncorrelated forecasts, it can work well empirically, as tested by Jay S. Holmen (1987). Granger-Ramanathan combination will thus be compared to the Simple Average Combination Model as the baseline. The weights of the forecasts for Granger-Ramanathan combination is as follows

$$w_m = \frac{\hat{\sigma}_m^{-2}}{\hat{\sigma}_1^{-2} + \hat{\sigma}_2^{-2} + ... + \hat{\sigma}_M^{-2}}$$

5. DISCUSSION

Thus far, we have taken a look at six models to forecast the volatility of Tesla stocks. In this section, we will be evaluating these out-of-sample forecasts produced by these models based on the Diebold Mariano EPA test with root mean squared loss and QLIKE loss.

For the GARCH(1, 1) and EGARCH(1, 1) models, we adopted a hybrid window method of window size 1200, 1196, and 1181 observations for 1, 5 and 20 step horizons respectively to estimate the model, reestimating the model every 300 observations as well as a fixed window to produce forecasts. The rolling window method was adopted in order to take into account the spikes of RV due to shocks as shown in the RV plot in Fig. 1

For HAR-RV and HAR-RAV models, we initially used a window of size 1198 to estimate the 1-step horizon for the first observation. For the subsequent observations, the window size expands recursively to include more observations. Each HAR-RV and HAR-RAV models is then reestimated to account for the increasing observations. Although using additional data can produce biased estimates,but the variance reduction from using additional data outweighs the biasedness. Similarly, 1194 and 1179 observations are used initially to estimate the 5-step horizon and the 20-step horizon with recursive windows. The difference in the initial number of observations between the GARCH variants and HAR variants can be attributed to the usage of lags in the HAR models.

5.1 GARCH(1, 1) & EGARCH(1, 1)

Diebold-Mariano Test under MSE loss			
Forecast Step (x-step ahead)	Coefficient	p-value	Significance at 5% level
1	$6.53 \text{x} 10^{-7}$	0.113	No
5	$-4.88 \text{x} 10^{-7}$	0.157	No
20	$-3.50 \text{x} 10^{-7}$	0.092	No

Diebold-Mariano Test under QLike loss			
Forecast Step (x-step ahead)	Coefficient	p-value	Significance at 5% level
1	-0.152	0.045	Yes
5	-0.140	0.061	No
20	-0.134	0.030	Yes

Under MSE loss, the Diebold-Mariano test does not reject the null that both models forecast that have equal performance for 1, 5 and 20 steps ahead at a 5% significance level. Under the QLIKE loss function, the Diebold-Mariano test concludes that the GARCH(1,1) model outperforms the EGARCH(1,1) model. However, this result is only significant at a 5% level for the 1-step ahead and 20-step ahead forecasts. Thus, it would not be worth choosing EGARCH for variance forecasting given the poor performance of its forecasts and the additional computation it requires.

5.2 GARCH(1, 1) & HAR-RV

Diebold-Mariano Test under MSE loss			
Forecast Step (x-step ahead)	Coefficient	p-value	Significance at 5% level
1	$6.81x10^{-7}$	0.038	Yes
5	$2.42x10^{-7}$	0.132	No
20	$1.85x10^{-7}$	0.129	No

Diebold-Mariano Test under QLike loss			
Forecast Step (x-step ahead)	Coefficient	p-value	Significance at 5% level
1	0.104	0.000	Yes
5	0.0990	0.000	Yes
20	0.0481	0.011	Yes

Under MSE loss, the Diebold-Mariano test suggests that Under the QLIKE loss function, the HAR-RV model beats out the GARCH(1,1) model yet again at every forecast horizon. However, we are now able to reject all the null hypotheses at a 5% level.

5.3 GARCH(1, 1) & HAR-RAV

Diebold-Mariano Test under MSE loss			
Forecast Step (x-step ahead)	Coefficient	p-value	Significance at 5% level
1	6.23×10^{-7}	0.015	Yes
5	3.47×10^{-7}	0.002	Yes
20	2.45×10^{-7}	0.017	Yes

Under MSE, the Diebold-Mariano test concludes that HAR-RAV outperforms GARCH(1,1) at 1,5 and 20 step forecast horizons. Furthermore, they are all significant at a 5% level.

Diebold-Mariano Test under QLike loss			
Forecast Step (x-step ahead)	Coefficient	p-value	Significance at 5% level
1	-0.107	0.485	No
5	0.102	0.000	Yes
20	0.0584	0.001	Yes

Under the QLIKE loss function, we see that HAR-RAV outperforms GARCH(1,1) in a 5 and 20 step forecast horizon while being significant at a 5% level. However, we note that the HAR-RAV seems to fall short at a 1 step forecast horizon. This arises due to an error we encountered, which we had to force the algorithm through. We will elaborate on this in the limitations section later on.

5.4 GARCH(1, 1) & Simple Average Model

Diebold-Mariano Test under MSE loss			
Forecast Step (x-step ahead)	Coefficient	p-value	Significance at 5% level
1	$4.40\text{x}10^{-7}$	0.003	Yes
5	$2.16\text{x}10^{-7}$	0.000	Yes
20	$1.37\text{x}10^{-7}$	0.107	No

Under MSE loss, the Diebold-Mariano test concludes that Simple Average outperforms GARCH(1,1) at 1 and 5 step forecast horizons.However, this result is only significant at a 5% level for the 1-step ahead and 20-step ahead forecasts..

Diebold-Mariano Test under QLike loss			
Forecast Step (x-step ahead)	Coefficient	p-value	Significance at 5% level
1	0.0868	0.000	Yes
5	0.0668	0.000	Yes
20	0.0233	0.103	No

Under the QLike loss function, the Simple Average Model Combination forecast beats out the GARCH(1,1) model at every forecast step we have examined above.

5.5 GARCH(1, 1) & Granger-Ramanathan Combination

Diebold-Mariano Test under MSE loss			
Forecast Step (x-step ahead)	Coefficient	p-value	Significance at 5% level
1	$6.94\text{x}10^{-7}$	0.029	Yes
5	$3.47\text{x}10^{-7}$	0.002	Yes
20	$2.45\text{x}10^{-7}$	0.016	Yes

Under the MSE loss function, the Granger-Ramanathan Combination forecast beats out the GARCH(1,1) model at every forecast step we have examined above.

Diebold-Mariano Test under QLike loss			
Forecast Step (x-step ahead)	Coefficient	p-value	Significance at 5% level
1	0.113	0.000	Yes
5	0.102	0.000	Yes
20	0.0581	0.001	Yes

Under the QLike loss function, the Granger-Ramanathan Combination forecast beats out the GARCH(1,1) model at every forecast step we have examined above.

5.6 Simple Average Model and Granger-Ramanathan Combination

Diebold-Mariano Test under MSE loss			
Forecast Step (x-step ahead)	Coefficient	p-value	Significance at 5% level
1	$2.54x10^{-7}$	0.158	No
5	$1.31x10^{-7}$	0.105	No
20	$1.08x10^{-7}$	0.020	Yes

Under the MSE loss function, the Granger-Ramanathan Combination forecast beats out the Simple Average Combination Model at every forecast step we have examined above. However, this result is only significant at the 20 step ahead forecast horizon.

Diebold-Mariano Test under QLike loss			
Forecast Step (x-step ahead)	Coefficient	p-value	Significance at 5% level
1	0.0261	0.000	Yes
5	0.0356	0.004	Yes
20	0.0349	0.000	Yes

Under the QLike loss function, the Granger-Ramanathan Combination forecast beats out the Simple Average Combination Model at every forecast step we have examined above.

6. CONCLUSION

Overall, we have concluded that, most interestingly, the Granger-Ramanathan combination model beats out both the benchmark forecasting models for this dataset at a 1,5 and 20 steps ahead forecast horizon under both MSE and QLike Loss functions, while for single model forecast, the HAR-RAV model is better than GARCH(1, 1) in the case of 5 and 20 step horizons under both MSE and QLIKE loss.

Given a longer timeframe, it may have been beneficial to conduct this test on multiple datasets of various stock volatilities in order to eliminate any bias toward a particular stock, in this case Tesla Inc.

As GARCH and EGARCH are estimated using Maximum Likelihood Estimation, it is important to note that the coefficients may not be fully robust. Additionally, the RV 20-step ahead forecast produced has converged to the unconditional variance in some cases. As such, it may be useful to consider forecast horizons lower than 20.

We encountered an issue while running the Diebold-Mariano test under the QLIKE loss function between GARCH(1,1) and HAR-RAV, in which we encountered null values for the QLIKE function. This is likely caused by an undefined log term arising due to negative out-of-sample forecast values.

So far in this paper, we managed to consider linear combinations of forecasts, and those methods have brought great improvements to the forecasts. Beyond these, it would be worth studying non-linear combinations of forecasts to further beat the standards set by GARCH(1, 1) and the Granger-Ramanathan combined model. In addition, there has been an increasing interest in applying machine learning in forecasting daily volatility, as demonstrated by Zhang et al.(2022). They noted that neural networks dominate linear regressions and are able to uncover complex latent interactions among the variables. Given the increase in computational capability recently, it would also be intriguing to explore their usage in forecasting volatility in the future and model combinations.

REFERENCES

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196. https://doi.org/10.1093/jjfinec/nbp001

Diebold, F. X. (2008). *Elements of forecasting*. South-Western/Cengage Learning.

Forsberg, L., & Ghysels, E. (2006). Why do absolute returns predict volatility so well? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.929494

Granger, C. W., & Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, *3*(2), 197–204. https://doi.org/10.1002/for.3980030207

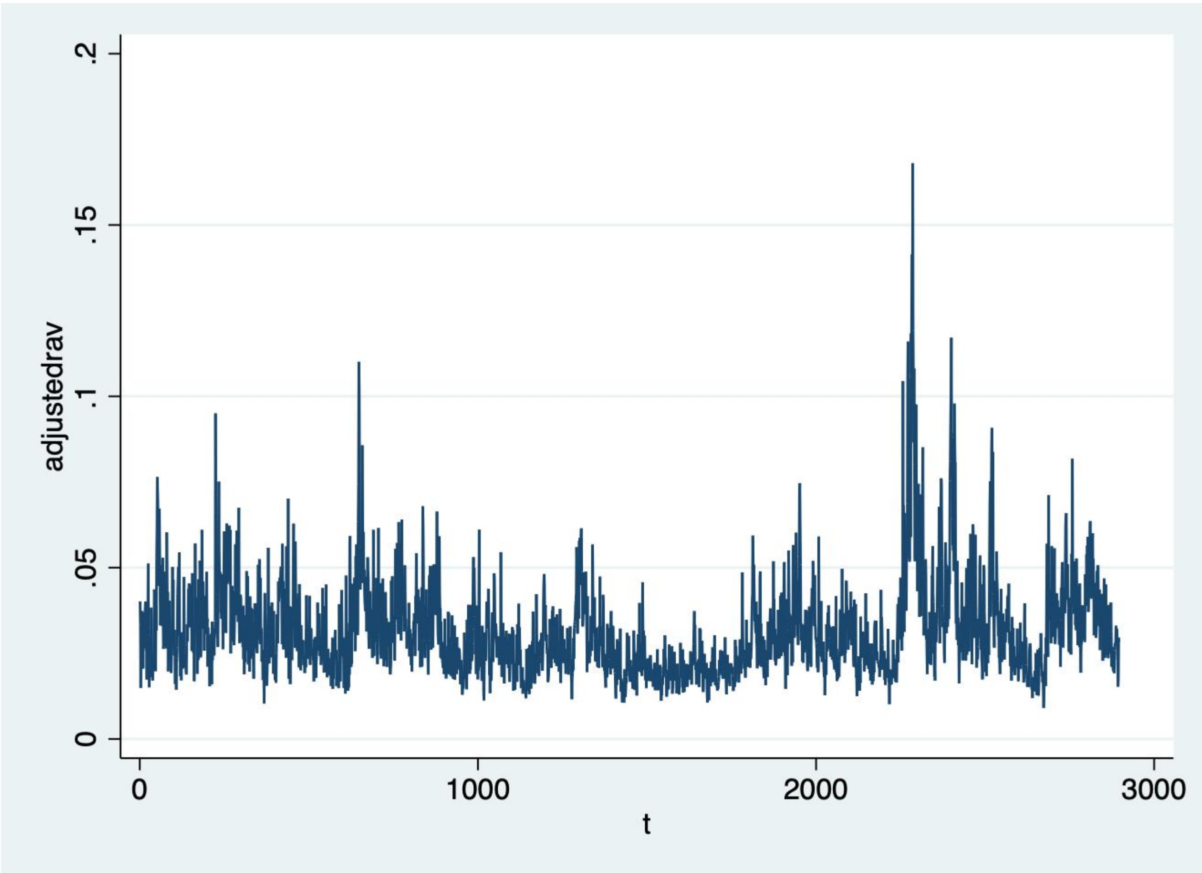
Jay S. Holmen (1987). A note on the value of combining short-term earnings forecasts: A test of Granger and Ramanathan, International Journal of Forecasting, https://doi.org/10.1016/0169-2070(87)90005-7

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, *160*(1), 246–256. https://doi.org/10.1016/j.jeconom.2010.03.034

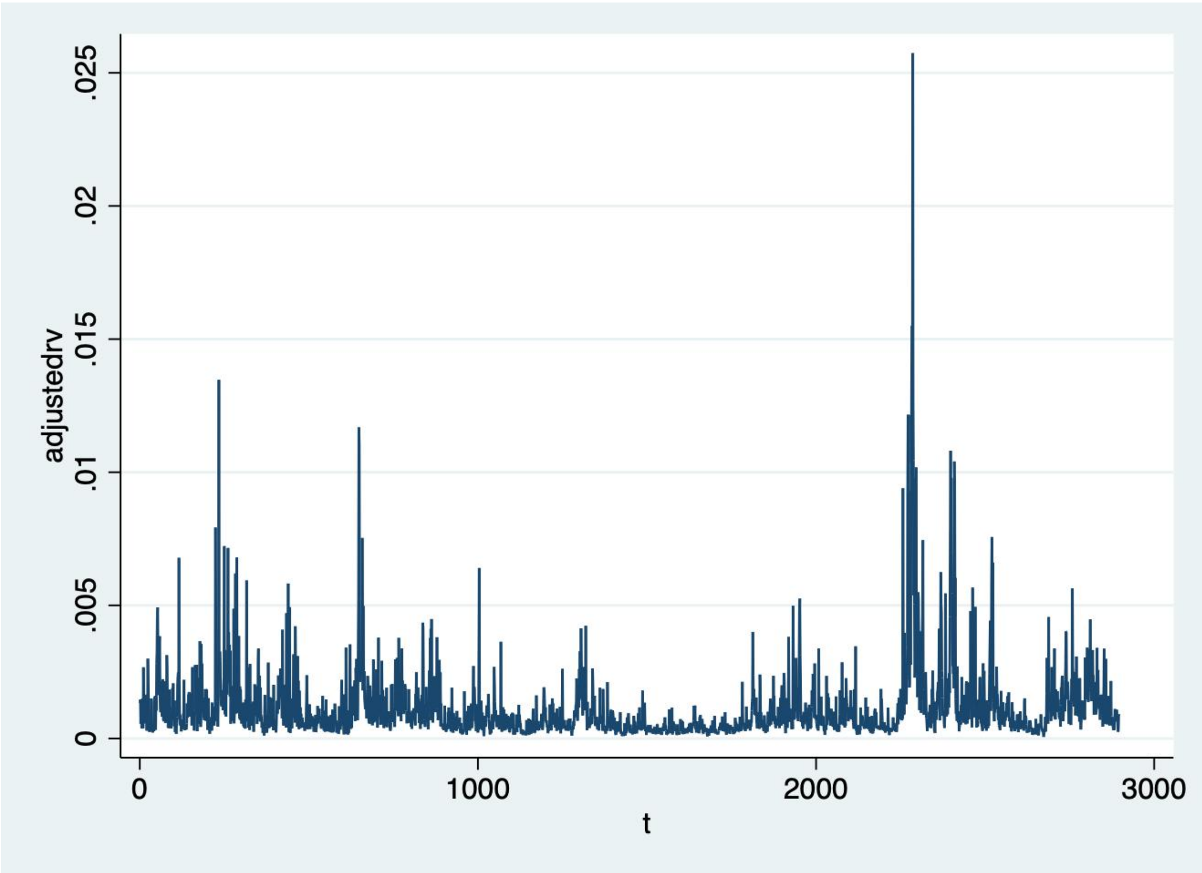
Zhang, Chao and Zhang, Yihuang and Cucuringu, Mihai and Qian, Zhongmin, Volatility Forecasting with Machine Learning and Intraday Commonality (May 1, 2022). https://ssrn.com/abstract=4022147

APPENDIX

Adjusted RAV Graph



Adjusted RV



1-Step Ahead Forecast Model's RMSE

Results

```
. *Compare RMSE's across all forecasts including combined forecast for 1-Step Ah
> ead: + Tze Hao Model

. sum adjustedrav_rmse_1step adjustedrv_rmse_1step rmse_garch_h1 rmse_egarch_h1
```

Variable	Obs	Mean	Std. dev.	Min	Max
adjustedra..	2,896	.0010676	0	.0010676	.0010676
adjustedrv..	2,896	.0010402	0	.0010402	.0010402
rmse_garch~1	2,896	.0013273	0	.0013273	.0013273
rmse_egarc~1	2,896	.001554	0	.001554	.001554

.

<u>5-Step Ahead Forecast Model's RMSE</u>

```
. *Compare RMSE's across all forecasts including combined forecast for 5-Step Ah
> ead:

. sum adjustedrav_rmse_5step adjustedrv_rmse_5step rmse_egarch_h5 rmse_garch_h5
```

Variable	Obs	Mean	Std. dev.	Min	Max
adjustedra..	2,896	.001312	0	.001312	.001312
adjustedrv..	2,896	.0013514	0	.0013514	.0013514
rmse_egarc~5	2,896	.0015986	0	.0015986	.0015986
rmse_garch~5	2,896	.0014378	0	.0014378	.0014378

<u>20-Step Ahead Forecast Model's RMSE</u>

```
. *Compare RMSE's across all forecasts including combined forecast for 20-Step A
> head:

. sum adjustedrav_rmse_20step adjustedrv_rmse_20step rmse_egarch_h20 rmse_garch_
> h20
```

Variable	Obs	Mean	Std. dev.	Min	Max
adjustedra..	2,896	.0014836	0	.0014836	.0014836
adjustedrv..	2,896	.0015039	0	.0015039	.0015039
rmse_egar~20	2,896	.0016718	0	.0016718	.0016718
rmse_garc~20	2,896	.0015636	0	.0015636	.0015636

.

<u>All Combined 1-step Ahead RMSE (Simple Average and Granger Ramanathan)</u>

```
.
. *Now we are down to three forecasts(With coeffieciemt all +ve). Use weights to
>  generate combined forecasts:

. gen ycombine_1step = adjustedravfithar_1step*0.2572083 + adjustedrvfithar_1ste
> p*0.7136624  + gar_varh1*0.0291293
(1,222 missing values generated)

. egen ecombine_1step = mean((actrv-ycombine_1step)^2)

. gen rmsecombine_1step =sqrt(ecombine_1step)

.
. *Examine RMSE's of 1-Step of Simple Average and Granger-Ramanathan so far:

. sum rmsecomb_1 rmsecombine_1step
```

Variable	Obs	Mean	Std. dev.	Min	Max
rmsecomb_1	2,896	.0011501	0	.0011501	.0011501
rmsecombin~p	2,896	.0010339	0	.0010339	.0010339

```
.
```

Results

```
. cnsreg actrv adjustedravfithar_1step adjustedrvfithar_1 gar_varh1, constraints
> (1) noconstant

Constrained linear regression                       Number of obs =  1,674
                                                    Root MSE      = 0.0010

 ( 1)  adjustedravfithar_1step + adjustedrvfithar_1step + gar_varh1 = 1
```

actrv	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
adjustedrav..	.2572083	.0813084	3.16	0.002	.0977313	.4166853
adjustedrvf..	.7136624	.06809	10.48	0.000	.5801117	.847213
gar_varh1	.0291293	.0337179	0.86	0.388	-.0370045	.0952631

Combined 5 step ahead

```
. *** Combined 5-Step Ahead Forecaet***

.
. *Unconstrained regression first to see which forecasts are collinear. If find
> any, drop them(For 5-step ahead): Add Tze Hao Model

. reg actrv adjustedravfithar_5step adjustedrvfithar_5step egar_varh5 gar_varh5,
>  noconstant
```

Source	SS	df	MS
Model	.003169957	4	.000792489
Residual	.002846167	1,670	1.7043e-06
Total	.006016124	1,674	3.5939e-06

Number of obs	=	1,674
F(4, 1670)	=	465.00
Prob > F	=	0.0000
R-squared	=	0.5269
Adj R-squared	=	0.5258
Root MSE	=	.00131

actrv	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
adjustedrav..	1.291841	.1472772	8.77	0.000	1.002974	1.580709
adjustedrvf..	-.2976209	.1202904	-2.47	0.013	-.5335567	-.0616851
egar_varh5	-.2698671	.0622923	-4.33	0.000	-.3920462	-.1476879
gar_varh5	.2034957	.0755561	2.69	0.007	.055301	.3516904

```
.
. *gar_varh5 contains negative weight, so we remove it, keeping adjustedravfitha
> r_5step only

. reg actrv adjustedravfithar_5step

      Source |       SS           df       MS      Number of obs   =     1,674
-------------+----------------------------------   F(1, 1672)      =    669.69
       Model |  .001154051         1  .001154051   Prob > F        =    0.0000
    Residual |  .002881285     1,672  1.7233e-06   R-squared       =    0.2860
-------------+----------------------------------   Adj R-squared   =    0.2856
       Total |  .004035335     1,673  2.4120e-06   Root MSE        =    .00131

------------------------------------------------------------------------------
        actrv | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
adjustedrav.. |   1.014752   .0392123    25.88   0.000     .9378412    1.091662
        _cons |  -.0000189   .0000535    -0.35   0.724    -.0001237     .000086
------------------------------------------------------------------------------

.
. *With only one forecasts, we will use adjustedravfithar_5step to forecast only
> :

. gen ycombine_5step= adjustedravfithar_5step
(1,222 missing values generated)

. egen ecombine_5step = mean((actrv-ycombine_5step)^2)

. gen rmsecombine_5step=sqrt(ecombine_5step)

.
. *Examine RMSE's of 5-Step of Simple Average and Granger-Ramanathan so far:

. sum rmsecomb_5 rmsecombine_5step

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
  rmsecomb_5 |      2,896     .001361           0    .001361    .001361
rmseco~5step |      2,896     .001312           0    .001312    .001312
```

```
.
. *We found 2(adjustedravfithar_5step,adjustedrvfithar_5step), so we remove both
>  and reg again

. reg actrv adjustedravfithar_5step gar_varh5, noconstant

      Source |       SS           df       MS      Number of obs   =     1,674
-------------+----------------------------------   F(2, 1672)      =    909.44
       Model |  .003134625         2  .001567313   Prob > F        =    0.0000
    Residual |  .002881499     1,672  1.7234e-06   R-squared       =    0.5210
-------------+----------------------------------   Adj R-squared   =    0.5205
       Total |  .006016124     1,674  3.5939e-06   Root MSE        =    .00131

------------------------------------------------------------------------------
       actrv | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
adjustedrav..|   1.004373   .0608867    16.50   0.000     .8849503    1.123795
    gar_varh5|  -.0006757   .0545812    -0.01   0.990    -.1077304     .106379
------------------------------------------------------------------------------
```

Combined 20 step

```
. *Unconstrained regression first to see which forecasts are collinear. If find
> any, drop them(For 20-step ahead): Add Tze Hao Model

. reg actrv adjustedravfithar_20step adjustedrvfithar_20step egar_varh20 gar_var
> h20, noconstant

      Source |       SS           df       MS      Number of obs   =     1,674
-------------+----------------------------------   F(4, 1670)      =    282.96
       Model |  .002430285         4  .000607571   Prob > F        =    0.0000
    Residual |  .003585839     1,670  2.1472e-06   R-squared       =    0.4040
-------------+----------------------------------   Adj R-squared   =    0.4025
       Total |  .006016124     1,674  3.5939e-06   Root MSE        =    .00147

------------------------------------------------------------------------------
       actrv | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
adjustedrav.. |   2.202767   .2855977     7.71   0.000     1.642599    2.762934
adjustedrvf.. |  -1.163791   .2423047    -4.80   0.000    -1.639044   -.6885379
  egar_varh20 |  -.4443564   .0733637    -6.06   0.000    -.5882508   -.3004619
   gar_varh20 |   .3382157   .1043403     3.24   0.001     .1335642    .5428672
------------------------------------------------------------------------------

.
. *We find 2 models with negative weights, so we remove the two and regress agai
> n with the remaining

. reg actrv adjustedravfithar_20step gar_varh20, noconstant

      Source |       SS           df       MS      Number of obs   =     1,674
-------------+----------------------------------   F(2, 1672)      =    529.06
       Model |  .002331699         2   .00116585   Prob > F        =    0.0000
    Residual |  .003684425     1,672  2.2036e-06   R-squared       =    0.3876
-------------+----------------------------------   Adj R-squared   =    0.3868
       Total |  .006016124     1,674  3.5939e-06   Root MSE        =    .00148

------------------------------------------------------------------------------
       actrv | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
adjustedrav.. |   .9648108   .1051105     9.18   0.000     .7586488    1.170973
   gar_varh20 |   .0246183    .083211     0.30   0.767    -.1385905    .1878271
------------------------------------------------------------------------------
```

```
.
. *The remaining 2 models has positive weight, so we set the constraint and esti
> mate for the weights

. constraint 2 adjustedravfithar_20step+gar_varh20=1

. cnsreg actrv adjustedravfithar_20step gar_varh20, constraints(2) noconstant

Constrained linear regression                        Number of obs =  1,674
                                                     Root MSE      = 0.0015

 ( 1)  adjustedravfithar_20step + gar_varh20 = 1

------------------------------------------------------------------------------
       actrv | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
adjustedrav..|   .9877314   .0723478    13.65   0.000     .8458297    1.129633
  gar_varh20 |   .0122686   .0723478     0.17   0.865    -.129633     .1541703
------------------------------------------------------------------------------


.
. *Now we are down to two forecasts. Use weights to generate combined forecasts:

. gen ycombine_20step= adjustedravfithar_20step*0.9648108  + gar_varh20*0.024618
> 3
(1,222 missing values generated)

. egen ecombine_20step = mean((actrv-ycombine_20step)^2)

. gen rmsecombine_20step=sqrt(ecombine_20step)

.
. *Examine RMSE's of 20-Step Simple Average and Granger-Ramanathan so far:

. sum rmsecomb_20 rmsecombine_20step

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
 rmsecomb_20 |      2,896    .0015197           0   .0015197   .0015197
rmseco~0step |      2,896    .0014836           0   .0014836   .0014836
```

Forecast errors

```
.
. sum fe_garch_h1 fe_rav_h1
```

Variable	Obs	Mean	Std. dev.	Min	Max
fe_garch_h1	1,675	-.0002584	.0013023	-.0041893	.02208
fe_rav_h1	1,674	-5.73e-06	.0010679	-.0035471	.0183925

```
. sum fe_garch_h5 fe_rav_h5
```

Variable	Obs	Mean	Std. dev.	Min	Max
fe_garch_h5	1,675	-.0002854	.0014096	-.0033642	.022781
fe_rav_h5	1,674	-2.79e-06	.0013124	-.005122	.0216383

```
. sum fe_garch_h20 fe_rav_h20
```

Variable	Obs	Mean	Std. dev.	Min	Max
fe_garch_h20	1,675	-.000271	.0015404	-.0040979	.0240624
fe_rav_h20	1,674	.0000196	.0014839	-.0043873	.0239651

```
. *examine bias and RMSE

. sum fe_garch_h1 fe_egarch_h1
```

Variable	Obs	Mean	Std. dev.	Min	Max
fe_garch_h1	1,675	-.0002584	.0013023	-.0041893	.02208
fe_egarch_h1	1,675	-1.74e-06	.0015544	-.006534	.0248793

```
. sum fe_garch_h5 fe_egarch_h5
```

Variable	Obs	Mean	Std. dev.	Min	Max
fe_garch_h5	1,675	-.0002854	.0014096	-.0033642	.022781
fe_egarch_h5	1,675	-.0000435	.0015985	-.003623	.0249744

```
. sum fe_garch_h20 fe_egarch_h20
```

Variable	Obs	Mean	Std. dev.	Min	Max
fe_garch_h20	1,675	-.000271	.0015404	-.0040979	.0240624
fe_egarch~20	1,675	-.0001417	.0016662	-.004741	.0249634

```
.
```

Newey test

```
Results
. newey d_egarch_1, lag(17)

Regression with Newey-West standard errors      Number of obs     =      1,675
Maximum lag = 17                                F(  0,       1674) =          .
                                                Prob > F          =          .

------------------------------------------------------------------------------
             |             Newey-West
  d_egarch_1 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _cons |  -6.53e-07   4.11e-07    -1.59   0.113    -1.46e-06    1.54e-07
------------------------------------------------------------------------------

. newey d_egarch_5, lag(18)

Regression with Newey-West standard errors      Number of obs     =      1,675
Maximum lag = 18                                F(  0,       1674) =          .
                                                Prob > F          =          .

------------------------------------------------------------------------------
             |             Newey-West
  d_egarch_5 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _cons |  -4.88e-07   3.45e-07    -1.42   0.157    -1.16e-06    1.88e-07
------------------------------------------------------------------------------

. newey d_egarch_20, lag(17)

Regression with Newey-West standard errors      Number of obs     =      1,675
Maximum lag = 17                                F(  0,       1674) =          .
                                                Prob > F          =          .

------------------------------------------------------------------------------
             |             Newey-West
 d_egarch_20 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _cons |  -3.50e-07   2.07e-07    -1.69   0.092    -7.57e-07    5.69e-08
------------------------------------------------------------------------------
```

Results

```
. newey d_rv_1, lag(11)

Regression with Newey-West standard errors      Number of obs     =      1,674
Maximum lag = 11                                F(  0,       1673) =          .
                                                Prob > F          =          .

------------------------------------------------------------------------------
             |             Newey-West
      d_rv_1 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _cons |   6.81e-07   3.28e-07     2.08   0.038     3.75e-08    1.32e-06
------------------------------------------------------------------------------

. newey d_rv_5, lag(11)

Regression with Newey-West standard errors      Number of obs     =      1,674
Maximum lag = 11                                F(  0,       1673) =          .
                                                Prob > F          =          .

------------------------------------------------------------------------------
             |             Newey-West
      d_rv_5 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _cons |   2.42e-07   1.61e-07     1.51   0.132    -7.32e-08    5.58e-07
------------------------------------------------------------------------------

. newey d_rv_20, lag(11)

Regression with Newey-West standard errors      Number of obs     =      1,674
Maximum lag = 11                                F(  0,       1673) =          .
                                                Prob > F          =          .

------------------------------------------------------------------------------
             |             Newey-West
     d_rv_20 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _cons |   1.85e-07   1.22e-07     1.52   0.129    -5.37e-08    4.23e-07
------------------------------------------------------------------------------
```

Results

```
. newey d_rav_1, lag(11)

Regression with Newey-West standard errors      Number of obs     =      1,674
Maximum lag = 11                                F(  0,        1673) =          .
                                                Prob > F          =          .

------------------------------------------------------------------------------
             |             Newey-West
     d_rav_1 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _cons |   6.23e-07   2.57e-07     2.42   0.015     1.19e-07    1.13e-06
------------------------------------------------------------------------------

. newey d_rav_5, lag(11)

Regression with Newey-West standard errors      Number of obs     =      1,674
Maximum lag = 11                                F(  0,        1673) =          .
                                                Prob > F          =          .

------------------------------------------------------------------------------
             |             Newey-West
     d_rav_5 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _cons |   3.47e-07   1.11e-07     3.12   0.002     1.29e-07    5.65e-07
------------------------------------------------------------------------------

. newey d_rav_20, lag(11)

Regression with Newey-West standard errors      Number of obs     =      1,674
Maximum lag = 11                                F(  0,        1673) =          .
                                                Prob > F          =          .

------------------------------------------------------------------------------
             |             Newey-West
    d_rav_20 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _cons |   2.45e-07   1.02e-07     2.39   0.017     4.42e-08    4.46e-07
------------------------------------------------------------------------------
```

Expected loss

```
. sum l_garch_h1 l_rv_h1
```

Variable	Obs	Mean	Std. dev.	Min	Max
l_garch_h1	1,675	.2703101	.3004827	9.71e-10	4.083507
l_rv_h1	1,674	.1655819	.3177366	1.55e-09	7.467969

```
. sum l_garch_h5 l_rv_h1
```

Variable	Obs	Mean	Std. dev.	Min	Max
l_garch_h5	1,675	.3425437	.4211223	1.72e-06	5.537941
l_rv_h1	1,674	.1655819	.3177366	1.55e-09	7.467969

```
. sum l_garch_h20 l_rv_h1
```

Variable	Obs	Mean	Std. dev.	Min	Max
l_garch_h20	1,675	.3967052	.6160393	4.16e-07	11.62007
l_rv_h1	1,674	.1655819	.3177366	1.55e-09	7.467969

```
.
```

```
. sum l_garch_h1 l_rav_h1
```

Variable	Obs	Mean	Std. dev.	Min	Max
l_garch_h1	1,675	.2703101	.3004827	9.71e-10	4.083507
l_rav_h1	1,648	.3746671	5.952335	2.75e-07	239.3777

```
. sum l_garch_h5 l_rav_h5
```

Variable	Obs	Mean	Std. dev.	Min	Max
l_garch_h5	1,675	.3425437	.4211223	1.72e-06	5.537941
l_rav_h5	1,674	.2402989	.428853	1.19e-06	5.857606

```
. sum l_garch_h20 l_rav_h20
```

Variable	Obs	Mean	Std. dev.	Min	Max
l_garch_h20	1,675	.3967052	.6160393	4.16e-07	11.62007
l_rav_h20	1,674	.3385465	.7306667	2.60e-07	11.76728

```
.
```

Newey QLIKE

```
Results
. newey egar_h1, lag(11)

Regression with Newey-West standard errors      Number of obs     =      1,675
Maximum lag = 11                                F(  0,       1674) =          .
                                                Prob > F          =          .

------------------------------------------------------------------------------
             |             Newey-West
     egar_h1 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _cons |   -.151502   .0755748    -2.00   0.045     -.299733   -.003271
------------------------------------------------------------------------------

. newey egar_h5, lag(11)

Regression with Newey-West standard errors      Number of obs     =      1,675
Maximum lag = 11                                F(  0,       1674) =          .
                                                Prob > F          =          .

------------------------------------------------------------------------------
             |             Newey-West
     egar_h5 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _cons |  -.1401338   .0747926    -1.87   0.061    -.2868307     .006563
------------------------------------------------------------------------------

. newey egar_h20, lag(11)

Regression with Newey-West standard errors      Number of obs     =      1,675
Maximum lag = 11                                F(  0,       1674) =          .
                                                Prob > F          =          .

------------------------------------------------------------------------------
             |             Newey-West
    egar_h20 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _cons |    -.13371    .061453    -2.18   0.030    -.2542428   -.0131772
------------------------------------------------------------------------------
```

Results

```
. newey rv_h1, lag(11)

Regression with Newey-West standard errors      Number of obs     =      1,674
Maximum lag = 11                                F(  0,        1673) =          .
                                                Prob > F          =          .

------------------------------------------------------------------------------
             |             Newey-West
       rv_h1 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _cons |   .1048853   .0138439     7.58   0.000     .0777321    .1320384
------------------------------------------------------------------------------

. newey rv_h5, lag(11)

Regression with Newey-West standard errors      Number of obs     =      1,674
Maximum lag = 11                                F(  0,        1673) =          .
                                                Prob > F          =          .

------------------------------------------------------------------------------
             |             Newey-West
       rv_h5 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _cons |   .0990234   .0156005     6.35   0.000     .0684249     .129622
------------------------------------------------------------------------------

. newey rv_h20, lag(11)

Regression with Newey-West standard errors      Number of obs     =      1,674
Maximum lag = 11                                F(  0,        1673) =          .
                                                Prob > F          =          .

------------------------------------------------------------------------------
             |             Newey-West
      rv_h20 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _cons |   .0481388   .0188026     2.56   0.011     .0112597    .0850178
------------------------------------------------------------------------------
```

Results

```
. newey rav_h1, lag(11) force

Regression with Newey-West standard errors      Number of obs     =      1,648
Maximum lag = 11                                F(  0,       1647) =          .
                                                Prob > F          =          .
```

rav_h1	Coefficient	Newey-West std. err.	t	P>|t|	[95% conf.	interval]
_cons	**-.1074367**	**.1537049**	**-0.70**	**0.485**	**-.4089143**	**.1940408**

```
. newey rav_h5, lag(11)

Regression with Newey-West standard errors      Number of obs     =      1,674
Maximum lag = 11                                F(  0,       1673) =          .
                                                Prob > F          =          .
```

rav_h5	Coefficient	Newey-West std. err.	t	P>|t|	[95% conf.	interval]
_cons	**.1024423**	**.0183143**	**5.59**	**0.000**	**.066521**	**.1383635**

```
. newey rav_h20, lag(11)

Regression with Newey-West standard errors      Number of obs     =      1,674
Maximum lag = 11                                F(  0,       1673) =          .
                                                Prob > F          =          .
```

rav_h20	Coefficient	Newey-West std. err.	t	P>|t|	[95% conf.	interval]
_cons	**.0583546**	**.0180111**	**3.24**	**0.001**	**.023028**	**.0936811**