Forecasting the Realized Variance

We have discussed several ways of using high-frequency observations of stock prices to estimate the integrated variance of a stock. One of the estimators for integrated variance is the realized variance. The realized variance converges to the integrated variance as the sampling frequency increases, and its asymptotic variance is small when compared to other estimators, like the bipower variance. For this reason, the realized variance of a stock is often taken as a good measure of its integrated variance. We now turn to forecasting a stock's variance. We will do so by using the realized variance as a proxy for the integrated variance, and develop models for forecasting RV.

1 AR Models

Let's consider autoregressive (AR) models to forecast the realized variance. The simplest is the $AR(1)$ model:

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + u_t$$

We estimate the β parameters using data for a period $t = S, S+1, ..., T$ by simple OLS regression to get $\hat{\beta}_0$ and $\hat{\beta}_1$. The in-sample fitted values are

$$\widehat{RV}_t = \hat{\beta}_0 + \hat{\beta}_1 RV_{t-1}.$$

If we want to do out of sample forecasting, the one-step-ahead forecast is

$$\widehat{RV}_{T+1} = \hat{\beta}_0 + \hat{\beta}_1 RV_T$$

For a multistep ahead implied forecast apply the chain rule of forecasting

$$\widehat{RV}_{T+j+1} = \hat{\beta}_0 + \hat{\beta}_1 \widehat{RV}_{T+j}, \quad j = 1, 2, \dots$$

Evidently, an $AR(2)$ model is

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-2} + u_t$$

The same logic for estimation and forecasting applies to the $AR(2)$ model.

We know that volatility exhibits a high degree of persistence and it's likely that RV_t is better forecast by using more lags, $RV_{t-1}, RV_{t-3}, \dots$. That makes us think of the model with J lags:

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-2} + \cdots + \beta_J RV_{t-J} + u_t$$

where u_t is the error term. But to capture long-range dependence might entail $J = 10$, $J = 20$, or higher. Estimating such a large number of coefficients could entail a lot of estimation error and lead to bad forecasting properties.

2 HAR Models

In Corsi (2009), Professor Corsi argues that for forecasting purposes all that might matter is the average level of volatility over the previous week and month. So we define the variables

$$RV_{t-1}^{w} = \frac{1}{5}\sum_{j=1}^{5} RV_{t-j}$$

and

$$RV_{t-1}^{m} = \frac{1}{22}\sum_{j=1}^{22} RV_{t-j}$$

Recall that there are usually 22 business days in a month, which is why we average over the previous 22 instead of 30. Using these variables we write a restricted version of the $AR(J)$ model as:

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_w RV_{t-1}^{w} + \ \beta_m RV_{t-1}^{m} + u_t$$

which is called the HAR model. Note that the HAR model above has as many as $J = 22$ lags as in $AR(J)$ model, but by restricting the form of the dependence there are only three coefficients to estimate, a great simplification for forecasting purposes.

3 Examples

Using 5-min and 2013−2016 data for the stock Bank of America (BAC) an OLS regression gives the estimated $AR(1)$:

	β_0	β_1
coefficient:	0.9196	0.4780
standard deviation:	(0.0711)	(0.0278)
R^2:	0.23	

The estimated HAR model is:

	β_0	β_1	β_w	β_m
coefficient:	0.3499	0.2978	0.2186	0.2840
standard deviation:	(0.1016)	(0.0367)	(0.0647)	(0.0724)
R^2:	0.28			

4 No-Change (White Noise) Model

Rather than estimating parameters, another often-used forecasting rule is just to use the preceding period's value as in

$$\widehat{RV}_{t+1} = RV_t.$$

Evidently, the no change forecast results from an $AR(1)$ constrained as $\beta_0 = 0$ and $\beta_1 = 1$. The implicit model is

$$RV_t = RV_{t-1} + u_t$$

This model is often called the random walk model.

5 Forecast Evaluation

Suppose we have data $(RV_t)_{t=0}^T$ and we want to forecast RV_{T+1}. We now have three possibilities:

$$\begin{aligned} \text{No Change}: \widehat{RV}_{T+1} &= RV_T \\ AR(1): \widehat{RV}_{T+1} &= \hat{\beta}_0 + \hat{\beta}_1 RV_T \\ HAR: \widehat{RV}_{T+1} &= \hat{\beta}_0 + \hat{\beta}_1 RV_T + \hat{\beta}_w RV_T^w + \hat{\beta}_m RV_T^m \end{aligned}$$

How do we evaluate these three competing models under consideration. The standard strategy is rolling regressions and quasi-out-of-sample forecasting.

We will let $[S, T]$ denote the endpoints of our subsample of data, e.g., $[S, T]$ is a window of a fixed width. (Think of $1,000$ observations or four years.) We start with the window $[S_1, T_1]$, and using all of the data for $t = S_1, S_1 + 1, \ldots, T_1$ (including the high frequency within day to form RV) we estimate by OLS the parameters of the models above. We then form the forecast errors:

$$\begin{aligned} \text{No Change}: \quad & e_1^{NC} = RV_{T_1+1} - RV_{T_1} \\ AR(1): \quad & e_1^{AR1} = RV_{T_1+1} - \hat{\beta}_0 - \hat{\beta}_1 RV_{T_1} \\ HAR: \quad & e_1^{HAR} = RV_{T_1+1} - \hat{\beta}_0 - \hat{\beta}_1 RV_T - \hat{\beta}_w RV_T^w - \hat{\beta}_m RV_T^m \end{aligned}$$

We then advance each endpoint by one day: $S_1 \to S_2 = S_1 + 1,\ T_1 \to T_2 = T_1 + 1$, and repeat the above. In general, we keep increasing S and T by one unit and form the error at each step j:

$$\begin{aligned} \text{No Change}: \quad & e_j^{NC} = RV_{T_j+1} - RV_{T_j} \\ AR(1): \quad & e_j^{AR1} = RV_{T_j+1} - \hat{\beta}_{0j} - \hat{\beta}_{1j} RV_{T_j} \\ HAR: \quad & e_j^{HAR} = RV_{T_j+1} - \hat{\beta}_{0j} - \hat{\beta}_{1j} RV_{T_j} - \hat{\beta}_{wj} RV_{T_j}^w - \hat{\beta}_{mj} RV_{T_j}^m \end{aligned}$$

The conventional metric for evaluating forecasts is the mean squared error (MS):

$$\begin{aligned} \text{No Change}: \quad MSE^{NC} &= \frac{1}{J}\sum_{j=1}^{J} (e_j^{NC})^2 \\ AR(1): \quad MSE^{AR1} &= \frac{1}{J}\sum_{j=1}^{J} (e_j^{AR1})^2 \\ HAR: \quad MSE^{HAR} &= \frac{1}{J}\sum_{j=1}^{J} (e_j^{HAR})^2 \end{aligned}$$

6 RQ Models

Recall the AR1 model:

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + u_t$$

Bollerslev, Patton, and Quaedvlieg (2016) note that the above model is somewhat implausible because β_1 varies with the measurement error in RV_{t-1}. We should really write

$$RV_t = \beta_0 + \beta_{1,t-1} RV_{t-1} + u_t$$

The coefficient $\beta_{1,t-1}$ is small when the measurement error is large and vice versa. However, from the high frequency theory we know something about the magnitude of the measurement error. Specifically,

$$\hat{\sigma}_{e_{t-1}} = \text{constant} \times \widehat{QIV}_{t-1}^{1/2},$$

and $\beta_{1,t-1}$ should be small when $\hat{\sigma}_{e_{t-1}}$ is large. So Bollerslev, Patton, and Quaedvlieg (2016) propose

$$\beta_{1,t-1} = \beta_1 + \beta_{1Q}\widehat{QIV}_{t-1}^{1/2}$$

with the expectation that $\beta_{1Q} < 0$. The linear functional form is used just to keep things simple, which is generally important for forecasting. Putting the last two equations together gives

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_{1Q}\widehat{QIV}_{t-1}^{1/2} RV_{t-1} + u_t$$

The preceding is just an AR1 model with a second term to adjust for the measurement error in RV_{t-1}.

Evidently, the HAR model can be adjusted via the QIV_{t-1} adjustment. As suggested by Bollerslev, Patton, and Quaedvlieg (2016) we consider the model

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_{1Q}\widehat{QIV}_{t-1}^{1/2} RV_{t-1} + \beta_w RV_{t-1}^w + \beta_m RV_{t-1}^m + u_t$$

There are even more variations one can consider, but if we stop here we now have four forecasting models:

$$\begin{aligned}
AR1 : \widehat{RV}_{T+1} &= \hat{\beta}_0 + \hat{\beta}_1 RV_T \\
ARQ1 : \widehat{RV}_{T+1} &= \hat{\beta}_0 + \hat{\beta}_1 RV_T + \hat{\beta}_{1Q}\widehat{QIV}_T^{1/2} RV_T \\
HAR : \widehat{RV}_{T+1} &= \hat{\beta}_0 + \hat{\beta}_1 RV_T + \hat{\beta}_w RV_T^w + \hat{\beta}_m RV_T^m \\
HARQ1 : \widehat{RV}_{T+1} &= \hat{\beta}_0 + \hat{\beta}_1 RV_T + \hat{\beta}_{1Q}\widehat{QIV}_T^{1/2} RV_T + \hat{\beta}_w RV_T^w + \hat{\beta}_m RV_T^m
\end{aligned}$$

These are the four models you evaluate as forecasting models for RV for the next project.

A caveat of the RQ correction is that the model might generate occasional "insane" forecasts due to extreme outliers in $\widehat{QIV}$. Thus we need to add a "sanity" filter. If $\widehat{RV} > \max RV$ or $\widehat{RV} < \min RV$, where $\max RV$, $\min RV$ are the max and min of RV_t over the sample window, then just set $\widehat{RV} = meanRV$ where $meanRV$ is the average RV over the sample window.

References

Bollerslev, Tim, Andrew J. Patton, and Rogier Quaedvlieg (2016). "Exploiting the errors: A simple approach for improved volatility forecasting". In: *Journal of Econometrics* 192, pp. 1–18.

Corsi, Fulvio (2009). "A simple approximate long-memory model of realized volatility". In: *Journal of Financial Econometrics* 7.2, pp. 174–196. URL: `https://doi.org/10.1093/jjfinec/nbp001`.