Amath 546/Econ 589
Univariate GARCH Models: Advanced Topics

Eric Zivot

April 29, 2013

Lecture Outline

- The Leverage Effect
- Asymmetric GARCH Models
- Forecasts from Asymmetric GARCH Models
- GARCH Models with Non-normal Errors
- Long Memory GARCH Models
- Evaluating GARCH Forecasts

Asymmetric Leverage Effects and News Impact

- In the basic GARCH model, since only squared residuals ϵ_{t-i}^2 enter the conditional variance equation, the signs of the residuals or shocks have no effect on conditional volatility.

- A stylized fact of financial volatility is that bad news (negative shocks) tends to have a larger impact on volatility than good news (positive shocks). That is, volatility tends to be higher in a falling market than in a rising market. Black (1976) attributed this effect to the fact that bad news tends to drive down the stock price, thus increasing the leverage (i.e., the debt-equity ratio) of the stock and causing the stock to be more volatile. Based on this conjecture, the asymmetric news impact on volatility is commonly referred to as the "leverage effect".

Testing for Asymmetric Effects on Conditional Volatility

- A simple diagnostic for uncovering possible asymmetric leverage effects is the sample correlation between r_t^2 and r_{t-1}. A negative value of this correlation provides some evidence for potential leverage effects.

- Other simple diagnostics result from estimating the following test regression

$$\hat{\epsilon}_t^2 = \beta_0 + \beta_1 \hat{w}_{t-1} + \xi_t,$$

where $\hat{w}_{t-1}$ is a variable constructed from $\hat{\epsilon}_{t-1}$ and the sign of $\hat{\epsilon}_{t-1}$. A significant value of β_1 indicates evidence for asymmetric effects on conditional volatility.

- Let S_{t-1}^{-} denote a dummy variable equal to unity when $\hat{\epsilon}_{t-1}$ is negative, and zero otherwise. Engle and Ng consider three tests for asymmetry.
  - Setting $\hat{w}_{t-1} = S_{t-1}^{-}$ gives the Sign Bias test;
  - Setting $\hat{w}_{t-1} = S_{t-1}^{-}\hat{\epsilon}_{t-1}$ gives the Negative Size Bias test;
  - Setting $\hat{w}_{t-1} = S_{t-1}^{+}\hat{\epsilon}_{t-1}$ gives the Positive Size Bias test.

EGARCH Model

Define $h_t = \ln(\sigma_t^2)$ and $\epsilon_t = \sigma_t z_t$ where $z_t \sim iid\ (0, 1)$. Nelson's *exponential* GARCH model is then

$$h_t = a_0 + \sum_{i=1}^{p} a_i \frac{|\epsilon_{t-i}| + \gamma_i \epsilon_{t-i}}{\sigma_{t-i}} + \sum_{j=1}^{q} b_j h_{t-j}$$

- Variance is always positive because $\sigma_t^2 = \exp(h_t)$

- Total effect of positive shocks (good news) to ϵ_{t-i}

$$(1 + \gamma_i)|\epsilon_{t-i}|$$

- Total effect of negative shocks (bad news) to ϵ_{t-i}

$$(1 - \gamma_i)|\epsilon_{t-i}|$$

- *Leverage effect* implies that $\gamma_i < 0$

- EGARCH is covariance stationary provided $b(1) = \sum_{j=1}^{q} b_j < 1$. Here, $b(1)$ is called the *persistence*.

Remark: The EGARCH model in the rugarch package is specified slightly differently

$$h_t = a_0 + \sum_{i=1}^{p} (a_i z_{t-i} + \gamma_i \left(|z_{t-i}| - E[|z_{t-i}|\right)) + \sum_{j=1}^{q} b_j h_{t-j}$$

- a_i captures the sign effect: leverage effect $\Rightarrow a_i < 0$

- γ_i captures the size effect: bigger γ_i implies a larger leverage effect. Hence, $\gamma_i > 0$

TGARCH/GJR Model

Zakoian's *threshold* GARCH (aka GJR - Glosten, Jagannathan, and Runkle) model is

$$\begin{aligned} \sigma_t^2 &= a_0 + \sum_{i=1}^{p} a_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \gamma_i S_{t-i} \epsilon_{t-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-j}^2 \\ S_{t-i} &= \begin{cases} 1 & \text{if } \epsilon_{t-i} < 0 \\ 0 & \text{if } \epsilon_{t-i} \geq 0 \end{cases} \end{aligned}$$

- When ϵ_{t-i} is positive, the total effects are $a_i \epsilon_{t-i}^2$
- when ϵ_{t-i} is negative, the total effects are $(a_i + \gamma_i)\epsilon_{t-i}^2$

- *Leverage effect* implies that $\gamma_i > 0$

- TGARCH/GJR is covariance stationary provided the persistence $\sum_{i=1}^{p}(a_i + \gamma_i/2) + \sum_{j=1}^{q}\beta_j < 1$

PGARCH Model (aka APARCH)

Ding, Granger and Engle's power GARCH model for $d > 0$

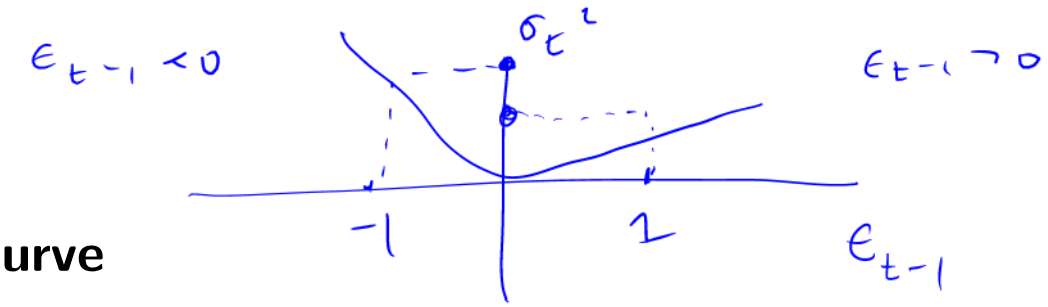
$$\sigma_t^d = a_0 + \sum_{i=1}^{p} a_i(|\epsilon_{t-i}| - \gamma_i \epsilon_{t-i})^d + \sum_{j=1}^{q} b_j \sigma_{t-j}^d$$

- Leverage effect implies that $\gamma_i > 0$ (if $\epsilon_{t-i} < 0$ and $\gamma_i > 0$ then $-\gamma_i \epsilon_{t-i} > 0$ so negative returns increase σ_t^d)
- $d = 2$ gives a regular GARCH model with leverage effects
- $d = 1$ gives a model for σ_t and is more robust to outliers than when $d = 2$

- d can be fixed at a particular value or estimated by mle

- Condition for stationarity is complicated (see rugarch package documentation) and depends on b_j and γ_j.

News Impact Curve



Engle and Ng propose the use of the news impact curve to evaluate asymmetric GARCH models:

> The news impact curve is the functional relationship between conditional variance at time t and the shock term (error term) at time $t-1$, holding constant the information dated $t-2$ and earlier, and with all lagged conditional variance evaluated at the level of the unconditional variance.

News impact curves can be easily constructed for all types of GARCH models.

Forecasts from Asymmetric GARCH(1,1) Models

$\epsilon_t = z_t \sigma_t$

$z_t \sim N(0,1)$

- Consider the TGARCH(1,1) model at time T

$$\begin{aligned} \sigma_T^2 &= a_0 + a_1 \epsilon_{T-1}^2 + \gamma_1 S_{T-1} \epsilon_{T-1}^2 + b_1 \sigma_{T-1}^2 \\ S_T &= 1 \text{ if } \epsilon_T < 0; 0 \text{ otherwise} \end{aligned}$$

- Assume that ϵ_t has a symmetric distribution about zero. The forecast for $T+1$ based on information at time T is

$$E[\sigma_{T+1}^2 | I_T] = a_0 + a_1 \epsilon_T^2 + \gamma_1 S_T \epsilon_T^2 + b_1 \sigma_T^2,$$

where it assumed that ϵ_T^2, S_T and σ_T^2 are known. Hence, the TGARCH(1,1) forecast for $T+1$ will be different than the GARCH(1,1) forecast if $S_T = 1$ ($\epsilon_T < 0$).

- The forecast at $T+2$ is

$$E[\sigma^2_{T+2}|I_T] = a_0 + a_1 E[\epsilon^2_{T+1}|I_T] + \gamma_1 E[S_{T+1}\epsilon^2_{T+1}|I_T] + b_1 E[\sigma^2_{T+1}|I_T]$$
$$= a_0 + \left(\frac{\gamma_1}{2} + a_1 + b_1\right) E[\sigma^2_{T+1}|I_T],$$

which follows since S_{T+1} is independent of ϵ^2_{T+1} and z_t has a symmetric distribution about zero:

$$E[S_{T+1}\epsilon^2_{T+1}|I_T] = E[S_{T+1}|I_T]E[\epsilon^2_{T+1}|I_T] = \frac{1}{2}E[\sigma^2_{T+1}|I_T]$$

Notice that the asymmetric impact of leverage is present even if $S_T = 0$.

- By recursive substitution for the forecast at $T+h$ is

$$E[\sigma^2_{T+h}|I_T] = a_0 + \left(\frac{\gamma_1}{2} + a_1 + b_1\right)^{h-1} E[\sigma^2_{T+1}|I_T],$$

which is similar to the GARCH(1,1) forecast.

- The mean reverting form is

$$E[\sigma^2_{T+h}|I_T] - \bar{\sigma}^2 = \left(\frac{\gamma_1}{2} + a_1 + b_1\right)^{h-1} \left(E[\sigma^2_{T+h}|I_T] - \bar{\sigma}^2\right)$$

  where $\bar{\sigma}^2 = a_0/(1 - \frac{\gamma_1}{2} - a_1 - b_1)$ is the long run variance.

- Forecasting algorithms for σ^d_{T+h} in the PGARCH$(1, d, 1)$ and for $\ln \sigma^2_{T+h}$ in the EGARCH(1,1) follow in a similar manner.

GARCH Models with Non-Normal Errors

In the GARCH model with normal errors, $\epsilon_t = \sigma_t z_t$ and $z_t \sim iid\ N(0,1)$

- Often the estimated standardized residuals $\hat{z}_t = \frac{\hat{\epsilon}_t}{\hat{\sigma}_t}$ from a GARCH model with normal errors still has fat and/or asymmetric tails. This suggests using a *standardized* fat-tailed and/or asymmetric error distribution for z_t instead of $N(0,1)$.

- The most common fat-tailed error distributions for fitting GARCH models are: the *Student's t distribution*; the *double exponential distribution*; and the *generalized error distribution*. Another fat-tailed distribution implemented in rugarch is the *generalized hyperbolic distribution*.

- The most common standardized fat-tailed and asymmetric distribution is the *skewed-t*. Another fat-tailed and asymmetric distribution implemented in rugarch is the *generalized hyperbolic skew Student distribution*.

GARCH with Student-t errors (most common non-normal GARCH model)

Let u_t be Student-t random variable degrees of freedom parameter v and scale parameter s_t. Then

$$\begin{aligned} f(u_t) &= \frac{\Gamma[(\nu+1)/2]}{(\pi\nu)^{1/2}\Gamma(\nu/2)} \frac{s_t^{-1/2}}{[1+u_t^2/(s_t\nu)]^{(\nu+1)/2}} \\ \text{var}(u_t) &= \frac{s_t\nu}{\nu-2}, \; v>2. \end{aligned}$$

If ϵ_t in the GARCH model is Student-t with $E[\epsilon_t^2|I_{t-1}] = \sigma_t^2$ then set

$$s_t = \frac{\sigma_t^2(\nu-2)}{\nu}$$

to create a standardized Student-t distribution for z_t.

Generalized Error Distribution

Nelson suggested using the *generalized error distribution* (GED) with parameter $v > 0$. If u_t is distributed GED with parameter v then

$$f(u_t) = \frac{\nu \exp[-(1/2)|u_t/\lambda|^\nu]}{\lambda \cdot 2^{(\nu+1)/\nu}\Gamma(1/\nu)}$$

where

$$\lambda = \left[\frac{2^{-2/\nu}\Gamma(1/\nu)}{\Gamma(3/\nu)}\right]^{1/2}$$

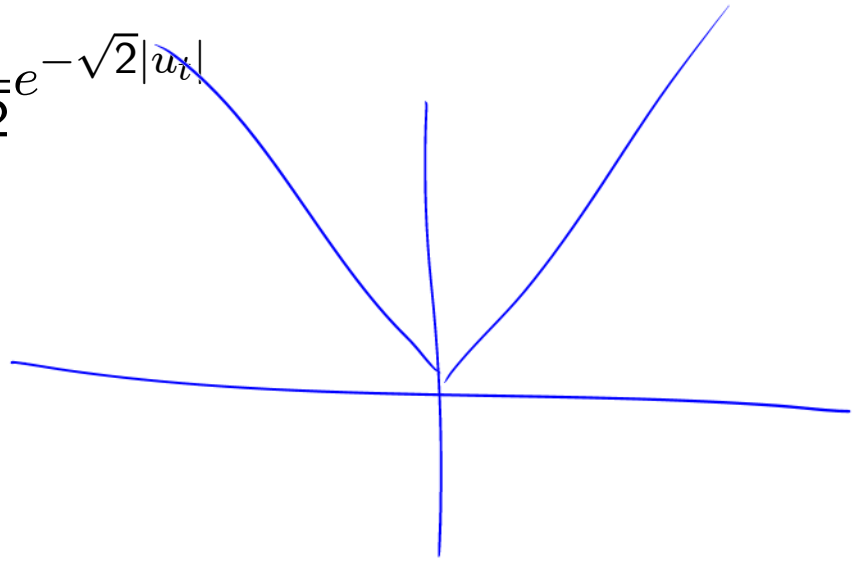
- $v = 2$ gives the normal distribution

- $0 < v < 2$ gives a distribution with fatter tails than normal

- $v > 2$ gives a distribution with thinner tails than normal

- $v = 1$ gives the *double exponential distribution*

$$f(u_t) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|u_t|}$$



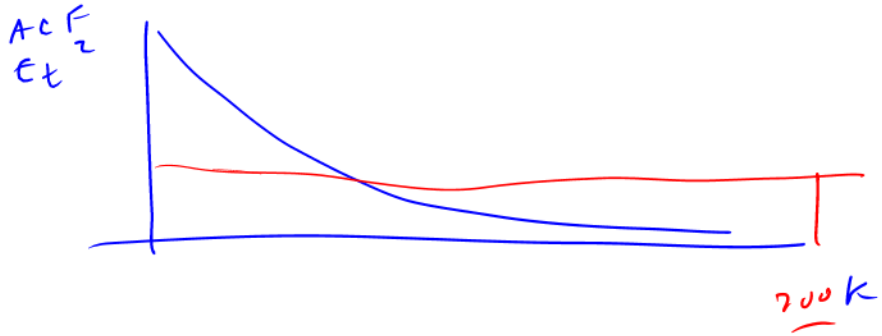
Skewed Student-t Distribution

There are several definitions of the Skewed Student-t distribution (e.g. Azzalini and Capitanio, Fernandez and Steel, etc.). In their scaled form (mean zero and unit variance), all versions have

- degrees of freedom parameter $v > 0$ controlling tail-thickness relative to normal

- skew (asymmetry) parameter ξ such that $\xi < 0$ gives negative skew (long left tail) and $\xi > 0$ gives long right tail.

Long Memory GARCH Models



- If returns follow a GARCH(p, q) model, then the autocorrelations of the squared and absolute returns should decay exponentially.

- However, the SACF of r_t^2 and $|r_t|$ often appear to decay much more slowly. This is evidence of so-called *long memory* behavior.

- Formally, a stationary process has long memory or long range dependence if its autocorrelation function behaves like

$$\rho(k) \to C_\rho k^{2d-1} \text{ as } k \to \infty,$$

where C_ρ is a positive constant, and d is a real number between 0 and $\frac{1}{2}$. Thus the autocorrelation function of a long memory process decays slowly at a hyperbolic rate.

Long Memory GARCH Models

- Long memory behavior can be built into the conditional variance equation in a variety of ways
  - Bollerslev's fractionally integrated GARCH model (FIGARCH)
  - Engle's two component GARCH model
- Estimation of long memory GARCH models is very difficult and it is seldom used in practice (academics love it because it is complicated)

Integrated GARCH Model

$$\sigma_t^2 = a_0 + a_1\epsilon_{t-1}^2 + b_1\sigma_{t-1}^2$$

- The high persistence often observed in fitted GARCH(1,1) models suggests that volatility might be nonstationary implying that $a_1 + b_1 = 1$, in which case the GARCH(1,1) model becomes the integrated GARCH(1,1) or IGARCH(1,1) model.

- In the IGARCH(1,1) model the unconditional variance is not finite and so the model does not exhibit volatility mean reversion. However, it can be shown that the model is strictly stationary provided $E[\ln(a_1 z_t^2 + b_1)] < 0$.

- IGARCH(1,1) is equivalent to the riskMetrics EWMA model when $a_0 = 0$, $a_1 = 1 - \lambda$ and $b_1 = \lambda$ where $\lambda = 0.94$.

- Diebold and Lopez (1996) argued against the IGARCH specification for modeling highly persistent volatility processes for two reasons.
  - First, they argue that unconditional variance should be finite
  - Second, they argue that the observed convergence toward normality of aggregated returns is inconsistent with the IGARCH model.
  - Third, they argue that observed IGARCH behavior may result from mis-specification of the conditional variance function. For example, ignored structural breaks or regime switching in the unconditional variance can result in IGARCH behavior.

Evaluating Volatility Predictions

- GARCH models are often judged by their out-of-sample forecasting ability

- This forecasting ability can be measured using
  - traditional forecast error metrics such as MSE
  - Specific economic considerations such as value-at-risk violations, option pricing accuracy, or portfolio performance.

- Out-of-sample forecasts for use in model comparison are typically computed using one of two methods.

  - *Recursive forecasts*: An initial sample using data from $t = 1, \ldots, T$ is used to estimate the models, and $h-$step ahead out-of-sample forecasts are produced starting at time T. The sample is increased by one, the models are re-estimated, and $h-$step ahead forecasts are produced starting at $T + 1$.

  - *Rolling forecasts*. An initial sample using data from $t = 1, \ldots, T$ is used to determine a window width T, to estimate the models, and to form $h-$step ahead out-of-sample forecasts starting at time T. Then the window is moved ahead one time period, the models are re-estimated using data from $t = 2, \ldots, T + 1$, and $h-$step ahead out-of-sample forecasts are produced starting at time $T + 1$.

Traditional Forecast Evaluation Statistics

- Let $E_i[\sigma^2_{T+h}|I_T]$ denote the $h-$step ahead forecast of σ^2_{T+h} at time T from GARCH model i using either recursive or rolling methods.

- Define the corresponding forecast error as $e_{i,T+h|T} = E_i[\sigma^2_{T+h}|I_T] - \sigma^2_{T+h}$.

- Common forecast evaluation statistics

$$\begin{aligned}
\text{MSE}_i &= \frac{1}{N}\sum_{j=T+1}^{T+N} e^2_{i,j+h|j}, \text{MAE}_i = \frac{1}{N}\sum_{j=T+1}^{T+N} \left|e_{i,j+h|j}\right|, \\
\text{MAPE}_i &= \frac{1}{N}\sum_{j=T+1}^{T+N} \frac{\left|e_{i,j+h|j}\right|}{\sigma_{j+h}}.
\end{aligned}$$

- The model which produces the smallest values of the forecast evaluation statistics is judged to be the best model.

- Of course, the forecast evaluation statistics are random variables and a formal statistical procedure should be used to determine if one model exhibits superior predictive performance.

Diebold-Mariano Tests for Predictive Accuracy

- Let $\{e_{1,j+h|j}\}_{T+1}^{T+N}$, and $\{e_{2,j+h|j}\}_{T+1}^{T+N}$ denote forecast errors from two different GARCH models.

- The accuracy of each forecast is measured by a particular loss function $L(e_{i,T+h|T})$, $i = 1, 2$.

  - squared error loss function: $L(e_{i,T+h|T}) = \left(e_{i,T+h|T}\right)^2$; absolute error loss function $L(e_{i,T+h|T}) = \left|e_{i,T+h|T}\right|$.

- The Diebold-Mariano (DM) test is based on the loss differential

$$d_{T+h} = L(e_{1,T+h|T}) - L(e_{2,T+h|T}).$$

- The null of equal predictive accuracy is $H_0 : E[d_{T+h}] = 0$.

- The DM test statistic is

$$\mathsf{S} = \frac{\bar{d}}{\left(\widehat{\mathsf{avar}}(\bar{d})\right)^{1/2}}, \; \bar{d} = N^{-1} \sum_{j=T+1}^{T+N} d_{j+h}$$

- DM recommend using the Newey-West estimate for $\widehat{\mathsf{avar}}(\bar{d})$ because the sample of loss differentials $\{d_{j+h}\}_{T+1}^{T+N}$ are serially correlated for $h > 1$.

- Under the null of equal predictive accuracy,

$$\mathsf{S} \sim N(0, 1)$$

Hence, the DM statistic can be used to test if a given forecast evaluation statistic (e.g. MSE_1) for one model is statistically different from the forecast evaluation statistic for another model (e.g. MSE_2).

Mincer-Zarnowitz Forecasting Regression

- Forecasts are also often judged using the forecasting regression

$$\sigma^2_{T+h} = \alpha + \beta E_i[\sigma^2_{T+h}|I_T] + e_{i,T+h}.$$

- Unbiased forecasts have $\alpha = 0$ and $\beta = 1$, and accurate forecasts have high regression R^2 values.

- In practice, the forecasting regression suffers from an errors-in-variables problem when estimated GARCH parameters are used to form $E_i[\sigma^2_{T+h}|I_T]$ and this creates a downward bias in the estimate of β. As a result, attention is more often focused on the R^2.

Fundamental Problem with Evaluating Volatility Forecasts

- An important practical problem with applying forecast evaluations to volatility models is that the $h-$step ahead volatility σ^2_{T+h} is not directly observable.

- Typically, ϵ^2_{T+h} (or just the squared return) is used to proxy σ^2_{T+h} since

$$E[\epsilon^2_{T+h}|I_T] = E[z^2_{T+h}\sigma^2_{T+h}|I_T] = E[\sigma^2_{T+h}|I_T]$$

- ϵ^2_{T+h} is a very noisy proxy for σ^2_{T+h} since $\mathsf{var}(\epsilon^2_{T+h}) = E[\sigma^4_{T+h}](\kappa - 1)$, where κ is the fourth moment of z_t, and this causes problems for the interpretation of the forecast evaluation metrics.

- Many empirical papers have evaluated the forecasting accuracy of competing GARCH models using ϵ^2_{T+h} as a proxy for σ^2_{T+h}. Poon (2005) gave a comprehensive survey.

  - The typical findings are that the forecasting evaluation statistics tend to be large, the forecasting regressions tend to be slightly biased, and the regression R^2 values tend to be very low (typically below 0.1).

  - In general, asymmetric GARCH models tend to have the lowest forecast evaluation statistics. The overall conclusion, however, is that GARCH models do not forecast very well.

- Andersen and Bollerslev (1998) provided an explanation for the apparent poor forecasting performance of GARCH models when ϵ^2_{T+h} is used as a proxy for σ^2_{T+h}.

- For the GARCH(1,1) model in which z_t has finite kurtosis κ, they showed that the population R^2 value in the forecasting regression with $h = 1$ is equal to

$$R^2 = \frac{a_1^2}{1 - b_1^2 - 2a_1 b_1},$$

  and is bounded from above by $1/\kappa$. Assuming $z_t \sim N(0, 1)$, this upper bound is $1/3$. With a fat-tailed distribution for z_t the upper bound is smaller.

- Hence, very low R^2 values are to be expected even if the true model is a GARCH(1,1).

- Moreover, Hansen and Lund (2004) found that the substitution of ϵ^2_{T+h} for σ^2_{T+h} in the evaluation of GARCH models using the DM statistic can result in inferior models being chosen as the best with probability one. These results indicate that extreme care must be used when interpreting forecast evaluation statistics and tests based on ϵ^2_{T+h}.

Using Realized Variance to Evaluate Volatility Forecasts

- If high frequency intraday data are available, then instead of using ϵ^2_{T+h} to proxy σ^2_{T+h} Andersen and Bollerslev (1998) suggested using the so-called realized variance

$$RV^m_{t+h} = \sum_{j=1}^{m} r^2_{t+h,j},$$

where $\{r_{T+h,1}, \ldots, r_{T+h,m}\}$ denote the squared intraday returns at sampling frequency $1/m$ for day $T+h$.

  - For example, if prices are sampled every 5 minutes and trading takes place 24 hours per day then there are $m = 288$ 5-minute intervals per trading day.

- Under certain conditions, RV_{t+h}^m is a consistent estimate of σ_{T+h}^2 as $m \to \infty$. As a result, RV_{t+h}^m is a much less noisy estimate of σ_{T+h}^2 than ϵ_{T+h}^2 and so forecast evaluations based on RV_{t+h}^m are expected to be much more accurate than those based on ϵ_{T+h}^2.

  – For example, in evaluating GARCH(1,1) forecasts for the Deutschemark-US daily exchange rate, Andersen and Bollerslev reported R^2 values of $0.047, 0.331$ and 0.479 using ϵ_{T+1}^2, RV_{T+1}^{24} and RV_{T+1}^{288}, respectively.