Financial Data Analysis

Variance Forecasting and Realized Volatility

Summer 2014

July 15, 2014
Variance Forecasts

• Suppose we have estimated a GARCH(1,1) model, \( \sigma_t^2 = \omega + \alpha \epsilon_{t+1}^2 + \beta \sigma_{t-1}^2 \).

• At time \( t \), our GARCH–implied estimate ("forecast") for the conditional variance in \( t + 1 \) would be

\[
\hat{\sigma}_{t+1}^2 = \hat{\omega} + \hat{\alpha} \hat{\epsilon}_t^2 + \hat{\beta} \hat{\sigma}_t^2. 
\]

(1)

• How would we check whether the GARCH model actually provides good volatility forecasts?
Variance Forecasts

• Suppose we can calculate out–of–sample one–step variance forecasts $\hat{\sigma}_t^2$ of the form (1) for periods $t = 1, \ldots, T$.

• A popular method for assessing the forecasting performance of volatility models is by means of the Mincer–Zarnowitz regression, given by

\[
\tilde{\sigma}_t^2 = a_0 + a_1 \hat{\sigma}_t^2 + u_t, \quad t = 1, \ldots, T, \tag{2}
\]

where $\tilde{\sigma}_t^2$ is a proxy for the ex–post volatility (which is not observable).

• If our model for the conditional variance is correctly specified (and parameters are known), and if the volatility proxy is unbiased for the true (but unobservable) variance, i.e., $E_{t-1}(\tilde{\sigma}_t^2) = \sigma_t^2$, then the population values of the parameters in (2) are $a_0 = 0$ and $a_1 = 1$.

---

Variance Forecasts

• Mincer–Zarnowitz regression:

\[ \tilde{\sigma}^2_t = a_0 + a_1 \tilde{\sigma}^2_t + u_t, \quad t = 1, \ldots, T. \] (3)

• The $R^2$ (coefficient of determination) of the regression (3) is often used to measure and compare the predictive ability of volatility models.

• Since volatility is not an observable quantity, the question arises of how to choose $\tilde{\sigma}^2_t$ in (3).

• In early work, researchers used daily squared returns $r^2_t$ as a proxy for the ex–post volatility, i.e., $\tilde{\sigma}^2_t = r^2_t$ in (3).\(^2\)

\(^2\)Or squared residuals if the conditional mean is not assumed to be practically zero.
Variance Forecasts

- The $R^2$ from these regressions (with $\tilde{\sigma}_t^2 = r_t^2$) is typically low, which gave rise to concerns about the predictive accuracy (and thus practical value) of standard volatility models such as GARCH.

- To illustrate, consider daily S&P 500 returns from January 2001 to December 2010, $T = 2481$ daily observations.
daily S&P 500 returns, January 2001 – December 2010
Variance Forecasts

- First estimate a standard Gaussian GARCH(1,1) process,\(^3\)

\[
\epsilon_t = \sigma_t \eta_t, \quad \eta_t \overset{iid}{\sim} \mathcal{N}(0, 1), \quad \sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (4)
\]

over the first 1000 observations.

- The estimated model is (standard errors in parentheses)

\[
\sigma_t^2 = 0.010_{(0.0072)} + 0.072_{(0.0182)} \epsilon_{t-1}^2 + 0.921_{(0.0204)} \sigma_{t-1}^2, \quad (5)
\]

which we shall use to compute one–step ahead variance forecasts for the days 1001 to 2481, i.e., 1481 one–step ahead out–of–sample variance forecasts.\(^4\)

\(^3\)A constant mean term has been incorporated into the model.  
\(^4\)In practice, we would regularly update the parameter estimates.
Variance Forecasts

• The Mincer–Zarnowitz regression yields (standard errors in parentheses)

\[ r_t^2 = 0.087 + 0.992 \hat{\sigma}_{t-1}^2 + \hat{u}_t, \quad R^2 = 0.247. \] (6)

• (For many series, the \( R^2 \) is actually considerably lower than this.)

• Andersen and Bollerslev (1998)\(^5\) were the first to demonstrate that a low \( R^2 \) is likely due to the fact that the daily squared return is a very noisy proxy for the (unobservable) variance (which is actually what is relevant for economic decisions).

• Note that we basically use just a single observation \( (r_t) \) to estimate \( \sigma_t^2 \).

The squared returns are more noisy than the one-step GARCH variances.
Variance Forecasts

• To illustrate, consider the \( R^2 \) that we would expect (i.e., the population \( R^2 \)) if the true data generating process would be GARCH(1,1) and parameters are known, i.e., the population \( R^2 \) of the regression

\[
   r_t^2 = a_0 + a_1 \sigma_{t-1}^2 + u_t, \tag{7}
\]

where

\[
   r_t = \eta_t \sigma_t, \quad \eta_t \overset{iid}{\sim} (0, 1), \quad \sigma_{t-1}^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{8}
\]

and, with \( E(\eta_t^4) = \kappa \),

\[
   \kappa \alpha^2 + 2\alpha \beta + \beta < 1, \tag{9}
\]

i.e., a finite fourth moment.
Variance Forecasts

- Recall that in a bivariate regression such as (2) the $R^2$ is the coefficient of correlation squared, i.e., in (7),

$$R^2 = \frac{\{\text{Cov}(r^2_t, \sigma^2_t)\}^2}{\text{Var}(r^2_t) \text{Var}(\sigma^2_t)}. \tag{10}$$

- Now since

$$\text{Cov}(r^2_t, \sigma^2_t) = \text{Cov}(\eta^2_t \sigma^2_t, \sigma^2_t) = \mathbb{E}(\eta^2_t \sigma^4_t) - \mathbb{E}(\eta^2_t \sigma^2_t) \mathbb{E}(\sigma^2_t) = \text{Cov}(\sigma^2_t, \sigma^2_t) = \text{Var}(\sigma^2_t),$$

we have that $R^2$ is ratio of the variability of the conditional variance and that of the squared return,

$$R^2 = \frac{\text{Var}(\sigma^2_t)}{\text{Var}(r^2_t)}. \tag{11}$$
Variance Forecasts

- Then

\[
\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2
\]

\[
\text{Var}(\sigma_t^2) = \alpha^2 \text{Var}(r_t^2) + \beta^2 \text{Var}(\sigma_t^2) + 2\alpha\beta \text{Cov}(r_t^2, \sigma_t^2)
\]

\[
= \alpha^2 \text{Var}(r_t^2) + (\beta^2 + 2\alpha\beta) \text{Var}(\sigma_t^2),
\]

i.e.,

\[
R^2 = \frac{\alpha^2}{1 - \beta^2 - 2\alpha\beta}. \quad (12)
\]

- E.g., an upper bound can be obtained by the assumption of a finite fourth moment, i.e.,

\[
1 > \kappa \alpha^2 + 2\alpha\beta + \beta^2 \Rightarrow \frac{1}{\kappa} > \frac{\alpha^2}{1 - 2\alpha\beta - \beta^2} = R^2; \quad (13)
\]

e.g., when \( \eta_t \sim N(0, 1) \), \( R^2 \) can never be larger than \( \frac{1}{3} \) (and even smaller with fat-tailed innovations).
\[ R^2 = \frac{\alpha^2}{1 - 2\alpha \beta - \beta^2} \]
Variance Forecasts

For the estimated model in (5),

\[
\hat{\alpha} = 0.0718, \quad \hat{\beta} = 0.9209, \quad (14)
\]

so the fourth moment would be finite, and the estimated population \( R^2 \) would be

\[
\hat{R}^2 = \frac{0.0718^2}{1 - 2 \cdot 0.0718 \cdot 0.9209 - 0.9209^2} = 0.262. \quad (15)
\]
Variance Forecasts

• A better measure for the ex-post volatility is based on intraday returns and referred to as realized variance.

• Suppose there are $n$ intraday observations during day $t$, then the $i$th intraday return is

$$r_{t,i} = \log P_{t,i} - \log P_{t,i-1}, \quad i = 1, \ldots, n,$$

(16)

where $P_{it}$ is the $i$th price observation on day $t$, and the daily return

$$r_t = \sum_{i=1}^{n} r_{i,t}.$$  

(17)
Variance Forecasts

- The daily realized variance ($RV_{t,n}$) for day $t$ is defined as\(^6\)

$$RV_{t,n} = \sum_{i=1}^{n} r_{t,i}^2$$  \hspace{1cm} (18)

which, as $n$ increases, provides a more and more accurate estimator of the variance, provided certain conditions on the data generating process apply.\(^7\)

---

\(^6\) The mean of intraday returns is basically zero.

\(^7\) In practice, there are many data problems caused by, e.g., market microstructure effects, which have to be resolved; for discussion and further references, see, e.g., Chapter 12 of Taylor (2005): *Asset Price Dynamics, Volatility, and Prediction*; and Chapter 5 of Christoffersen (2012): *Elements of Financial Risk Management*. 
Squared returns are more noisy than realized variances.
Stylized features of realized volatility

1. Distribution of standardized returns:

Recall that standardizing by the GARCH–implied conditional variances $\hat{\sigma}_{t}^{2}$ produces standardized residuals $\hat{\eta}_{t} = \frac{\hat{\epsilon}_{t}}{\hat{\sigma}_{t}}$ which display significant excess kurtosis.

In contrast, standardizing by the square root of realized variance produces almost $N(0, 1)$ residuals, i.e., calculating

$$\tilde{\eta}_{t} = \frac{\hat{\epsilon}_{t}}{\sqrt{RV_{t,n}}}.$$  (19)
GARCH-standardized residuals

RV-standardized residuals

- kernel
- \(- N(0,1)\) density
Stylized features of realized volatility

Table 1: Approximate normality of returns divided by the square root of realized variance

<table>
<thead>
<tr>
<th>skewness</th>
<th>kurtosis</th>
<th>Jarque–Bera</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.134***</td>
<td>3.06</td>
<td>7.87**</td>
</tr>
</tbody>
</table>

Asterisks ** and *** indicate significance at the 5% and 1% levels, respectively.

- The explanation is that RV$_{t,n}$ is an *ex–post* measure, i.e., it is known only at the end of day $t$,$^8$ whereas the GARCH volatilities are *ex–ante* measures.

$^8$Thus, the normality of standardized residuals has no immediate value for prediction.
Stylized features of realized volatility

2. The distribution of $\log RV_{t,n}$ is approximately normal.

3. Since $RV_{t,n}$ is less noisy than $r_t^2$, a larger part of it is predictable. This is reflected in higher autocorrelations with slow decay.\(^9\)

The next slide shows that autocorrelations of both squared returns and realized variance.

\(^9\)This is often modeled by means of long memory processes, see Taylor (2005), Sec. 12.9.5.
sample autorcorrelations of squared returns

sample autorcorrelations of realized variances
Mincer–Zarnowitz regression with realized variance

- We can replace the squared return in (2) with realized variance, i.e., run the regression

\[ \text{RV}_{t,n} = a_0 + a_1 \hat{\sigma}_t^2 + u_t, \]  \hspace{2cm} (20)

where \( \hat{\sigma}_t^2 \) is the conditional variance implied by the GARCH model.

- For the S&P 500 returns, regression (20) produces a coefficient of determination \( R^2 = 0.477 \).

- That is, approximately 50% of the variation in realized variance can be predicted by the conditional GARCH–variances.
Forecasting with an asymmetric GARCH model

- The GARCH model considered so far assumes that the conditional variance $\sigma_t^2$ depends only on the magnitude and not on the sign of past shocks.

- However, stock market variance tends to exhibit an asymmetric pattern in that it reacts more strongly to negative than to positive news (shocks).

- This is sometimes referred to as the leverage effect.
Forecasting with an asymmetric GARCH model

- This can be made visible by means of both squared returns and realized volatility, defining the correlation functions\(^{10}\)

\[
L_1(\tau) = \text{Corr}(\epsilon_{t-\tau}, \epsilon_t^2)
\]

(21)

and

\[
L_2(\tau) = \text{Corr}(\epsilon_{t-\tau}, \text{RV}_t).
\]

(22)

- As shown on the next slide, the presence of a leverage effect is more clear–cut on the basis of \(L_2(\tau)\) based on \(\text{RV}_t\).

\(^{10}\)We could also look at \(\text{Corr}(\epsilon_{t-\tau}, |\epsilon_t|)\).
\[ L_1(\tau) = \text{Corr}(\varepsilon_{t-\tau}, \varepsilon_t^2) \]

\[ L_2(\tau) = \text{Corr}(\varepsilon_{t-\tau}, \text{RV}_t) \]
Forecasting with an asymmetric GARCH model

- The asymmetric GARCH model proposed by Glosten, Jagannathan and Runkle (1993), referred to as GJR–GARCH, models the conditional variance as
  \[
  \sigma_t^2 = \omega + (\alpha + \theta I_{t-1})\epsilon_{t-1}^2 + \beta \sigma_{t-1}^2,
  \]
  where the indicator variable
  \[
  I_{t-1} = \begin{cases} 
  1 & \text{if } \epsilon_{t-1} < 0 \\
  0 & \text{if } \epsilon_{t-1} \geq 0 
  \end{cases}
  \]
  - Clearly \( \theta > 0 \) implies that conditional variance reacts more strongly to negative than to positive shocks of the same magnitude.
Forecasting with an asymmetric GARCH model

- If we fit this model (with Gaussian innovations) to the S&P 500 returns over the first 1000 observations (retaining the remaining observations for out-of-sample forecasting), we get

\[
\sigma_t^2 = 0.010 + (0.000 + 0.107 I_{t-1}) \epsilon_{t-1}^2 + 0.938 \sigma_{t-1}^2,
\]

i.e., positive shocks are estimated to have *no impact* on conditional volatility.
Forecasting with an asymmetric GARCH model

- Doing the forecast exercise with both squared returns and realized variance as variance proxies, the results in Table 2 are obtained.

Table 2: Forecasting S&P 500 variances: \( R^2 \) from Mincer–Zarnowitz regressions

<table>
<thead>
<tr>
<th>proxy</th>
<th>( r_t^2 )</th>
<th>( RV_t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH</td>
<td>0.247</td>
<td>0.477</td>
</tr>
<tr>
<td>GJR–GARCH</td>
<td>0.279</td>
<td>0.534</td>
</tr>
</tbody>
</table>
Can we do better?

- Yes. Namely, provided high-quality intraday data are available, we can model $RV_t$ (more specifically, $\log RV_t$) directly, e.g., using (long memory) ARMA-type processes.

- This typically gives rise to improved variance (and covariance) forecasts and can be used in many applications.\(^\text{11}\)

- In some applications, such as Value-at-Risk measurement, however, the specification of the conditional distribution tends to be of much greater importance than that of the volatility dynamics.

\(^{11}\)See, e.g., the references in Footnote 7 and the references therein.
Realized GARCH

- As suggested by Engle (2002),\(^\text{12}\) we can also extend the GARCH model with RV as an explanatory variable, i.e.,

\[
\begin{align*}
    r_t &= \mu_t + \epsilon_t \\
    \epsilon_t &= \eta_t \sigma_t, \quad \eta_t \overset{iid}{\sim} N(0, 1) \\
    \sigma^2_t &= \omega + \alpha \epsilon^2_{t-1} + \beta \sigma^2_{t-1} + \gamma RV_{t-1}.
\end{align*}
\]

- Structure (23)–(25) can be estimated via conditional maximum likelihood as before.

- Doing so for the S&P 500 returns 2001–2010, we obtain estimates

\[
\sigma^2_t = 0.012 + 0.000 \frac{\epsilon^2_{t-1}}{0.0099} + 0.661 \frac{\sigma^2_{t-1}}{0.0591} + 0.329 \frac{RV_{t-1}}{0.0597},
\]

i.e., the squared lagged shock loses its explanatory power when realized variance is added to the variance equation.

Realized GARCH

• A drawback of model (23)–(25) is that the dynamics of $RV_t$ are not explicitly specified, so that multi-period-ahead forecasts cannot be calculated.

• A more general model which accounts for this weakness is discussed in Hansen et al. (2012).\textsuperscript{13}