More GARCH, More Volatility

(Tsay: Chapters 3.5–3.8)

Isabel Casas

SYDDANSK UNIVERSITET

- Estimating GARCH by MLE
- GARCH forecasting
- Extensions of GARCH
  - GARCH-M
  - Asymmetric GARCH
  - Forecasting
  - Long memory
  - Evaluating volatility predictions

Estimating GARCH by MLE

Consider estimating the model

$$\begin{aligned} r_t =& \mu_t + \epsilon_t = \mathbf{X_t}\beta + \epsilon_t \\ \epsilon_t =& z_t \sigma_t, \qquad z_t \sim IIDN(0,1) \\ \sigma_t^2 =& a_0 + a(L)\epsilon_t^2 + b(L)\sigma_t^2 \end{aligned}$$

Result: The regression parameters β and GARCH parameters $\gamma = (a_0, a_1, \ldots, a_p, b_1, \ldots, b_q)'$ can be estimated separately because the information matrix for $\theta = (\beta', \gamma')'$ is block diagonal.

Estimating GARCH by MLE

Step 1 Estimate β by OLS ignoring ARCH errors and form residuals $\hat{\epsilon}_t = r_t - \mathbf{X}\beta$

Step 2 Estimate GARCH process for residuals $\hat{\epsilon}_t$ by MLE.

Warning: Block diagonality of information matrix fails if

- pdf of z_t is not a symmetric density
- $\boldsymbol{\beta}$ and γ are not variation free; e.g. GARCH-M model

GARCH likelihood function under normality

Assume $\mu_t = 0$. Let $\boldsymbol{\theta} = (a_0, a_1, \ldots, a_p, b_1, \ldots, b_q)'$ denote the parameters to be estimated. Since $\epsilon_t = z_t \sigma_t$

$$f(\epsilon_t | F_{t-1}; \theta) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{ -\frac{\epsilon_t^2}{2\sigma_t^2} \right\}$$

GARCH likelihood function under normality

For a sample of size T, the prediction error decomposition gives

$$f_\epsilon(\epsilon_t, \ldots, \epsilon_T; \boldsymbol{\theta}) = \prod_{t=p+1}^{T} f_{\epsilon_t|F_{t-1}}(\epsilon_t|\epsilon_{t-1}, \ldots \epsilon_1; \boldsymbol{\theta}) \underbrace{f_{\epsilon_1,\ldots,\epsilon_p}(\epsilon_1, \ldots, \epsilon_p; \boldsymbol{\theta})}_{\text{no closed form}}$$

$$= \left(\prod_{t=p+1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{ -\frac{\epsilon_t^2}{2\sigma_t^2} \right\} \right) f_{\epsilon_1,\ldots,\epsilon_p}(\epsilon_1, \ldots, \epsilon_p; \boldsymbol{\theta})$$

where $\sigma_t^2 = a_0 + a(L)\epsilon_t^2 + b(L)\sigma_t^2$ may be evaluated recursively given starting values for σ_t^2. The log-likelihood function is

GARCH likelihood function under normality

The exact log-likelihood function

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{(T-m+1)}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=p+1}^{T}\left[\ln(\sigma_t^2) + \frac{\epsilon_t^2}{\sigma_t^2}\right] + \log(f_{\epsilon_1,\ldots,\epsilon_p}(\epsilon_1,\ldots,\epsilon_p;\boldsymbol{\theta}))$$

GARCH likelihood function under normality

- Problem: the marginal density for the initial values does not have a closed form expression so exact mle is not possible.
- In practice: initial values $(\epsilon_1, \ldots, \epsilon_p)$ are set equal to zero and the marginal density $f(\epsilon_1, \ldots, \epsilon_p; \theta)$ is ignored.
- This is conditional mle.

Practical issues

- Starting values for the model parameters $a_i : (i = 0, \ldots, p)$ and $b_j : (j = 1, \ldots, q)$ need to be chosen and an initialization of ϵ_t^2 and σ_t^2 must be supplied.
- Zero values are often given for $a_i : i > 1$ and b_j
- a_0 is set equal to the unconditional variance of r_t
- For the initial values of σ_t^2, a popular choice is

$$\sigma_t^2 = \epsilon_t^2 = \frac{1}{T} \sum_{s=p+1}^{T} y_s^2, \quad t < p$$

Practical issues

- Once the log-likelihood is initialized, it can be maximized using numerical optimization techniques. The most common method is based on a Newton-Raphson iteration of the form

$$\hat{\theta}_{n+1} = \hat{\theta}_n - \lambda_n \mathbf{H}(\hat{\theta}_n)^{-1} s(\hat{\theta}_n)$$

- For GARCH models, the BHHH algorithm is often used. This algorithm approximates the Hessian matrix using only first derivative information

$$-\mathbf{H}(\theta) \approx \mathbf{B}(\theta) = \sum_{t=1}^{T} \frac{\partial l_t}{\partial \theta} \frac{\partial l_t}{\partial \theta'}$$

- Under suitable regularity conditions, the ML estimates are consistent and asymptotically normally distributed and an estimate of the asymptotic covariance matrix of the ML estimates is constructed from an estimate of the final Hessian matrix from the optimization algorithm used.

Numerical Accuracy of GARCH Estimates

- GARCH estimation is widely available in a number of commercial software packages (e.g. EVIEWS, GAUSS, MATLAB, Ox, RATS, S-PLUS, TSP) and there are also a few free open source implementations. (Even Excel!)
- Starting values, optimization algorithm choice, and use of analytic or numerical derivatives, and convergence criteria all influence the resulting numerical estimates of the GARCH parameters.
- The GARCH log-likelihood function is not always well behaved, especially in complicated models with many parameters, and reaching a global maximum of the log-likelihood function is not guaranteed using standard optimization techniques. Poor choice of starting values can lead to an ill-behaved log-likelihood and cause convergence problems.

Numerical Accuracy of GARCH Estimates

- In many empirical applications of the GARCH(1,1) model, the estimate of a_1 is close to zero and the estimate of b_1 is close to unity. This situation is of some concern since the GARCH parameter b_1 becomes unidentified if $a_1 = 0$, and it is well known that the distribution of ML estimates can become ill-behaved in models with nearly unidentified parameters.
- Ma, Nelson and Startz (2007) studied the accuracy of ML estimates of the GARCH parameters a_0, a_1 and b_1 when a_1 is close to zero. They found that the estimated standard error for b_1 is spuriously small and that the t-statistics for testing hypotheses about the true value of b_1 are severely size distorted. They also showed that the concentrated loglikelihood as a function of b_1 exhibits multiple maxima.

Numerical Accuracy of GARCH Estimates

- To guard against spurious inference they recommended comparing estimates from pure ARCH(p) models, which do not suffer from the identification problem, with estimates from the GARCH(1,1). If the volatility dynamics from these models are similar then the spurious inference problem is not likely to be present.

Quasi-Maximum Likelihood Estimation

- The assumption of conditional normality is not always appropriate.
- However, even when normality is inappropriately assumed, maximising the Gaussian log-likelihood results in quasi-maximum likelihood estimates (QMLEs) that are consistent and asymptotically normally distributed provided the conditional mean and variance functions of the GARCH model are correctly specified.
- An asymptotic covariance matrix for the QMLEs that is robust to conditional non-normality is estimated using

$$\mathbf{H}(\hat{\boldsymbol{\theta}}^{\mathbf{QML}})^{-1}\mathbf{B}(\hat{\boldsymbol{\theta}}^{\mathbf{QML}})\mathbf{H}(\hat{\boldsymbol{\theta}}^{\mathbf{QML}})^{-1}$$

where $\hat{\boldsymbol{\theta}}^{QML}$ denotes the QMLE of $\boldsymbol{\theta}$, and is often called the "sandwich" estimator.

Determining lag length

- Use model selection criteria (AIC or BIC)
- For GARCH(p, q) models, those with $p, q \leq 2$ are typically selected by AIC and BIC.
- Low order GARCH(p,q) models are generally preferred to a high order ARCH(p) for reasons of parsimony and better numerical stability of estimation (high order GARCH(p, q) processes often have many local maxima and minima).
- For many applications, it is hard to beat the simple GARCH(1,1) model.

SYDDANSK UNIVERSITET

Model Diagnostics

Correct model specification implies

$$\frac{\hat{\epsilon}_t}{\hat{\sigma}_t} \sim IIDN(0,1)$$

- Test for normality — Jarque-Bera, QQ-plot
- Test for serial correlation — Ljung-box, SACF, SPACF
- Test for ARCH effects — serial correlation in squared standardized residuals, LM test for ARCH

Returns forecast for a GARCH model

- Suppose one is interested in forecasting future values of r_t in the standard GARCH model.
- For simplicity assume that $E[r_{t+1}|F_t] = c \Rightarrow r_t = c + \sigma_t z_t$.
- The h-step-ahead forecast of $r^*_{T+h|T} = c$ which does not depend on the GARCH parameters
- The forecast error is $\epsilon_{T+h} = \sigma_{T+h} z_{t+h}$
- The conditional variance of this forecast error is then

$$Var(\epsilon_{T+h}) = E[\sigma^2_{T+h}|F_T] = \sigma^{2*}_{T+h}$$

  which does depend on the GARCH parameters.
- Therefore, in order to produce confidence bands for the h-step-ahead forecast of the returns, the h-step-ahead volatility forecast is needed.

SYDDANSK UNIVERSITET

Volatility forecast for a GARCH(1,1)

It is similar to the ARMA model.

- 1-step-ahead forecast (all known)

$$\sigma^{2*}_{T+1} = a_0 + a_1\epsilon^2_T + b_1\sigma^2_T$$

- 2-step-ahead forecast: $\epsilon^2_t = \sigma^2_t z^2_t$

$$\sigma^{2*}_{T+2} = a_0 + (a_1 + b_1)\sigma^{2*}_{T+1}$$

- h-step-ahead forecast

$$\begin{aligned}\sigma^{2*}_{T+h} =& a_0 + (a_1 + b_1)\sigma^{2*}_{T+h-1} \\ =& \frac{a_0(1 - (a_1 + b_1)^{h-1})}{1 - a_1 - b_1} + (a_1 + b_1)^{h-1}\sigma^{2*}_{T+1}\end{aligned}$$

Volatility forecast for a GARCH(1,1)

If $(a_1 + b_1) < 1$, then

$$\sigma^{2*}_{T+h} \to \frac{a_0}{1 - a_1 - b_1}$$

when $h \to \infty$

The multistep-ahead volatility forecasts of a GARCH(1,1) converge to the unconditional variance of ϵ_t as the forecast increases to infinity, provided $Var(\epsilon_t)$ exists.

Volatility forecast for a GARCH(1,1)

Remarks

- The forecast of volatility (standard deviation) is defined as

$$E[\sigma_{T+k}|F_T] \neq (E[\sigma^2_{T+k}|F_T])^{1/2} \text{ (by Jensen's inequality)}$$

- Standard errors for $E[\sigma_{T+k}|F_T]$ are not available in closed form but may be computed using simulation methods.

SYDDANSK UNIVERSITET

Forecasting the Volatility of Multiperiod Returns

- Let $r_t = \log(P_t) - \log(P_{t-1})$. The GARCH forecasts are for daily volatility at different horizons h.
- For risk management and option pricing with stochastic volatility, volatility forecasts are needed for multiperiod returns.
- With continuously compounded returns, the h-day return between days T and $T+h$ is simply the sum of h single day returns

$$r_{T+h}(h) = \sum_{j=1}^{h} r_{T+j}$$

Forecasting the Volatility of Multiperiod Returns

- Assuming returns are uncorrelated, the conditional variance of the h-period return is then

$$
\begin{aligned}
Var(r_{T+h}(h)|F_T) =& \sigma_T^2(h) = \sum_{j=1}^{h} var(r_{T+j}|F_T) \\
=& E[\sigma_{T+1}^2|F_T] + \ldots + E[\sigma_{T+h}^2|F_T]
\end{aligned}
$$

- If returns have constant variance $\bar{\sigma}^2$, then $\sigma_T^2(h) = h\bar{\sigma}^2$ and $\sigma_T(h) = \sqrt{h}\bar{\sigma}$.
- This is known as the "square root of time" rule as the h-day volatility scales with $\sqrt{h}$.
- In this case, the h-day variance per day, $\sigma_T^2(h)/h$, is constant.

Forecasting the Volatility of Multiperiod Returns

- If returns are described by a GARCH model then the square root of time rule does not necessarily apply.
- Plugging the GARCH(1,1) model forecasts for $E[\sigma^2_{T+1}|F_T], \ldots, E[\sigma^2_{T+h}|F_T]$ into $var(r_{T+h}(h)|F_T)$ gives:

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

- For the GARCH(1,1) process the square root of time rule only holds if $E[\sigma^2_{T+1}] = \bar{\sigma}^2$. Whether $\sigma^2_T(h)$ is larger or smaller than $h\bar{\sigma}^2$ depends on whether $E[\sigma^2_{T+1}]$ is larger or smaller than $\bar{\sigma}^2$.
- Term structure of volatility is a plot of $\sigma^2_T(h)/h$ versus h

GARCH-in-Mean (GARCH-M)

Modern finance theory suggests that volatility may be related to risk premia on assets.

The GARCH-M model allows time-varying volatility to be realted to expected returns

$$r_t = c + \alpha\sigma_t^2 + \epsilon_t \quad \epsilon_t \sim GARCH$$

- α is the risk premium parameter
- $\alpha > 0$ indicates taht the return is positively related to its volatility
- Other risk premium specifications:
  - $r_t = c + \alpha\sigma_t + \epsilon_t$
  - $r_t = c + \alpha\ln(\sigma_t^2) + \epsilon_t$

Temporal Aggregation

- Volatility clustering and non-Gaussian behavior in financial returns is typically seen in weekly, daily or intraday data. The persistence of conditional volatility tends to increase with the sampling frequency.
- For GARCH models there is no simple aggregation principle that links the parameters of the model at one sampling frequency to the parameters at another frequency. This occurs because GARCH models imply that the squared residual process follows an ARMA type process with MDS innovations which is not closed under temporal aggregation.

Temporal Aggregation

- The practical result is that GARCH models tend to be fit to the frequency at hand. This strategy, however, may not provide the best out-of-sample volatility forecasts. For example,Martens (2002) showed that a GARCH model fit to S&P 500 daily returns produces better forecasts of weekly and monthly volatility than GARCH models fit to weekly or monthly returns, respectively.

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 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 ϵ_{t-1} and the sign of ϵ_{t-1}. A significant value of β_1 indicates evidence for asymmetric effects on conditional volatility.

Testing 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; and
    - Setting $\hat{w}_{t-1} = S^+_{t-1}\hat{\epsilon}_{t-1}$ gives the Positive Size Bias test.

EGARCH (p, q) Model

Nelson (1991) proposes the exponential GARCH model.

- This model explains asymmetric effects between positive and negative returns
- Define $h_t = \ln(\sigma_t^2)$

$$
\begin{aligned}
r_t =& \mu_t + \sigma_t z_t \\
h_t =& a_0 + \sum_{i=1}^{q} a_i g(z_{t-i}) + \sum_{j=1}^{q} b_j h_{t-j} \\
g(z_t) =& \theta z_t + \gamma[|z_t| - E(|z_t|)]
\end{aligned}
$$

EGARCH (p, q) Model

- Polynomials $1 + b(L)$ and $1 - a(L)$ have zeros outside the unit circle and have no common factors
- Variance is always positive because $\sigma_t^2 = \exp(h_t)$
- A positive ϵ_{t-i} contributes with $a_i(1 + \gamma_i)|z_{t-i}|$ to the log volatility
- A negative gives $a_i(1 - \gamma_i)|z_{t-i}|$
- We expect a negative γ_i in real applications. Why?

Example EGARCH(1,1)

$$
\begin{aligned}
r_t =& \mu_t + \epsilon_t \\
\epsilon_t =& \sigma_t z_t \qquad z_t \sim N(0,1) \\
h_t =& a_0 + a_1(\theta z_t + \gamma[|z_t| - E(|z_t|)]) + b_1 h_{t-1}
\end{aligned}
$$

- If $\gamma < 0$, this means the levearage effect of ϵ_{t-1}

Syddansk Universitet

Example EGARCH(1,1)

- The application is run on the monthly GM simple returns, file "m-gmsp500.txt"
- The R function *egarch* fits an EGARCH (p,q) with a normal or a GED distributed innovations
- Their formula is:

$$
\begin{aligned}
h_t = &\beta_0 + \sum_{j=1}^{q} \eta_j \epsilon_{t-j} \\
&+ \sum_{j=1}^{q} \gamma_j (|\epsilon_{t-j}| - E(|\epsilon_{t-j}|) + \sum_{i=1}^{p} \beta_i h_{t-i}
\end{aligned}
$$

- include.shape=T implies that the GED is used as the distribution of innovations

Example EGARCH(1,1)

```
> options(width = 55)
> library(egarch)
> gm <- read.table("../data/m-gmsp500.txt",
+     h = T)[, 2]
> gmln = log(gm + 1)
> gm.egarch = egarch(gmln, order = c(1, 1),
+     include.mu = T, include.shape = T)
```

SYDDANSK UNIVERSITET

Example EGARCH(1,1)

```
> print(gm.egarch)

$mu
       mu
0.0685876

$beta
     beta0      beta1
0.05807339 1.00828524

$eta
     eta1
0.0965128

$gamma
   gamma1
0.2498223

$nu
      nu
1.705952

$ics
      AIC       BIC      HQIC
-2.442644 -2.403979 -2.427705

attr(,"class")
[1] "egarch"
```

Example EGARCH(1,1)

$$r_t = 0.0686 + \epsilon_t \qquad z_t \sim GED(\nu = 1.706)$$

$$\begin{aligned} h_t =& 0.0581 + 0.0965 z_{t-1} \\ &+ 0.2498(|z_t| - E(|z_t|)) + 1.0083 h_{t-1} \end{aligned}$$

SYDDANSK UNIVERSITET

TGARCH/GJR Model

Zakoian (1994) introduces the Threshold GARCH (aka GJR - Glosten, Jagannathan, and Runkle , 1993). A TGARCH(p,q)

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

$$S_{t-i} = \begin{cases} 1 & \text{if} \quad \epsilon_{t-i} < 0 \\ 0 & \text{if} \quad \epsilon_{t-i} \geq 0 \end{cases}$$

- The model uses zero as the threshold to separate impacts, but other thresholds can be used
- $\delta = 1$ TGARCH
- $\delta = 2$ GJR

TGARCH/GJR Model

- When ϵ_{t-i} is positive, the total effects are $a_i\epsilon_{t-i}^{\delta}$
- When ϵ_{t-i} is negative, the total effects are $(a_i+\gamma_i)\epsilon_{t-i}^{\delta}$
- Leverage effect implies that $\gamma_i > 0$
- TGARCH/GJR is covariance stationary provided $\sum_{i=1}^{p}(ai+\gamma_i/2)+\sum_{j=1}^{q} b_j < 1$

Example GJR(1,1)

Monthly GM log returns as like in the previous example.

```
> library(fGarch)
> gm.GJR = garchFit(gmln ~ 1 + aparch(1, 1),
+      trace = F, include.shape = T, cond.dist = "ged",
+      delta = 2, leverage = T, include.delta = 2)
> coef = round(coef(gm.GJR), 4)
> coef

     mu   omega  alpha1  gamma1   beta1   shape
-0.0164  0.0217  0.0972  0.0960  0.7948  3.9983
```

$$r_t = -0.0164 + \epsilon_t \qquad z_t \sim GED(\nu = 3.9983)$$

$$\sigma_t^2 = 0.0217 + (0.0972 + 0.096 S_{t-1})\epsilon_{t-1}^2 + 0.7948\sigma_{t-1}^2$$

A-PARCH Model

Ding, Granger and Engle (1993) Asymmetric Power ARCH model:

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

- $a_0 > 0$, $d > 0$, $-1 < \gamma_i < 1$
- Leverage effect implies that $\gamma_i < 0$
- $d = 2$, $\gamma_i = 0$, $b_j = 0 \Rightarrow$ ARCH
- $d = 2$, $\gamma_i = 0 \Rightarrow$ GARCH
- $d = 2 \Rightarrow$ GJR , $d = 1 \Rightarrow$ TGARCH
- $d = 1$, $\gamma_i = 0 \Rightarrow$ Taylor/Schwert's GARCH for standard deviation
- d can be fixed at a particular value or estimated by MLE

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.*

Forecasts from Asymmetric GARCH(1,1) Models

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

$$\sigma_T^2 = a_0 + a_1\epsilon_{T-1}^2 + \gamma_1 S_{T-1}\epsilon_{T-1}^2 + b_1\sigma_{T-1}^2$$

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

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

  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 $S_T = 1$ ($\epsilon_T < 0$).

Forecasts from Asymmetric GARCH(1,1) Models

- The forecast at $T+2$ is

$$\begin{aligned} E[\sigma^2_{T+2}|F_T] =& a_0 + a_1 E[\epsilon^2_{T+1}|F_T] + \gamma_1 E[S_{T+1}\epsilon^2_{T+1}|F_T] + b_1 E[\sigma^2_{T+1}|F_T \\ =& a_0 + (\frac{\gamma_1}{2} + a_1 + b_1) E[\sigma^2_{T+1}|F_T] \end{aligned}$$

  which follows since

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

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

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

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

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

Forecasts from Asymmetric GARCH(1,1) Models

- The mean reverting form is

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

where $\bar{\sigma}^2 = a_0/(1 - \frac{\gamma_1}{2} - a_1 - b_1)$ is the long run variance.
- Forecasting algorithm in the EGARCH(1,1) follow in a similar manner.

GARCH Models with Non-Normal Errors

- Often the standardized residuals from a GARCH model with Gaussian errors still has fat tails. This suggests using a fat-tailed error distribution instead.
- The most common fat-tailed error distributions for fitting GARCH models are: the Student-t distribution; the double exponential distribution; and the generalized error distribution.

GARCH with Student-t errors

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

$$f(u_t) = \frac{\Gamma[(\nu+1)/2] s_t^{-1/2}}{(\phi\nu)^{1/2}\Gamma(\nu/2)[1+u_t^2/(s_t\nu)]^{(\nu+1)/2}}$$
$$Var(u_t) = \frac{s_t\nu}{\nu-2}, \qquad \nu > 2$$

If u_t in GARCH model is Student-t with $E[u_t^2|F_{t-1}] = \sigma_t^2$ then,

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

Generalised Error Distribution

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

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

where

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

- $\nu = 2$ gives the normal distribution
- $0 < \nu < 2$ gives a distribution with fatter tails than normal
- $\nu > 2$ gives a distribution with thinner tails than normal
- $\nu = 1$ gives the double exponential distribution

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

$$\rho_k = C_\rho k^{2d-1} \quad \text{as } k \to \infty$$

  where C_ρ is a positive constant, and d is a real number between 0 and $1/2$.

Thus the autocorrelation function of a long memory process decays slowly at a hyperbolic rate.

Integrated GARCH Model

- 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 $E[\ln(a_1 z_t^2 + b_1)] < 0$.

Integrated GARCH Model

- Diebold and Lopez (1996) argued against the IGARCH specification for modeling highly persistent volatility processes for two reasons
  1. the observed convergence toward normality of aggregated returns is inconsistent with the IGARCH model.
  2. the observed IGARCH behavior may result from misspecification of the conditional variance function. For example, a two components structure or ignored structural breaks 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 as well as with 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.

Evaluating Volatility Predictions

1. **Recursive forecasts**: An initial sample using data from $t = 1, \dots, T$ 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 $T + 1$.
2. **Rolling forecasts**. An initial sample using data from $t = 1, \dots, T$ 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 $t = 2, \dots, T + 1$, and h-step-ahead out-of-sample forecasts are produced starting at time $T + 1$.

Traditional Forecast Evaluation Statistics

- Let $E_i[\sigma^2_{T+h}|F_T]$ 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 $e_{i,T+h|T} = E_i[\sigma^2_{T+h}] - \sigma^2_{T+h}$
- Common forecast evaluation statistics

$$MSE_i = \frac{1}{N} \sum_{j=T+1}^{N} e^2_{i,j+h|j}$$

$$MAE_i = \frac{1}{N} \sum_{j=T+1}^{N} |e_{i,j+h|j}|$$

$$MAPE_i = \frac{1}{N} \sum_{j=T+1}^{N} \frac{|e_{i,j+h|j}|}{\sigma_{j+h}}$$

Traditional Forecast Evaluation Statistics

- 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 $L(e_{i,T+h|T})$, $i = 1, 2$. Common choices:
  - squared error loss function: $L(e_{i,T+h|T}) = (e_{i,T+h|T})^2$
  - absolute error loss function: $L(e_{i,T+h|T}) = |e_{i,T+h|T}|$
- The Diebold-Mariano (DM) test is based on the loss differential

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

Diebold-Mariano Tests for Predictive Accuracy

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

$$S = \frac{\bar{d}}{(\widehat{\bar{d}})^{1/2}}, \qquad \bar{d} = \frac{1}{N} \sum_{j=T+1}^{T+N} d_{j+h}$$

- Diebold and Mariano recommend using the Newey-West estimate for $AVar(d)$ 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,

$$S \sim N(0, 1)$$

Diebold-Mariano Tests for Predictive Accuracy

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

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

- 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 $E_i[\sigma^2_{T+h}|F_T]$ 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

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

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

Fundamental Problem with Evaluating Volatility Forecasts

- 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.

Fundamental Problem with Evaluating Volatility Forecasts

- 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).

Fundamental Problem with Evaluating Volatility Forecasts

- 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.

SYDDANSK UNIVERSITET

Using Realized Variance to Evaluate Volatility Forecasts

- Under certain conditions, RV^m_{t+h} is a consistent estimate of σ^2_{t+h} as $m \to \infty$. As a result, RV^m_{T+h} is a much less noisy estimate of σ^2_{T+h} than ϵ^2_{T+h} and so forecast evaluations based on the realised variance are expected to be much more accurate than those based on the squared residuals.
- 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^2_{T+1}, RV^{24}_{T+1} and $RV^2_{T+1}88$, respectively.