Financial Data Analysis

GARCH Models, Part II

Summer 2014

July 15, 2014
GARCH Models

- In practice, pure ARCH($q$) processes are rarely used.
- The reason is that, for an adequate fit, a large number $q$ of lags is usually called for, thus requiring estimation of a large number of parameters.
Illustration: Fitting ARCH Models

- To illustrate, consider the daily DAX returns from January 2000 to October 2011, $T = 3081$ daily returns.

- Fit ARCH($q$) models of the form

  \[ r_t = c + \epsilon_t \]
  \[ \epsilon_t = \eta_t \sigma_t, \quad \eta_t \overset{iid}{\sim} N(0, 1), \]
  \[ \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2, \]
  \[ \omega > 0, \quad \alpha_i \geq 0, \quad i = 1, \ldots, q. \]

- Denote parameter estimates by $\hat{c}, \hat{\omega}, \hat{\alpha}_1, \ldots, \hat{\alpha}_q$. 
Illustration: Fitting ARCH Models

Table 1: ARCH($q$) parameter estimates

<table>
<thead>
<tr>
<th>$q =$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}_1$</td>
<td>0.317</td>
<td>0.205</td>
<td>0.133</td>
<td>0.070</td>
<td>0.050</td>
<td>0.042</td>
<td>0.039</td>
<td>0.041</td>
<td>0.031</td>
<td>0.022</td>
</tr>
<tr>
<td>$\hat{\alpha}_2$</td>
<td>0.392</td>
<td>0.323</td>
<td>0.281</td>
<td>0.233</td>
<td>0.206</td>
<td>0.171</td>
<td>0.153</td>
<td>0.137</td>
<td>0.133</td>
<td></td>
</tr>
<tr>
<td>$\hat{\alpha}_3$</td>
<td>0.268</td>
<td>0.222</td>
<td>0.185</td>
<td>0.170</td>
<td>0.167</td>
<td>0.158</td>
<td>0.155</td>
<td>0.151</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\alpha}_4$</td>
<td></td>
<td>0.192</td>
<td>0.179</td>
<td>0.177</td>
<td>0.156</td>
<td>0.149</td>
<td>0.134</td>
<td>0.125</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\alpha}_5$</td>
<td></td>
<td></td>
<td>0.174</td>
<td>0.150</td>
<td>0.138</td>
<td>0.129</td>
<td>0.122</td>
<td>0.100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\alpha}_6$</td>
<td></td>
<td></td>
<td></td>
<td>0.121</td>
<td>0.097</td>
<td>0.089</td>
<td>0.072</td>
<td>0.069</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\alpha}_7$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.111</td>
<td>0.108</td>
<td>0.095</td>
<td>0.091</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\alpha}_8$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.055</td>
<td>0.048</td>
<td>0.043</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\alpha}_9$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.083</td>
<td>0.086</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\alpha}_{10}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.080</td>
<td></td>
</tr>
<tr>
<td>$\sum_{i=1}^{q} \hat{\alpha}_i$</td>
<td>0.317</td>
<td>0.597</td>
<td>0.724</td>
<td>0.765</td>
<td>0.821</td>
<td>0.867</td>
<td>0.878</td>
<td>0.882</td>
<td>0.876</td>
<td>0.899</td>
</tr>
</tbody>
</table>
Illustration: Fitting ARCH Models

• The residuals and model–implied conditional variances,

\[
\begin{align*}
\hat{e}_t &= r_t - \hat{c} \\
\hat{\sigma}_t^2 &= \hat{\omega} + \hat{\alpha}_1 \hat{e}_{t-1}^2 + \cdots + \hat{\alpha}_q \hat{e}_{t-q}^2,
\end{align*}
\]

and the standardized residuals

\[
\hat{\eta}_t = \frac{\hat{e}_t}{\hat{\sigma}_t}, \quad t = q + 1, \ldots, T. \quad (1)
\]

• For a correctly specified model, sequence (1) should behave similar to the iid innovation sequence \( \{\eta_t\} \); in particular, it should not display significant ARCH effects.

• We may look at the autocorrelations of \(|\hat{\eta}_t|\) for various model orders \(q\).
Same exercise for the S&P 500 (ACF of $|\hat{\eta}_t|$ defined in (1)): 

![Graphs showing ACF for different values of q (1 to 9)]
GARCH Models

- A first step towards a more parsimonious model is to impose some weighting scheme on the coefficients of an (potentially very high-order) ARCH-type process.

- The simplest example is an equally weighted rolling window estimator based on a window length $n$.

- Here we simply estimate the variance at time $t$ as the average of the previous $n$ shocks, i.e.,

$$
\sigma_t^2 = \frac{1}{n} \sum_{i=1}^{n} \epsilon_{t-i}^2 \quad (2)
$$

where $\epsilon_t = r_t - \mu_t$ (return minus its conditional mean, i.e., the unexpected “shock”).
GARCH Models

- It is plausible that more recent observations (shocks) contain more information about the next day’s (or week’s) variance than older observations.

- A natural way to incorporate this is to place more weight on the more recent observations.

- That is, with a weighting parameter

\[ 0 < \lambda < 1, \quad (3) \]

we replace (2) with

\[ \sigma^2_t = \frac{1 - \lambda}{1 - \lambda^n} \sum_{i=1}^{n} \lambda^{i-1} \epsilon_{t-i}^2. \quad (4) \]
GARCH Models

• More weight on recent observations:

\[ \sigma_t^2 = \frac{1 - \lambda}{1 - \lambda^n} \sum_{i=1}^{n} \lambda^{i-1} \epsilon_{t-i}^2. \] (5)

• In (5), the factor

\[ \frac{1 - \lambda}{1 - \lambda^n} \] (6)

is introduced to make the weights summing up to one, since

\[ \sum_{i=1}^{n} \lambda^{i-1} = \sum_{i=0}^{n-1} \lambda^i = \frac{1 - \lambda^n}{1 - \lambda}. \] (7)
GARCH Models

- With the weights (impact of past shocks) decaying to zero geometrically and sufficiently large $n$, the finite–order moving window estimator (4) can be viewed as an approximation to an infinite–order estimator, where $n \to \infty$,

$$\sigma_t^2 = (1 - \lambda) \sum_{i=1}^{\infty} \lambda^{i-1} \epsilon_{t-i}^2.$$  \hspace{1cm} (8)
GARCH Models

- With lag operator notation, $L^i x_t = x_{t-i}$, (8) can be written as

$$
\sigma_t^2 = (1 - \lambda) \sum_{i=1}^{\infty} \lambda^{i-1} \epsilon_{t-i}^2
$$

$$
= (1 - \lambda) \sum_{i=1}^{\infty} \lambda^{i-1} L^{i-1} \epsilon_{t-1}^2
\quad (\epsilon_{t-i}^2 = L^{i-1} \epsilon_{t-1}^2)
$$

$$
= (1 - \lambda) \sum_{i=0}^{\infty} (\lambda L)^i \epsilon_{t-1}^2
$$

$$
\sigma_t^2 = \frac{(1 - \lambda) \epsilon_{t-1}^2}{1 - \lambda L}.
$$

(9)
GARCH Models

- Equation (9) can be written

\[
(1 - \lambda L)\sigma_t^2 = (1 - \lambda)\epsilon_{t-1}^2 \\
\sigma_t^2 - \lambda\sigma_{t-1}^2 = (1 - \lambda)\epsilon_{t-1}^2 \\
\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1 - \lambda)\epsilon_{t-1}^2, \tag{10}
\]

which provides a simple updating scheme of the conditional variance.

- Specification (8) or, equivalently, (10) is known as the exponentially weighted moving average (EWMA), popularized by RiskMetrics of J.P. Morgan (occasionally also referred to as the RiskMetrics model.\(^1\))

- Often \(\lambda\) is fixed at 0.94 for daily data.\(^2\)


\(^2\) See the document cited in Footnote 1 for further details.
GARCH Models

- The EWMA approach has several drawbacks.

- First, when viewed as a statistical model for the variance (an ARCH(∞) model) rather than just a heuristic updating rule, its multi-period forecasts have some undesirable properties (as will be seen below).³

---

GARCH Models

- Second (and related to the first issue), since the “model” uses only one parameter, the dynamics of the conditional volatility are somewhat restricted:

\[
\sigma_t^2 = (1 - \lambda)\epsilon_{t-1}^2 + \lambda(1 - \lambda)\epsilon_{t-2}^2 + \lambda^2(1 - \lambda)\epsilon_{t-3}^2 + \lambda^3(1 - \lambda)\epsilon_{t-4}^2 + \cdots + (1 - \lambda)\lambda^\tau \epsilon_{t-\tau-1}^2 + \cdots
\]

- From this equation, we observe that
  - Parameter \((1 - \lambda)\) is a \textit{reaction parameter} which measures the immediate impact of a unit shock on the next period's variance.
  - The impact of past shocks on future variances decays with rate \(\lambda\), hence \(\lambda\) is a \textit{memory parameter}.
We may want to disentangle reaction and memory parameters.

A more reasonable specification appears to be a generalization of (10) as follows:

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

\[ \omega > 0, \quad \alpha \geq 0, \quad 1 > \beta \geq 0. \]

Enriched with a (conditional) distributional assumption for \( \epsilon_t \), this is a GARCH(1,1) model, referring to generalized ARCH.

The restrictions (12) guarantee a positive conditional variance.
• Using lag operator notation, we can write

\[(1 - \beta L) \sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2.\]  

(13)

• Assuming stationarity, we can invert the lag polynomial \(1 - \beta L\) and write this as ARCH(\(\infty\)),

\[
\begin{align*}
\sigma_t^2 &= \frac{\omega}{1 - \beta} + \frac{\alpha \epsilon_{t-1}^2}{1 - \beta L} \\
&= \frac{\omega}{1 - \beta} + \alpha \sum_{i=1}^{\infty} \beta^{i-1} \epsilon_{t-i}^2 \\
&= \frac{\omega}{1 - \beta} + \alpha \epsilon_{t-1}^2 + \alpha \beta \epsilon_{t-2}^2 + \cdots + \alpha \beta^\tau \epsilon_{t-\tau-1}^2 + \cdots
\end{align*}
\]

(14)
(15)
(16)

\(^4\text{Cf. Francq and Zakoïan (2010). _GARCH Models._ Wiley, p. 42. Note that stationarity requires inter alia that } \beta_1 < 1.\)
This shows that $\alpha$ and $\beta$ are the reaction and memory parameters, respectively (disentangled).
GARCH Models

- The results for the ARCH\(q\) suggest that the model is covariance stationary if the sum of ARCH coefficients

\[
\sum_{i=0}^{\infty} \alpha \beta^{i-1} = \frac{\alpha}{1 - \beta} < 1 \quad (\beta < 1) \quad \alpha + \beta < 1. \quad (17)
\]

- The GARCH(1,1), with only three parameters, can thus be viewed as an ARCH(\(\infty\)) with geometrically declining lag structure.

- A declining lag structure seems reasonable, since then more recent observations contain more information about the next period’s conditional variance.

- This explains why in practice a simple GARCH(1,1) tends to exhibit a better fit than even high–order (pure) ARCH processes.
GARCH($p, q$) Models

- The GARCH($p, q$) model\textsuperscript{5} generalizes the ARCH($q$) to

\begin{align*}
\epsilon_t &= \sigma_t \eta_t, \quad \eta_t \sim \text{iid } N(0, 1) \\
\sigma_t^2 &= \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2.
\end{align*}

(18) \hspace{1cm} (19)

- A GARCH(0, $q$) process is ARCH($q$).

---

GARCH\((p, q)\) Models

- GARCH\((p, q)\) model:

\[
\begin{align*}
\epsilon_t &= \sigma_t \eta_t, \quad \eta_t \overset{iid}{\sim} N(0, 1) \\
\sigma_t^2 &= \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2.
\end{align*}
\] (20)

- To make sure that the conditional variance is positive for all \(t\), we may impose

\[
\omega > 0; \quad \alpha_i \geq 0, \quad i = 1, \ldots, q; \quad \beta_i \geq 0, \quad i = 1, \ldots, p.
\] (22)

- In the general case, these conditions are sufficient but can be weakened for models where (at least) one of the orders (i.e., \(p\) and/or \(q\)) is larger than unity (see Appendix A).\(^6\)

GARCH\(_{(p, q)}\) Models

- For the most frequently applied GARCH(1,1) process, however, the nonnegativity constraints \(\omega > 0, \alpha_1, \beta_1 \geq 0\) are necessary.

- Conditions (22) are also necessary for guaranteeing a positive conditional variance in (practically less relevant) pure ARCH processes \((p = 0)\).
Covariance stationarity and unconditional variance

• As for the pure ARCH specification, a GARCH process is uncorrelated, i.e., $E(\varepsilon_t \varepsilon_{t-\tau}) = 0$ for $\tau \neq 0$.

• Assuming stationarity, we can calculate the unconditional variance $E(\sigma_t^2) = E(\varepsilon_t^2)$ of the process as

$$E(\sigma_t^2) = \omega + \sum_{i=1}^{q} \alpha_i E(\varepsilon_{t-i}^2) + \sum_{i=1}^{p} \beta_i E(\sigma_{t-i}^2) \quad (23)$$

$$= \omega + \sum_{i=1}^{q} \alpha_i E(\eta_{t-i}^2 \sigma_{t-i}^2) + \sum_{i=1}^{p} \beta_i E(\sigma_{t-i}^2) \quad (24)$$

$$= \omega + \sum_{i=1}^{q} \alpha_i E(\eta_{t-i}^2) \mathbb{E}(\sigma_{t-i}^2) + \sum_{i=1}^{p} \beta_i E(\sigma_{t-i}^2) \quad (25)$$

$$= \omega + \left( \sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i \right) E(\sigma_t^2). \quad (26)$$
Covariance stationarity and unconditional variance

- That is,

\[
E(\epsilon_t^2) = E(\sigma_t^2)
= \frac{\omega}{1 - \sum_{i=1}^{q} \alpha_i - \sum_{i=1}^{p} \beta_i}
\]  

(27)  

(28)

provided the (covariance) stationarity condition

\[
\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1
\]

(29)

is satisfied.\(^7\)

\(^7\)More precisely, condition (29) is the covariance stationarity condition when the nonnegativity restrictions (22) are imposed; otherwise we need to check the roots of the associated characteristic equation

\[
z^m - \sum_{i=1}^{m} (\alpha_i + \beta_i)z^{m-i} = 0,
\]

(30)

where \(m = \max\{p, q\}\), \(\alpha_i = 0\) for \(i > q\), and \(\beta_i = 0\) for \(i > p\).
GARCH(1,1): Some Examples

• Consider various simulated GARCH(1,1) processes

\[
\begin{align*}
\epsilon_t &= \eta_t \sigma_t, \quad \eta_t \overset{iid}{\sim} N(0, 1) \\
\sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2.
\end{align*}
\]

(31) (32)

• Processes with the same \( \omega \) and \( \alpha_1 + \beta_1 \) have the same unconditional variance, but the behavior is not “symmetric” in \( \alpha_1 \) and \( \beta_1 \): A large \( \beta_1 \) makes volatility more persistent, whereas a large \( \alpha_1 \) induces sudden changes in volatility in response to shocks.
\omega = 0.01, \alpha_1 = 0.99, \beta_1 = 0
$\omega = 0.01, \alpha_1 = 0.3, \beta_1 = 0.69$
\( \omega = 0.01, \alpha_1 = 0.1, \beta_1 = 0.89 \)
ARMA representation and correlation structure

• To characterize the correlation structure of the squared process (provided \(E(\epsilon_t^4)\) exists), define the prediction error

\[ u_t = \epsilon_t^2 - E(\epsilon_t^2 | I_{t-1}) = \epsilon_t^2 - \sigma_t^2 = \sigma_t^2 (\eta_t^2 - 1), \quad (33) \]

where the information set at time \(t\), \(I_t = \{\epsilon_s : s \leq t\}\).

• \(u_t = \epsilon_t^2 - \sigma_t^2 = (\eta_t^2 - 1)\sigma_t^2\) is white noise but not strict white noise, since it is uncorrelated but not independent.

• Substituting (33) for \(\sigma_t^2\) into (21) results in

\[ \epsilon_t^2 = \omega + \max\{p,q\} \sum_{i=1}^{\max\{p,q\}} (\alpha_i + \beta_i)\epsilon_{t-i}^2 - \sum_{i=1}^{p} \beta_i u_{t-i} + u_t, \quad (34) \]

where \(\alpha_i = 0\) for \(i > q\) and \(\beta_i = 0\) for \(i > p\).
**ARMA representation and correlation structure**

- Equation (34) is an ARMA\( (\max\{p, q\}, p) \) representation for the squared process \( \{\epsilon_t^2\} \), which characterizes its autocorrelations.

- Note that an ARCH\( (q) \) (where \( p = 0 \)) thus implies an AR\( (q) \) for \( \epsilon_t^2 \), hence GARCH generalizes ARCH just as ARMA generalizes AR.

- (Also note that the ACF of the squared process is only well-defined if the fourth moment is finite, i.e., \( \mathbb{E}(\epsilon_t^4) < \infty \); see below for the GARCH(1,1) model.)
ARMA representation and correlation structure

- For example, the ARMA(1,1) representation of the GARCH(1,1) process is
  \[ \epsilon_t^2 = \omega + (\alpha_1 + \beta_1)\epsilon_{t-1}^2 + u_t - \beta_1 u_t. \]  
  (35)

- Recall that the ACF of the ARMA(1,1) process
  \[ Y_t = \phi Y_{t-1} + \theta \epsilon_{t-1} + \epsilon_t \]
  is
  \[ \text{Corr}(Y_t, Y_{t-\tau}) = \phi^{\tau-1}(\phi + \theta)(1 + \phi \theta) \frac{1}{1 + 2 \theta \phi + \theta^2}. \]
ARMA representation and correlation structure

• Plugging in $\alpha_1 + \beta_1$ for $\phi$ and $-\beta_1$ for $\theta$ gives the ACF of the squares of a GARCH(1,1) process as

$$\phi(\tau) = \text{Corr}(e^2_t, e^2_{t-\tau})$$

$$= \frac{E(e^2_t e^2_{t-\tau}) - E(e^2_t)}{E(e^4_t) - E(e^2_t)}$$

$$= (\alpha_1 + \beta_1)^{\tau-1} \frac{\alpha_1 (1 - \alpha_1 \beta_1 - \beta_1^2)}{1 - 2\alpha_1 \beta_1 - \beta_1^2},$$

provided the fourth moment is finite (see below for the condition for this to be the case).

• The ACF decays at rate $\alpha_1 + \beta_1$, which is often taken as a measure of persistence in volatility.

• If $\alpha_1$ is small, $\beta_1$ is large, and $\alpha_1 + \beta_1$ is close to unity (as is often the case for estimated models), then the ACF starts at low values but decreases slowly, cf. the next slide.
ACF of $\epsilon_t^2$ with $\alpha = 0.085$ and $\beta = 0.9$
The GARCH(1, 1) process

- The GARCH(1,1) process is often found to provide an adequate description of the conditional volatility process.

- To find the moments of the marginal distribution of this process, it is convenient to write

\[
\sigma^2_t = \omega + \alpha_1 \epsilon^2_{t-1} + \beta_1 \sigma^2_{t-1}
\]

\[
= \omega + (\alpha_1 \eta^2_{t-1} + \beta_1) \sigma^2_{t-1}.
\]
The GARCH(1, 1) process

- Recall that $\sigma_t^2$ is independent of $\{\eta_{t+\tau} : \tau \geq 0\}$.

- Thus
  \[
  E[(\alpha_1 \eta_t^2 + \beta_1)^m \sigma_t^{2m}] = E[(\alpha_1 \eta_t^2 + \beta_1)^m] E(\sigma_t^{2m}).
  \] (36)

- We have, under normality of $\eta_t$,
  \[
  E(\alpha_1 \eta_t^2 + \beta_1) = \alpha_1 + \beta_1
  \]
  \[
  E[(\alpha_1 \eta_t^2 + \beta_1)^2] = E[(\alpha_1 \eta_t^2 + \beta_1)^2]
  \]
  \[
  = E(\alpha_1^2 \eta_t^4 + 2\alpha_1 \beta_1 \eta_t^2 + \beta_1^2)
  \]
  \[
  = 3\alpha_1^2 + 2\alpha_1 \beta_1 + \beta_1^2.
  \]
The GARCH(1, 1) process

- Since we have seen that

\[ E(\sigma_t^2) = \frac{\omega}{1 - \alpha_1 - \beta_1}, \]

we get, assuming finiteness of the fourth moment,

\[
E(\sigma_t^4) = E[(\omega + (\alpha_1 \eta_{t-1}^2 + \beta_1)\sigma_{t-1}^2)^2] \\
= \omega^2 + 2\omega E(\alpha_1 \eta_{t-1}^2 + \beta_1) E(\sigma_t^2) + E[(\alpha_1 \eta_{t-1}^2 + \beta_1)^2] E(\sigma_t^4) \\
= \omega^2 + \frac{2\omega^2(\alpha_1 + \beta_1)}{1 - \alpha_1 - \beta_1} + (3\alpha_1^2 + 2\alpha_1 \beta_1 + \beta_1^2) E(\sigma_t^4) \\
= \frac{\omega^2(1 + \alpha_1 + \beta_1)}{1 - \alpha_1 - \beta_1} + (3\alpha_1^2 + 2\alpha_1 \beta_1 + \beta_1^2) E(\sigma_t^4) \\
E(\sigma_t^4) = \frac{\omega^2(1 + \alpha_1 + \beta_1)}{(1 - \alpha_1 - \beta_1)(1 - 3\alpha_1^2 - 2\alpha_1 \beta_1 - \beta_1^2)},
\]

where \( 3\alpha_1^2 + 2\alpha_1 \beta_1 + \beta_1^2 < 1 \) is the condition for the finiteness of the fourth moment.
The GARCH(1, 1) process

• With standard normally distributed $\eta_t$, i.e, $E(\eta_t^4) = 3$,

$$E(\epsilon_t^4) = E(\eta_t^4 \sigma_t^4) = E(\eta_t^4) E(\sigma_t^4) = 3 E(\sigma_t^4),$$

and the kurtosis is then

$$\frac{E(\epsilon_t^4)}{E^2(\epsilon_t^2)} = \frac{3(1 - \alpha_1 - \beta_1)(1 + \alpha_1 + \beta_1)}{1 - 3\alpha_1^2 - 2\alpha_1\beta_1 - \beta_1^2}$$

$$= \frac{3}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2}$$

$$= 3 + \frac{6\alpha_1^2}{1 - 3\alpha_1^2 - 2\alpha_1\beta_1 - \beta_1^2}.$$

• Note that the denominator is positive, since $1 - 3\alpha_1^2 - 2\alpha_1\beta_1 - \beta_1^2 > 0$ is required for the fourth moment (and hence the kurtosis) to be well-defined.
The GARCH(1, 1) process

- As in the ARCH model, we have excess kurtosis of the unconditional distribution, even with conditional normality.
**ACF of the squares of ARCH(1) and GARCH(1,1)**

- In the ARCH(1), $\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2$, the ACF of the squares has been seen to be
  \[
  \text{Corr}(\epsilon_t^2, \epsilon_{t-\tau}^2) = \rho(\tau) = \alpha^\tau, \quad \tau \geq 1, \quad \alpha < 1/\sqrt{3},
  \]
  i.e., the ACF at lag 1 and the decay pattern (the persistence) are both given by $\alpha$.

- The GARCH(1,1) displays considerably greater flexibility in modeling the correlation structure.

- To illustrate, consider a Gaussian GARCH (i.e., *conditional normality*) with
  \[
  3\alpha^2 + 2\alpha\beta + \beta^2 < 1.
  \]
  \[\text{(38)}\]
ACF of the squares of ARCH(1) and GARCH(1,1)

- Recall the ARMA(1,1) representation of the squared GARCH(1,1),

\[ \varepsilon_t^2 = \omega + \phi \varepsilon_{t-1}^2 + u_t - \theta u_{t-1}, \]  

(39)

where the persistence parameter

\[ \phi = \alpha + \beta, \]  

(40)

and

\[ \theta = -\beta \leq 0, \]  

(41)

and thus the ACF at lag 1,

\[ \rho(1) = \rho_1 = \frac{(\phi + \theta)(1 + \phi \theta)}{1 + 2\phi \theta + \theta^2}. \]  

(42)
ACF of the squares of ARCH(1) and GARCH(1,1)

• To figure out possible combinations of $\rho_1$ and $\phi$, fix both $\rho_1$ and $\phi$ and find a value of $\theta$ such that (42) holds.

• If $\phi = \rho_1$, then $\theta = 0$, and $\rho_1 = \phi = \alpha$ (ARCH(1)). Thus consider the case $\phi \neq \rho_1$, and check that always $\phi > \rho_1$ in this case.

• This gives rise to a quadratic in $\theta$,

\[ \theta^2 + b\theta + 1 = 0, \quad b = \frac{\phi^2 + 1 - 2\rho_1\phi}{\phi - \rho_1}. \quad (43) \]

• Since $b > 2$ in (43), we have solutions of (43) as

\[ \tilde{\theta}_{1/2} = \frac{-b \pm \sqrt{b^2 - 4}}{2}. \]
ACF of the squares of ARCH(1) and GARCH(1,1)

\[ \frac{1}{2} \left( -b - \sqrt{b^2 - 4} \right) \] exceeds one in magnitude, and therefore the only possible candidate is \(^8\)

\[ \theta = \frac{-b + \sqrt{b^2 - 4}}{2}. \] (44)

• The particular combination of \( \varrho_1 \) and \( \phi = \alpha + \beta \) is then feasible if \( \theta \) in (44) is such that

\[ \phi^2 + 2(\phi + \theta)^2 = (\alpha + \beta)^2 + 2\alpha^2 = 3\alpha^2 + 2\alpha\beta + \beta^2 < 1. \]

---

\(^8\)These calculations can also be used to compute a simple moment estimator for the GARCH(1,1), which may serve, e.g., as a starting value for maximum likelihood estimation, see Kristensen and Linton (2006): A closed–form estimator for the GARCH(1,1) model. *Econometric Theory* 22, 323-337.
ACF of the squares of ARCH(1) and GARCH(1,1)

• We can thus identify all feasible combinations of $\varrho_1$ and $\phi = \alpha + \beta$.

• E.g., if we fix $\varrho_1 = 0.3$, $\phi = \alpha + \beta = 0.99$, we get $b = 2.0088$, and $\beta = -\theta = 0.9103$, and thus $\alpha = \phi - \beta = 0.0797$.

• This has

$$3\alpha^2 + 2\alpha\beta + \beta^2 = 0.9928 < 1.$$ 

• See next slide, where the possible combinations for the ARCH(1) are just given by the lower boundary of the set of possible combinations of $\varrho_1$ and persistence $(\alpha + \beta)$ of the GARCH(1,1).
feasible combinations of lag-one ACF and persistence in the GARCH(1,1), $\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$

possible combinations of $\rho_1$ and $\alpha_1 + \beta_1$
Strict stationarity and the marginal distribution of the GARCH(1,1) process

• This is analogous to the discussion of the ARCH(1) process.

• From the same reasoning as for the ARCH(1), unconditional moment $E(\epsilon_t^{2m})$, $m \in \mathbb{N}$, of the GARCH(1,1) is finite if and only if

$$E[(\alpha_1 \eta_t^2 + \beta_1)^m] = \sum_{i=1}^{m} \binom{m}{i} \alpha_1^i \kappa_i \beta_1^{m-i} < 1,$$

where $\kappa_i = E(\eta_t^{2i})$ depends on the distribution of $\eta_t$, e.g., $\kappa_i = \prod_{j=1}^{i} (2j - 1)$ for $\eta_t \sim \mathcal{N}(0,1)$ (cf. the discussion of the ARCH(1) process).

• For $m = 1$ and $m = 2$, with $\eta_t \overset{iid}{\sim} \mathcal{N}(0,1)$, condition (45) becomes $\alpha_1 + \beta_1 < 1$ and $3\alpha_1^2 + 2\alpha_1 \beta_1 + \beta_1^2 < 1$, respectively, as above.
Strict stationarity and the marginal distribution of the GARCH(1,1) process

- As in the ARCH case, condition (45) will eventually be violated for some $m$, so the GARCH process cannot have finite moments of all orders.

- Even with normal (i.e., very light-tailed) innovations, the marginal distribution of the GARCH process has power tails in the sense that\footnote{This is in sharp contrast to linear time series models, where normal (Gaussian) input leads to normal output.}

\[
\Pr(|\epsilon_t| > x) \approx cx^{-\gamma} \quad \text{as} \quad x \to \infty, \tag{46}
\]

where $\gamma$ is the \textit{tail index}, which can be shown to be given by the unique positive solution of

\[
E[(\alpha_1 \eta_t^2 + \beta_1)^{\gamma/2}] = 1. \tag{47}
\]

- Moments of order $m$ exist only for $m < \gamma$. 
Strict stationarity and the marginal distribution of the GARCH(1,1) process

- For some series it has been found that \( \alpha + \beta \geq 1 \), so that the second moment would not exist, and the estimated model not be covariance stationary.

- As for the ARCH(1), such a model may still be (strictly) stationary.

- The case \( \alpha_1 + \beta_1 = 1 \) (\( \gamma = 2 \)) is often referred to as IGARCH (integrated GARCH).
Strict stationarity and the marginal distribution of the GARCH(1,1) process

The condition for strict stationarity of the GARCH(1,1) model has been shown to be\(^{10}\)

\[
E\{\log (\alpha_1 \eta_t^2 + \beta_1)\} < 0.
\]  
(48)

This is weaker than \(\alpha_1 + \beta_1 < 1\), since (Jensen’s inequality)

\[
E\{\log (\alpha_1 \eta_t^2 + \beta_1)\} < \log (E(\alpha_1 \eta_t^2 + \beta_1)) = \log (\alpha_1 + \beta_1).
\]  
(49)

E.g., the IGARCH process with \(\alpha_1 + \beta_1 = 1\) is strictly stationary with infinite variance.

Note that, in contrast to the covariance stationarity condition \(\alpha_1 + \beta_1 < 1\), (48) is not symmetric in \(\alpha_1\) and \(\beta_1\) (see also the figure below). It also depends on the distribution of \(\eta_t\).

---

Strict stationarity and the marginal distribution of the GARCH(1,1) process

• Among other things, that a GARCH(1,1) with $\alpha_1 + \beta_1 > 1$ does not exhibit “explosive volatility behavior” as long as (48) is satisfied.

• This is in contrast to the AR(1) process, where $|\phi| > 1$ in

$$y_t = c + \phi y_{t-1} + \epsilon_t$$

(50)

implies explosive behavior.

• On the next slide, the parameter restrictions for strict stationarity are shown along with those for weak stationarity and finiteness of the fourth moment for Gaussian GARCH(1,1), i.e., with $\eta_t \overset{iid}{\sim} N(0, 1)$.

• E.g., in the ARCH(1), where $\beta = 0$, $\alpha_1$ can be as large as 3.56, and still the process is stationary (although extremely (and unrealistically) fat tailed.)
The stationarity/moment-finiteness regions are between the axes and the indicated lines.
Forecast Variance

• Forecasting the (conditional) distribution of future returns is one of the main goals of specifying GARCH models.

• Under the assumption that $\eta_t \sim N(0,1)$, the conditional distribution of $\epsilon_{t+1}$, given $I_t = \{\epsilon_t, \epsilon_{t-1}, \ldots\}$, is normal with mean zero and variance $\sigma_{t+1}^2$, i.e.,

$$\epsilon_{t+1}|I_t \sim N(0, \sigma_{t+1}^2). \quad (51)$$

• Due to the nonlinearity of the GARCH process, the $d$–step distributions $(d \geq 2)$ are not normal.\(^{11}\)

• Thus, multi–step distributions are generally not available in closed–form, but can be evaluated by simulation.

• However, interest often centers on specific moments (such as the variance) of the conditional distribution.

\(^{11}\)i.e., the distribution of $\epsilon_{t+d}$, $d \geq 2$, given $I_t$. 
Forecast Variance

• That is, we want to compute

\[ \text{Var}(\epsilon_{t+d}|I_t) = \text{Var}_t(\epsilon_{t+d}) = E_t(\epsilon_{t+d}^2) \]  \hspace{1cm} (52)

\[ = E_t(\eta_{t+d}^2 \sigma_{t+d}^2) \]  \hspace{1cm} (53)

\[ = E_t(\eta_{t+d}) E_t(\sigma_{t+d}^2) \]  \hspace{1cm} (54)

\[ = E_t(\sigma_{t+d}^2), \]  \hspace{1cm} (55)

where \( E_t \) denotes an expectation on the basis of the information up to time \( t \).

• For convenience, write

\[ \sigma_t^2(d) := E_t(\sigma_{t+d}^2) = E_t(\epsilon_{t+d}^2). \]  \hspace{1cm} (56)

• \( d \) is the forecast horizon, and \( t \) is the forecast origin.
Forecast Variance

• Then, for \( d = 1 \),

\[
\sigma^2_t(1) = \sigma^2_{t+1},
\]

(57)

since \( \sigma^2_{t+1} \) is determined by information up to time \( t \).

• For \( d = 2 \),

\[
\sigma^2_t(2) = E_t(\sigma^2_{t+2}) = E_t(\omega + \alpha_1 \epsilon^2_{t+1} + \beta_1 \sigma^2_{t+1}) = \omega + \alpha_1 E_t(\epsilon^2_{t+1}) + \beta_1 E_t(\sigma^2_{t+1}) = \omega + (\alpha_1 + \beta_1) E_t(\sigma^2_{t+1}) = \omega + (\alpha_1 + \beta_1) \sigma^2_{t+1}.
\]

(58)

(59)

(60)

(61)

(62)
Forecast Variance

- More generally, for \( d \geq 2 \),

\[
\sigma_t^2(d) = \mathbb{E}_t(\sigma_{t+d}^2) = \mathbb{E}_t(\omega + \alpha_1 \epsilon_{t+d-1}^2 + \beta_1 \sigma_{t+d-1}^2) \\
= \omega + \alpha_1 \mathbb{E}_t(\epsilon_{t+d-1}^2) + \beta_1 \mathbb{E}_t(\sigma_{t+d-1}^2) \\
= \omega + \alpha_1 \mathbb{E}_t(\epsilon_{t+d-1}^2) + \beta_1 \mathbb{E}_t(\sigma_{t+d-1}^2) \\
= \omega + (\alpha_1 + \beta_1) \mathbb{E}_t(\sigma_{t+d-1}^2) \\
= \omega + (\alpha_1 + \beta_1) \sigma_t^2(d-1).
\] (63)

- Equation (63), i.e.,

\[
\sigma_t^2(d) = \omega + (\alpha_1 + \beta_1) \sigma_t^2(d-1), \quad d \geq 2,
\] (64)

allows to compute multi-step (conditional) forecast error variances via a simple first-order recursive scheme, with starting value \( \sigma_t^2(1) = \sigma_{t+1}^2 \).
Forecast Variance

- Solving recursion (64) (with starting value $\sigma^2_t(1) = \sigma^2_{t+1}$) shows that

$$\sigma^2_t(d) = \omega \sum_{i=0}^{d-2} (\alpha_1 + \beta_1)^i + (\alpha_1 + \beta_1)^{d-1}\sigma^2_{t+1}$$

$$= \frac{\omega [1 - (\alpha_1 + \beta_1)^{d-1}]}{1 - \alpha_1 - \beta_1} + (\alpha_1 + \beta_1)^{d-1}\sigma^2_{t+1}$$

$$= \frac{\omega}{1 - \alpha_1 - \beta_1} + (\alpha_1 + \beta_1)^{d-1} \left( \sigma^2_{t+1} - \frac{\omega}{1 - \alpha_1 - \beta_1} \right)$$

$$\sigma^2_t(d) = \frac{\omega}{1 - \alpha_1 - \beta_1} + (\alpha_1 + \beta_1)^{d-1} \left( \sigma^2_{t+1} - \frac{\omega}{1 - \alpha_1 - \beta_1} \right), \quad (65)$$

where

$$\overline{\sigma^2} = \mathbb{E}(\sigma^2_t) = \frac{\omega}{1 - \alpha_1 - \beta_1}.$$

- Equation (65) shows that, if $\alpha_1 + \beta_1 < 1$, the (conditional) forecast error variances converge to the long–run (unconditional) variance at rate $\alpha_1 + \beta_1$. 
Forecast Variance

• Compare this with the EWMA approach, where $\omega = 0$, $\alpha_1 = (1 - \lambda)$, and $\beta_1 = \lambda$, so (64) becomes

$$\sigma^2_{t+1} = \sigma^2_t(1) = \sigma^2_t(2) = \sigma^2_t(3) = \cdots = \sigma^2_t(d) = \cdots,$$

(66)

i.e., (conditional) forecast variances are fixed at the current volatility estimate.\(^{12}\)

---

\(^{12}\)The EWMA, when viewed as a statistical model, is an IGARCH process with zero intercept ($\omega = 0$).
The forecast variance calculated above refers to the error term $\epsilon_{t+d}$ and, for $d > 1$, is equal to the conditional variance of the return $r_{t+d}$ only in the absence of conditional mean dynamics.

To illustrate, consider the AR(1)–GARCH(1,1) process for return $r_t$,

\begin{align*}
  r_t &= \mu + \phi(r_{t-1} - \mu) + \epsilon_t \quad (67) \\
  \epsilon_t &= \eta_t \sigma_t, \quad \eta_t \overset{iid}{\sim} (0, 1) \quad (68) \\
  \sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (69)
\end{align*}
**Forecast Variance**

- From (67) and recursive substitution, $r_{t+d}$ can be written

\[
\begin{align*}
    r_{t+d} - \mu &= \phi(r_{t+d-1} - \mu) + \epsilon_{t+d} \\
    &= \phi^2(r_{t+d-2} - \mu) + \phi \epsilon_{t+d-1} + \epsilon_{t+d} \\
    &\vdots \\
    &= \phi^d(r_t - \mu) + \sum_{i=0}^{d-1} \phi^i \epsilon_{t+d-i}.
\end{align*}
\]

(70)

- From the previous development,

\[
\text{Var}_t(\epsilon_{t+d-i}) = \mathbb{E}_t(\epsilon_{t+d-i}^2) = \bar{\sigma}^2 + (\alpha_1 + \beta_1)^{d-i-1}(\sigma_{t+1}^2 - \bar{\sigma}^2),
\]

\[
i = 0, \ldots, d-1,
\]

(71)

where $\bar{\sigma}^2 = \mathbb{E}(\sigma_t^2) = \omega(1 - \alpha_1 - \beta_1)^{-1}$. 


Forecast Variance

• Combining (71) and (70) leads to the forecast variance

\[
\text{Var}_t(r_{t+d}) = \sum_{i=0}^{d-1} \phi^{2i} \text{Var}_t(\epsilon_{t+d-i}) \\
= \frac{\sigma^2}{1 - \phi^2} \left( 1 - \phi^{2d} \right) + \left( \sigma^2 - \overline{\sigma}^2 \right) (\alpha_1 + \beta_1)^{d-1} \sum_{i=0}^{d-1} \left( \frac{\phi^{2i}}{(\alpha_1 + \beta_1)^i} \right) \\
= \frac{\sigma^2}{1 - \phi^2} \left( 1 - \phi^{2d} \right) + \left( \sigma^2 - \overline{\sigma}^2 \right) (\alpha_1 + \beta_1)^d - \phi^{2d} (\phi^2 \neq \alpha_1 + \beta_1).
\]

• This converges to the unconditional return variance as the forecast horizon \(d\) increases,

\[
\lim_{d \to \infty} \text{Var}_t(r_{t+d}) = \text{Var}(r_t) = \frac{\overline{\sigma}^2}{1 - \phi^2} = \frac{\omega}{(1 - \alpha_1 - \beta_1)(1 - \phi^2)}. \quad (72)
\]
Forecast Variance

- The expressions on the previous slide can be compared with those implied by a linear AR(1) model, where $\epsilon_t \sim iid (0, \sigma^2)$.

- For these models, we have seen that

$$\text{Var}_t(r_{t+d}) = \text{Var}_t \left( \sum_{i=0}^{d-1} \phi^i \epsilon_{t+d-i} \right) = \frac{\sigma^2(1 - \phi^{2d})}{1 - \phi^2}, \quad (73)$$

i.e., conditional forecast error variances do only depend on the forecast horizon $d$ but not on the current process values.

- That is, the width of (conditional) prediction intervals is constant, for a given horizon.
Forecast Variance

- Consider various simulated GARCH(1,1) processes

\[ 
\begin{align*}
\epsilon_t &= \eta_t \sigma_t \\
\sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2,
\end{align*}
\]

(74)

(75)

with \( \eta_t \overset{iid}{\sim} \mathcal{N}(0, 1) \) along with one-step 95% prediction intervals (red lines).
$\omega = 1$, $\alpha_1 = \beta_1 = 0$, red line is one–step 95% prediction interval
ω = .05, α_1 = 0.9, β_1 = 0, red line is one-step 95% prediction interval
$\omega = .05$, $\alpha_1 = 0.2$, $\beta_1 = 0.7$, red line is one-step 95% prediction interval
\(\omega = .05, \alpha_1 = 0.1, \beta_1 = 0.89,\) red line is one-step 95% prediction interval
Forecast Variance

- Finally, suppose we use a GARCH model for daily data to calculate forecasts for *cumulative* (weekly/monthly/...) returns.

- Multi-period (cumulative) log-returns are sums of single-period returns, and so we are interested in expressions of the form

\[ \text{Var}_t(r_{t+1} + r_{t+2} + \cdots + r_{t+D}), \]  

(76)

which can be calculated using the reasoning above.

- If the returns are correlated as in (67), then clearly calculation of (76) involves both variance as well as covariance terms.

Estimation

• GARCH models are most frequently estimated by conditional maximum likelihood.

• To illustrate, suppose we want to estimate an AR(1)–GARCH(1,1) model for returns \( r_t \).

• That is, the conditional mean of the time series is described by an AR(1), and the conditional variance is driven by GARCH(1,1).

• If we assume conditional normality, the model is

\[
\begin{align*}
  r_t &= c + \phi r_{t-1} + \epsilon_t, \quad |\phi| < 1 \\
  \epsilon_t &= \eta_t \sigma_t, \quad \eta_t \overset{iid}{\sim} N(0,1) \\
  \sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\
  \omega &> 0, \quad \alpha, \beta \geq 0.
\end{align*}
\]

(77) \hspace{1cm} (78) \hspace{1cm} (79) \hspace{1cm} (80)
Estimation

- The parameter vector is $\theta = (c, \phi, \omega, \alpha, \beta)$.

- We observe a stretch of length $T$, $\{r_t\}_{t=1}^T$, and a pre-sample value $r_0$.

- Want to calculate the log-likelihood function for a given value $\hat{\theta}$.

- From the ARMA part, we calculate the residuals

$$\hat{\epsilon}_t = \hat{\epsilon}_t(\hat{\theta}) = r_t - \hat{c} - \hat{\phi}r_{t-1}, \quad t = 1, \ldots, T. \quad (81)$$
Estimation

- The conditional log-likelihood function, \( \log L(\theta) \), is then given by

\[
\log L(\hat{\theta}) = \sum_{t=1}^{T} \ell_t(\hat{\theta}).
\]  

(82)

where, under conditional normality,

\[
\ell_t(\hat{\theta}) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log \hat{\sigma}_t^2 - \frac{1}{2} \frac{\hat{\epsilon}_t^2}{\hat{\sigma}_t^2}, \quad t = 1, \ldots, T.
\]  

(83)

- In (83), from the GARCH part,

\[
\hat{\sigma}_t^2 = \hat{\omega} + \hat{\alpha}_1 \hat{\epsilon}_{t-1}^2 + \hat{\beta}_1 \hat{\sigma}_{t-1}^2, \quad t = 1, \ldots, T.
\]  

(84)
Estimation

- To start the GARCH recursion (84), we need initial values $\hat{\sigma}_0^2$ and $\hat{\epsilon}_0^2$.

- One possibility is to set these equal to their unconditional values estimated from the sample at hand, i.e.,

$$\hat{\sigma}_0^2 = \hat{\epsilon}_0^2 = \frac{1}{T} \sum_{t=1}^{T} \hat{\epsilon}_t^2,$$

(85)

with $\hat{\epsilon}_t$, $t = 1, \ldots, T$, given by (81).

- In practice, GARCH models are typically applied to sufficiently long time series, so that the choice of the initialization has negligible impact on the results.

- We then maximize (82) with respect to $\theta$ to obtain the maximum likelihood estimate (MLE) $\hat{\theta}_{ML}$.
Estimation

- Analogous to standard large sample theory for the MLE, the MLE is consistent and asymptotically normal, i.e.,

\[ \sqrt{T}(\hat{\theta}_{ML} - \theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1}), \tag{86} \]

where \( \theta_0 \) is the true parameter value, and the Fisher information matrix (stationarity assumed)

\[ I(\theta_0) = -E \left[ \frac{\partial^2 l_t(\theta_0)}{\partial \theta \partial \theta'} \right], \tag{87} \]

which can be consistently estimated via

\[ I(\hat{\theta}_{ML}) = -\frac{1}{T} \frac{\partial^2 \log L(\hat{\theta}_{ML})}{\partial \theta \partial \theta'} = -\frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 l_t(\hat{\theta}_{ML})}{\partial \theta \partial \theta'}, \tag{88} \]

i.e., the negative of the Hessian matrix of the log–likelihood function, evaluated at the MLE (divided by \( T \)).
Estimation

• Inference (e.g., calculation of standard errors) is then based on

\[ \hat{\theta}_{ML} \approx \text{Normal}(\theta_0, T^{-1} I(\hat{\theta}_{ML})^{-1}). \] (89)

• The derivatives in (88) can be calculated analytically or numerically.

• Under relatively weak assumptions, using the Gaussian likelihood, the parameters of the volatility process can be consistently estimated even if the innovations are not normally distributed. This is known as quasi–ML (QML) estimation and requires adjustment of the standard errors.\(^\text{13}\)

Illustration Fitting GARCH Models

• To illustrate typical results, we fit model\textsuperscript{14}

\[
\begin{align*}
\epsilon_t & = \eta_t \sigma_t, \quad \eta_t \overset{iid}{\sim} N(0, 1) \\
\sigma_t^2 & = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2
\end{align*}
\]

(91) (92)

to various stock index series, as shown in Table 2.

\textsuperscript{14}Here a constant conditional mean is assumed for simplicity; in practice, we would check for significant dynamics in the conditional mean first. They are usually weak and may be captured by a low–order ARMA structure, typically AR(1) or MA(1).
Table 2: Basic statistical properties of daily stock index returns, January 2000–October 2011 ($T = 3081$)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CAC 40</td>
<td>−0.008</td>
<td>2.44</td>
<td>0.027</td>
<td>7.91</td>
<td>3092.4***</td>
<td>551.7***</td>
</tr>
<tr>
<td>DAX 30</td>
<td>−0.004</td>
<td>2.64</td>
<td>0.004</td>
<td>7.30</td>
<td>2376.8***</td>
<td>573.6***</td>
</tr>
<tr>
<td>FTSE 100</td>
<td>0.006</td>
<td>1.70</td>
<td>−0.144</td>
<td>8.98</td>
<td>4598.5***</td>
<td>709.7***</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>0.002</td>
<td>1.84</td>
<td>−0.163</td>
<td>10.6</td>
<td>7396.1***</td>
<td>782.3***</td>
</tr>
</tbody>
</table>

“Var.” is the variance, and “Skew.” and “Kurt.” refer to the moment–based sample skewness and kurtosis coefficients, respectively. “JB” is the Jarque–Bera test for normality, and “LM–ARCH(10)” is the Lagrange multiplier test with 10 lags. Asterisks *** indicate significance at 1% level.
Illustration: Fitting GARCH Models

- Parameter estimates are reported in Table 3.

Table 3: GARCH(1,1) estimates for various stock return series, approx. 1990–2010

<table>
<thead>
<tr>
<th>Series</th>
<th>$\hat{\omega}$</th>
<th>$\hat{\alpha}_1$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\alpha}_1 + \hat{\beta}_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAC 40</td>
<td>0.020</td>
<td>0.098</td>
<td>0.896</td>
<td>0.994</td>
</tr>
<tr>
<td></td>
<td>(0.0051)</td>
<td>(0.0103)</td>
<td>(0.0103)</td>
<td>(0.0044)</td>
</tr>
<tr>
<td>DAX 30</td>
<td>0.024</td>
<td>0.099</td>
<td>0.893</td>
<td>0.992</td>
</tr>
<tr>
<td></td>
<td>(0.0055)</td>
<td>(0.0101)</td>
<td>(0.0102)</td>
<td>(0.0045)</td>
</tr>
<tr>
<td>FTSE 100</td>
<td>0.012</td>
<td>0.108</td>
<td>0.888</td>
<td>0.996</td>
</tr>
<tr>
<td></td>
<td>(0.0033)</td>
<td>(0.0113)</td>
<td>(0.0111)</td>
<td>(0.0047)</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>0.013</td>
<td>0.081</td>
<td>0.912</td>
<td>0.993</td>
</tr>
<tr>
<td></td>
<td>(0.0029)</td>
<td>(0.0083)</td>
<td>(0.0086)</td>
<td>(0.0036)</td>
</tr>
</tbody>
</table>

Estimates of the mean parameter $\mu$ in (90) are not reported. Standard errors are given in parentheses. (Note that the standard error of $\hat{\alpha}_1 + \hat{\beta}_1$ is calculated as $\sqrt{\text{Var}(\hat{\alpha}_1) + \text{Var}(\hat{\beta}_1) + 2\text{Cov}(\hat{\alpha}_1, \hat{\beta}_1)}$.)

- All the fitted models imply a very high degree of persistence in volatility, $\alpha + \beta$. 
Basic diagnostics

- Diagnostics can be based on the sequence of standardized residuals,

\[
\hat{\eta}_t = \frac{\hat{\epsilon}_t}{\hat{\sigma}_t}, \quad t = 1, \ldots, T,
\]

(93)

where

\[
\hat{\sigma}_t^2 = \hat{\omega} + \hat{\alpha}_1 \hat{\epsilon}_t^2 + \hat{\beta}_1 \hat{\sigma}_{t-1}^2,
\]

(94)

and \( \hat{\epsilon}_t \) is the residual from a model for the conditional mean.\(^{15}\)

- This sequence should behave \textit{(very) similar to} an iid sequence from the innovation distribution that has been assumed for \( \{\eta_t\} \).

- They are only \textit{(very) similar to} (even for a correctly specified model) because \( \hat{\eta}_t \) is not equal to \( \eta_t \): The former is based on \textit{estimated parameters}.

\(^{15}\) In the simplest case, as in our example, the conditional mean is constant, i.e., \( \hat{\epsilon}_t = r_t - \hat{\mu} \).
Basic diagnostics

• The GARCH model should capture the conditional heteroskedasticity.

• Thus, sequence (93) should not display significant conditional heteroskedasticity.

• A useful check can be provided by plotting the SACF of the absolute or squared residuals.

• The fact that we only have $\{\hat{\eta}_t\}$ rather than $\{\eta_t\}$ makes the derivation of test statistics with known asymptotic distribution more complicated.
Basic diagnostics

• However, a common approach is to compute a Ljung–Box–type statistic for the sample autocorrelations of the squared residuals,

\[
\tilde{Q} = T \sum_{\tau=1}^{K} \hat{\rho}_{\tilde{\eta}_t^2}(\tau),
\]

(95)

and compare it with the critical value of the \( \chi^2(K - p - q) \) distribution, where \( p \) and \( q \) are the orders of the fitted GARCH model.

• See Bollerslev and Mikkelsen (1996) for some justification and evidence in favor of this procedure.\(^{16}\)

• For an overview over specification tests for GARCH models, see Teräsvirta et al. (2010): Modelling Nonlinear Economic Time Series, Oxford University Press, Chapter 8.

CAC 40: ACF of $\{\hat{\eta}_t^2\}$

DAX 30: ACF of $\{\hat{\eta}_t^2\}$

FTSE 100: ACF of $\{\hat{\eta}_t^2\}$

S&P 500: ACF of $\{\hat{\eta}_t^2\}$
DAX 30 returns

simulated path of GARCH(1,1) model fitted to DAX 30 returns
Basic diagnostics

• In many applications of GARCH models (e.g., in risk management), appropriate specification of the conditional distribution is also important.

• If the innovations have been assumed normal, we can apply normality tests to (93).

• A qq–plot or a kernel density estimate of the standardized residuals \( \hat{\eta}_t \) against the hypothesized innovation distribution can also be informative.

• Table 4 shows the kurtosis of the returns and the standardized residuals (93).
Basic diagnostics

Table 4: Kurtosis of returns of standardized residuals (93)

<table>
<thead>
<tr>
<th>Series</th>
<th>returns</th>
<th>residuals (93)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAC 40</td>
<td>7.91</td>
<td>3.89</td>
</tr>
<tr>
<td>DAX 30</td>
<td>7.30</td>
<td>3.91</td>
</tr>
<tr>
<td>FTSE 100</td>
<td>8.98</td>
<td>3.60</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>10.6</td>
<td>4.46</td>
</tr>
</tbody>
</table>

- Although part of the excess kurtosis has been captured by the GARCH process, a Jarque–Bera test still rejects normality at any reasonable level for all series.

- In summary, the simple normal GARCH(1,1) appears to do a good job in filtering out the conditional heteroskedasticity, but typically is not sufficient to capture the full amount of excess kurtosis.

- Kernel densities of $\hat{\eta}_t$ against the standard normal pdf convey the same message.
Alternative Innovation Distributions

- In financial applications, it is often of interest to know the entire conditional distribution rather than just the first two moments.

- Thus it appears reasonable to replace the normal distribution of $\eta_t$ in the GARCH(1,1) with a more flexible alternative that allows for conditional leptokurtosis.

- Conditional ML estimation works as before, but with a different density function.

- Two popular candidates in this regard are the
  - Student’s $t$
  - Generalized Error Distribution (GED).
Alternative Innovation Distributions

- The unit–variance versions\textsuperscript{17} of these are given by

\[ f_t(\eta_t; \nu) = \frac{\Gamma \left( \frac{\nu+1}{2} \right)}{\Gamma(\nu/2)\sqrt{(\nu - 2)\pi}} \left( 1 + \frac{\eta_t^2}{\nu - 2} \right)^{-\frac{\nu+1}{2}} , \quad (96) \]

and

\[ f_{GED}(\eta_t; p) = \frac{\lambda p}{2^{1/p+1}\Gamma(1/p)} \exp \left\{ -\frac{\lambda|\eta_t|^p}{2} \right\} , \quad (97) \]

where \( \lambda = 2^{1/p} \sqrt{\Gamma(3/p)/\Gamma(1/p)} \).

- These have been standardized so that they have unit variance.

- Parameters \( \nu \) and \( p \) govern the degree of conditional leptokurtosis in (96) and (97), respectively.

- Note that (97) nests the normal distribution for \( p = 2 \).

\textsuperscript{17}Sometimes unstandardized innovation distributions (with variance different from one) are used, see Appendix B.
Table 5: GARCH(1,1) estimates with nonnormal innovation distributions

<table>
<thead>
<tr>
<th>Series</th>
<th>$\hat{\omega}$</th>
<th>$\hat{\alpha}_1$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\nu}$</th>
<th>$\hat{\alpha}_1 + \hat{\beta}_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAC 40</td>
<td>0.016 (0.0052)</td>
<td>0.089 (0.0107)</td>
<td>0.906 (0.0107)</td>
<td>11.5 (2.1221)</td>
<td>0.995 (0.0046)</td>
</tr>
<tr>
<td>DAX</td>
<td>0.018 (0.0056)</td>
<td>0.093 (0.0109)</td>
<td>0.902 (0.0108)</td>
<td>11.0 (2.0897)</td>
<td>0.995 (0.0048)</td>
</tr>
<tr>
<td>FTSE</td>
<td>0.012 (0.0037)</td>
<td>0.104 (0.0124)</td>
<td>0.891 (0.0121)</td>
<td>11.2 (2.1608)</td>
<td>0.995 (0.0052)</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>0.008 (0.0031)</td>
<td>0.081 (0.0103)</td>
<td>0.918 (0.0099)</td>
<td>6.81 (0.8842)</td>
<td>0.999 (0.0045)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Series</th>
<th>$\hat{\omega}$</th>
<th>$\hat{\alpha}_1$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{p}$</th>
<th>$\hat{\alpha}_1 + \hat{\beta}_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAC 40</td>
<td>0.018 (0.0055)</td>
<td>0.093 (0.0112)</td>
<td>0.901 (0.0112)</td>
<td>1.56 (0.0582)</td>
<td>0.994 (0.0048)</td>
</tr>
<tr>
<td>DAX</td>
<td>0.020 (0.0060)</td>
<td>0.097 (0.0116)</td>
<td>0.898 (0.0115)</td>
<td>1.47 (0.0571)</td>
<td>0.995 (0.0050)</td>
</tr>
<tr>
<td>FTSE</td>
<td>0.013 (0.0038)</td>
<td>0.107 (0.0129)</td>
<td>0.888 (0.0127)</td>
<td>1.54 (0.0592)</td>
<td>0.995 (0.0053)</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>0.010 (0.0033)</td>
<td>0.081 (0.0106)</td>
<td>0.915 (0.0104)</td>
<td>1.29 (0.0481)</td>
<td>0.997 (0.0044)</td>
</tr>
</tbody>
</table>

Estimates of the mean parameter $\mu$ in (90) are not reported. Standard errors are given in parentheses. (Note that the standard error of $\hat{\alpha}_1 + \hat{\beta}_1$ is calculated as $\sqrt{\text{Var}(\hat{\alpha}_1) + \text{Var}(\hat{\beta}_1) + 2\text{Cov}(\hat{\alpha}_1, \hat{\beta}_1)}$.)
Alternative Innovation Distributions

Table 6: Maximized log–likelihood values

<table>
<thead>
<tr>
<th></th>
<th>CAC 40</th>
<th>DAX</th>
<th>FTSE</th>
<th>S&amp;P500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>-5151.6</td>
<td>-5249.7</td>
<td>-4497.3</td>
<td>-4588.4</td>
</tr>
<tr>
<td>Student’s $t$</td>
<td>-5128.8</td>
<td>-5227.2</td>
<td>-4479.0</td>
<td>-4536.0</td>
</tr>
<tr>
<td>GED</td>
<td>-5128.6</td>
<td>-5217.9</td>
<td>-4473.8</td>
<td>-4517.8</td>
</tr>
</tbody>
</table>

Differences in log–likelihood

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Student’s $t$–Normal</td>
<td>22.76</td>
<td>22.54</td>
<td>18.36</td>
<td>52.38</td>
</tr>
<tr>
<td>GED–Normal</td>
<td>22.98</td>
<td>31.85</td>
<td>23.48</td>
<td>70.59</td>
</tr>
</tbody>
</table>

• Since the normal is strictly included in the GED (for $p = 2$), standard tests can be used to discriminate between conditional normality and conditional GED.

• This is not the case for Student’s $t$, where the normal arises as a limit case.
Note: AIC and BIC

- To compare (nonnested) models, information criteria such as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are occasionally used.

- They are given by

\[
AIC = -2 \log L + 2K, \quad (98)
\]

and

\[
BIC = -2 \log L + K \log T, \quad (99)
\]

respectively, where \( L \) is the value of the maximized likelihood, and \( K \) is the number of parameters of a model, and \( T \) is the sample size.

- Whereas increasing the order of the model will always increase the likelihood, the second component is a penalty factor for inclusion of additional parameters.

- Smaller values of AIC and BIC are preferred, with the BIC being more conservative (i.e., preferring more parsimonious models).
Asymmetric GARCH Models

• The basic 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 react more strongly to bad news than to good news, which is often referred to as the leverage effect.

• To illustrate, we may define the leverage effect at lag $\tau$ as

\[
L(\tau) = \text{Corr}(\epsilon_{t-\tau}, |\epsilon_t|).
\]  

(100)
leverage for the CAC 40

leverage for the DAX 30

leverage for the FTSE 100

leverage for the S&P 500
The first model that has been put forward is the Asymmetric GARCH (AGARCH) of Engle (1990), which specifies the conditional variance as

\[ \sigma_t^2 = \omega + \alpha (\epsilon_{t-1} - \theta)^2 + \beta \sigma_{t-1}^2 \]  \hspace{1cm} (101)

\[ = \omega + \alpha \theta^2 + \alpha \epsilon_{t-1}^2 - 2\alpha \theta \epsilon_{t-1} + \beta \sigma_{t-1}^2. \]  \hspace{1cm} (102)

In model (101), the conditional variance, as a function of \( \epsilon_{t-1} \), has its minimum at \( \theta \) rather than at zero.

Thus, if \( \theta > 0 \), negative shocks will have a greater impact on the conditional variance than positive shocks of the same magnitude.

(102) shows that, if \( \alpha + \beta < 1 \), the unconditional variance of this process is

\[ E(\sigma_t^2) = \frac{\omega + \alpha \theta^2}{1 - \alpha - \beta}. \]  \hspace{1cm} (103)
The asymmetric GARCH model proposed by Glosten, Jagannathan and Runkle (1993), referred to as *GJR–GARCH*, models the conditional variance as

\[ \sigma_t^2 = \omega + (\alpha + \theta S_{t-1}) \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \]

where

\[ S_{t-1} = \begin{cases} 
1 & \text{if } \epsilon_{t-1} < 0 \\
0 & \text{if } \epsilon_{t-1} \geq 0
\end{cases} \]

Clearly \( \theta > 0 \) implies that the change in the next period’s variance is negatively correlated with today’s return.

If the innovation density is symmetric (e.g., normal or Student’s t), the unconditional variance is

\[ E(\sigma_t^2) = \frac{\omega}{1 - \alpha - \theta/2 - \beta}. \]
News Impact Curve

- To analyze the asymmetric response of the variance in different GARCH specifications, Engle and Ng (1993) defined the new impact curve (NIC).

- This is defined as the functional relationship

$$\sigma_t^2 = \sigma_t^2(\epsilon_{t-1}),$$

with the lagged variance $\sigma_{t-1}^2$ evaluated at its unconditional value.

- For example, for the standard symmetric GARCH(1,1) model, we have

$$\sigma_t^2(\epsilon_{t-1}) = A + \alpha \epsilon_{t-1}^2,$$

where

$$A = \omega + \beta \sigma^2, \quad \sigma^2 = \frac{\omega}{1 - \alpha - \beta}.$$

- This is a symmetric function of $\epsilon_{t-1}$. 
News Impact Curve

- Asymmetries may be introduced in various ways: Compared to the standard GARCH, we can change either the position of the slope of the NIC (or both).

- For example, the AGARCH captures asymmetry by allowing its NIC to be centered at a positive $\epsilon_{t-1}$, since

$$
\sigma_t^2(\epsilon_{t-1}) = A + \alpha(\epsilon_{t-1} - \theta)^2,
$$

where

$$
A = \omega + \beta \sigma^2, \quad \sigma^2 = \frac{\omega + \alpha \theta^2}{1 - \alpha - \beta}.
$$
\( \omega = 0.025, \alpha = 0.075, \beta = 0.9 \)
News Impact Curve

• The GJR captures the asymmetry in the impact of news on volatility via a steeper slope for negative than for positive shocks, i.e.,

\[
\sigma^2_t(\epsilon_{t-1}) = A + \begin{cases} 
(\alpha + \theta)\epsilon^2_{t-1} & \text{if } \epsilon_{t-1} < 0 \\
\alpha\epsilon^2_{t-1} & \text{if } \epsilon_{t-1} \geq 0,
\end{cases}
\]

but the NIC of the GJR is still centered at zero, i.e., \(\sigma^2_t(\epsilon_{t-1})\) is minimized for \(\epsilon_{t-1} = 0\).

• The estimates reported on the following pages are based on normal innovations; clearly nonnormal distributions allowing for fat tails and asymmetries would be considered in practice.
Table 7: Asymmetric GARCH(1,1) estimates for various stock return series, January 1990 to October 2009

<table>
<thead>
<tr>
<th>Series</th>
<th>$\hat{\omega}$</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\beta}$</th>
<th>$\hat{\theta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAC 40</td>
<td>0.0000</td>
<td>0.0890</td>
<td>0.8865</td>
<td>0.7237</td>
</tr>
<tr>
<td></td>
<td>(0.0123)</td>
<td>(0.0114)</td>
<td>(0.0141)</td>
<td>(0.0955)</td>
</tr>
<tr>
<td>DAX</td>
<td>0.0000</td>
<td>0.0906</td>
<td>0.8832</td>
<td>0.7592</td>
</tr>
<tr>
<td></td>
<td>(0.0115)</td>
<td>(0.0108)</td>
<td>(0.0130)</td>
<td>(0.0929)</td>
</tr>
<tr>
<td>FTSE</td>
<td>0.0000</td>
<td>0.0916</td>
<td>0.8878</td>
<td>0.5427</td>
</tr>
<tr>
<td></td>
<td>(0.0067)</td>
<td>(0.0115)</td>
<td>(0.0133)</td>
<td>(0.0764)</td>
</tr>
<tr>
<td>S&amp;P500</td>
<td>0.0000</td>
<td>0.0820</td>
<td>0.8960</td>
<td>0.5995</td>
</tr>
<tr>
<td></td>
<td>(0.0065)</td>
<td>(0.0094)</td>
<td>(0.0114)</td>
<td>(0.0756)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Series</th>
<th>$\hat{\omega}$</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\beta}$</th>
<th>$\hat{\theta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAC 40</td>
<td>0.0231</td>
<td>0.0000</td>
<td>0.9096</td>
<td>0.1579</td>
</tr>
<tr>
<td></td>
<td>(0.0043)</td>
<td>(0.0106)</td>
<td>(0.0106)</td>
<td>(0.0174)</td>
</tr>
<tr>
<td>DAX</td>
<td>0.0273</td>
<td>0.0000</td>
<td>0.9058</td>
<td>0.1608</td>
</tr>
<tr>
<td></td>
<td>(0.0047)</td>
<td>(0.0100)</td>
<td>(0.0100)</td>
<td>(0.0175)</td>
</tr>
<tr>
<td>FTSE</td>
<td>0.0159</td>
<td>0.0000</td>
<td>0.9085</td>
<td>0.1574</td>
</tr>
<tr>
<td></td>
<td>(0.0030)</td>
<td>(0.0096)</td>
<td>(0.0103)</td>
<td>(0.0171)</td>
</tr>
<tr>
<td>S&amp;P500</td>
<td>0.0143</td>
<td>0.0000</td>
<td>0.9219</td>
<td>0.1338</td>
</tr>
<tr>
<td></td>
<td>(0.0028)</td>
<td>(0.0106)</td>
<td>(0.0111)</td>
<td>(0.0150)</td>
</tr>
</tbody>
</table>
### Table 8: Maximized log–likelihood values

<table>
<thead>
<tr>
<th></th>
<th>CAC 40</th>
<th>DAX</th>
<th>FTSE</th>
<th>S&amp;P 500</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GARCH</strong></td>
<td>−5151.6</td>
<td>−5249.7</td>
<td>−4497.3</td>
<td>−4588.4</td>
</tr>
<tr>
<td><strong>AGARCH</strong></td>
<td>−5090.4</td>
<td>−5186.7</td>
<td>−4450.5</td>
<td>−4536.1</td>
</tr>
<tr>
<td><strong>GJR–GARCH</strong></td>
<td>−5079.6</td>
<td>−5180.9</td>
<td>−4437.3</td>
<td>−4519.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Differences in log–likelihood</th>
<th>AGARCH – GARCH</th>
<th>61.2</th>
<th>63.0</th>
<th>46.8</th>
<th>52.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>GJR – GARCH</td>
<td>72.0</td>
<td>68.8</td>
<td>60.0</td>
<td>68.7</td>
<td></td>
</tr>
</tbody>
</table>

- Note that in the estimated GJR model, positive shocks have no impact on future volatility (i.e., \( \hat{\alpha} = 0 \)).

- It is not possible, however, that good news (i.e., positive shocks) actually decrease future volatility, since this would require \( \alpha < 0 \) and \( \sigma_t^2 > 0 \) for all \( t \) would no longer be guaranteed.

- The EGARCH model is more flexible in this regard.
NICs for CAC 40

\[ \sigma_t^2 = \text{NICs for CAC 40} \]

\[ \text{GARCH} \]
\[ \text{AGARCH} \]
\[ \text{GJR–GARCH} \]
NICs for DAX 30

GARCH
AGARCH
GJR-GARCH

\( \sigma_t^2 \)

\( \varepsilon_{t-1} \)
NICs for FTSE 100

GARCH
AGARCH
GJR−GARCH
NICs for S&P 500

- GARCH
- AGARCH
- GJR-GARCH
Asymmetric GARCH Models III: EGARCH

- The Exponential GARCH process of Nelson (1991)\(^{18}\) is another popular GARCH specification.

- The EGARCH(1,1) model for the conditional variance of \(\epsilon_t\) is

\[
\log \sigma^2_t = \omega + g(\eta_{t-1}) + \beta \log \sigma^2_{t-1}, \tag{104}
\]

where \(\eta_t = \epsilon_t / \sigma_t\), and

\[
g(\eta_{t-1}) = \theta \eta_{t-1} + \alpha (|\eta_{t-1}| - \mathbb{E}(|\eta_{t-1}|)). \tag{105}
\]

For example, if \(\eta_t\) is standard normally distributed, then\(^{19}\)

\[
\mathbb{E}(|\eta_{t-1}|) = \sqrt{\frac{2}{\pi}} \approx 0.7979.
\]


\(^{19}\)With \(\phi(z) = e^{-z^2/2} / \sqrt{2\pi}\) being the pdf of the standard normal, and \(\phi'(z) = -z\phi(z)\), by symmetry of \(\phi(z)\), \(\mathbb{E}(|z|) = \int_{-\infty}^{\infty} |z| \phi(z) dz = 2 \int_{0}^{\infty} z\phi(z) dz = -2 \phi(z)|_{-\infty}^{\infty} = 2\phi(0) = 2 / \sqrt{2\pi} = \sqrt{2/\pi}\).
• Note that specification (104)–(105) does not require any parameter constraints to make sure that the conditional variance remains positive.
Asymmetric GARCH Models III: EGARCH

- **EGARCH:** \( \log \sigma_t^2 = \omega + g(\eta_{t-1}) + \beta \log \sigma_{t-1}^2 \), where \( \eta_t = \epsilon_t / \sigma_t \), and
  \[
g(\eta_{t-1}) = \theta \eta_{t-1} + \alpha (|\eta_{t-1}| - \mathbb{E}(|\eta_{t-1}|)).
  \] (106)

- The term \( g(\eta_t) \) is iid with zero mean, so that (104) is an AR(1) process for \( \log \sigma_t^2 \) which is stationary when \( |\beta| < 1 \).

- Regarding the asymmetric response to shocks,
  - for \( \eta_t > 0 \), \( g(\eta_t) \) is linear with slope \( \alpha + \theta \)
  - for \( \eta_t < 0 \), \( g(\eta_t) \) is linear with slope \( \theta - \alpha \)

- Thus, a variety of asymmetric response patterns are possible, e.g.,
  - if \( \theta = \alpha \), we have a response only to positive shocks, whereas
  - for \( \theta = -\alpha \), we have a response only to negative shocks.
  - If \( \alpha + \theta < 0 \), the slope is negative for positive shocks.
Asymmetric GARCH Models III: EGARCH

Table 9: EGARCH(1,1) estimates

<table>
<thead>
<tr>
<th>Series</th>
<th>$\hat{\omega}$</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\beta}$</th>
<th>$\hat{\theta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAC 40</td>
<td>0.0096 (0.0032)</td>
<td>0.1009 (0.0125)</td>
<td>0.9816 (0.0026)</td>
<td>-0.1417 (0.0113)</td>
</tr>
<tr>
<td>DAX 30</td>
<td>0.0138 (0.0033)</td>
<td>0.1216 (0.0135)</td>
<td>0.9793 (0.0028)</td>
<td>-0.1287 (0.0108)</td>
</tr>
<tr>
<td>FTSE 100</td>
<td>0.0016 (0.0025)</td>
<td>0.1063 (0.0133)</td>
<td>0.9849 (0.0024)</td>
<td>-0.1281 (0.0097)</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>0.0044 (0.0025)</td>
<td>0.0960 (0.0120)</td>
<td>0.9827 (0.0025)</td>
<td>-0.1269 (0.0103)</td>
</tr>
</tbody>
</table>

- The Model is

\[
\log \sigma_t^2 = \omega + g(\eta_{t-1}) + \beta \log \sigma_{t-1}^2,
\]

where $\eta_t = \epsilon_t / \sigma_t$ is assumed to have a standard normal distribution, and

\[
g(\eta_{t-1}) = \theta \eta_{t-1} + \alpha(\eta_{t-1} - E(\eta_{t-1})).
\]
\[ g(\eta_{t-1}) = \theta \eta_{t-1} + \alpha (|\eta_{t-1}| - \sqrt{\frac{2}{\pi}}) \]
Asymmetric GARCH Models

Table 10: Maximized log–likelihood values

<table>
<thead>
<tr>
<th></th>
<th>CAC 40</th>
<th>DAX</th>
<th>FTSE</th>
<th>S&amp;P 500</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH</td>
<td>−5151.6</td>
<td>−5249.7</td>
<td>−4497.3</td>
<td>−4588.4</td>
</tr>
<tr>
<td>AGARCH</td>
<td>−5090.4</td>
<td>−5186.7</td>
<td>−4450.5</td>
<td>−4536.1</td>
</tr>
<tr>
<td>GJR–GARCH</td>
<td>−5079.6</td>
<td>−5180.9</td>
<td>−4437.3</td>
<td>−4519.7</td>
</tr>
<tr>
<td>EGARCH</td>
<td>−5062.7</td>
<td>−5179.2</td>
<td>−4425.9</td>
<td>−4525.0</td>
</tr>
</tbody>
</table>
Asymmetric GARCH Models: Exercise

• Consider the AGARCH model

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

where \( \tilde{\omega} = \omega + \alpha \theta^2. \)

• Show that

\[ E(\epsilon_t^2) = \frac{\tilde{\omega}}{1 - \alpha - \beta}, \quad E(\epsilon_t^4) = 3 \frac{\tilde{\omega}^2(1 + \alpha + \beta) + 4 \alpha^2 \theta^2 \tilde{\omega}}{(1 - \alpha - \beta)(1 - 3\alpha^2 - 2\alpha\beta - \beta^2)}, \]

and thus the kurtosis

\[ [E^2(\epsilon_t^2)]^{-1} E(\epsilon_t^4) = \frac{3(1 - (\alpha + \beta)^2)}{1 - (\alpha + \beta)^2 - 2\alpha^2} + \frac{12\alpha^2 \theta^2 / E(\epsilon_t^2)}{1 - (\alpha + \beta)^2 - 2\alpha^2}. \quad (107) \]

Compare with the symmetric GARCH(1,1), where \( \theta = 0. \)
ARCH–M

• In the finance literature, a link is often made between the expected return and the risk of an asset.

• Investors are willing to hold risky assets only if their expected return compensate for the risk.

• A model that incorporates this link is the GARCH–in–mean or GARCH–M model, which can be written as

\[ r_t = c + \delta g(\sigma_t^2) + \epsilon_t, \]

where \( \epsilon_t \) is a GARCH error process, and \( g \) is a known function such as \( g(\sigma_t^2) = \sigma_t^2 \), \( g(\sigma_t^2) = \sigma_t \), or \( g(\sigma_t^2) = \log(\sigma_t^2) \).

• If \( \delta > 0 \) and \( g \) is monotonically increasing, then the term \( \delta g(\sigma_t^2) \) can be interpreted as a risk premium that increases expected returns if conditional volatility \( \sigma_t^2 \) is high.
Appendix A: Note on the nonnegativity conditions (22)

• To see that the nonnegativity conditions can be weakened for higher-order models, consider the GARCH(1,2).

• Using lag–operator notation, write the model as ARCH(∞),

\[
\begin{align*}
\sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \beta_1 \sigma_{t-1}^2 \\
(1 - \beta_1 L) \sigma_t^2 &= \omega + (\alpha_1 L + \alpha_2 L^2) \epsilon_t^2 \\
\sigma_t^2 &= \frac{\omega}{1 - \beta_1} + \left( \frac{\alpha_1 L}{1 - \beta_1 L} + \frac{\alpha_2 L^2}{1 - \beta_1 L} \right) \epsilon_t^2 \\
&= \frac{\omega}{1 - \beta_1} + \left( \alpha_1 \sum_{i=1}^{\infty} \beta_1^{i-1} L^i + \alpha_2 \sum_{i=2}^{\infty} \beta_1^{i-2} L^i \right) \epsilon_t^2 \\
&= \frac{\omega}{1 - \beta_1} + \alpha_1 \epsilon_{t-1}^2 + \sum_{i=2}^{\infty} (\alpha_1 \beta_1^{i-1} + \alpha_2 \beta_1^{i-2}) \epsilon_{t-i}^2 \\
&= \frac{\omega}{1 - \beta_1} + \alpha_1 \epsilon_{t-1}^2 + \sum_{i=2}^{\infty} \beta_1^{i-2} (\alpha_1 \beta_1 + \alpha_2) \epsilon_{t-i}^2.
\end{align*}
\]
Appendix A: Note on the nonnegativity conditions (22)

- That is, inversion of $1 - \beta_1 L$ gives the ARCH($\infty$) representation $\sigma_t^2 = \omega/(1 - \beta_1) + \sum_{i=1}^{\infty} \psi_i \epsilon_{t-i}^2$ with coefficients $\psi_i$.

- The ARCH($\infty$) coefficients are given by

  $$\begin{align*}
  \psi_1 &= \alpha_1 \\
  \psi_k &= \beta_1^{k-2}(\alpha_1 \beta_1 + \alpha_2), \quad k \geq 2.
  \end{align*}$$

- Clearly $\sigma_t^2$ will remain positive for all $t$ if $\omega > 0$ as well as $\psi_k \geq 0$ for all $k$.

- This gives rise to the set of conditions

  $$\begin{align*}
  \omega &> 0 \\
  \alpha_1 &\geq 0 \\
  \beta_1 &\geq 0 \\
  \alpha_1 \beta_1 + \alpha_2 &\geq 0.
  \end{align*}$$
Appendix A: Note on the nonnegativity conditions (22)

- Hence $\alpha_2$ may be negative if $\alpha_1 > 0$ and $\beta_1 > 0$.
Appendix B: Non–normal innovations with non–unit variance

• It is not necessary that the innovation distribution in a GARCH model has unit variance.

• For example, often the $t$ distribution is used in the form as it appears in standard sampling theory, i.e.,

$$f_t(\eta_t; \nu) = \frac{\Gamma \left( \frac{\nu+1}{2} \right)}{\Gamma(\nu/2) \sqrt{\nu\pi}} \left(1 + \frac{\eta_t^2}{\nu}\right)^{-(\nu+1)/2}, \quad (108)$$

which has variance (provided $\nu > 2$)

$$\mathbb{E}(\eta^2_t) = \frac{\nu}{\nu - 2}. \quad (109)$$

• Using conditional distributions with variance different from unity will not change the fit of the model, but we have to account for the change in scale when we check the (covariance) stationarity condition, calculate unconditional variance etc.
Note: Non–normal innovations with non–unit variance

• To illustrate, in the GARCH\((p, q)\),

\[
\begin{align*}
\epsilon_t &= \eta_t \sigma_t \\
\sigma_t^2 &= \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2,
\end{align*}
\]

(110)

to find the unconditional variance, we take expectations on both sides,

\[
E(\sigma_t^2) = \omega + \sum_{i=1}^{q} \alpha_i E(\epsilon_{t-i}^2) + \sum_{i=1}^{p} \beta_i E(\sigma_{t-i}^2).
\]
Appendix B: Non–normal innovations with non–unit variance

• If the innovations \( \eta_t \) in (110) have unit variance, \( E(\eta_t^2) = 1 \), it follows that \( E(\epsilon_t^2) = E(\eta_t^2 \sigma_t^2) = E(\eta_t^2) E(\sigma_t^2) = E(\sigma_t^2) \), and so

\[
E(\epsilon_t^2) = E(\sigma_t^2) = \frac{\omega}{1 - \sum_i \alpha_i - \sum_i \beta_i},
\]

(111)

provided the second–order stationarity condition

\[
\sum_i \alpha_i + \sum_i \beta_i < 1
\]

(112)

is satisfied.
Appendix B: Non–normal innovations with non–unit variance

- If, in general,
  \[ E(\eta_t^2) = \kappa_2 \neq 1, \]
  then
  \[ E(\sigma_t^2) = \omega + \sum_{i=1}^{q} \alpha_i E(e_{t-i}^2) + \sum_{i=1}^{p} \beta_i E(\sigma_{t-i}^2) \]
  \[ = \omega + \sum_{i=1}^{q} \alpha_i E(\eta_t^2 \sigma_t^2) + \sum_{i=1}^{p} \beta_i E(\sigma_t^2) \]
  \[ = \omega + \left( \sum_{i=1}^{q} \alpha_i \kappa_2 + \sum_{i=1}^{p} \beta_i \right) E(\sigma_t^2). \]
Appendix B: Non–normal innovations with non–unit variance

• Hence,

\[ E(\sigma^2_t) = \frac{\omega}{1 - \kappa_2 \sum_{i=1}^{q} \alpha_i - \sum_{i=1}^{p} \beta_i} \]

provided the covariance stationarity condition

\[ \kappa_2 \sum_{i} \alpha_i + \sum_{i} \beta_i < 1 \]

is satisfied.

• Moreover, the unconditional variance of \( \epsilon_t \) is

\[ E(\epsilon^2_t) = E(\eta^2_t \sigma^2_t) = \kappa_2 E(\sigma^2_t) = \frac{\kappa_2 \omega}{1 - \kappa_2 \sum_i \alpha_i - \sum_i \beta_i}. \]