Financial Data Analysis

GARCH Models, Part I

May 26, 2010
Several Stylized Facts

- Returns usually show no or only little autocorrelation.
- Volatility appears to be autocorrelated (volatility clusters).
- Normality is rejected in favor of a leptokurtic (fat-tailed) distribution.
Volatility Modeling and the Stylized Facts

- Consider the following model for returns $r_t$,

\begin{align}
    r_t &= \mu_t + \epsilon_t \\
    \epsilon_t &= \eta_t \sigma_t, \quad \eta_t \overset{iid}{\sim} N(0, 1),
\end{align}

where we assume that the innovation sequence $\eta_t$ is also independent of $\sigma_t$.

- $\mu_t$ in (1) is the conditional mean of $r_t$ conditional on the information up to time $t - 1$. This may, for example, be constant or described by a low–order ARMA process.

- We are interested in the error term described by the second line of (1).

- If $\sigma_t^2$ depends on information available at time $t - 1$, then $\sigma_t^2$ is the conditional variance of $\epsilon_t$ (and thus also $r_t$).

- Denote the information available up to time $t$ by $I_t$; $I_t$ typically consists of the past history of the process, $\{\epsilon_s : s \leq t\}$.
• Then we can also write

\[ \epsilon_t | I_{t-1} \sim N(0, \sigma_t^2), \]  

(2)
i.e., \( \epsilon_t \) is conditionally normally distributed with variance \( \sigma_t^2 \).

• However, if the conditional variance is time–varying (which is the case we are interested in), the unconditional distribution of \( \epsilon_t \) will not be normal.

• To illustrate, consider the marginal kurtosis of \( \epsilon_t \), assuming \( \epsilon_t \) is stationary with finite fourth moment,

\[
\text{kurtosis}(\epsilon_t) = \frac{E(\epsilon_t^4)}{E^2(\epsilon_t^2)} = \frac{E(\eta_t^4 \sigma_t^2)}{E^2(\eta_t^2 \sigma_t^2)} = \frac{E(\eta_t^4) E(\sigma_t^4)}{E^2(\eta_t^2) E^2(\sigma_t^2)} \\
= \frac{E(\eta_t^4)}{E^2(\eta_t^2)} \frac{E(\sigma_t^4)}{E^2(\sigma_t^2)} > 3, \tag{3}
\]

since

\[
E(\sigma_t^4) > E^2(\sigma_t^2) \quad (E(X^2) > E^2(X)). \tag{4}
\]
• An interpretation of (3) results from noting that

\[
\frac{E(\sigma_t^4)}{E^2(\sigma_t^2)} = 1 + \frac{E(\sigma_t^4) - E^2(\sigma_t^2)}{E^2(\sigma_t^2)} = 1 + \frac{\text{Var}(\sigma_t^2)}{E^2(\sigma_t^2)}.
\]

• Thus, for a given level of the *unconditional variance* \(E(\sigma_t^2) = E(\epsilon_t^2)\), the kurtosis of the marginal distribution of \(\epsilon_t\) is increasing in the variability of the conditional variance.

• If \(\text{Var}(\sigma_t^2)\) is large, then \(\sigma_t^2\) will often be considerably smaller (larger) than \(E(\sigma_t^2)\), giving rise to high peaks and thick tails of the marginal distribution, respectively.

• Thus, even with normal innovations (conditional normality), time–varying conditional volatility may account for at least part of the leptokurtosis observed in financial return series.
• A further property of the error process is uncorrelatedness,

\[ E(\epsilon_t \epsilon_{t-\tau}) = E(\eta_t \eta_{t-\tau} \sigma_t \sigma_{t-\tau}) = E(\eta_t)E(\eta_{t-\tau} \sigma_t \sigma_{t-\tau}) = 0. \]

• Absolute values and squares will in general be correlated, however.

• Thus, at least in principle, a process of the form (1) is capable of reproducing several of the properties typically detected in financial returns.
The ARCH Process

• Engle (1982) introduced the class of **autoregressive conditional heteroskedastic (ARCH)** models,¹ where (1) is specified as

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

which is referred to as ARCH(\(q\)).

• Shocks drive the variance.

---

• It has been shown that the ARCH process generates marginal distributions that (asymptotically) decay as a power law, i.e., for some $\gamma > 0$,

$$\Pr(|r_t| > x) \simeq cx^{-\gamma}, \quad \text{as } x \to \infty,$$

so that moments of $\{\epsilon_t\}$ exist only of order smaller than $\gamma$.

• For example, by taking expectations in (5),

$$E(\sigma^2_t) = E(\epsilon_t^2) = \omega + \sum_{i=1}^{q} \alpha_i E(\epsilon_{t-i}^2),$$

we get, for the unconditional variance,

$$E(\sigma^2_t) = E(\epsilon_t^2) = \frac{\omega}{1 - \alpha_1 - \alpha_2 - \cdots - \alpha_q}.$$

• This makes sense only if

$$\sum_{i=1}^{q} \alpha_i < 1,$$  \hspace{1cm} (6)
which turns out to be the condition for the finiteness of the variance in the ARCH(q) model, and is often referred to as the stationarity condition.

- Several further properties of the model can best be illustrated by means of the ARCH(1) specification, given by

$$
\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2.
$$

(7)

- We first calculate the fourth moment of the process,

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

(8)
\[
\sigma_t^4 = (\omega + \alpha_1 \epsilon_{t-1}^2)^2 = \omega^2 + 2\omega \alpha_1 \epsilon_{t-1}^2 + \alpha_1^2 \epsilon_{t-1}^4
\]

\[
E(\sigma_t^4) = \omega^2 + 2\omega \alpha_1 E(\epsilon_t^2) + \alpha_1^2 E(\epsilon_t^4)
\]

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

\[
E(\sigma_t^4) = \frac{1}{1 - 3\alpha_1^2} \left[ \omega^2 + \frac{2\omega^2 \alpha_1}{1 - \alpha_1} \right] = \frac{\omega^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)},
\]

which makes sense only if \(3\alpha^2 < 1\), which is the condition for the finiteness of the fourth moment.

- In this case, from (8)

\[
E(\epsilon_t^4) = \frac{3\omega^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)}, \quad (9)
\]

and the kurtosis of the unconditional distribution is, with \(E(\epsilon_t^2) = \)
\[ \omega/(1 - \alpha_1), \]

\[
\frac{E(\epsilon_t^4)}{E^2(\epsilon_t^2)} = \frac{3\omega^2(1 + \alpha_1)(1 - \alpha_1)^2}{\omega^2(1 - \alpha_1)(1 - 3\alpha_1^2)} \\
= \frac{3(1 - \alpha_1)(1 + \alpha_1)}{1 - 3\alpha_1^2} \\
= \frac{3(1 - \alpha_1^2)}{1 - 3\alpha_1^2} > 3. 
\]

- The ACF of the squared process,

\[
\varrho(\tau) = \text{Corr}(\epsilon_t^2, \epsilon_{t-\tau}^2) = \frac{E(\epsilon_t^2 \epsilon_{t-\tau}^2) - E^2(\epsilon_t^2)}{E(\epsilon_t^4) - E^2(\epsilon_t^2)}, \tag{10}
\]

which is well-defined for \(3\alpha_1^2 < 1\), is also of interest.
\[
E(e_t^2 e_{t-\tau}^2) = E(e_t^2 \eta_t^2 (\omega + \alpha_1 e_{t-1}^2)) = \omega E(e_t^2) + \alpha_1 E(e_{t-\tau}^2 e_{t-1}^2) = E^2(e_t^2)(1 - \alpha_1) + \alpha_1 E(e_{t-\tau}^2 e_{t-1}^2) = E^2(e_t^2) + \alpha_1 [E(e_{t-\tau}^2 e_{t-1}^2) - E^2(e_t^2)]
\]

\[
E(e_t^2 e_{t-\tau}^2) - E^2(e_t^2) = \alpha_1 [E(e_{t-\tau}^2 e_{t-1}^2) - E^2(e_t^2)],
\]

which implies \( \varrho(\tau) = \alpha_1 \varrho(\tau - 1). \)

- For \( \tau = 1 \), we have

\[
E(e_t^2 e_{t-1}^2) - E^2(e_t^2) = \alpha_1 [E(e_t^4) - E^2(e_t^2)],
\]

so

\[
\varrho(\tau) = \alpha^\tau. \quad (11)
\]
GARCH Models

- In practice, pure ARCH(\(q\)) processes are rarely used, since for an adequate fit a large number of lags is usually required.

- A more parsimonious formalization is provided by the Generalized ARCH (GARCH) process, which was published independently by Bollerslev (1986) and Taylor (1986).\(^2\)

- The GARCH(\(p, q\)) model generalizes (5) to

  \[
  \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2. \tag{12}
  \]

- To make sure that the variance is positive, Bollerslev (1986) imposed that

  \[
  \omega > 0; \quad \alpha_i \geq 0, i = 1, \ldots, q; \quad \beta_i \geq 0, i = 1, \ldots, p. \tag{13}
  \]

---

• These conditions are sufficient but can be substantially weakened for models where one of the orders is larger than unity. Conditions (13) are necessary and sufficient for guaranteeing a positive variance process in pure ARCH processes and the GARCH(1,1) process, however.

• Similar to the ARCH(q) process, we can calculate the unconditional variance of process as

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

provided the (covariance) stationarity condition

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

is satisfied.

• To characterize the correlation structure of the squared process, define the prediction error

\[
u_t = \epsilon_t^2 - E(\epsilon_t^2|I_{t-1}) = \epsilon_t^2 - \sigma_t^2.
\]
• $u_t = \epsilon_t^2 - \sigma_t^2 = (\eta_t^2 - 1)\sigma_t^2$ is white noise but not strict white noise, since it is uncorrelated but not independent.

• Substituting (17) for $\sigma_t^2$ into (12) results in

$$\epsilon_t^2 = \omega + \sum_{i=1}^{\max\{p,q\}} (\alpha_i + \beta_i)\epsilon_{t-i}^2 - \sum_{i=1}^{p} \beta_i u_{t-i} + u_t, \quad (17)$$

where $\alpha_i = 0$ for $i > q$ and $\beta_i = 0$ for $i > p$.

• Equation (17) is an ARMA$(\max\{p, q\}, p)$ representation for the squared process $\{\epsilon_t^2\}$, which characterizes its autocorrelations.

• The ARMA representation can also be used to explicitly calculate the autocorrelations.

• For example, the ARMA(1,1) representation of the GARCH(1,1) process is

$$\epsilon_t^2 = \omega + (\alpha_1 + \beta_1)\epsilon_{t-1}^2 + u_t - \beta_1 u_t. \quad (18)$$
Recall that the ACF of the ARMA(1,1) process

\[ Y_t = \phi Y_{t-1} + \theta \epsilon_{t-1} + \epsilon_t \]

is

\[ \text{Corr}(Y_t, Y_{t-\tau}) = \phi^{\tau-1} \frac{(\phi + \theta)(1 + \phi \theta)}{1 + 2\theta \phi + \theta^2}. \]

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

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

provided the fourth moment is finite.

The GARCH(1,1) process is most often applied in practice.
To find the moments of this process, it is convenient to write

\[
\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 = \omega + (\alpha_1 \eta_{t-1}^2 + \beta_1) \sigma_{t-1}^2
\]

\[
= \omega + c_{t-1} \sigma_{t-1}^2, \quad c_t = \alpha_1 \eta_t^2 + \beta_1.
\]

Since \( E(\sigma_t^2) = \frac{\omega}{1 - \alpha - \beta} = \omega/(1 - E(c_t)) \),

\[
E(\sigma_t^4) = \omega^2 + 2\omega E(c_t) E(\sigma_t^2) + E(c_t^2) E(\sigma_t^4)
\]

\[
E(\sigma_t^4) = \frac{\omega^2 (1 + E(c_t))}{(1 - E(c_t))(1 - E(c_t^2))}
\]

\[
= \frac{\omega^2 (1 + \alpha_1 + \beta_1)}{(1 - \alpha_1 - \beta_1)(1 - 3\alpha_1^2 - 2\alpha_1\beta_1 - \beta_1^2)},
\]

where \( E(c_t^2) = 3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1 \) is the condition for the existence of the fourth moment.
The kurtosis is then

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

\[
= 3 + \frac{6\alpha_1^2}{1 - 3\alpha_1^2 - 2\alpha_1 \beta_1 - \beta_1^2}.
\]
To see why a GARCH(1,1) fits better then even a high–order ARCH($q$), we substitute recursively for the lagged variances,

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

\[
= \omega (1 + \beta_1) + \alpha_1 \sum_{i=1}^{2} \beta_1^{i-1} \epsilon^2_{t-i} + \beta_1^2 \sigma^2_{t-2}
\]

\[
\vdots
\]

\[
= \omega \sum_{i=0}^{\tau-1} \beta_1^i + \alpha_1 \sum_{i=1}^{\tau} \beta_1^{i-1} \epsilon^2_{t-i} + \beta_1^\tau \sigma^2_{t-\tau}
\]

\[
= \frac{\omega}{1 - \beta_1} + \alpha_1 \epsilon^2_{t-1} + \alpha_1 \sum_{i=1}^{\infty} \beta_1^i \epsilon^2_{t-1-i}, \quad (19)
\]

provided $\beta_1 < 1$, so that $\beta_1^\tau \sigma^2_{t-\tau} \xrightarrow[\tau \to \infty]{} 0$.

This shows that GARCH(1,1) is ARCH($\infty$) with geometrically declining lag structure, i.e., $\sigma^2_t = \tilde{\omega} + \sum_{i=1}^{\infty} \psi_i \epsilon^2_{t-i}$, with $\psi_i = \alpha_1 \beta_1^{i-1}$. 


• The declining lag structure is reasonable as it implies that the impact of more recent shocks on the current variance is larger than that of earlier shocks.

• The ARCH(∞) representation (19) shows that $\alpha_1$ can be interpreted as a reaction parameter, as it measures the reactiveness of the conditional variance to a shock in the previous period, i.e., the immediate impact of a unit shock on the next period’s conditional variance.

• Parameter $\beta_1$, on the other hand, is a persistence parameter which measures the memory in the variance process. E.g., if $\beta_1$ is small, $\beta_{1}^{i}$ tends to zero very rapidly with $i$, and the direct impact of a shock on future conditional variances dies out soon.
Testing for GARCH

- If it is assumed that the conditional mean of the returns is not constant, then these test have to be applied to the residuals of a regression or ARMA model.

- The Ljung–Box–Pierce statistic for the autocorrelations of the squares,

\[ Q^* = T(T + 2) \sum_{\tau=1}^{K} \frac{\hat{\rho}_\tau^2}{T - \tau} \xrightarrow{as} \chi^2(K). \]  

- Engle (1982) derived a Lagrange multiplier test which works as follows.

- Run the regression

\[ \epsilon_t^2 = b_0 + b_1 \epsilon_{t-1}^2 + \cdots + b_q \epsilon_{t-q}^2 + u_t. \]  

- Under \( H_0 \) of no ARCH (conditional homoskedasticity), the test statistic

\[ LM = TR^2 \xrightarrow{as} \chi^2(q), \]
where $T$ is the sample size and $R^2$ is the coefficient of determination obtained from the regression (21).
Estimation

- GARCH models are usually estimated by conditional maximum likelihood.

- The parameter vector is \( \theta = (\omega, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p) \).

- Under conditional normality, we have for a sample of \( T \) observations,

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

  where

  \[
  \ell_t(\theta) = -\frac{1}{2} \log \sigma_t^2 - \frac{1}{2} \frac{\epsilon_t^2}{\sigma_t^2},
  \]

  where \( \sigma_t^2 \) is generated by the GARCH recursion.

- To start the recursion, we need pre-sample values \( \sigma_{0}^2, \ldots, \sigma_{-p+1}^2 \) and \( \epsilon_{0}^2, \ldots, \epsilon_{-q+1}^2 \).
• Bollerslev (1986) proposes to estimate their unconditional values from the sample, i.e., to set

\[ \sigma_0^2 = \cdots = \sigma_{-p+1}^2 = \epsilon_0^2 = \cdots = \epsilon_{-q+1}^2 = \frac{1}{T} \sum_{t=1}^{T} \epsilon_t^2. \]  

(23)

• To see what we typically get for a GARCH(1,1), the model

\[ r_t = \mu + \epsilon_t, \]

where \( \epsilon_t \) is GARCH(1,1), was fitted to several stock return series.
Table 1: 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>S&amp;P 500</td>
<td>0.0077 (0.0017)</td>
<td>0.0655 (0.0067)</td>
<td>0.9284 (0.0072)</td>
<td>0.9939</td>
</tr>
<tr>
<td>DAX</td>
<td>0.0355 (0.0053)</td>
<td>0.0918 (0.0089)</td>
<td>0.8910 (0.0099)</td>
<td>0.9828</td>
</tr>
<tr>
<td>FTSE</td>
<td>0.0113 (0.0025)</td>
<td>0.0856 (0.0081)</td>
<td>0.9059 (0.0087)</td>
<td>0.9915</td>
</tr>
<tr>
<td>CAC 40</td>
<td>0.0290 (0.0054)</td>
<td>0.0851 (0.0085)</td>
<td>0.9001 (0.0097)</td>
<td>0.9852</td>
</tr>
</tbody>
</table>

- Diagnostics can be based on the sequence

\[
\hat{\eta}_t = \frac{\epsilon_t}{\hat{\sigma}_t}, \quad t = 1, \ldots, T. \tag{24}
\]

Table 2: Kurtosis of raw returns and residuals (24)

<table>
<thead>
<tr>
<th></th>
<th>S&amp;P 500</th>
<th>DAX</th>
<th>FTSE</th>
<th>CAC 40</th>
</tr>
</thead>
<tbody>
<tr>
<td>raw returns</td>
<td>12.1307</td>
<td>8.0553</td>
<td>9.6318</td>
<td>7.8069</td>
</tr>
<tr>
<td>residuals (24)</td>
<td>4.8993</td>
<td>9.6475</td>
<td>3.8232</td>
<td>4.9332</td>
</tr>
</tbody>
</table>
- The unintuitive number for the DAX is due to the Gorbatschow-Putsch in August 1991.
Standard Normal Quantiles
Quantiles of Input Sample
S&P residuals

Standard Normal Quantiles
Quantiles of Input Sample
DAX residuals

Standard Normal Quantiles
Quantiles of Input Sample
FTSE residuals

Standard Normal Quantiles
Quantiles of Input Sample
CAC residuals
Note on the nonnegativity conditions (13)

- We can use lag–operator notation to write the GARCH model as

\[ \beta(L)\sigma_t^2 = \omega + \alpha(L)e_t^2, \]

where

\[
\begin{align*}
\beta(L) & = 1 - \beta_1 L - \beta_2 L^2 - \cdots - \beta_p L^p \\
\alpha(L) & = \alpha_1 L + \alpha_2 L^2 + \cdots + \alpha_q L^q.
\end{align*}
\]

Inverting gives the ARCH(\(\infty\))\(^3\)

\[
\sigma_t^2 = \frac{\omega}{1 - \sum_i \beta_i} + \frac{\alpha(L)}{\beta(L)}e_t^2 = \frac{\omega}{1 - \sum_i \beta_i} + \sum_{i=1}^{\infty} \psi_i e_{t-i}^2.
\]

- For \(\sigma_t^2\) to remain positive with probability 1, we observe that it is

\(^3\)This requires that \(\beta(z)\) has all roots outside the unit circle.
necessary and sufficient that

\[
\frac{\omega}{1 - \sum_i \beta_i} > 0, \quad \psi_i \geq 0 \text{ for all } i.
\]

- Nelson and Cao (1992) showed that these restrictions are weaker than (13) except for the pure ARCH\((q)\) and the GARCH\((1,1)\).

- The simplest case is the GARCH\((1,2)\),

\[
(1 - \beta_1 L)\sigma_t^2 = \omega + (\alpha_1 L + \alpha_2 L^2)\epsilon_t^2
\]

\[
= \sigma_t^2 = \frac{\omega}{1 - \beta_1} + \left(\alpha_1 \sum_{i=0}^\infty \beta_i L^{i+1} + \alpha_2 \sum_{i=0}^\infty \beta_i L^{i+2}\right) \epsilon_t^2
\]

\[
= \frac{\omega}{1 - \beta_1} + \alpha_1 \epsilon_{t-1}^2 + (\alpha_1 \beta_1 + \alpha_2) \epsilon_{t-2}^2 + (\alpha_1 \beta_1^2 + \alpha_2 \beta_1) \epsilon_{t-3}^2 + \cdots
\]

Thus

\[
\psi_1 = \alpha_1
\]

\[
\psi_k = \beta_1^{k-2}(\alpha_1 \beta_1 + \alpha_2), \quad k \geq 2.
\]
• This gives rise to the set of necessary and sufficient conditions

\[
\begin{align*}
\omega & > 0 \\
\alpha_1 & \geq 0 \\
1 & > \beta_1 \geq 0 \\
\alpha_1 \beta_1 + \alpha_2 & \geq 0.
\end{align*}
\]

• \(\alpha_2\) may thus be negative.