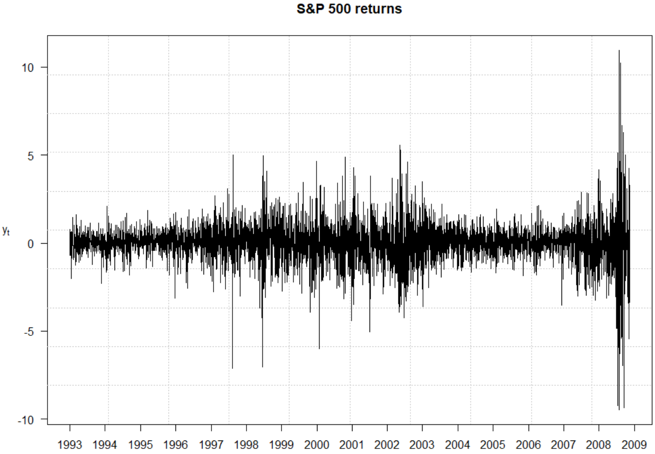
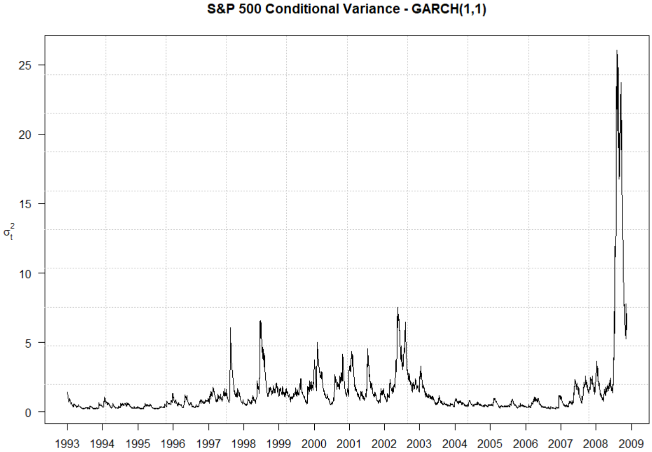
Univariate Volatility Modeling

Leopoldo Catania
Aarhus University and CREATES
leopoldo.catania@econ.au.dk

S&P500 Returns



S&P 500 Conditional Variance - GARCH(1,1)



Engle (1982)

Econometrica, Vol. 50, No. 4 (July, 1982)

AUTOREGRESSIVE CONDITIONAL HETEROSCEDASTICITY WITH ESTIMATES OF THE VARIANCE OF UNITED KINGDOM INFLATION[1]

BY ROBERT F. ENGLE

Traditional econometric models assume a constant one-period forecast variance. To generalize this implausible assumption, a new class of stochastic processes called autoregressive conditional heteroscedastic (ARCH) processes are introduced in this paper. These are mean zero, serially uncorrelated processes with nonconstant variances conditional on the past, but constant unconditional variances. For such processes, the recent past gives information about the one-period forecast variance.

A regression model is then introduced with disturbances following an ARCH process. Maximum likelihood estimators are described and a simple scoring iteration formulated. Ordinary least squares maintains its optimality properties in this set-up, but maximum likelihood is more efficient. The relative efficiency is calculated and can be infinite. To test whether the disturbances follow an ARCH process, the Lagrange multiplier procedure is employed. The test is based simply on the autocorrelation of the squared OLS residuals.

This model is used to estimate the means and variances of inflation in the U.K. The ARCH effect is found to be significant and the estimated variances increase substantially during the chaotic seventies.

GARCH models: a class of models

- The family of ARCH-GARCH models has been introduced by Engle (1982) and Bollerslev (1986) to capture the features of the financial log-returns.
- The idea is to assume that the variance at time t can be represented as a function of past observations, $\mathcal{F}_{t-1} = \{y_{t-s}, s = 1, 2, \dots\}$.
- The original ARCH(1) model of Engle (1982) assumes that:

$$r_t = \mu_t + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t$$
$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2$$

  where $\mathbb{E}[z_t|\mathcal{F}_{t-1}] = 0$ and $\mathbb{E}[z_t^2|\mathcal{F}_{t-1}] = 1$ and ϵ_{t-1}^2 is called the 'forcing variable' of the volatility process.
- An ARCH(q) model can be easily specified as $\sigma_t^2 = \omega + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2$

ARCH model

Assume that $\mu_t = \mu = 0$ such that $r_t = \varepsilon_t = \sigma_t z_t$. The unconditional variance of an ARCH(q) model is:

$$\begin{aligned} \mathbb{E}[r_t^2] &= \mathbb{E}[\sigma_t^2 z_t^2] \\ &= \mathbb{E}[\sigma_t^2] = \frac{\omega}{1 - \sum_{j=1}^{1} \alpha_j} \end{aligned}$$

In the simpler ARCH(1) case, assuming gaussianity for z_t, the forth moment is

$$\begin{aligned} \mathbb{E}[r_t^4] &= \mathbb{E}[\sigma_t^4 z_t^4] \\ &= \mathbb{E}[\sigma_t^4]\mathbb{E}[z_t^4] = 3\frac{\omega^2(1+\alpha)}{(1-\alpha)(1-3\alpha^2)} \end{aligned}$$

such that the kurtosis coefficient is

$$\kappa = 3\frac{1-\alpha^2}{1-3\alpha^2} > 3$$

AR(p) representation

Let $v_t = r_t^2 - \sigma_t^2$, such that $\sigma_t^2 = r_t^2 - v_t$. We can write

$$\sigma_t^2 = \omega + \sum_{j=1}^{q} \alpha_j r_{t-j}^2$$

$$r_t^2 - v_t = \omega + \sum_{j=1}^{q} \alpha_j r_{t-j}^2$$

$$r_t^2 = \omega + \sum_{j=1}^{q} \alpha_j r_{t-j}^2 + v_t,$$

where we know that $\mathbb{E}[v_t] = 0$ for all t and $\mathbb{E}[v_t^2] = 2\mathbb{E}[\sigma_t^4]$, i.e. $\{v_t\}$ is a martingale difference sequence with constant variance. So the ARCH(p) model has an AR(p) representation for r_t^2.

GARCH models: a class of models

- Bollerslev (1986) has generalized the ARCH model to GARCH as follow:

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

- Note that now the volatility also linearly depends from its past value.

- GARCH(p, q) is constructed as follow:

$$\sigma_t^2 = \omega + \sum_{j=1}^{q} \alpha_j r_{t-j}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2,$$

Properties of the model: stationarity

The unconditional variance of r_t is

$$Var(r_t) = E(\sigma_t^2) = \omega + \alpha E(r_{t-1}^2) + \beta E(\sigma_{t-1}^2)$$

under the assumption of weak stationarity (that we will check later) we have $E(\sigma_{t-1}^2) = E(r_{t-1}^2) = E(\sigma_t^2)$ and we note that $Var(r_t)$ has a parametric expression

$$Var(r_t) = \frac{\omega}{1 - \alpha - \beta}$$

Condition for weak stationarity is $\alpha + \beta < 1$. Conditions for positivity are $\omega > 0$, $\alpha > 0$ and $\beta > 0$, see Nelson and Cao (1992).

Properties of the model: Excess unconditional kurtosis

From the previous result, it follows that the second moment of returns ($\mu = 0$) is

$$E(r_t^2) = \frac{\omega}{1 - \alpha - \beta}$$

The fourth moment is

$$\begin{aligned} E(r_t^4) &= E(z_t^4)E(\sigma_t^4) \\ &= 3\omega^2(1+\alpha+\beta)[(1-\alpha-\beta)(1-\beta^2-2\alpha\beta-3\alpha^2)]^{-1} \end{aligned}$$

Such that the kurtosis

$$\begin{aligned} K(r_t) &= \frac{E(r_t^4)}{E(r_t^2)^2} \\ &= \frac{3[(1+\alpha+\beta)(1-\alpha-\beta)]}{1-\beta^2-2\alpha\beta-3\alpha^2} > 3 \end{aligned}$$

which is larger than 3.

ARMA representation

Consider the GARCH(1,1) model. Let $v_t = r_t^2 - \sigma_t^2$, such that $\sigma_t^2 = r_t^2 - v_t$. We can write

$$
\begin{aligned}
\sigma_t^2 &= \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2 \\
r_t^2 - v_t &= \omega + \alpha r_{t-1}^2 + \beta (r_{t-1}^2 - v_{t-1}) \\
r_t^2 &= \omega + (\alpha + \beta) r_{t-1}^2 - \beta v_{t-1} + v_t,
\end{aligned}
$$

where we know that $\mathbb{E}[v_t] = 0$ for all t and $\mathbb{E}[v_t^2] = const$, i.e. $\{v_t\}$ is a martingale difference sequence with constant variance. So the GARCH(1, 1) model has an ARMA(1, 1) representation for r_t^2.

It can be shown that GARCH(p,q) has an ARMA(max(p,q),p) representation for r_t^2.

ARCH(∞) representation

The GARCH(1,1) model has an ARCH(∞) representation. To see this write:

$$
\begin{aligned}
\sigma_t^2 &= \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2 \\
\sigma_t^2 (1 - \beta L) &= \omega + \alpha r_{t-1}^2 \\
\sigma_t^2 &= \frac{\omega}{1-\beta} + \alpha \sum_{s=0}^{\infty} \beta^s r_{t-s-1}^2 \\
\sigma_t^2 &= \omega^* + \sum_{s=0}^{\infty} \phi_s r_{t-s-1}^2,
\end{aligned}
$$

where $\omega^* = \frac{\omega}{1-\beta}$ and $\phi_s = \alpha\beta^s$.

Stationarity of GARCH(1,1)

Assume that r_t follows the GARCH(1,1) process:

$$
\begin{aligned}
r_t &= \sigma_t z_t \\
\sigma_t^2 &= \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2,
\end{aligned}
$$

We have that $\{r_t\}$ is a Martingale difference sequence: $\mathbb{E}[r_t|\mathcal{F}_{t-1}] = 0$ for all t, where $\mathcal{F}_{t-1}$ represents the whole history of the process up to time $t-1$.

Since $\mathbb{E}[r_t|\mathcal{F}_{t-1}] = 0$ for all t, we also have that $\mathbb{E}[r_t] = 0$ for all t, such that $Var(r_t) = \mathbb{E}[r_t^2] = \mathbb{E}[\sigma_t^2]$. To have weak stationarity we need $\mathbb{E}[\sigma_t^2] = \sigma^2 < \infty$ for all t. We have that

$$
\mathbb{E}[\sigma_t^2] = \frac{\omega}{1 - \alpha - \beta},
$$

such that $\alpha + \beta < 1$ is required for weak stationarity of r_t (proof later).

Stationarity of GARCH(1,1)

Many people believe that since GARCH(1,1) has an ARMA(1,1) representation for r_t^2 the condition $\alpha + \beta < 1$ is also necessary to have strong stationarity of r_t.
This is wrong!

The reason is that GARCH processes are thick tailed (we saw evidence for this looking at the kurtosis coefficient) such that the conditions for weak stationarity are often more stringent than the conditions for strict stationarity.

A GARCH(1,1) model can be written as:

$$\sigma_t^2 = \omega \left[1 + \sum_{k=1}^{\infty} \prod_{i=1}^{k} (\beta + \alpha z_{t-i}^2) \right]$$

Note that we wrote σ_t^2 only as a function of the z_t variables. Thus, we can now formally derive

$$\mathbb{E}[\sigma_t^2] = \omega \left[1 + \sum_{k=1}^{\infty} \mathbb{E} \left[\prod_{i=1}^{k} (\beta + \alpha z_{t-i}^2) \right] \right] = \frac{\omega}{1 - \alpha - \beta}$$

Stationarity of GARCH(1,1)

Indeed, by substituting r_{t-1}^2 with its definition $r_{t-1}^2 = \sigma_{t-1}^2 z_{t-1}^2$ we obtain:

$$
\begin{aligned}
\sigma_t^2 &= \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2 \\
&= \omega + \alpha \sigma_{t-1}^2 z_{t-1}^2 + \beta \sigma_{t-1}^2 \\
&= \omega + \sigma_{t-1}^2 (\alpha z_{t-1}^2 + \beta),
\end{aligned}
$$

which holds for all t, i.e. $\sigma_{t-s}^2 = \omega + \sigma_{t-s-1}^2 (\alpha z_{t-s-1}^2 + \beta)$. By repeated substitutions we obtain

$$
\sigma_t^2 = \omega \left[1 + \sum_{k=1}^{\infty} \prod_{i=1}^{k} (\beta + \alpha z_{t-i}^2) \right].
$$

Stationarity of GARCH(1,1)

Nelson (1990) shows that when $\omega > 0$ we have $\sigma_t^2 < \infty$ almost surely and the joint process $\{r_t, \sigma_t^2\}$ is strictly stationary if and only if $\mathbb{E}[\ln(\beta + \alpha z_t^2)] < 0$. Note that, by Jensen inequality we have

$$\mathbb{E}[\ln(\beta + \alpha z_t^2)] < \ln(\mathbb{E}[(\beta + \alpha z_t^2)]) = \ln(\alpha + \beta).$$

Exploiting this, we have the (at first glance counterintuitive) result that when $\alpha + \beta = 1$ the model is strictly stationary but not weakly stationary.

Indeed, the condition $\mathbb{E}[\ln(\beta + \alpha z_t^2)]$ is weaker than $\alpha + \beta < 1$.

As a byproduct we have that the ARCH(1) model with $\alpha = 1$ is also strictly stationary but not weekly stationary.

Stationarity of GARCH(1,1): about $\mathbb{E}[\ln(\beta + \alpha z_t^2)]$

So far we have assumed that the process goes back to the infinite past. In reality this is not true such that there is always a starting point $0 < \sigma_0^2 < \infty$. Let's define:

- σ_t^2 as the process initialized at time $t = 0$ at σ_0^2, and
- ${}_u\sigma_t^2$ as the process that goes back to the infinite past (u for unconditional).

It is easy to show that

$$\sigma_t^2 = \sigma_0^2 \prod_{i=1}^{t} (\beta + \alpha z_{t-i}^2) + \omega \left[1 + \sum_{k=1}^{t-1} \prod_{i=1}^{k} (\beta + \alpha z_{t-i}^2) \right]$$

$${}_u\sigma_t^2 = \omega \left[1 + \sum_{k=1}^{\infty} \prod_{i=1}^{k} (\beta + \alpha z_{t-i}^2) \right]$$

Stationarity of GARCH(1,1): about $\mathbb{E}[\ln(\beta + \alpha z_t^2)]$

Note that the following theorems assume that $\mathbb{E}[\ln(\beta + \alpha z_t^2)]$ exists.

Theorem (1 of Nelson 1990)

If $\omega = 0$

(a) ${}_u\sigma_t^2 = 0$ *almost surely (a.s.) for all* t.

(b) $\sigma_t^2 \to \infty$ *a.s. if and only if (iff)* $\mathbb{E}[\ln(\beta + \alpha z_t^2)] > 0$.

(c) $\sigma_t^2 \to 0$ *a.s. iff* $\mathbb{E}[\ln(\beta + \alpha z_t^2)] < 0$.

(d) *If* $\mathbb{E}[\ln(\beta + \alpha z_t^2)] = 0$, $\log(\sigma_t^2)$ *is a driftless random walk (RW) such that* $\limsup \ln(\sigma_t^2) = +\infty$ *and* $\liminf \ln(\sigma_t^2) = -\infty$

Proof.

Point (a) is straightforward. Points (b)-(c) follows by computing the log and realizing that $\ln(\sigma_t^2)$ is a RW with drift $\mathbb{E}[\ln(\beta + \alpha z_t^2)]$ and by direct application of the strong law of large numbers for iid random variables. Point (d) is proven by an application of Corollary 6.1.1 and Theorem 6.1.4 of Stout (1974). □

Stationarity of GARCH(1,1): about $\mathbb{E}[\ln(\beta + \alpha z_t^2)]$

Theorem (2 of Nelson 1990)

If $\omega > 0$ *and* $\mathbb{E}[\ln(\beta + \alpha z_t^2)] \geq 0$

(a) $\sigma_t^2 \to \infty$ *a.s.*

(b) ${}_u\sigma_t^2 \to \infty$ *a.s. for all* t.

If $\omega > 0$ *and* $\mathbb{E}[\ln(\beta + \alpha z_t^2)] < 0$

(c) $\frac{\omega}{1-\beta} \leq {}_u\sigma_t^2 < \infty$ *for all* t *a.s.*

(d) ${}_u\sigma_t^2 - \sigma_t^2 \to 0$ *a.s.*

We need to be in this situation to ensure good results.

Proof.

To prove point (a) note that:

$$\sigma_t^2 = \sigma_0^2 \prod_{i=1}^{t}(\beta + \alpha z_{t-1}^2) + \omega \left[1 + \sum_{k=1}^{t-1} \prod_{i=1}^{k}(\beta + \alpha z_{t-i}^2)\right]$$

$$\geq \omega \sup_{1 \leq k \leq t-1} \prod_{i=1}^{k}(\beta + \alpha z_{t-i}^2)$$

such that

$$\log \sigma_t^2 \geq \log \omega + \sup_{1 \leq k \leq t-1} \sum_{i=1}^{k} \log(\beta + \alpha z_{t-i}^2) \to \infty,$$

because from Theorem 1 we have:

$$\sum_{i=1}^{k} \log(\beta + \alpha z_{t-i}^2) \to \infty \quad \textit{iff} \quad \mathbb{E}[\ln(\beta + \alpha z_t^2)] > 0$$

$$\lim_{t \to \infty} \sup_{1 \leq k \leq t-1} \sum_{i=1}^{k} \log(\beta + \alpha z_{t-i}^2) = \infty \quad \textit{iff} \quad \mathbb{E}[\ln(\beta + \alpha z_t^2)] = 0$$

Proof.

To prove point (b) we use similar arguments to:

$$
\begin{aligned}
{}_u\sigma_t^2 &= \omega\left[1 + \sum_{k=1}^{\infty}\prod_{i=1}^{k}(\beta + \alpha z_{t-i}^2)\right] \\
&\geq \omega \sup_{1\leq k\leq\infty}\prod_{i=1}^{k}(\beta + \alpha z_{t-i}^2)
\end{aligned}
$$

□

Proof.

To prove the lower bound of point (c) we note that:

$$ {}_u\sigma_t^2 = \omega \left[1 + \sum_{k=1}^{\infty} \prod_{i=1}^{k} (\beta + \alpha z_{t-i}^2) \right] $$

and its infimum is achieved for $z_{t-1}^2 = 0$ for all t, such that

$$ \inf {}_u\sigma_t^2 = \omega \left[1 + \sum_{k=1}^{\infty} \prod_{i=1}^{k} \beta \right] = \omega \left[1 + \sum_{k=1}^{\infty} \beta^k \right] = \omega \sum_{k=0}^{\infty} \beta^k = \frac{\omega}{1-\beta}, $$

since $0 < \beta < 1$. The proof of the upper bound is more intricate and relies on the fact that each element $\prod_{i=1}^{k} (\beta + \alpha z_{t-i}^2)$ is $O(\exp(-\lambda k))$ for some $\lambda > 0$. □

Proof.

To prove point (d) we note that

$$\sigma_t^2 - {}_u\sigma_t^2 = \sigma_0^2 \prod_{i=1}^{t} (\beta + \alpha z_{t-i}^2) - \omega \left[1 + \sum_{k=t}^{\infty} \prod_{i=1}^{k} (\beta + \alpha z_{t-i}^2) \right]$$

the first term goes to zero according to Theorem 1. It can be shown that the second term also goes to zero because each element $\prod_{i=1}^{k}(\beta + \alpha z_{t-i}^2)$ is $O(\exp(-\lambda k))$ for some $\lambda > 0$. □

Forecasting Volatility: GARCH(1,1)

Forecasting with a GARCH(1,1) (Engle and Bollerslev 1986).
For the GARCH(1,1) we have:

$$\sigma_t^2 = \omega + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

and

$$\begin{aligned} E_t[\sigma_{t+k}^2] &= \sum_{i=0}^{k-2} [(\alpha_1 + \beta_1)^i \omega] + (\alpha_1 + \beta_1)^{k-1} \sigma_{t+1}^2 \\ &= \sigma^2 [1 - (\alpha_1 + \beta_1)^{k-1}] + (\alpha_1 + \beta_1)^{k-1} \sigma_{t+1}^2 \\ &= \sigma^2 + (\alpha_1 + \beta_1)^{k-1} [\sigma_{t+1}^2 - \sigma^2], \end{aligned}$$

where $\sigma^2 = \frac{\omega}{1-\alpha-\beta}$. When $\alpha_1 + \beta_1 < 1$, i.e. the process is stationary, $E_t[\sigma_{t+k}^2] \to \sigma^2$ as $k \to \infty$.

Forecasting Volatility

For the GARCH(p,q) we have:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i r_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$

we can write the process in two parts, before and after time t:

$$\sigma_{t+k}^2 = \omega + \sum_{i=1}^{n} [\alpha_i r_{t+k-i}^2 + \beta_i \sigma_{t+k-i}^2] + \sum_{i=k}^{m} [\alpha_i r_{t+k-i}^2 + \beta_i \sigma_{t+k-i}^2]$$

where $m = max[p; q]$, and $n = min[m; k-1]$. It follows that

$$E_t[\sigma_{t+k}^2] = \omega + \sum_{i=1}^{n} [(\alpha_i + \beta_i) E_t(\sigma_{t+k-i}^2)] + \sum_{i=k}^{m} [\alpha_i r_{t+k-i}^2 + \beta_i \sigma_{t+k-i}^2]$$

Testing for ARCH effects

- A widely used test for ARCH effects is the Lagrange-Multiplier (LM) test of Engle (1982).
- Null hypothesis: $r_t|\mathcal{F}_{t-1} \sim N(0, \sigma^2)$
- Alternative hypothesis: $r_t \sim$ ARCH(q)
- The LM test exploits the following auxiliary regression

$$r_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + ... + \alpha_1 r_{t-q}^2 + \nu_t$$

  and test that $\alpha_1 = ... = \alpha_q = 0$ vs $\alpha_0 > 0, \alpha_1 > 0, ..., \alpha_q > 0$.
- The test statistic is a $T \times R^2$ type (it is an OLS regression) and the asymptotic distribution is $\chi^2(q)$.
- One-sided test proposed by Demos and Sentana (1998).

Non-stationary volatility: the IGARCH

The IGARCH(1,1) is characterized by

$$\alpha_1 + \beta_1 = 1$$

$$\begin{aligned}\sigma_t^2 &= \omega + \alpha_1 r_{t-1}^2 + (1-\alpha_1)\sigma_{t-1}^2 \\ &= \omega + \sigma_{t-1}^2 + \alpha_1(r_{t-1}^2 - \sigma_{t-1}^2)\end{aligned}$$

In this case, the conditional variance k steps in the future is:

$$E_t[\sigma_{t+k}^2] = (k-1)\omega + \sigma_{t+1}^2 \tag{1}$$

Riskmetrics (1996): $\alpha_1 = 0.06$, $\beta_1 = 0.94$ and $\omega = 0$.

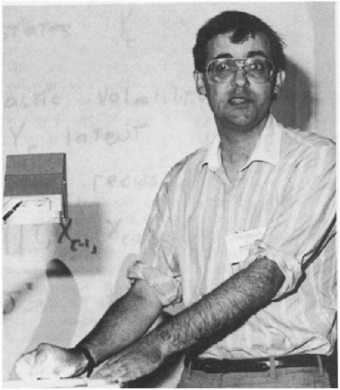
Stationarity of IGARCH(1,1)

IGARCH(1,1) with $\omega > 0$ is strictly stationary because $\mathbb{E}[\ln(1 - \alpha + \alpha z_t^2)] < 0$ (Remember the result for GARCH(1,1)).

We have the following behaviour of the IGARCH(1,1) process (Nelson, 1990):

- In the IGARCH(1,1) model with $\omega = 0$ (the "Riskmetrics approach"), σ_t^2 collapses to zero almost surely.
- In the IGARCH(1,1) model with $\omega > 0$, σ_t^2 is strictly stationary and thus does not behave like a random walk! (Remember that random walk diverge almost surely).

Nelson (1991)



Econometrica, Vol. 59, No. 2 (March, 1991), 347–370

CONDITIONAL HETEROSKEDASTICITY IN ASSET RETURNS: A NEW APPROACH

BY DANIEL B. NELSON[1]

GARCH models have been applied in modelling the relation between conditional variance and asset risk premia. These models, however, have at least three major drawbacks in asset pricing applications: (i) Researchers beginning with Black (1976) have found a negative correlation between current returns and future returns volatility. GARCH models rule this out by assumption. (ii) GARCH models impose parameter restrictions that are often violated by estimated coefficients and that may unduly restrict the dynamics of the conditional variance process. (iii) Interpreting whether shocks to conditional variance "persist" or not is difficult in GARCH models, because the usual norms measuring persistence often do not agree. A new form of ARCH is proposed that meets these objections. The method is used to estimate a model of the risk premium on the CRSP Value-Weighted Market Index from 1962 to 1987.

E-GARCH

- There exists a negative correlation between stock returns and changes in returns volatility, i.e. volatility tends to rise in response to "bad news", (excess returns lower than expected) and to fall in response to "good news" (excess returns higher than expected).

- The GARCH models are not able to explain the observed covariance between r_t^2 and r_{t-j}.

- GARCH models essentially specify the behavior of the square of the data. In this case a few large observations can dominate the sample.

E-GARCH

In the EGARCH(p,q) model (Exponential GARCH(p,q)) put forward by Nelson the σ_{t-1}^2 process depends on both size and the sign of lagged residuals. This is the first example of asymmetric model:

$$\log(\sigma_t^2) = \omega + \sum_{i=1}^{p} \beta_i \log(\sigma_{t-i}^2) + \sum_{i=1}^{q} \alpha_i [\phi z_{t-i} + \psi(|z_{t-i}| - E|z_{t-i}|)].$$

Note that $\alpha_1 \equiv 1$, $E(|z_t|) = (2/\pi)^{1/2}$ when $z_t \sim NID(0,1)$ and for any value of the parameters α_i, ω_i and β_i we have $\sigma_t > 0$. Let's define

$$g\left(z_t\right) \equiv \phi z_t + \psi\left[\left|z_t\right| - E\left|z_t\right|\right]$$

by construction $\{g\left(z_t\right)\}_{t=-\infty}^{\infty}$ is a zero-mean, i.i.d. random sequence.

The EGARCH(p,q) Model

- The components of $g\left(z_t\right)$ are ϕz_t and $\psi\left[\left|z_t\right| - E\left|z_t\right|\right]$, each with mean zero.
- If the distribution of z_t is symmetric, the components are orthogonal, but not independent.
- Over the range $0 < z_t < \infty$, $g\left(z_t\right)$ is linear in z_t with slope $\psi + \phi$, and over the range $-\infty < z_t \leq 0$, $g\left(z_t\right)$ is linear with slope $\psi - \phi$.
- The term $\psi\left[\left|z_t\right| - E\left|z_t\right|\right]$ represents a magnitude effect.

To see the different slope of $g(z_t)$ note that $z_t = sgn(z_t)|z_t|$, where sgn is the sign function[1] and

$$\begin{aligned} g(z_t) &= \phi z_t + \psi(|z_t| - E|z_t|) \\ &= \phi sgn(z_t)|z_t| + \psi|z_t| - \psi E|z_t|) \\ &= |z_t|(\phi sgn(z_t) + \psi) - \psi E|z_t| \end{aligned}$$

[1] $sgn(x) = 1$ if $x > 0$ and $sgn(x) = -1$ if $x < 0$.

The EGARCH(p,q) Model

- If $\psi > 0$ and $\phi = 0$, the innovation in $\ln\left(\sigma_{t+1}^2\right)$ is positive (negative) when the magnitude of z_t is larger (smaller) than its expected value.
- If $\psi = 0$ and $\phi < 0$, the innovation in conditional variance is now positive (negative) when returns innovations are negative (positive).
- A negative shock to the returns which would increase the debt to equity ratio and therefore increase uncertainty of future returns could be accounted for when $\alpha_i > 0$ and $\phi < 0$.

The EGARCH(p,q) Model

Nelson assumes that z_t has a GED distribution (exponential power family). The density of a GED random variable normalized is:

$$f\left(z;v\right)=\frac{v\exp\left[-\left(\frac{1}{2}\right)\left|z/\lambda\right|^{v}\right]}{\lambda 2^{(1+1/v)}\Gamma\left(1/v\right)}-\infty<z<\infty,0<v\leq\infty$$

The EGARCH(p,q) Model

where $\Gamma(\cdot)$ is the gamma function, and

$$\lambda \equiv \left[2^{(-2/\upsilon)}\Gamma(1/\upsilon)/\Gamma(3/\upsilon)\right]^{1/2}$$

υ is a tail thickness parameter.

	z's distribution
$\upsilon = 2$	standard normal distribution
$\upsilon < 2$	thicker tails than the normal
$\upsilon = 1$	double exponential distribution
$\upsilon > 2$	thinner tails than the normal
$\upsilon = \infty$	uniformly distributed on $\left[-3^{1/2}, 3^{1/2}\right]$

With this density, $E|z_t| = \dfrac{\lambda 2^{1/\upsilon}\Gamma(2/\upsilon)}{\Gamma(1/\upsilon)}$.

More on $E\left|z_t\right|$

The term $E\left|z_t\right|$ depends on the distributional assumption of z_t. Some examples are

- If $z_t \sim N(0,1)$ then $E\left|z_t\right| = \sqrt{\frac{2}{\pi}}$
- If $z_t \sim GED(0,1,\upsilon)$ then $E\left|z_t\right| = \dfrac{\lambda 2^{1/\upsilon}\Gamma\left(2/\upsilon\right)}{\Gamma\left(1/\upsilon\right)}$
- If $z_t \sim \mathcal{T}(0,1,\nu)$ then $E\left|z_t\right| = \frac{2\sqrt{\nu-2}\Gamma((\nu+1)/2)}{(\nu-1)\Gamma(\nu/2)\sqrt{\pi}}$

In general, if a closed form is not available, it can be evaluated by numerical integration

$$E\left|z_t\right| = \int_{-\infty}^{\infty} |z_t| p_z(z_t) dz_t$$

or by simulation

$$E\left|z_t\right| \approx \frac{1}{B}\sum_{i=1}^{B} |z_t^{(i)}|,$$

where $z_t^{(i)}$, $i = 1, \ldots, B$ are B independent draws from the distribution of z_t.

Volatiltiy prediction with EGARCH

Consider an EGARCH(1,1)

$$\begin{aligned} r_t &= \sigma_t z_t \\ \log \sigma_t^2 &= \omega + \beta \log \sigma_{t-1}^2 + \phi z_{t-1} + \psi(|z_{t-1}| - \mathbb{E}|z_{t-1}|) \end{aligned}$$

As in GARCH, the one step ahead prediction ($h = 1$) is deterministic and given by

$$\begin{aligned} \sigma_{t+1}^2 &= \exp\left[\omega + \beta \log \sigma_t^2 + \phi z_t + \psi(|z_t| - \mathbb{E}|z_t|)\right] \\ &= \exp(\omega)\sigma_t^{2\beta} \exp(\phi z_t + \psi(|z_t| - \mathbb{E}|z_t|)) \end{aligned}$$

Volatiltiy prediction with EGARCH

The h step ahead prediction for the log-variance $\widehat{\log \sigma}^2_{t+h|t} = \mathbb{E}[\log \sigma^2_{t+h}|\mathcal{F}_t]$ is

$$\widehat{\log \sigma}^2_{t+h|t} = \omega \sum_{i=0}^{h-2} \beta^i + \beta^{h-1} \log \sigma^2_{t+1}$$

However, note that by the Jensen's inequality

$$\mathbb{E}_t[\sigma^2_{t+h}] > \exp(\mathbb{E}[\log(\sigma^2_{t+h})]),$$

such that, if we wrongly compute the h step ahead variance prediction by taking $\exp(\widehat{\log \sigma}^2_{t+h|t})$, we will always (for $h > 1$) underestimate the future volatility!

In practice we simulate the process - especially when the distribution is different than that of the Normal.

Volatiltiy prediction with EGARCH

To compute the h step ahead prediction of the EGARCH(1,1) model we need to solve

$$\mathbb{E}_t[\sigma^2_{t+h}] = \mathbb{E}_t\left[\exp\left[\omega + \beta\log\sigma^2_{t+h-1} + \phi z_{t+h-1} + \psi(|z_{t+h-1}| - \mathbb{E}|z_{t+h-1}|)\right]\right]$$

First note that for $h \geq 2$

$$\begin{aligned}\log\sigma^2_{t+h} &= \omega + \beta\log\sigma^2_{t+h-1} + g(z_{t+h-1})\\ &= \omega\sum_{i=0}^{h-2}\beta^i + \beta^{h-1}\log\sigma^2_{t+1} + \sum_{i=0}^{h-2}\beta^i g(z_{t+h-i-1}),\end{aligned}$$

thus

$$\sigma^2_{t+h} = \exp\left(\omega\sum_{i=0}^{h-2}\beta^i\right)\sigma_{t+1}^{2\beta^{h-1}}\prod_{i=0}^{h-2}\exp(\beta^i g(z_{t+h-i-1})),$$

where $g(z_t) = \phi z_t + \psi(|z_t| - \mathbb{E}|z_t|)$.

Volatiltiy prediction with EGARCH

It follows that the variance prediction is given by:

$$\hat{\sigma}^2_{t+h|t} = \exp\left(\omega \sum_{i=0}^{h-2} \beta^i\right) \sigma_{t+1}^{2\beta^{h-1}} \mathbb{E}_t \left[\prod_{i=0}^{h-2} \exp(\beta^i g(z_{t+h-i-1}))\right],$$

however, since z_t is iid we have that

this assumption is truly required.

$$\mathbb{E}_t \left[\prod_{i=0}^{h-2} \exp(\beta^i g(z_{t+h-i-1}))\right] = \prod_{i=0}^{h-2} \mathbb{E}_t \left[\exp(\beta^i g(z_{t+h-i-1}))\right]$$

and that

Conditional expectations are equal to unconditional expectations with an iid process.

$$\mathbb{E}_t \left[\exp(\beta^i g(z_{t+h-i-1}))\right] = \mathbb{E}\left[\exp(\beta^i g(z_{t+h-i-1}))\right] = \mathbb{E}\left[\exp(\beta^i g(z_t))\right]$$

Volatiltiy prediction with EGARCH

The computation of $\mathbb{E}[\exp(bg(z_t))]$ depends on the distributional assumption of z_t. If $z_t \sim N(0,1)$ we have

$$\mathbb{E}[e^{bg(z_t)}] = e^{-b\psi\sqrt{\frac{2}{\pi}}}\left[\Phi(b\psi + b\phi)\exp\left(\frac{b^2(\psi+\phi)^2}{2}\right) + \Phi(b\psi - b\phi)\exp\left(\frac{b^2(\psi-\phi)^2}{2}\right)\right]$$

where $\Phi(\cdot)$ is the cdf of a standard Gaussian distribution, i.e. $\Phi(z) = P(z_t \leq z)$.

If $z_t \sim GED(0,1,\upsilon)$ this expression is more intricate and can be found in the appendix of Nelson (1991).

Volatiltiy prediction with EGARCH

So, the h-step ahead variance prediction in the EGARCH(1,1) model with Gaussian shocks is

$$\begin{aligned}
\hat{\sigma}^2_{t+h|t} &= \exp\left(\omega \sum_{i=0}^{h-2} \beta^i\right) \sigma_{t+1}^{2\beta^{h-1}} \mathbb{E}_t\left[\prod_{i=0}^{h-2} \exp(\beta^i g(z_{t+h-i-1}))\right] \\
&= \exp\left[\left(\omega - \psi\sqrt{\frac{2}{\pi}}\right) \sum_{i=0}^{h-2} \beta^i\right] \sigma_{t+1}^{2\beta^{h-1}} \prod_{i=0}^{h-2}\left[\Phi(\beta^{2i}(\psi+\phi)) \exp\left(\frac{\beta^{2i}(\psi+\phi)^2}{2}\right) + \right. \\
&\qquad \left. + \Phi(\beta^{2i}(\psi-\phi)) \exp\left(\frac{\beta^{2i}(\psi-\phi)^2}{2}\right)\right]
\end{aligned}$$

GJR model

The Glosten - Jagannathan - Runkle model (1993):

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{i=1}^{q} \left(\alpha_i r_{t-i}^2 + \gamma_i S_{t-i}^- r_{t-i}^2 \right)$$

gamma is usually positive.

where

$$S_t^- = \begin{cases} 1 & if \quad r_t < 0 \\ 0 & if \quad r_t \geq 0 \end{cases}$$

Pretty much GARCH with an indicator-fucntion.

If $p = q = 1$, we have

$$\sigma_t^2 = \omega + r_{t-1}^2 (\alpha + \gamma S_{t-1}^-) + \beta \sigma_{t-1}^2$$

The unconditional variance is

$$\mathbb{E}[\sigma_t^2] = \frac{\omega}{1 - \alpha - \kappa\gamma - \beta},$$

where $\kappa = P(z_t < 0)$ such that $\kappa = \frac{1}{2}$ if z_t is symmetrically distributed. Necessary conditions for covariance stationarity are $\omega > 0$ and $0 < \alpha + \kappa\gamma + \beta < 1$

ARCH-in-mean

The ARCH-in-mean model of Engle et al. (1987) was designed to capture the so-called volatility risk premium. It assumes that the volatility enters linearly the conditional mean of the returns. The model is formulated as:

$$r_t = \delta\phi(\sigma_t) + \varepsilon_t, \qquad \varepsilon_t = \sigma_t z_t$$
$$\sigma_t^2 = \text{some ARCH specification}$$

The function $\phi(\cdot)$ is usually chosen between

$$\phi(\sigma_t) = \sigma_t$$
$$\phi(\sigma_t) = \sigma_t^2$$
$$\phi(\sigma_t) = \log \sigma_t$$

The News Impact Curve

- The news have asymmetric effects on volatility.
- In the asymmetric volatility models good news and bad news have different predictability for future volatility.
- The news impact curve characterizes the impact of past return shocks on the return volatility which is implicit in a volatility model. It has been introduced by Engle and Ng (1993).
- Holding constant the information dated $t-2$ and earlier, we can examine the implied relation between r_{t-1} and σ_t^2, with $\sigma_{t-i}^2 = \sigma^2 \;\; i = 1, \dots, p$.
- This impact curve relates past return shocks (news) to current volatility.
- This curve measures how new information is incorporated into volatility estimates.

For the GARCH model the News Impact Curve (NIC) is centered on $r_{t-1} = 0$.
GARCH(1,1):

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

The news impact curve has the following expression:

$$\sigma_t^2 = A + \alpha r_{t-1}^2$$

$$A \equiv \omega + \beta \sigma^2$$

In the case of EGARCH model the curve has its minimum at $r_{t-1} = 0$ and is exponentially increasing in both directions but with different paramters.
EGARCH(1,1):

$$\ln(\sigma_t^2) = \omega + \beta \ln\left(\sigma_{t-1}^2\right) + \phi z_{t-1} + \psi\left(|z_{t-1}| - E\,|z_{t-1}|\right)$$

where $z_t = r_t/\sigma_t$. The news impact curve is

$$\sigma_t^2 = \begin{cases} A \exp\left[\dfrac{\phi+\psi}{\sigma} r_{t-1}\right] & \textit{for} \quad r_{t-1} > 0 \\ A \exp\left[\dfrac{\phi-\psi}{\sigma} r_{t-1}\right] & \textit{for} \quad r_{t-1} < 0 \end{cases}$$

$$\begin{aligned} A &\equiv \sigma^{2\beta} \exp\left[\omega - \psi E\,|z_{t-1}|\right] \\ \phi &< 0 \qquad \psi + \phi > 0 \end{aligned}$$

- The EGARCH allows good news and bad news to have different impact on volatility, while the standard GARCH does not.
- The EGARCH model allows big news to have a greater impact on volatility than GARCH model. EGARCH would have higher variances in both directions because the exponential curve eventually dominates the quadrature.

The Asymmetric GARCH(1,1) (Engle, 1990)

$$\sigma_t^2 = \omega + \alpha \left(r_{t-1} + \gamma\right)^2 + \beta \sigma_{t-1}^2$$

the NIC is

$$\sigma_t^2 = A + \alpha \left(r_{t-1} + \gamma\right)^2$$

$$A \equiv \omega + \beta \sigma^2$$

$$\omega > 0, 0 \leq \beta < 1, \sigma > 0, 0 \leq \alpha < 1.$$

is asymmetric and centered at $r_{t-1} = -\gamma$.

The Glosten-Jagannathan-Runkle model

$$\sigma_t^2 = \omega + \alpha r_t^2 + \beta \sigma_{t-1}^2 + \gamma S_{t-1}^- r_{t-1}^2$$

$$S_{t-1}^- = \begin{cases} 1 & if \quad r_{t-1} < 0 \\ 0 & otherwise \end{cases}$$

The NIC is

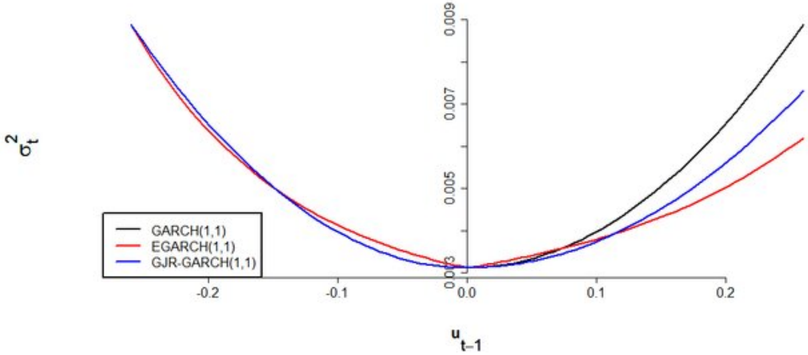
$$\sigma_t^2 = \begin{cases} A + \alpha r_{t-1}^2 & if \quad r_{t-1} > 0 \\ A + (\alpha + \gamma) r_{t-1}^2 & if \quad r_{t-1} < 0 \end{cases}$$

$$A \equiv \omega + \beta \sigma^2$$

$$\omega > 0, 0 \leq \beta < 1, \sigma > 0, 0 \leq \alpha < 1, \alpha + \beta < 1$$

is centered at $r_{t-1} = -\gamma$.

News Impact Curve



Likelihood function

Assume the model:

$$r_t = \sigma_t z_t \quad z_t \overset{iid}{\sim} p_z(z_t; \eta)$$
$$\sigma_t^2 = \sigma^2(\psi, r_{1:t-1})$$

Let $\theta = (\psi, \eta)$ be the vector of fixed parameters to be estimated. The likelihood function is

$$\begin{aligned}
\log L_T(\theta | y_{1:T}) &= \sum_{t=1}^{T} \log p_r(r_t | \mathcal{F}_{t-1}) \\
&= \sum_{t=1}^{T} \log p_z(z_t(\psi); \eta) - \frac{1}{2} \log \sigma_t^2(\psi), \\
&= \sum_{t=1}^{T} l_t(\theta)
\end{aligned}$$

r_1

this step is explained in the video.

where $l_t(\theta) = \log p_z(z_t(\psi); \eta) - \frac{1}{2} \log \sigma_t^2(\psi)$ and $z_t(\theta) = \frac{r_t}{\sigma_t(\psi)}$.

Likelihood function

Consider the Gaussian case where

$$p_z(z_t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z_t^2}{2}},$$

note that in this case $\theta = \psi$ and $\eta = \{\emptyset\}$.
The log likelihood becomes

$$l_t(\theta) \propto -\frac{z_t(\theta)^2}{2} - \frac{1}{2}\log\sigma_t^2(\theta),$$

where "$\propto$" means "proportional to" a constant $(-\frac{1}{2}\log 2\pi)$.

Likelihood function: Gradient and Hessian

The gradient (or score vector) is

$$\begin{aligned}\frac{\partial l_t(\theta)}{\partial\theta} &= -\frac{1}{2}\frac{\partial z_t(\theta)^2}{\partial\theta} - \frac{1}{2}\frac{\partial \log\sigma_t^2(\theta)}{\partial\theta}\\ &= \frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{1}{2\sigma_t^2(\theta)}\left[\frac{r_t^2}{\sigma_t^2(\theta)} - 1\right].\end{aligned}$$

The Hessian matrix is

$$\begin{aligned}\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'} &= \left[\frac{r_t^2}{\sigma_t^2(\theta)} - 1\right]\frac{\partial}{\partial\theta'}\left(\frac{1}{2}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{1}{2\sigma_t^2(\theta)}\right) - \frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{1}{2\sigma_t^2(\theta)}\frac{r_t^2}{\sigma_t^4(\theta)}\frac{\partial\sigma_t^2(\theta)}{\partial\theta'}\\ &= \left[\frac{r_t^2}{\sigma_t^2(\theta)} - 1\right]\frac{\partial}{\partial\theta'}\left(\frac{1}{2}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{1}{2\sigma_t^2(\theta)}\right) - \frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{\partial\sigma_t^2(\theta)}{\partial\theta\prime}\frac{r_t^2}{2\sigma_t^6(\theta)}\end{aligned}$$

Likelihood function: Gradient and Hessian

Note that the expected value of the score is zero:

$$
\begin{aligned}
\mathbb{E}\left[\frac{\partial l_t(\theta)}{\partial \theta}\right] &= \mathbb{E}\left[\frac{\partial \sigma_t^2(\theta)}{\partial \theta}\frac{1}{2\sigma_t^2(\theta)}\left(\frac{r_t^2}{\sigma_t^2(\theta)} - 1\right)\right] \\
&= \mathbb{E}\left[\frac{\partial \sigma_t^2(\theta)}{\partial \theta}\frac{1}{2\sigma_t^2(\theta)}\left(z_t^2 - 1\right)\right] \\
&= \mathbb{E}\left[\frac{\partial \sigma_t^2(\theta)}{\partial \theta}\frac{1}{2\sigma_t^2(\theta)}\right]\mathbb{E}\left[z_t^2 - 1\right] = 0
\end{aligned}
$$

where we exploited the fact that $r_t = \sigma_t(\theta)z_t$ (in the second line), the independence of the z_t (to split the expectation in the third line), and the fact that $\mathbb{E}\left[z_t^2\right] = 1$.

Likelihood function: Gradient and Hessian

From ML theory we know that the Fisher information matrix is related to the expected value of the Hessian matrix in this way:

$$\mathcal{I}(\theta) = -\mathbb{E}\left[\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}\right]$$

such that, for a Gaussian likelihood we have

$$\begin{aligned}
\mathcal{I}(\theta) &= -\mathbb{E}\left[\left[\frac{r_t^2}{\sigma_t^2(\theta)} - 1\right]\frac{\partial}{\partial\theta'}\left(\frac{1}{2}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{1}{2\sigma_t^2(\theta)}\right) - \frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{\partial\sigma_t^2(\theta)}{\partial\theta\prime}\frac{r_t^2}{2\sigma_t^6(\theta)}\right] \\
&= -\mathbb{E}\left[\left(z_t^2 - 1\right)\frac{\partial}{\partial\theta'}\left(\frac{1}{2}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{1}{2\sigma_t^2(\theta)}\right)\right] + \mathbb{E}\left[\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{\partial\sigma_t^2(\theta)}{\partial\theta\prime}\frac{z_t^2}{2\sigma_t^4(\theta)}\right] \\
&= \mathbb{E}\left[\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{\partial\sigma_t^2(\theta)}{\partial\theta\prime}\frac{1}{2\sigma_t^4(\theta)}\right]
\end{aligned}$$

Likelihood function: Gradient and Hessian

For a GARCH(1,1) model we have

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

Jondeau et al. (2007) propose to estimate $\mathcal{I}(\theta)$ as

$$\widehat{\mathcal{I}}(\theta) = \frac{1}{2T} \sum_{t=1}^{T} \frac{\hat{\xi}_t' \hat{\xi}_t}{\hat{\sigma}_t^4},$$

where $\hat{\xi}_t = (1, r_{t-1}^2, \sigma_{t-1}^2)'$.[2] However, it is easily seen that

$$\begin{aligned}
\frac{\partial \sigma_t^2(\theta)}{\partial \omega} &= 1 + \beta \frac{\partial \sigma_{t-1}^2(\theta)}{\partial \omega} \\
\frac{\partial \sigma_t^2(\theta)}{\partial \alpha} &= r_{t-1}^2 + \beta \frac{\partial \sigma_{t-1}^2(\theta)}{\partial \alpha} \\
\frac{\partial \sigma_t^2(\theta)}{\partial \beta} &= \sigma_{t-1}^2(\theta) + \beta \frac{\partial \sigma_{t-1}^2(\theta)}{\partial \beta}
\end{aligned}$$

[2] Note a typo in their derivations at page 90-91. Instead of σ_t^4 they write σ_t^2 which is wrong.

Regression GARCH

For models of the type

$$r_t = X_t'\delta + \varepsilon_t, \qquad \varepsilon_t = \sigma_t z_t$$
$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2$$

it can be shown that

$$\mathbb{E}\left[\frac{\partial^2 l_t(\theta;\delta)}{\partial\theta\partial\delta'}\right] = 0,$$

where $\theta = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)'$. This result implies that the Fisher information matrix is block diagonal and that the two set of parameters δ and θ can be estimated in two steps.

Two words of caution:

1) For other ARCH models (like EGARCH) this is not true. The condition need to be checked on a case by case basis.
2) For the ARCH-in-mean specification the condition is never satisfied.

Quasi-maximum likelihood estimation

Quasi-maximum likelihood estimation (QML) = the method based on the maximization of the log likelihood assuming conditional normality.

- Thus, we do as if the conditionally standardized process z_t follows a normal distribution.
- Even if the normality assumption does not hold (i.e., the true distribution is not conditionally normal), the estimator - then called quasi-maximum likelihood estimator - is consistent and asymptotically normal

Quasi-maximum likelihood estimation

Under regularity conditions, the QML estimator is asymptotically normal distributed with

$$\sqrt{T}\left(\widehat{\varphi}_n - \varphi_0^*\right) \xrightarrow{d} N\left(0, A^{-1}BA^{-1}\right)$$

The matrices A and B are, respectively, equal to:

$$A = -\frac{1}{T}E_0\left[\frac{\partial^2 \log L\left(\varphi\right)}{\partial\varphi\partial\varphi'}\right]$$

$$B = \frac{1}{T}E_0\left[\frac{\partial \log L\left(\varphi\right)}{\partial\varphi}\frac{\partial \log L\left(\varphi\right)}{\partial\varphi'}\right]$$

The matrices A and B are not, in general, equal when specification errors are present. Thus comparing estimates of the matrices A and B can be useful for detecting specification errors.

What we have learned in this lecture?

1) Several ARCH specifications exist which are able to represent two of the main features of financial returns: i) volatility clustering, and ii) excess of kurtosis in the unconditional distribution ($\kappa > 3$).
2) The methodology to study the properties of an ARCH model: i) strong stationarity, ii) weak stationarity, iii) ARMA representations, and iv) how to make predictions.
3) The effect of the initialization of the volatility in the properties of the GARCH and IGARCH models.
4) The ARCH, GARCH, IGARCH, EGARCH, GJRGARCH models, and the ARCH-in-mean specification.
5) The asymmetric reaction of volatility to past positive/negative returns (leverage effect), and the News Impact Curve.
6) Model estimation via the Maximum Likelihood Estimator.

Which tools we used from math/stat/econometrics?

1) Probability concepts:
   i) conditional/unconditional/joint distributions,
   ii) taking the expectation of a random variable,
   iii) moments of a random variable,
   iv) independence of random variables,
   v) transformation of a random variable,
   vi) strong/weak stationarity,
   vii) the strong law of large numbers for iid random variables.

2) Econometrics tools:
   i) Maximum Likelihood Estimation,
   ii) OLS,
   iii) hypothesis testing via the Lagrange Multiplier test,
   iv) linear models: Random Walks and ARMA.

3) Math tools:
   i) convergence of geometric series,
   ii) computing derivatives: gradient and Hessian of a function,
   iii) maximization of a function,
   iv) Jensen's inequality.

References I

Engle, R. F., Lilien, D. M., and Robins, R. P. (1987). Estimating time varying risk premia in the term structure: The arch-m model. *Econometrica: journal of the Econometric Society*, pages 391–407.

Engle, R. F. and Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The journal of finance*, 48(5):1749–1778.

Jondeau, E., Poon, S.-H., and Rockinger, M. (2007). *Financial modeling under non-Gaussian distributions*. Springer Science & Business Media.

Nelson, D. B. (1990). Stationarity and persistence in the garch (1, 1) model. *Econometric theory*, 6(3):318–334.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, pages 347–370.

Stout, W. F. (1974). Almost sure convergence. Technical report.