Financial Data Analysis

GARCH Models, Part I: ARCH

Summer 2014

July 15, 2014
ARCH/GARCH Models

- Have a look at the daily DAX returns from January 2000 to October 2011.
SACF of DAX returns

SACF of squared DAX returns
Autocorrelations of the DAX

- Raw returns show no or only little autocorrelation.

- Volatility appears to be autocorrelated (volatility clusters), as indicated by significant and slowly decaying autocorrelations of absolute and squared returns.

- We might ask: Is there any chance that a linear time series model can reproduce such a pattern?
Autocorrelations of returns

- Consider the case of a *Gaussian linear process*.\(^1\)

- This case is particularly tractable, since we know that all the marginal and joint distributions are likewise normal.

- That is, the joint distribution of \(y_t\) and \(y_{t-\tau}\) (assuming zero unconditional mean for simplicity) is

\[
\begin{pmatrix}
  y_t \\
  y_{t-\tau}
\end{pmatrix}
\sim
\mathcal{N}
\left(\begin{bmatrix}
  0 \\
  0
\end{bmatrix},
\begin{bmatrix}
  \gamma(0) & \gamma(\tau) \\
  \gamma(\tau) & \gamma(0)
\end{bmatrix}\right),
\]

where \(\gamma(\tau)\) is the autocovariance at lag \(\tau\), i.e., the covariance between \(y_t\) and \(y_{t-\tau}\) for any \(t\).

Autocorrelations of returns

- The moment generating function of a bivariate zero–mean normal random vector \((x, y)’\) with covariance matrix

\[
H = \begin{bmatrix}
\sigma_1^2 & \sigma_{12} \\
\sigma_{12} & \sigma_2^2
\end{bmatrix}
\]  \hspace{1cm} (2)

is, with \(t = [t_1, t_2]’\),

\[
m(t) = \mathbb{E}(e^{t_1x + t_2y})
= \exp \left\{ \frac{1}{2} t’ H t \right\}
= \exp \left\{ \frac{1}{2} \left( t_1^2 \sigma_1^2 + 2t_1 t_2 \sigma_{12} + t_2^2 \sigma_2^2 \right) \right\}.
\]
Autocorrelations of returns

• Tedious calculations show that\(^2\)

\[
\text{E}(x^2y^2) = \frac{\partial^4 m(0)}{\partial t_1^2 \partial t_2^2} = \sigma_1^2 \sigma_2^2 + 2\sigma_{12}^2
\]

\[
= \sigma_1^2 \sigma_2^2 \left(1 + 2 \frac{\sigma_{12}^2}{\sigma_1^2 \sigma_2^2}\right)
\]

\[
= \sigma_1^2 \sigma_2^2 (1 + 2\rho^2),
\]

where \(\rho\) is the coefficient of correlation between \(x\) and \(y\).

\(^2\)An alternative and perhaps less tedious derivation of (3) is in Appendix A.
Autocorrelations of returns

- Hence, as due to normality

\[ \text{Var}(x^2) = \text{E}(x^4) - \text{E}^2(x^2) = 3\sigma_1^4 - \sigma_1^4 = 2\sigma_1^4, \quad \text{Var}(y^2) = 2\sigma_2^4, \]

we have

\[
\text{Corr}(x^2, y^2) = \frac{\text{Cov}(x^2, y^2)}{\sqrt{\text{Var}(x^2)} \sqrt{\text{Var}(y^2)}} = \frac{\text{E}(x^2 y^2) - \text{E}(x^2) \text{E}(y^2)}{\sqrt{\text{Var}(x^2)} \sqrt{\text{Var}(y^2)}}
\]

\[
= \frac{\sigma_1^2 \sigma_2^2 (1 + 2\rho^2) - \sigma_1^2 \sigma_2^2}{2\sigma_1^2 \sigma_2^2}
\]

\[
= \rho^2, \quad \text{that is,}
\]

\[
\text{Corr}(x^2, y^2) = [\text{Corr}(x, y)]^2.
\]

- A Gaussian linear process cannot reproduce the above pattern.
**Stylized Facts**

- The distribution of returns is not normal but characterized by excess kurtosis.

- Raw returns show no or only little autocorrelation.

- Volatility appears to be autocorrelated (volatility clusters), as indicated by significant and slowly decaying autocorrelations of absolute and squared returns.
The ARCH\((q)\) Process

- Engle (1982) introduced the class of autoregressive conditional heteroskedastic (ARCH) models,\(^3\) where the return \(r_t\) is specified as

\[
    r_t = \mu_t + \epsilon_t
\]

\[
    \epsilon_t = \eta_t \sigma_t, \quad \eta_t \sim \text{iid } N(0,1),
\]

\[
    \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2,
\]

\[
    \omega > 0, \quad \alpha_i \geq 0, \quad i = 1, \ldots, q,
\]

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

- \(\mu_t\) in (4) is the (conditional) mean of \(r_t\), e.g., just a constant or a low–order ARMA process.

- We are interested in the error term \(\epsilon_t\) described by (5)–(7).

---

The ARCH\((q)\) Process

• \(\sigma_t^2\) in (6) is the *conditional variance* of \(\epsilon_t\) (and hence also of \(r_t\)), given (conditional on) the information up to time \(t - 1\), denoted by \(I_{t-1}\), i.e.,

\[
I_t = \{\epsilon_t, \epsilon_{t-1}, \ldots\} = \{\epsilon_s : s \leq t\}.
\] (8)

• The process is driven by the iid *innovation sequence* \(\{\eta_t\}\), and
  
  – \(\sigma_t^2\) is independent of \(\{\eta_{t+\tau} : \tau \geq 0\}\) (since \(\sigma_t^2\) is determined using only the process values up to time \(t - 1\)).
  
  – \(\epsilon_t^2\) is independent of \(\{\eta_{t+\tau} : \tau > 0\}\).

• For the above interpretation of \(\sigma_t^2\), since \(\sigma_t^2\) is known (nonrandom) conditional on \(I_{t-1}\),

\[
\text{Var}_{t-1}(\epsilon_t) = \text{Var}(\epsilon_t | I_{t-1}) = E_{t-1}(\epsilon_t^2)
\] (9)

\[
= E_{t-1}(\eta_t^2 \sigma_t^2)
\] (10)

\[
= \sigma_t^2 E_{t-1}(\eta_t^2) = \sigma_t^2 \underbrace{E(\eta_t^2)}_{=1} = \sigma_t^2.
\] (11)
The ARCH(\(q\)) Process

- Conditions (7) make sure that \(\sigma_t^2\) will remain positive for all \(t\) (as is required for the variance).

- Alternatively, we can write (5) as

\[
\epsilon_t|I_{t-1} \sim N(0, \sigma_t^2), \tag{12}
\]

i.e., \(\epsilon_t\) is conditionally normally distributed with variance \(\sigma_t^2\).
The ARCH($q$) Process: White Noise

• Now assume that $\{\epsilon_t\}$ is covariance stationary.

• The ARCH process $\epsilon_t$ is white noise, since, for $\tau > 0$,

\[
\text{Cov}(\epsilon_t, \epsilon_{t-\tau}) = \mathbb{E}(\epsilon_t \epsilon_{t-\tau}) \\
= \mathbb{E}(\eta_t \sigma_t \eta_{t-\tau} \sigma_{t-\tau}) \\
= \mathbb{E}(\eta_t) \mathbb{E}(\sigma_t \eta_{t-\tau} \sigma_{t-\tau}) \\
= 0,
\]

where the third equality follows from independence of $\sigma_t^2$ and $\eta_t$.

• The sign of $\epsilon_t = \eta_t \sigma_t$ is the sign of $\eta_t$, which is not predictable.

• The ARCH process is not independent (strong) white noise, however, since $\{\epsilon_t\}$ displays dependencies in higher moments.
Properties of ARCH processes: Covariance stationarity and unconditional variance

• Let us find the unconditional variance of $\epsilon_t$, as described by (5)–(7).\footnote{Note that this is not the unconditional variance of $r_t$ when the conditional mean of $r_t$ is not constant, e.g., a low–order ARMA process.}

• Taking (unconditional) expectations in (6), assuming stationarity,

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

$$= \omega + \sum_{i=1}^{q} \alpha_i E(\eta^2_{t-i} \sigma^2_{t-i})$$

$$= \omega + \sum_{i=1}^{q} \alpha_i E(\eta^2_{t-i}) E(\sigma^2_{t-i})$$

$$= \omega + \sum_{i=1}^{q} \alpha_i E(\sigma^2_{t-i}).$$
The ARCH($q$) Process

• Hence, assuming weak stationarity,

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

• Since the variance must be positive, this makes sense only if

$$\sum_{i=1}^{q} \alpha_i < 1,$$  \hspace{1cm} (13)

which turns out to be the condition for the finiteness of the variance (covariance stationarity) in the ARCH($q$) model.\(^5\)

\(^5\)We have already seen that the autocorrelations are all zero.
The ARCH($q$) Process

- Why even consider that $\eta_t$ may be normally distributed?

- Didn’t we observe blatant deviations from normality in asset returns, so that this *must* be an inappropriate and potentially harmful assumption?

- *A priori*, this is not clear, since ARCH is a nonlinear time series model.

- In nonlinear models, an iid normal innovation series may still result in non–normal unconditional distributions.

- That is, the one–step conditional distribution may be normal, as in (12), multi–step and *unconditional* distributions can be radically different.\(^6\)

---

\(^6\)It is clear that this does *not* imply that conditional normality is an appropriate assumption. The appropriate distribution for the innovations $\{\eta_t\}$ is an empirical issue, and we shall see that in many applications conditional normality is indeed *not* appropriate and should be replaced with a fatter tailed and/or asymmetric density.
The ARCH(\(q\)) Process

- To illustrate, consider the conditional *two-step* ahead distribution, i.e., the distribution of \(\epsilon_t\), given only the information up to time \(t - 2\).

- At time \(t - 1\), \(\sigma^2_t\) is deterministically known, and conditional normality of the one-step ahead conditional distribution follows from (5), i.e., \(\epsilon_t = \eta_t \sigma_t\).

- At time \(t - 2\), however, \(\sigma^2_t\) is a random variable, since \(\eta_{t-1}\) (and hence \(\epsilon_{t-1}\)) has not yet been realized.

- We have, considering an ARCH(1),

\[
\epsilon_t = \eta_t \sigma_t = \eta_t \sqrt{\omega + \alpha \epsilon^2_{t-1}} \quad (14)
\]

\[
= \eta_t \sqrt{\omega + \alpha \eta^2_{t-1} \sigma^2_{t-1}}, \quad (15)
\]

a nonlinear function of \(\eta_{t-1}\) and \(\eta_t\) (at time \(t - 2\)).
The ARCH(q) Process

• Going on that way,

\[ \epsilon_t = \eta_t \sqrt{\omega + \alpha \eta_{t-1}^2 \sigma_{t-1}^2} \]  
\[ = \eta_t \sqrt{\omega + \alpha \eta_{t-1}^2 (\omega + \alpha \eta_{t-2}^2 \sigma_{t-2}^2)} \]  
\[ = \text{etc.} \]

we see that \( \epsilon_t \) is a nonlinear function of the (current and past values of the) iid innovation sequence \( \{\eta_t\} \).

• Hence the \( \tau \)-step (\( \tau \geq 2 \)) and unconditional distributions are not normal.

• To illustrate, consider the kurtosis (standardized fourth moment) of the unconditional distribution.
Properties of ARCH processes: Excess kurtosis and thick tails

• In view of the stylized facts, the marginal kurtosis of $\epsilon_t$ described by (5)–(7) is of interest.

• Assuming $\{\epsilon_t\}$ is stationary with finite fourth moment,

$$\text{kurtosis}(\epsilon_t) = \frac{\text{E}(\epsilon_t^4)}{\text{E}^2(\epsilon_t^2)}$$

$$= \frac{\text{E}(\eta_t^4 \sigma_t^4)}{\text{E}^2(\eta_t^2 \sigma_t^2)}$$

$$= \frac{\text{E}(\eta_t^4) \text{E}(\sigma_t^4)}{\text{E}^2(\eta_t^2) \text{E}^2(\sigma_t^2)}$$

$$= 3 \frac{\text{E}(\sigma_t^4)}{\text{E}^2(\sigma_t^2)}$$

$$> 3. \quad (19)$$
Properties of ARCH processes: Excess kurtosis and thick tails

- The second line follows from the definition of the process, the third line uses independence of $\eta_t$ and $\sigma_t^2$, and the last line uses the fact that, due to normality,

$$\frac{E(\eta_t^4)}{E^2(\eta_t^2)} = \text{kurtosis}(\eta_t) = 3.$$  \hfill (20)

- Moreover,

$$E(\sigma_t^4) > E^2(\sigma_t^2)$$  \hfill (21)

follows from Jensen’s inequality or simply the well-known formula\(^7\)

$$\text{Var}(X) = E(X^2) - E^2(X) > 0.$$  \hfill (22)

\(^7\)For any nondegenerate random variable $X$. 
Properties of ARCH processes: Excess kurtosis and thick tails

• An interpretation of (19) 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/thick tails of the marginal distribution, respectively.
Properties of ARCH processes: Excess kurtosis and thick tails

• The kurtosis of $\epsilon_t$ is

$$kurtosis(\epsilon_t) = kurtosis(\eta_t) \times \left[ 1 + \frac{\text{Var}(\sigma_t^2)}{\text{E}^2(\sigma_t^2)} \right].$$

(23)

• Hence, even with conditional normality, conditional heteroskedasticity (e.g., ARCH effects) captures at least part of the observed excess kurtosis in financial return series.
For further discussion, let us consider the simplest case, i.e., the Gaussian ARCH(1) process, given by

\[
\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 = \omega + (\alpha_1 \eta_{t-1}^2) \sigma_{t-1}^2 \quad \eta_t \stackrel{iid}{\sim} N(0, 1).
\]  
(24)

The covariance stationarity condition is just

\[
\alpha_1 < 1,
\]  
(25)

and then

\[
E(\epsilon_t^2) = E(\sigma_t^2) = \frac{\omega}{1 - \alpha_1}.
\]  
(26)
The ARCH(1) process

- For the unconditional kurtosis, square both sides of (24),

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

\[
= \omega^2 + 2\omega \alpha_1 \eta_{t-1}^2 \sigma_{t-1}^2 + \alpha_1^2 \eta_{t-1}^4 \sigma_{t-1}^4.
\]

- By standard normality of \(\eta_t\), \(E(\eta_t^4) = 3\), and so (recalling independence of \(\sigma_t^2\) and \(\eta_t\)),

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

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

\Rightarrow 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)}. \]
The ARCH(1) process

- The last equation makes sense only if $3\alpha^2 < 1$, which is the condition for the finiteness of the fourth moment.

- In this case,

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

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

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

- The excess kurtosis result in the last equation is in accordance with (19).

- Clearly finiteness of the fourth moment implies finiteness of the second moment, and so the condition for the latter, $\alpha_1 < 1$, is weaker than for the former, $\alpha_1 < \sqrt{1/3} \approx 0.5774$.

- Hence, in sharp contrast to a linear time series model driven by Gaussian innovations, the ARCH process may be covariance stationary but the fourth and higher-order moments may not exist finite.
The ARCH(1) process

- To gain further insight into the unconditional distribution of the ARCH process, let $m \in \mathbb{N}$. Then

$$
\sigma_t^{2m} = (\omega + \alpha_1 \eta_{t-1}^2 \sigma_{t-1}^2)^m = \sum_{i=0}^{m} \binom{m}{i} \omega^{m-i} \alpha_1^i \eta_{t-1}^{2i} \sigma_{t-1}^{2i}.
$$

(27)

- Assuming the $m$th unconditional moment of $\sigma_t^2$ (and hence the $(2m)$th moment of $\epsilon_t$) is finite, and taking (unconditional) expectations,

$$
E(\sigma_t^{2m}) = \sum_{i=0}^{m} \binom{m}{i} \omega^{m-i} \alpha_1^i E(\eta_t^{2i}) E(\sigma_t^{2i}).
$$

(28)
The ARCH(1) process

- Solving for $E(\sigma^2_t)^8$

$$E(\sigma^2_t) = \frac{1}{1 - E(\eta^2_t) \alpha^m_1} \sum_{i=0}^{m-1} \binom{m}{i} \omega^{m-i} \alpha^i_1 E(\eta^2_i) E(\sigma^2_i). \quad (29)$$

- Since all the terms in the sum are positive, this can be the case only if the product term in front of the sum is likewise positive, i.e.,

$$E(\eta^2_t) \alpha^m_1 < 1. \quad (30)$$

- (30) turns out to be the condition for the existence (finiteness) of the $(2m)$th moment of the ARCH(1) process $\{\epsilon_t\}$.

8The $(2m)$th moment of $\epsilon_t$ would then be $E(\epsilon^2_t) = E(\eta^2_t \sigma^2_t) = E(\eta^2_t) E(\sigma^2_t)$. 

28
The ARCH(1) process

- To make further use of (30), consider the case $\eta_t \overset{iid}{\sim} N(0, 1)$.

- We first need an expression for the even moments of $\eta_t$ (odd moments are zero by symmetry around 0).

- It is easily checked that

$$\phi'(x) = -x\phi(x), \quad \text{where} \quad \phi(x) = (2\pi)^{-1/2}e^{-x^2/2}, \quad (31)$$

and integration by parts leads to ($m \in \mathbb{N}$)

$$E(\eta_t^{2m}) = \int_{-\infty}^{\infty} \eta^{2m} \phi(\eta) d\eta = \int_{-\infty}^{\infty} \eta^{2m-1}(\eta \phi(\eta)) d\eta$$

$$= -\eta^{2m-1} \phi(\eta) \bigg|_{-\infty}^{\infty} + (2m - 1) \int_{-\infty}^{\infty} \eta^{2(m-1)} \phi(\eta) d\eta.$$ (The first term in the last line is zero since $\lim_{\eta \to \infty} \eta^{2m-1} \phi(\eta) = \lim_{\eta \to \infty} \eta^{2m-1} \exp(-\eta^2/2)/\sqrt{2\pi} = 0.$)
The ARCH(1) process

• Thus, we have the recursion

\[ E(\eta_t^{2m}) = (2m - 1) E(\eta_t^{2(m-1)}), \quad (32) \]

and with starting value

\[ E(\eta_t^2) = 1, \quad (33) \]

it follows

\[ E(\eta_t^{2m}) = \prod_{i=1}^{m} (2i - 1), \quad (34) \]

e.g., \( E(\eta_t^4) = 3, E(\eta_t^6) = 5 \cdot 3 = 15. \)

• Condition (30), for \( \eta_t \overset{iid}{\sim} N(0, 1), \) becomes

\[ \alpha_1^m \prod_{i=1}^{m} (2i - 1) < 1. \quad (35) \]
The ARCH(1) process

• Note that
\[ \alpha_1^m \prod_{i=1}^{m} (2i - 1) = \prod_{i=1}^{m} \alpha_1(2i - 1). \quad (36) \]

• As long as \( \alpha_1 > 0 \), product terms \( \alpha_1(2i - 1) \) are increasing in \( i \) and will eventually become larger than 1.

• Therefore, (36) cannot be below unity for all \( m \), i.e., the marginal distribution of the ARCH process cannot have finite moments of all orders, even with rather light–tailed (such as normal) innovations.

\[ ^9 \text{As long as } \alpha_1 > 0. \text{ Otherwise } (\alpha_1 = 0) \text{ we have an iid process with conditional homoskedasticity.} \]
The ARCH(1) process

• It follows that the unconditional distribution must have tails which decay slower than exponential (even with conditional normality).

• It turns out that the marginal distribution of the ARCH process has polynomial (power) tails of the form

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

for some \( \gamma > 0 \), the tail index.\(^\text{10}\)

• The tail index \( \gamma \) is given by the unique positive solution of the equation

\[ h(\gamma) = E(c_T^{\gamma/2}) = E[(\alpha_1 \eta_t^2)^{\gamma/2}] = 1. \quad (37) \]

• Moments of \( \epsilon_t \) exist only for orders smaller than \( \gamma \).

The ARCH(1) process: Weak and strict stationarity

- A consequence of the polynomial tails of the ARCH process is that the covariance stationary condition is actually stronger than the condition for strict stationarity.

- E.g., for the ARCH(1) process, the condition for strict stationarity turns out to be

\[
E[\log(\alpha_1 \eta_t^2)] = \int_{-\infty}^{\infty} \log(\alpha_1 \eta_t^2) \frac{e^{-\eta_t^2/2}}{\sqrt{2\pi}} d\eta_t < 0. \tag{38}
\]

- By Jensen’s inequality,

\[
E[\log(\alpha_1 \eta_t^2)] < \log(E(\alpha_1 \eta_t^2)) = \log(\alpha_1 \underbrace{E(\eta_t^2)}_{=1}) = \log(\alpha_1), \tag{39}
\]

and hence covariance stationarity implies strict stationarity.
The ARCH(1) process

- If (38) holds but $\gamma \leq 2$ in (37) (i.e., $\alpha_1 \geq 1$), then the process is strictly stationary with infinite variance.

- E.g., in contrast to $|\phi_1| > 1$ in the AR(1) model $y_t = c + \phi_1 y_{t-1} + \epsilon_t$, this implies, among other things, that the ARCH(1) with $\alpha_1 > 1$ does not exhibit explosive behavior as long as (38) still holds.

- For the ARCH(1) with standard normal innovations $\eta_t$, (38) holds as long as $\alpha_1 < 3.56$.

- Processes with $\alpha_1 \geq 1$ but still satisfying (38)\(^{11}\) have very fat tails and occasionally produce rather extreme observations but do not explode.

\(^{11}\)i.e., $1 \leq \alpha_1 < 3.56$
simulated ARCH(1) process with $\omega = 0.05$, $\alpha_1 = 0.9$, $\eta_t \sim N(0,1)$
simulated ARCH(1) process with $\omega = 0.05$, $\alpha_1 = 1$, $\eta_t \sim N(0,1)$
simulated ARCH(1) process with $\omega = 0.05$, $\alpha_1 = 3$, $\eta_t \sim N(0,1)$
The ARCH(1) process

• Note that moment condition (30) as well as the condition for strict stationarity (38) depend on the conditional distribution in the ARCH model, i.e., the distribution of \( \{ \eta_t \} \).

• Normality has been assumed so far for convenience, but this assumption is not necessary and will have to be checked empirically in any application at hand.
The ARCH(1) process

- E.g., if $\eta_t$ is drawn from a *Laplace* distribution with unit variance, i.e.,

$$f(\eta) = \frac{1}{\sqrt{2}} \exp \left\{ -\sqrt{2}|\eta| \right\},$$

(40)

then straightforward integration shows

$$E(\eta_t^{2m}) = \frac{(2m)!}{2m}.$$  

(41)

- Condition (30) becomes

$$\alpha_1^m \frac{(2m)!}{2m} < 1,$$

(42)

and the normal density in (38) has to be replaced by (40).

---

12Repeated integration by parts shows $E(\eta^{2m}) = \frac{1}{\sqrt{2}} \int_{-\infty}^{\infty} \eta^{2m} e^{-\sqrt{2}|\eta|} d\eta = \sqrt{2} \int_0^{\infty} \eta^{2m} e^{-\sqrt{2}\eta} d\eta = 2m \int_0^{\infty} \eta^{2m-1} e^{-\sqrt{2}\eta} d\eta = \frac{2m(2m-1)}{2} \sqrt{2} \int_0^{\infty} \eta^{2(m-1)} e^{-\sqrt{2}\eta} d\eta = \frac{2m(2m-1)}{2} E(\eta^{2(m-1)})$. 

---
The ARCH(1) process

- E.g., conditions for covariance stationarity and a finite fourth moment in the ARCH(1) with \( \{\eta_t\} \) following (40) would be

\[
\alpha_1 < 1 \quad \text{(since the variance of } \eta_t \text{ is still unity)} \tag{43}
\]

and

\[
\alpha_1^2 \frac{4!}{2^2} = 6\alpha_1^2 < 1, \tag{44}
\]

respectively

- The fourth moment–condition (44) is stronger than \( 3\alpha_1^2 < 1 \) under normality since now the conditional distribution (distribution of \( \eta_t \)) is leptokurtic.
Laplace: \( f(x) = \frac{1}{\sqrt{2}} \exp\left\{-\sqrt{2}|x|\right\} \), Normal: \( f(x) = \exp\left\{-x^2/2\right\}/\sqrt{2\pi} \)
The ARCH(1) process: Student’s $t$ innovations

- A distribution frequently used in (G)ARCH models to account for conditional leptokurtosis is the unit–variance Student’s $t$, with density

$$f(\eta; \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\nu/2) \sqrt{\pi(\nu - 2)}} \left\{ 1 + \frac{\eta^2}{\nu - 2} \right\}^{-(\nu+1)/2}, \quad (45)$$

with degrees of freedom parameter $\nu > 2$ which is estimated along with the other parameters.

- Note that (45) differs slightly from the “standard” $t$ distribution; the version in (45) has been standardized so that $E(\eta_t^2) = 1$ irrespective of shape parameter $\nu$.\[^{13}\]

- For $\{\eta_t\}$ following a $t$ distribution as in (45), we have, e.g., $E(\eta_t^2) = 3\frac{\nu - 2}{\nu - 4}$ if $\nu > 4$, so that the fourth–moment condition in an ARCH(1) would become

$$3\frac{\nu - 2}{\nu - 4} \alpha_1^2 < 1. \quad (46)$$

\[^{13}\]This is also why we have to restrict $\nu > 2$ in this case.
Temporal properties of the ARCH process

- Stylized facts:
  (i) Returns show no or only little autocorrelation.
  (ii) Volatility appears to be autocorrelated (volatility clusters), as indicated by significant and slowly decaying autocorrelations of absolute (or squared) returns.

- We have already seen that the ARCH process is uncorrelated.

- There is no closed–from expression for the autocorrelations of the absolute ARCH process, but a characterization of the ACF of the squares of the ARCH process is relatively straightforward.

- Assume $\mathbb{E}(\epsilon_t^4) < \infty$ and define the autocorrelations of the squares,

$$
\varrho(\tau) = \text{Corr}(\epsilon_t^2, \epsilon_{t-\tau}^2) = \frac{\mathbb{E}(\epsilon_t^2 \epsilon_{t-\tau}^2) - \mathbb{E}^2(\epsilon_t^2)}{\mathbb{E}(\epsilon_t^4) - \mathbb{E}^2(\epsilon_t^2)},
$$

(47)

which for the ARCH(1) are well–defined for $3\alpha_1^2 < 1$. 


Temporal properties of the ARCH process

• Define the prediction error

\[ u_t = \epsilon_t^2 - \mathbb{E}(\epsilon_t^2 | I_{t-1}) = \epsilon_t^2 - \sigma_t^2 = (\eta_t^2 - 1)\sigma_t^2, \tag{48} \]

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

• \( \{ u_t \} \) is uncorrelated with zero mean, hence white noise\(^{14} \) (but not strict white noise).

• Substituting \( \epsilon_t^2 - u_t \) for \( \sigma_t^2 \) in the ARCH(\( q \)) equation,

\[ \epsilon_t^2 = \omega + \alpha_1\epsilon_{t-1}^2 + \alpha_2\epsilon_{t-2}^2 + \cdots + \alpha_q\epsilon_{t-q}^2 + u_t, \tag{49} \]

shows that the squared ARCH(\( q \)) process \( \{ \epsilon_t^2 \} \) has an AR(\( q \)) representation.

\(^{14}\)This follows from \( \mathbb{E}[(\eta_t^2 - 1)\sigma_t^2] = \mathbb{E}[(\eta_t^2 - 1)]\mathbb{E}(\sigma_t^2) = 0 \) and likewise \( \mathbb{E}[(\eta_t^2 - 1)(\eta_{t-\tau}^2 - 1)\sigma_t^2\sigma_{t-\tau}^2] = \mathbb{E}[(\eta_t^2 - 1)]\mathbb{E}[(\eta_{t-\tau}^2 - 1)\sigma_t^2\sigma_{t-\tau}^2] = 0. \)
Temporal properties of the ARCH process

• Thus the autocorrelations of the squared process \( \{\epsilon_t^2\} \) exhibit the same behavior as those of an AR(\( q \)) model.

• E.g., for the ARCH(1), this implies

\[
\rho(\tau) = \alpha^\tau. \tag{50}
\]

• As in the AR(1) process, the level and the decay of the ACF of the squares of the ARCH(1) are determined by just one parameter, \( \alpha_1 \).

• This may indicate that the ARCH(1) process may not be appropriate, since empirically the ACF of the squares tends to start relatively low but is very persistent (cf. the figure on the next slide).
Estimation of ARCH Models

• (G)ARCH models are typically estimated by conditional maximum likelihood.

• To illustrate, consider the simplest case, where the conditional mean is constant\(^{15}\) and the variance is driven by ARCH\((q)\).

• If we assume conditional normality, the model is

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

\(^{15}\)In case of a time–varying conditional mean \(\mu_t\) we would have \(\hat{\epsilon}_t = r_t - \hat{\mu}_t\) in (55), where \(\hat{\mu}_t\) would be a function of the parameters describing the conditional mean dynamics.
Estimation of ARCH Models

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

- We observe a stretch \( r_{-q+1}, \ldots, r_{-1}, r_0, r_1, \ldots, r_T \), where the first \( q \) observations are used as presample values for conditional maximum likelihood.

- Want to calculate the log–likelihood function for a given value \( \hat{\theta} = (\hat{c}, \hat{\omega}, \hat{\alpha}_1, \ldots, \hat{\alpha}_q) \).

- From the mean equation, we calculate the residuals

\[
\hat{\epsilon}_t = \hat{\epsilon}_t(\theta) = r_t - \hat{c}, \quad t = -q + 1, \ldots, T. \tag{55}
\]
Estimation of ARCH Models

- The conditional log–likelihood function is then given by

\[
\ell(\hat{\theta}) = \sum_{t=1}^{T} \ell_t(\hat{\theta}). \tag{56}
\]

where, under conditional normality,

\[
\ell_t(\hat{\theta}) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log \hat{\sigma}_t^2 - \frac{1}{2} \frac{\hat{\epsilon}_t^2}{\hat{\sigma}_t^2}, \quad t = 1, \ldots, T, \tag{57}
\]

and

\[
\hat{\sigma}_t^2 = \hat{\sigma}_t^2(\hat{\theta}) = \hat{\omega} + \hat{\alpha}_1 \hat{\epsilon}_{t-1}^2 + \hat{\alpha}_2 \hat{\epsilon}_{t-2}^2 + \cdots + \hat{\alpha}_q \hat{\epsilon}_{t-q}^2, \quad t = 1, \ldots, T. \tag{58}
\]
Estimation of ARCH Models

- The conditional log-likelihood (56) is then maximized numerically with respect to $\theta$ to obtain the maximum likelihood estimator (MLE) $\hat{\theta}_{ML}$.

- Under fairly weak conditions, the MLE is consistent and asymptotically normal, i.e.,
\[
\sqrt{T}(\hat{\theta}_{ML} - \theta_0) \xrightarrow{d} N(0, J(\theta_0)^{-1}),
\]
where $\theta_0$ is the true parameter value, and the Fisher information matrix
\[
J(\theta_0) = -E \left[ \frac{\partial^2 \ell_t(\theta_0)}{\partial \theta \partial \theta'} \right],
\]
which can be consistently estimated via
\[
\hat{J}(\hat{\theta}_{ML}) = -\frac{1}{T} \frac{\partial^2 \ell(\hat{\theta}_{ML})}{\partial \theta \partial \theta'} = -\frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 \ell_t(\hat{\theta}_{ML})}{\partial \theta \partial \theta'}.
\]
Estimation of ARCH Models

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

\[ \hat{\theta}_{ML}^{asy} \sim \text{Normal}(\theta_0, T^{-1}J(\hat{\theta}_{ML})^{-1}). \] (62)

- The derivatives in (61) 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.\(^\text{16}\)

Lagrange Multiplier (LM) Test for ARCH Errors$^{17}$

- Recall that an LM test in general is constructed as follows.

- Suppose that, in parametric statistical model with true parameter vector $\theta_0 \in \mathbb{R}^d$, we want to test the null hypothesis

$$H_0 : R\theta_0 = r,$$

(63)

where $R$ is a given $q \times d$ matrix of rank $q$ ($q \leq d$) and $r$ is a given $s \times 1$ vector.

- That is, we want to test $q$ linear hypotheses about the parameter vector $\theta_0$.

---

For example, if the unconstrained model is the ARCH\((q)\) model,

\[
\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \cdots + \alpha_q \epsilon_{t-q}^2,
\]

we have \(d = q + 1\), \(\theta = (\omega, \alpha_1, \ldots, \alpha_q)'\),

The null hypothesis of conditional homoskedasticity corresponds to \(\alpha_1 = \alpha_2 = \cdots = \alpha_q = 0\), i.e.,

\[
R = [0_{q \times 1}, I_q], \quad r = 0_{q \times 1},
\]

where \(I_q\) is the identity matrix of dimension \(q\).
Lagrange Multiplier (LM) Test for ARCH Errors

- Let the log–likelihood function be denoted by

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

(65)

where \( \ell_t(\theta) \) is the contribution (the conditional density) of observation \( t \), and

\[ \frac{\partial}{\partial \theta} \ell(\theta) = \sum_{t=1}^{T} \frac{\partial}{\partial \theta} \ell_t(\theta) = \begin{bmatrix} \frac{\partial \ell(\theta)}{\partial \theta_1} \\ \frac{\partial \ell(\theta)}{\partial \theta_2} \\ \vdots \\ \frac{\partial \ell(\theta)}{\partial \theta_d} \end{bmatrix}, \]  

(66)

the score vector.
Lagrange Multiplier (LM) Test for ARCH Errors

- The constrained MLE $\hat{\theta}_c$ is given by a stationary point of the Lagrangian

$$L(\theta, \lambda) = \ell(\theta) - \lambda'(R\theta - r),$$  

(67)

where $\lambda$ is the $q \times 1$ vector of Lagrange multipliers.

- Let the stationary point of (67) be $(\hat{\theta}_c, \hat{\lambda})$; this satisfies

$$\frac{\partial}{\partial \theta} \ell(\hat{\theta}_c) = R'\hat{\lambda}.$$  

(68)

- The Lagrange multipliers indicate how the value of the objective function changes when the respective constraint is weakened, i.e., they measure the severeness of a binding constraint.

- Thus, we want to test whether $\hat{\lambda}$ is statistically significantly different from zero.
Lagrange Multiplier (LM) Test for ARCH Errors

- The distributional result to accomplish this is that, under the null hypothesis (63), the LM test statistic

\[
\text{LM} = \frac{1}{T} \hat{\lambda}' R [J(\hat{\theta}_c)]^{-1} R' \hat{\lambda}
\]

\[= \frac{1}{T} \left( \frac{\partial}{\partial \theta} \ell(\hat{\theta}_c) \right)' [J(\hat{\theta}_c)]^{-1} \left( \frac{\partial}{\partial \theta} \ell(\hat{\theta}_c) \right) \xrightarrow{d} \chi^2(q), \quad (69)
\]

where \( q \) is the number of restrictions under (63), and the Fisher information matrix under the null, evaluated at the restricted MLE,

\[
J(\hat{\theta}_c) = -E \left( \frac{\partial^2 \ell_t(\hat{\theta}_c)}{\partial \theta \partial \theta'} \right). \quad (70)
\]

Lagrange Multiplier (LM) Test for ARCH Errors

- In the ARCH($q$) with $\eta_t \sim N(0, 1)$,

\[
\ell_t = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma_t^2 - \frac{1}{2} \frac{\epsilon_t^2}{\sigma_t^2}
\]  \hspace{1cm} (71)

\[
\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2
\]  \hspace{1cm} (72)

\[
\theta = (\omega, \alpha_1, \ldots, \alpha_q)'
\]  \hspace{1cm} (73)

\[
H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_q = 0.
\]  \hspace{1cm} (74)

- Note that

\[
\frac{\partial \sigma_t^2}{\partial \theta} = \begin{pmatrix} 1 \\ \epsilon_{t-1}^2 \\ \vdots \\ \epsilon_{t-q}^2 \end{pmatrix} =: z_t
\]  \hspace{1cm} (75)

- In practice, $\{\epsilon_t\}$ is not observed and replaced with the residuals $\{\hat{\epsilon}_t\}$ from an estimated mean equation (e.g., a constant, ARMA, linear regression).
Lagrange Multiplier (LM) Test for ARCH Errors

\[
\frac{\partial \ell_t}{\partial \theta} = -\frac{1}{2\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta} + \frac{1}{2\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \theta} = 1
\]

\[
\frac{1}{2\sigma_t^2} \left( \frac{\epsilon_t^2}{\sigma_t^2} - 1 \right) \begin{pmatrix}
1 \\
\epsilon_{t-1}^2 \\
\vdots \\
\epsilon_{t-q}^2
\end{pmatrix} = 1
\]

\[
\frac{1}{2\sigma_t^2} \left( \frac{\epsilon_t^2}{\sigma_t^2} - 1 \right) z_t
\]
Lagrange Multiplier (LM) Test for ARCH Errors

- Under the null hypothesis $\alpha_1 = \cdots = \alpha_q = 0$, $\sigma_t^2$ is a constant $\sigma_0^2$.

- The restricted MLE under the null hypothesis is

$$\hat{\omega}_c = \hat{\sigma}_0^2 = T^{-1} \sum_{t=1}^{T} \epsilon_t^2.$$  \hspace{1cm} (79)
Lagrange Multiplier (LM) Test for ARCH Errors

- Thus,

\[
\frac{\partial \ell(\hat{\theta}_c)}{\partial \theta} = \sum_{t=1}^{T} \frac{\partial \ell_t(\hat{\theta}_c)}{\partial \theta}
\]

\[(80)\]

\[
= \frac{1}{2\hat{\sigma}^2_0} \sum_{t=1}^{T} \left( \frac{\epsilon_t^2}{\hat{\sigma}^2_0} - 1 \right) z_t
\]

\[(81)\]

\[
= \frac{1}{2\hat{\sigma}^2_0} Z' f_0,
\]

\[(82)\]

where

\[
Z = \begin{pmatrix}
z_1 \\
\vdots \\
z_T
\end{pmatrix}, \quad f_0 = \left( \frac{\epsilon_1^2}{\hat{\sigma}^2_0} - 1 \quad \frac{\epsilon_2^2}{\hat{\sigma}^2_0} - 1 \quad \cdots \quad \frac{\epsilon_T^2}{\hat{\sigma}^2_0} - 1 \right)'.
\]

\[(83)\]
Lagrange Multiplier (LM) Test for ARCH Errors

- Note that, since the restricted MLE

\[ \hat{\omega}_c = \hat{\sigma}_0^2 = \frac{1}{T} \sum_{t=1}^{T} \epsilon_t^2, \]  

(84)

the mean of \( \{\epsilon_t^2/\hat{\sigma}_0^2\}_{t=1}^{T} \) is unity, so \( f_0 \) in (83) may be viewed as a vector of demeaned variables.
Lagrange Multiplier (LM) Test for ARCH Errors

• The (negative of the) second derivative matrix is

\[
- \frac{\partial^2 \ell_t}{\partial \theta \partial \theta'} = - \left( \frac{1}{2} \frac{1}{\sigma_t^4} \frac{\sigma_t^2}{\sigma_t^6} \right) \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta'}
\]

(85)

\[
= \frac{1}{2\sigma_t^4} \left( 2 \frac{\epsilon_t^2}{\sigma_t^2} - 1 \right) z_t z_t'
\]

(86)

\[
= \frac{1}{2\sigma_t^4} \left( 2 \frac{\eta_t^2 \sigma_t^2}{\sigma_t^2} - 1 \right) z_t z_t'
\]

(87)

\[
= \frac{1}{2\sigma_t^4} \left( 2 \eta_t^2 - 1 \right) z_t z_t'
\]

(88)

• Hence, since \( E(\eta_t^2) = 1 \), and \( \eta_t \) is independent of the variables in \( z_t \),

\[
J(\hat{\theta}_c) = -E \left( \frac{\partial^2 \ell_t(\hat{\theta}_c)}{\partial \theta \partial \theta'} \right) = \frac{1}{2\hat{\sigma}_0^4} E(z_t z_t').
\]

(89)
Lagrange Multiplier (LM) Test for ARCH Errors

- Quantity (89) can be consistently estimated via its sample analogue,

\[
\hat{J}(\hat{\theta}_c) = -\frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 \ell_t(\hat{\theta}_c)}{\partial \theta \partial \theta'} = \frac{1}{2\hat{\sigma}_0^4 T} \sum_{t=1}^{T} z_t z_t' = \frac{1}{2\hat{\sigma}_0^4 T} Z' Z,
\]

(90)

and the LM test statistic (69) becomes

\[
\text{LM} = \frac{1}{T} \left( \frac{\partial}{\partial \theta} \ell(\hat{\theta}_c) \right)' \left[ \hat{J}(\hat{\theta}_c) \right]^{-1} \left( \frac{\partial}{\partial \theta} \ell(\hat{\theta}_c) \right) \\
= \frac{1}{T} \left( \frac{1}{2\hat{\sigma}_0^2} Z' f_0 \right)' \left( \frac{1}{2\hat{\sigma}_0^4 T} Z' Z \right)^{-1} \left( \frac{1}{2\hat{\sigma}_0^2} Z' f_0 \right) \\
= \frac{1}{2} f_0' Z (Z' Z)^{-1} Z' f_0 \\
\xrightarrow{d} \chi^2(q).
\]

(91)
Lagrange Multiplier (LM) Test for ARCH Errors

- Under normality, we have $E(\eta_t^4) = 3$, and so, under the null,

\[
E \left( \frac{\varepsilon_t^2}{\sigma_0^2} - 1 \right)^2 = E \left( \frac{\sigma_0^2 \eta_t^2}{\sigma_0^2} - 1 \right)^2 = E(\eta_t^4 - 2\eta_t^2 + 1) = 3 - 2 + 1 = 2. \tag{94}
\]

- Thus

\[
\frac{1}{T} f_0' f_0 \xrightarrow{p} 2, \tag{95}
\]

and a particularly simple and intuitive asymptotically equivalent test statistic is

\[
LM^* = \frac{f_0' Z (Z' Z)^{-1} Z' f_0}{f_0' f_0 / T} = TR^2. \tag{96}
\]
Lagrange Multiplier (LM) Test for ARCH Errors

\[
\text{LM}^* = T \frac{f'_0 Z (Z' Z)^{-1} Z' f_0}{f'_0 f_0} = TR^2 \overset{d}{\to} \chi^2(q), \quad (97)
\]

where \( R^2 \) is the coefficient of determination obtained from the regression of \( \epsilon_t^2 \) on a constant and \( q \) lags,

\[
\epsilon_t^2 = b_0 + b_1 \epsilon_{t-1}^2 + b_2 \epsilon_{t-2}^2 + \cdots + b_q \epsilon_{t-q}^2 + v_t, \quad t = 1, \ldots, T. \quad (98)
\]

• (Note that, as discussed above, \( f_0 \) is in demeaned form and multiplying by a scalar does not change the \( R^2 \) of a regression.)

• The LM test for ARCH disturbances is typically calculated and reported in the \( R^2 \)-form (97).
Summary of ARCH LM-test in $R^2$–form

- ARCH($q$):

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

(99) \hspace{1cm} (100)

- Null hypothesis:

\[H_0 : \alpha_1 = \cdots = \alpha_q = 0\]  

(101)

- Under $H_0$,

\[T \cdot R^2 \overset{d}{\rightarrow} \chi^2(q),\]  

(102)

where $T$ is the sample size and $R^2$ is the coefficient of determination obtained from the regression of $\epsilon_t^2$ on a constant and $q$ lags,

\[\epsilon_t^2 = b_0 + b_1 \epsilon_{t-1}^2 + b_2 \epsilon_{t-2}^2 + \cdots + b_q \epsilon_{t-q}^2 + v_t, \quad t = 1, \ldots, T.\]  

(103)
Summary of ARCH LM-test in $R^2$–form

• In practice, the error terms are obtained as residuals from a (perhaps very simple) model for the conditional mean,

$$r_t = \mu_t + \epsilon_t.$$  \hspace{1cm} (104)

• With $\hat{\mu}_t$ being the estimated $\mu_t$ (the conditional mean at time $t$), the residual

$$\hat{\epsilon}_t = r_t - \hat{\mu}_t$$  \hspace{1cm} (105)

is used in regression (103).
Appendix A: Alternative derivation of (3)

• Assume $\sigma_1^2 = \sigma_2^2 = 1$ for simplicity. Then the joint density of $x$ and $y$ is bivariate standard normal with correlation $\rho$, i.e.,

$$f(x, y; \rho) = \frac{1}{2\pi \sqrt{1 - \rho^2}} \exp \left\{ -\frac{x^2 - 2\rho xy + y^2}{2(1 - \rho^2)} \right\}$$

$$= \frac{1}{\sqrt{2\pi}} \exp\{-x^2/2\} \frac{1}{\sqrt{2\pi \sqrt{1 - \rho^2}}} \exp \left\{ -\frac{(y - \rho x)^2}{2(1 - \rho^2)} \right\}$$

$$= f(x)f(y|x; \rho),$$

i.e.,

$$y|x \sim N(\rho x, 1 - \rho^2), \quad (106)$$

the distribution of $y$ conditional on $x$ is normal with mean $\rho x$ and variance $1 - \rho^2$. 
Appendix A: Alternative derivation of (3)

- The expectation

\[ E(x^2 y^2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^2 y^2 f(x, y; \rho) dy dx \]

\[ = \int \int x^2 y^2 f(x) f(y|x; \rho) dy dx \]

\[ = \int x^2 \left[ \int y^2 f(y|x; \rho) dy \right] f(x) dx \]

\[ = \int x^2 E(y^2|x) f(x) dx, \quad (107) \]

where

\[ E(y^2|x) = \int y^2 f(y|x; \rho) dy \quad (108) \]

is the expectation of \( y^2 \) conditional on \( x \) (note this depends on \( x \) and thus is a random variable).
Appendix A: Alternative derivation of (3)

- From (106),

\[ E(y^2|x) = E^2(y|x) + \text{Var}(y|x) = \rho^2 x^2 + 1 - \rho^2. \]  \hspace{1cm} (109)

- Finally, combining (107) and (109),

\[
E(x^2 y^2) = \int x^2 E(y^2|x) f(x) \, dx \tag{110}
\]

\[
= \int x^2 (\rho^2 x^2 + 1 - \rho^2) f(x) \, dx \tag{111}
\]

\[
= E_x(\rho^2 x^4 + x^2 - \rho^2 x^2) \tag{112}
\]

\[
= 3\rho^2 + 1 - \rho^2 \tag{113}
\]

\[
= 1 + 2\rho^2, \tag{114}
\]

since, by standard normality of \( x \), \( E(x^2) = 1, \ E(x^4) = 3 \). This is (3) with \( \sigma_1^2 = \sigma_2^2 = 1 \).