Notes on Time Series Models[1]

Antonis Demos

Athens University of Economics and Business

First version January 2007

This version January 2016

[1]These notes include material taught to MSc students at Athens University of Economics and Business since 1999. Please report all typos to: A. Demos, Athens University of Economics and Business, Dept. of International and European Economic Studies, 76 Patision Str., Athens 10434, Greece (E-mail: demos@aueb.gr, Tel: +30-210-8203451, Fax: +30-210-8214122)

Before moving to the pure time series models we shall need the notion of orthogonal projections. We need this to evaluate the partial correlation coefficients, especially for Moving Average models.

1 Projections (Orthogonal)

Assume the usual linear regression setup, i.e.

$$y = X\beta + u, \quad u|X \backsim D\left(0, \sigma^2 I_n\right)$$

where y is the $n \times 1$ vector of endogenous variables, X is the $n \times k$ matrix of weakly exogenous explanatory variables, β is the $k \times 1$ vector of mean parameters and u is the $n \times 1$ vector of errors.

When we estimate a linear regression model, we simply map the regressand y into a vector of fitted values $X\widehat{\beta}$ and a vector of residuals $\widehat{u} = y - X\widehat{\beta}$. Geometrically, these mappings are examples of orthogonal projections. A **projection** is a mapping that takes each point of E^n into a point in a subset of E^n, while leaving all the points of the subset unchanged, where E^n is the usual Euclidean vector space, i.e. the set of all vectors in R^n where the addition, the scalar multiplication and the inner product (hence the norm) are defined. Because of this invariance the subspace is called **invariant subspace** of the projection. An **orthogonal projection** maps any point into the point of the subspace that is closest to it. If a point is already in the invariant subspace, it is mapped into itself.

Algebraically, an orthogonal projection on to a given subspace can be performed by premultiplying the vector to be projected by a suitable **projection matrix**. In the case of OLS, the two projection matrices that yield the vector of fitted values and the vector of residuals, respectively, are

$$P_X = X\left(X^{\prime}X\right)^{-1}X^{\prime}$$

and

$$M_X = I_n - P_X = I_n - X\left(X^{/}X\right)^{-1}X^{/}.$$

To see this notice that the fitted values

$$\widehat{y} = X\widehat{\beta} = X\left(X^{/}X\right)^{-1}X^{/}y = P_X y.$$

Hence the P_X projection matrix project on to $S(X)$, i.e. the subspace of E^n spanned by the columns of X. Notice that for any vector $\alpha \in R^k$ the vector $X\alpha$ belongs to $S(X)$. As now $X\alpha \in S(X)$ then it should be the case, due to the invariance of P_X, that

$$P_X X\alpha = X\alpha.$$

But notice that

$$P_X X = X\left(X^{/}X\right)^{-1}X^{/}X = XI_k = X.$$

It is clear that when P_X is applied to y it yields the vector of fitted values. On the other hand the M_X projection matrix yields the vector of residuals as

$$M_X y = \left[I_n - X\left(X^{/}X\right)^{-1}X^{/}\right]y = y - P_X y = y - X\widehat{\beta} = \widehat{u}.$$

The image of M_X is $S^{\perp}(X)$, the orthogonal complement of the image of P_X. To see this, consider any vector $w \in S^{\perp}(X)$. It must satisfy the condition $X^{/}w = 0$, which implies that $P_X w = 0$, by the definition of P_X. Consequently, $(I_n - P_X)w = M_X w = w$ and $S^{\perp}(X)$ must be contained in the image of M_X, i.e. $S^{\perp}(X) \subseteq \mathrm{Im}(M_X)$. Now consider any image of M_X. It must take the form $M_X z$. But then

$$(M_X z)^{/} X = z^{/} M_X X = 0$$

as M_X symmetric. Hence $M_X z$ belongs to $S^{\perp}(X)$, for any z. Consequently, $\mathrm{Im}(M_X) \subseteq S^{\perp}(X)$ and hence the image of M_X coincides with $S^{\perp}(X)$.

For any matrix to represent a projection, it must be **idempotent**. This is because the vector image of a projection matrix is say $S(X)$, and then project it again, the second projection should have no effect, i.e. $P_X P_X z = P_X z$ for any z. It is easy to prove that this is the case with P_X and M_X, as

$$P_X P_X = P_X \quad and \quad M_X M_X = M_X.$$

By the definition of M_X it is obvious that

$$M_X = I_n - X\left(X^{/}X\right)^{-1}X^{/} = I_n - P_X \Rightarrow M_X + P_X = I_n,$$

and consequently for any vector $z \in E^n$ we have

$$M_X z + P_X z = z.$$

The pair of projections M_X and P_X are called **complementary projections**, since the sum $M_X z$ and $P_X z$ restores the original vector z.

Assume that we have the following linear regression model:

$$y = X\beta + \varepsilon$$

where y and ε are $N \times 1$, β is $k \times 1$,and X is $N \times k$.

For $k = 2$ and if the first variable is a constant we have that :

$$y_i = \beta_0 + x_i\beta_1 + \varepsilon_i \quad for \quad i = 1, 2, ..., N.$$

Now

$$\begin{aligned}
\widehat{\beta} &= \begin{pmatrix} \widehat{\beta_0} \\ \widehat{\beta_1} \end{pmatrix} = \left(X^{/}X\right)^{-1}X^{/}y = \begin{pmatrix} T & \sum x \\ \sum x & \sum x^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum y \\ \sum xy \end{pmatrix} \\
&= \frac{1}{T\sum x^2 - \left(\sum x\right)^2} \begin{pmatrix} \sum x^2 & -\sum x \\ -\sum x & T \end{pmatrix} \begin{pmatrix} \sum y \\ \sum xy \end{pmatrix} \\
&= \begin{pmatrix} \frac{\sum y \sum x^2 - \sum x \sum xy}{T\sum x^2 - \left(\sum x\right)^2} \\ \frac{T\sum xy - \sum x \sum y}{T\sum x^2 - \left(\sum x\right)^2} \end{pmatrix}.
\end{aligned}$$

Notice however that

$$
\begin{aligned}
T\sum x^2-\left(\sum x\right)^2 &= T\left[\sum x^2-T\left(\frac{\sum x}{T}\right)^2\right]=T\left[\sum x^2-T\left(\overline{x}\right)^2\right] \\
&= T\left[\sum\left(x^2-\overline{x}^2\right)\right]=T\left[\sum\left(x^2-2x\overline{x}+2x\overline{x}+\overline{x}^2-2\overline{x}^2\right)\right] \\
&= T\left[\sum\left(x-\overline{x}\right)^2+2\overline{x}\sum\left(x-\overline{x}\right)\right]=T\sum\left(x-\overline{x}\right)^2,
\end{aligned}
$$

and

$$T\sum xy-\sum x\sum y=T\left[\sum\left(x-\overline{x}\right)\left(y-\overline{y}\right)\right].$$

Hence

$$
\begin{aligned}
\begin{pmatrix}\widehat{\beta_0}\\ \widehat{\beta_1}\end{pmatrix} &= \begin{pmatrix}\frac{T\overline{y}\sum x^2-T\overline{x}\sum xy}{T\sum(x-\overline{x})^2}\\ \frac{\sum(x-\overline{x})(y-\overline{y})}{\sum(x-\overline{x})^2}\end{pmatrix}=\begin{pmatrix}\frac{\overline{y}\left(T\sum(x-\overline{x})^2+(\sum x)^2\right)-\overline{x}\{T[\sum(x-\overline{x})(y-\overline{y})]+\sum x\sum y\}}{T\sum(x-\overline{x})^2}\\ \frac{\sum(x-\overline{x})(y-\overline{y})}{\sum(x-\overline{x})^2}\end{pmatrix} \\
&= \begin{pmatrix}\overline{y}-\frac{[\sum(x-\overline{x})(y-\overline{y})]}{\sum(x-\overline{x})^2}\overline{x}\\ \frac{\sum(x-\overline{x})(y-\overline{y})}{\sum(x-\overline{x})^2}\end{pmatrix}=\begin{pmatrix}\overline{y}-\widehat{\beta_1}\overline{x}\\ \frac{\sum(x-\overline{x})(y-\overline{y})}{\sum(x-\overline{x})^2}\end{pmatrix}.
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
Var\left(\widehat{\beta}\right) &= Var\begin{pmatrix}\widehat{\beta_0}\\ \widehat{\beta_1}\end{pmatrix}=\sigma^2\left(X^{/}X\right)^{-1}=\sigma^2\begin{pmatrix}T & \sum x\\ \sum x & \sum x^2\end{pmatrix}^{-1} \\
&= \frac{\sigma^2}{T\sum x^2-\left(\sum x\right)^2}\begin{pmatrix}\sum x^2 & -\sum x\\ -\sum x & T\end{pmatrix}.
\end{aligned}
$$

Hence

$$
\begin{aligned}
Var\left(\widehat{\beta_0}\right) &= \sigma^2\frac{\sum x^2}{T\sum\left(x-\overline{x}\right)^2}=\sigma^2\frac{\sum\left(x-\overline{x}\right)^2+T\overline{x}^2}{T\sum\left(x-\overline{x}\right)^2}=\sigma^2\left[\frac{1}{T}+\frac{\overline{x}^2}{\sum\left(x-\overline{x}\right)^2}\right], \\
and\quad Var\left(\widehat{\beta_1}\right) &= \frac{\sigma^2}{\sum\left(x-\overline{x}\right)^2}.
\end{aligned}
$$

Notice that to estimate these quantities we need an estimator of σ^2. We can employ its unbiased estimator, i.e.

$$\widehat{\sigma^2}=\frac{1}{T-2}\sum e_t^2\quad where\quad e_t^2=y_t-\widehat{\beta_0}-\widehat{\beta_1}x_t.$$

Part I

Linear Dynamic Stationary Processes

2 Autoregressive Models

These classes of models are of the form:

$$y_t^* = \mu + \alpha_1 y_{t-1}^* + \alpha_2 y_{t-2}^* + ... + \alpha_p y_{t-p}^* + u_t$$

i.e. the current y_t^* depends, directly, on p lags. Examples for various values of p are the $AR(1)$,

$$y_t^* = \mu + \alpha_1 y_{t-1}^* + u_t$$

the $AR(3)$,

$$y_t^* = \mu + \alpha_1 y_{t-1}^* + \alpha_2 y_{t-2}^* + \alpha_3 y_{t-3}^* + u_t$$

etc.

2.1 Autoregressive of Order 1

This class of models is given by

$$y_t^* = \mu + \alpha y_{t-1}^* + u_t \tag{1}$$

where y_t^* is the observed process and u_t are $iid\ (0, \sigma^2)$.

Properties

Notice that

$$\begin{aligned} y_t^* &= \mu + \alpha y_{t-1}^* + u_t = \mu + \alpha\mu + \alpha^2 y_{t-2}^* + u_t + \alpha u_{t-1} = ... \\ &= \mu\left(1 + \alpha + ... + \alpha^{t-1}\right) + \alpha^t y_0^* + u_t + \alpha u_{t-1} + ... + \alpha^{t-1} u_1 \\ &= \mu \frac{\alpha^t - 1}{\alpha - 1} + \alpha^t y_0^* + u_t + \alpha u_{t-1} + ... + \alpha^{t-1} u_1. \end{aligned}$$

Hence

$$E\left(y_t^*\right) = \mu\frac{\alpha^t - 1}{\alpha - 1} + \alpha^t E\left(y_0^*\right).$$

It is clear that for stationarity we need $|\alpha| < 1$ so as $t \to \infty$

$$E\left(y_t^*\right) \to \mu\frac{1}{1-\alpha}$$

which is independent of t.

To find the variance of y_t^* notice that

$$Var\left(y_t^*\right) = E\left\{\left[y_t^* - E\left(y_t^*\right)\right]^2\right\} = E\left\{\left[y_t^* - \frac{\mu}{1-\alpha}\right]^2\right\}$$

and that

$$y_t^* - \frac{\mu}{1-\alpha} = \mu - \frac{\mu}{1-\alpha} + \alpha y_{t-1}^* + u_t = \alpha\left(y_{t-1}^* - \frac{\mu}{1-\alpha}\right) + u_t.$$

Hence, if we substitute $y_t = y_t^* - \frac{\mu}{1-\alpha}$, it follows that

$$y_t = \alpha y_{t-1} + u_t \quad and \quad E\left\{\left[y_t^* - \frac{\mu}{1-\alpha}\right]^2\right\} = E\left(y_t^2\right). \tag{2}$$

Now squaring and taking expectations we get

$$\begin{aligned} E\left(y_t^2\right) &= \alpha^2 E\left(y_{t-1}^2\right) + E\left(u_t^2\right) + 2\alpha E\left(y_{t-1}u_t\right) \Rightarrow \\ E\left(y_t^2\right) &= \alpha^2 E\left(y_{t-1}^2\right) + \sigma^2 \end{aligned}$$

as $E\left(u_t^2\right) = \sigma^2$, by assumption, and $E\left(y_{t-1}u_t\right) = 0$. To see that the last expectation is equal to zero, substitute backwards y_{t-1} to get

$$\begin{aligned} E\left(y_{t-1}u_t\right) &= E\left[\left(\alpha y_{t-2} + u_{t-1}\right)u_t\right] = E\left[\left(\alpha^2 y_{t-3} + \alpha u_{t-2} + u_{t-1}\right)u_t\right] = \ldots = \\ &= E\left[\left(u_{t-1} + \alpha u_{t-2} + \alpha^2 u_{t-3} + \ldots\right)u_t\right] \\ &= E\left(u_{t-1}u_t\right) + \alpha E\left(u_{t-2}u_t\right) + \alpha^2 E\left(u_{t-3}u_t\right) + \ldots \end{aligned}$$

and all expectations are zero as the $u_t\prime s$ are independent (as before we need $|\alpha| < 1$ so $\alpha^k y_{t-k} \to 0$ as $k \to \infty$).

Hence

$$\begin{aligned} E\left(y_t^2\right) &= \alpha^2 E\left(y_{t-1}^2\right)+\sigma^2 = \alpha^2\left[\alpha^2 E\left(y_{t-2}^2\right)+\sigma^2\right]+\sigma^2 = \\ &= \alpha^4 E\left(y_{t-2}^2\right)+\alpha^2\sigma^2+\sigma^2 = ... = \alpha^{2k}E\left(y_{t-k}^2\right)+\left(\alpha^{2(k-1)}+....+\alpha^2+1\right)\sigma^2 \\ &= \left(1+\alpha^2+...+\alpha^{2(k-1)}+\alpha^{2k}+....\right)\sigma^2 = \frac{\sigma^2}{1-\alpha^2} \end{aligned}$$

as for $k \to \infty$, $\alpha^{2k} \to 0$, and the sum in the parenthesis is an infinite sum of a declining geometric series (for $|\alpha| < 1$).

To find the **autocorrelation function** notice that

$$\begin{aligned} \rho_k &= \frac{\gamma_k}{\gamma_0} = \frac{Cov\left(y_t^*, y_{t-k}^*\right)}{Var\left(y_t^*\right)} = \frac{E\left\{\left[y_t^*-E\left(y_t^*\right)\right],\left[y_{t-k}^*-E\left(y_{t-k}^*\right)\right]\right\}}{E\left(y_t^2\right)} = \\ &= \frac{E\left\{\left[y_t^*-\frac{\mu}{1-\alpha}\right],\left[y_{t-k}^*-\frac{\mu}{1-\alpha}\right]\right\}}{E\left(y_t^2\right)} = \frac{E\left(y_t, y_{t-k}\right)}{E\left(y_t^2\right)} = \frac{E\left(y_t, y_{t-k}\right)}{\frac{\sigma^2}{1-\alpha^2}}. \end{aligned}$$

To find the numerator multiply (2) by y_{t-k} and take expectations to get

$$E\left(y_t y_{t-k}\right) = \alpha E\left(y_{t-1}y_{t-k}\right)+E\left(u_t y_{t-k}\right) = \alpha E\left(y_{t-1}y_{t-k}\right) \Rightarrow \gamma_k = \alpha\gamma_{k-1}.$$

This is a difference equation with stable solution if $|\alpha| < 1$ (which is assumed due to stationarity). Hence

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\alpha\gamma_0}{\gamma_0} = \alpha \quad and \quad \rho_k = \alpha^k$$

in general.

Notice that for the existence of $E\left(y_t^*\right)$, $Var\left(y_t^*\right)$, and γ_k the only requirement is that $|\alpha| < 1$. Furthermore, for this condition the three quantities are independent of t. Hence the only condition for second order stationarity is $|\alpha| < 1$.

The **partial autocorrelation** of order k, $\rho_k^\bullet$, is the autocorrelation of y_t^* with y_{t-k}^* after we have taken in to account the autocorrelations of y_t^* with y_{t-1}^*, y_{t-2}^*,.., y_{t-k+1}^*, i.e. is the theoretical coefficient of y_{t-k}^* in a regression of y_t^* on a constant, y_{t-1}^*, y_{t-2}^*,.., y_{t-k+1}^*, y_{t-k}^*, i.e. of the regression

$$y_t^* = c+\beta_1 y_{t-1}^*+\beta_2 y_{t-2}^*+...+\beta_{k-1}y_{t-k+1}^*+\rho_k^\bullet y_{t-k}^*+\varepsilon_t.$$

Of course for $k = 1$ we have that $\rho_1 = \rho_1^{\bullet} = \alpha$. In the case of the $AR(1)$ model is obvious that equation in (1) can be written as

$$y_t^* = \mu + \alpha y_{t-1}^* + 0 y_{t-2}^* + u_t$$

so that $\rho_2^{\bullet} = 0$, Also

$$y_t^* = \mu + \alpha y_{t-1}^* + 0 y_{t-2}^* + 0 y_{t-3}^* + u_t$$

so that $\rho_3^{\bullet} = 0$. Hence, we can write that, for any $k \geq 1$, $\rho_k^{\bullet} = 0$ as

$$y_t^* = \mu + \alpha y_{t-1}^* + 0 y_{t-2}^* + 0 y_{t-3}^* + \ldots + 0 y_{t-k+1}^* + 0 y_{t-k}^* + u_t.$$

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.893	0.893	321.64	0.000
		2	0.801	0.015	580.94	0.000
		3	0.746	0.136	806.25	0.000
		4	0.692	-0.003	1000.7	0.000
		5	0.638	-0.001	1166.2	0.000
		6	0.582	-0.030	1304.7	0.000
		7	0.533	0.001	1421.0	0.000
		8	0.493	0.014	1520.5	0.000
		9	0.457	0.014	1606.4	0.000
		10	0.421	-0.006	1679.7	0.000
		11	0.400	0.063	1745.9	0.000
		12	0.380	-0.000	1805.7	0.000

Figure 2.1: ACF of the $AR(1)$ model $y_t^* = 0.1 + 0.89 y_{t-1}^* + u_t$, and u_t are $iid\ N(0, 1.4)$.

Notice that any $ARMA(p, q)$ model has a unique pair of autocorrelation and partial autocorrelation functions. In other words the two functions define uniquely the order of an $ARMA(p, q)$ model. This procedure is what Box and Jenkins call Identification Procedure of an $ARMA(p, q)$ model. For the $AR(1)$ model ρ_k decreases with a power law and $\rho_k^{\bullet}$ is non-zero for $k = 1$ and drops to zero for $k \geq 2$.

The last property of $ARMA$ models is **invertibility**, i.e. the property of these models so that they can be written as pure $AR(\infty)$ or $MA(\infty)$ models. Specifically for the $AR(1)$ model we can investigate under what conditions is the model invertible. Let us define he lag operator L as follows:

$$L^k x_t = x_{t-k}$$

for any time series x_t, i.e. the lag operators lags any time series observation by so many periods as its exponent. The usual properties of powers apply to the lag operators, as well, i.e.

$$L^k L^m = L^{k+m}, \quad \left(L^k\right)^m = L^{km}, \quad \frac{L^k}{L^m} = L^{k-m}.$$

With this definition we can write the model in equation (1) as

$$\begin{aligned} y_t^* &= \mu + \alpha y_{t-1}^* + u_t = \mu + \alpha L y_t^* + u_t \\ &\Rightarrow y_t^* - \alpha L y_t^* = \mu + u_t \Rightarrow (1 - \alpha L)\, y_t^* = \mu + u_t \\ &\Rightarrow y_t^* = \frac{\mu}{1-\alpha} + \frac{1}{1-\alpha L} u_t. \end{aligned}$$

Notice that as μ is nonstochastic L is dropped. Now if $|\alpha| < 1$ then $\frac{1}{1-\alpha L}$ can be considered as the sum of an infinite declining geometric series with ratio αL. Consequently,

$$\begin{aligned} y_t^* &= \frac{\mu}{1-\alpha} + \frac{1}{1-\alpha L} u_t = \frac{\mu}{1-\alpha} + \left(1 + \alpha L + \alpha^2 L^2 + \alpha^3 L^3 + ...\right) u_t \\ &\Rightarrow y_t^* = \frac{\mu}{1-\alpha} + u_t + \alpha L u_t + \alpha^2 L^2 u_t + \alpha^3 L^3 u_t + ... \\ &\Rightarrow y_t^* = \frac{\mu}{1-\alpha} + u_t + \alpha u_{t-1} + \alpha^2 u_{t-2} + \alpha^3 u_{t-3} + ... \end{aligned}$$

which is an $MA(\infty)$. Hence the invertibility condition is the same, in this case, as the stationarity one, i.e. $|\alpha| < 1$.

Prediction (with known parameters)

If we had no information on the model and we wanted to predict the value of y^*_{t+1}, or if we believed that the process y^*_t is white noise, or if we wanted to unconditionally predict y^*_{t+1}, then we would employ the unconditional distribution of y^*_{t+1}, which is the same as the unconditional distribution of y^*_t To find this recall that we can substitute backwards to get

$$y^*_t = \mu\left(1 + \alpha + \alpha^2 + ...\right) + u_t + \alpha u_{t-1} + \alpha^2 u_{t-2}...$$

Let us assume that $u_t \overset{iid}{\sim} N\left(0, \sigma^2\right)$. Consequently, the distribution of y^*_t is Normal, as it is a linear combination of independent Normal random variables, with

$$E\left(y^*_t\right) = \frac{\mu}{1-\alpha}, \quad Var\left(y^*_t\right) = \frac{\sigma^2}{1-\alpha^2}.$$

Hence

$$y^*_t \sim N\left(\frac{\mu}{1-\alpha}, \frac{\sigma^2}{1-\alpha^2}\right).$$

In this set up the best prediction for y_{t+1} is its unconditional mean, i.e.

$$\widehat{y^*_{t+1}} = \frac{\mu}{1-\alpha},$$

with error variance

$$Var\left(y^*_{t+1} - \widehat{y^*_{t+1}}\right) = Var\left(y^*_{t+1} - \frac{\mu}{1-\alpha}\right) = \frac{\sigma^2}{1-\alpha^2}.$$

On the other hand, if we wanted to utilise the information that the true model is an $AR(1)$, then we would employ the conditional distribution of y^*_{t+1}, i.e. the distribution of y^*_{t+1} given the previous observations and the model. Again under the Normality of $u_t's$, the conditional distribution of y^*_{t+1}, conditional on the previous observation y^*_t, is Normal with mean $\mu + \alpha y^*_t$ and variance σ^2, i.e.

$$\begin{aligned} E\left(y^*_{t+1}|y^*_t\right) &= E\left[\left(\mu + \alpha y^*_t + u_{t+1}\right)|y^*_t\right] = \mu + \alpha y^*_t + E\left[u_{t+1}|y^*_t\right] = \mu + \alpha y^*_t \\ & \quad and \\ Var\left(y^*_{t+1}|y^*_t\right) &= E\left[y^*_{t+1} - E\left(y^*_{t+1}|y^*_t\right)\right]^2 = E\left(y^*_{t+1} - \mu - \alpha y^*_t\right)^2 = E\left(u_t\right)^2 = \sigma^2. \end{aligned}$$

and it follows that

$$y_{t+1}^*|y_t^* \sim N\left(\mu + \alpha y_t^*, \sigma^2\right).$$

Hence the conditional predictor of y_{t+1}^* is

$$\widehat{y_{t+1}^*} = \mu + \alpha y_t^*$$

with an error variance of σ^2 (the variance of u_{t+1})

$$Var\left(y_{t+1}^* - \widehat{y_{t+1}^*}\right) = Var\left(u_{t+1}\right) = \sigma^2.$$

Notice that the distribution is not the same for all t (as the $y_t\prime s$ have different values for different $t\prime s$ with probability 1). Consequently, the conditional predictor is better than the unconditional one as it has smaller error variance.

Prediction (with estimated parameters)

In practice, predictions are made with estimated parameters. Let us assume, for easy of calculations that $\mu = 0$, and that from a sample of T observations we have estimated α, say by regression. Then our prediction for y_{T+1} is

$$\widetilde{y_{T+1}} = \widehat{\alpha} y_T^*$$

where $\widehat{\mu}$ and $\widehat{\alpha}$ are the estimated parameters. We can decompose the prediction error of the above prediction as

$$y_{T+1} - \widetilde{y_{T+1}} = \left(y_{T+1} - \widehat{y_{T+1}^*}\right) + \left(\widehat{y_{t+1}^*} - \widetilde{y_{T+1}}\right).$$

The first term on the right hand side represents the prediction error when μ and α are known, and the second is coming from the estimation of the parameters. Substituting in the second parenthesis the values of $\widehat{y_{t+1}^*}$ and $\widetilde{y_{T+1}}$ we get:

$$y_{T+1} - \widetilde{y_{T+1}} = \left(y_{T+1} - \widehat{y_{T+1}^*}\right) + \left(\alpha - \widehat{\alpha}\right) y_T^*.$$

One can prove that

$$MSE\left(\widetilde{y_{T+1}}\right) = s^2\left(1 + \frac{\left(y_T^*\right)^2}{\sum\limits_{t=2}^{T}\left(y_{t-1}^*\right)^2}\right) \simeq \sigma^2 + \frac{\left(y_T^*\right)^2\left(1-\alpha^2\right)}{T}$$

where MSE is the Mean Square Error. Notice that MSE is bigger to $Var\left(y_{t+1}^* - \widehat{y_{t+1}^*}\right)$ by $\frac{\left(y_T^*\right)^2\left(1-\alpha^2\right)}{T}$.

Estimation

To estimate the $AR(1)$ model in (1) we can employ either the simple regression or the maximum likelihood.

Employing the formulae for the linear **regression** we get:

$$\widehat{\mu} = \overline{y^*}\left(1-\widehat{\alpha}\right) \quad and \quad \widehat{\alpha} = \frac{\sum_{t=2}^{T}\left(y_t^* - \overline{y^*}\right)\left(y_{t-1}^* - \overline{y^*}\right)}{\sum_{t=2}^{T}\left(y_{t-1}^* - \overline{y^*}\right)^2}.$$

Notice that the summation runs from $t = 2$. Hence in a sample of T observations we employ $T-1$ for the estimation of α. Furthermore, the unbiased estimator of σ^2 is

$$\widehat{\sigma^2} = \frac{1}{T-3}\sum_{t=2}^{T} e_t^2 \quad where \quad e_t^2 = y_t^* - \widehat{\mu} - \widehat{\alpha} y_{t-1}^*.$$

To find the **Maximum Likelihood Estimators**, let us denote by $L(\mu, \alpha, \sigma^2) = L\left(y_1^*, y_2^*, ..., y_T^*; \mu, \alpha, \sigma^2\right)$ the likelihood function for the random variables then it can be written as

$$\begin{aligned} L(\mu, \alpha, \sigma^2) &= L\left(y_1^*, y_2^*, ..., y_T^*; \mu, \alpha, \sigma^2\right) = \\ &= L\left(y_T^* | y_{T-1}^*, ..., y_2^*, y_1^*; \mu, \alpha, \sigma^2\right) L\left(y_{T-1}^*, ..., y_2^*, y_1^*; \mu, \alpha, \sigma^2\right) \end{aligned}$$

i.e. the joint Likelihood equals the conditional of the last observation y_T^* (on rest of the sample $y_{T-1}^*, ..., y_2^*, y_1^*$) times the marginal (of the rest of the sample).

Repeating the above procedure we have (dropping the parameters to conserve space):

$$L(\mu,\alpha,\sigma^2) = L\left(y_T^*|y_{T-1}^*,...,y_2^*,y_1^*\right) L\left(y_{T-1}^*|y_{T-2}^*...,y_2^*,y_1^*\right) ...L\left(y_2^*|y_1^*\right) L\left(y_1^*\right).$$

Now the conditional distribution of y_t^* is (see Prediction section):

$$y_t^*|y_{t-1}^* \sim N\left(\mu+\alpha y_{t-1}^*,\sigma^2\right).$$

Of course the same is true for the conditional, on all previous observations, distribution of y_t^*, i.e.

$$y_t^*|y_{t-1}^*,y_{t-2}^*,...,y_2^*,y_1^* \sim N\left(\mu+\alpha y_{t-1}^*,\sigma^2\right),$$

as y_t^* depends only on the previous observation y_{t-1}^*.

Hence the Likelihood can be written as:

$$L(\mu,\alpha,\sigma^2) = \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(y_t^*-\mu-\alpha y_{t-1}^*\right)}{\sigma^2}\right) L\left(y_1^*\right).$$

Now the distribution of y_1^* is different as there is no information in the sample for the previous observation. Consequently the distribution of y_1^* is the unconditional distribution of the $AR(1)$ model, i.e. (see Prediction section)

$$y_1^* \sim N\left(\frac{\mu}{1-\alpha},\frac{\sigma^2}{1-\alpha^2}\right).$$

Therefore, the Likelihood function is

$$L(\mu,\alpha,\sigma^2) = \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(y_t^*-\mu-\alpha y_{t-1}^*\right)}{\sigma^2}\right) \frac{1}{\sqrt{2\pi\frac{\sigma^2}{1-\alpha^2}}} \exp\left(-\frac{\left(y_1^*-\frac{\mu}{1-\alpha}\right)}{\frac{\sigma^2}{1-\alpha^2}}\right),$$

and the log-Likelihood is

$$\begin{aligned}\ell(\mu,\alpha,\sigma^2) &= -\frac{T-1}{2}\ln(2\pi) - \frac{T-1}{2}\ln\sigma^2 - \sum_{t=2}^{T}\frac{\left(y_t^*-\mu-\alpha y_{t-1}^*\right)^2}{2\sigma^2} \\ &\quad -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\sigma^2 - \frac{1}{2}\ln\left(1-\alpha^2\right) - \frac{\left(y_1^*-\frac{\mu}{1-\alpha}\right)^2}{2\frac{\sigma^2}{1-\alpha^2}}.\end{aligned}$$

The first order condition are given by:

$$\frac{\partial \ell}{\partial \mu} = \sum_{t=2}^{T} \frac{\left(y_t^* - \mu - \alpha y_{t-1}^*\right)}{\sigma^2} + \left\{ \frac{(1+\alpha)\left(y_1^* - \frac{\mu}{1-\alpha}\right)}{\sigma^2} \right\} = 0,$$

$$\frac{\partial \ell}{\partial \alpha} = \sum_{t=2}^{T} \frac{\left(y_t^* - \mu - \alpha y_{t-1}^*\right) y_{t-1}^*}{\sigma^2} + \left\{ \frac{\alpha}{1-\alpha^2} + \frac{\left(y_1^* - \frac{\mu}{1-\alpha}\right)\left[\frac{\mu(1+\alpha)}{1-\alpha} + \alpha\left(y_1^* - \frac{\mu}{1-\alpha}\right)\right]}{\sigma^2} \right\} = 0$$

and

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{T-1}{2\sigma^2} + \sum_{t=2}^{T} \frac{\left(y_t^* - \mu - \alpha y_{t-1}^*\right)^2}{2\sigma^4} + \left\{ -\frac{1}{2\sigma^2} + \frac{(1-\alpha^2)\left(y_1^* - \frac{\mu}{1-\alpha}\right)^2}{2\sigma^4} \right\} = 0,$$

where in curly brackets are the derivatives of the log-Likelihood of the first observation y_1^*.

Notice that the estimator that solve the above first order conditions are different from the estimators of the regression. These estimators are called Full Information Maximum Likelihood Estimators. However, if we drop the contribution of the first observation to the log-likelihood function, by assuming that y_1^* is constant, then solving the first order conditions we have:

$$\widehat{\mu} = \overline{y^*}\left(1 - \widehat{\alpha}\right), \quad \widehat{\alpha} = \frac{\sum_{t=2}^{T}\left(y_t^* - \overline{y^*}\right)\left(y_{t-1}^* - \overline{y^*}\right)}{\sum_{t=2}^{T}\left(y_{t-1}^* - \overline{y^*}\right)^2},$$

which are identical to the regression estimators. These estimators are called Conditional (on the first observation) Maximum Likelihood Estimators.

3 Moving Average Models

These classes of models are of the form:

$$y_t^* = \mu + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \ldots - \theta_q u_{t-q}$$

i.e. the current y_t^* depends, directly, on q lags of u_t. Examples for various values of q are the $MA(2)$,

$$y_t^* = \mu + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2}$$

the $MA(4)$,

$$y_t^* = \mu + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \theta_3 u_{t-3} - \theta_4 u_{t-4}$$

etc.

3.1 Moving Average of Order 1

These models are given by

$$y_t^* = \mu + u_t - \theta u_{t-1} \tag{3}$$

where y_t^* is the observed process and u_t are $iid\ (0, \sigma^2)$.

Properties

It easy to prove that, for any value of θ,

$$E(y_t^*) = E(\mu + u_t - \theta u_{t-1}) = \mu + E(u_t) - \theta E(u_{t-1}) = \mu,$$

and

$$Var(y_t^*) = E[y_t^* - E(y_t^*)]^2 = E[u_t - \theta u_{t-1}]^2 = \sigma^2 \left(1 + \theta^2\right)$$

as $E(u_t u_{t-1}) = E(u_t) E(u_{t-1}) = 0$ due to independence of the $u_t\prime s$. Furthermore, it easy to prove that the **autocorrelation** is given by

$$\rho_k = \begin{cases} -\frac{\theta}{1+\theta^2} & for \quad k = 1 \\ 0 \quad for \quad k > 1 \end{cases} .$$

as

$$\begin{aligned} \gamma_k &= Cov\left(y_t^*, y_{t-k}^*\right) = E\left[(y_t^* - \mu)\left(y_{t-k}^* - \mu\right)\right] = E\left[(u_t - \theta u_{t-1})(u_{t-k} - \theta u_{t-k-1})\right] \\ &= E(u_t u_{t-k}) - \theta E(u_t u_{t-k-1}) - \theta E(u_{t-1} u_{t-k}) + \theta^2 E(u_{t-1} u_{t-k-1}) \\ &= -\theta E(u_{t-1} u_{t-k}) + \theta^2 E(u_{t-1} u_{t-k-1}) \end{aligned}$$

for any positive k and due to independence of the $u_t's$. Hence for $k=1$ we have that

$$\gamma_1 = -\theta E\left(u_{t-1}^2\right) + \theta^2 E\left(u_{t-1}u_{t-3}\right) = -\theta\sigma^2$$

whereas, for $k > 1$

$$\gamma_k = 0.$$

Notice that $E\left(y_t^*\right)$, $Var\left(y_t^*\right)$, and γ_k are independent of t for any value of θ, i.e. the MA(1) process is 2^{nd} **order stationary** for any value of θ.

In terms of the **partial correlation** we know that

$$\rho_1 = \rho_1^\bullet.$$

Now to find $\rho_2^\bullet$ we have to find the theoretical coefficient of y_{t-2} of the regression (projection) of y_t on y_{t-1} and y_{t-2}, i.e.

$$y_t = \alpha y_{t-1} + \rho_2^\bullet y_{t-2},$$

where $y_t = y_t^* - \mu$ for all t. Hence

$$y_t = \begin{pmatrix} y_{t-1} & y_{t-2} \end{pmatrix} \begin{pmatrix} \alpha \\ \rho_2^\bullet \end{pmatrix}.$$

Now from the normal equations we have

$$E\left[\begin{pmatrix} y_{t-1} & y_{t-2} \end{pmatrix}\begin{pmatrix} y_{t-1} \\ y_{t-1} \end{pmatrix}\right]\begin{pmatrix} \alpha \\ \rho_2^\bullet \end{pmatrix} = E\left[\begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix} y_t\right]$$

and it follows

$$\begin{aligned} E\begin{bmatrix} (y_{t-1})^2 & y_{t-1}y_{t-2} \\ y_{t-1}y_{t-2} & (y_{t-2})^2 \end{bmatrix}\begin{pmatrix} \alpha \\ \rho_2^\bullet \end{pmatrix} &= E\begin{bmatrix} y_{t-1}y_t \\ y_{t-2}y_t \end{bmatrix} \\ \begin{bmatrix} \gamma_0 & \gamma_1 \\ \gamma_1 & \gamma_0 \end{bmatrix}\begin{pmatrix} \alpha \\ \rho_2^\bullet \end{pmatrix} &= \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} \end{aligned}$$

or dividing by γ_0

$$\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} \begin{pmatrix} \alpha \\ \rho_2^\bullet \end{pmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix}.$$

Hence

$$\rho_2^\bullet = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2} = -\frac{\rho_1^2}{1-\rho_1^2}, \qquad \alpha = \frac{\begin{vmatrix} \rho_1 & \rho_1 \\ \rho_2 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_1 (1-\rho_2)}{1-\rho_1^2},$$

as $\rho_2 = 0$. Substituting for $\rho_1 = -\frac{\theta}{1+\theta^2}$ we get

$$\rho_2^\bullet = -\frac{\left(-\frac{\theta}{1+\theta^2}\right)^2}{1-\left(-\frac{\theta}{1+\theta^2}\right)^2} = -\frac{\theta^2}{\left(1+\theta^2\right)^2 - \theta^2} = -\theta^2 \frac{1-\theta^2}{1-\theta^6} \quad \alpha = -\frac{\theta\left(1+\theta^2\right)}{\left(1+\theta^2\right)^2-\theta^2}.$$

With the same logic, the normal equations for $\rho_3^\bullet$ are

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{bmatrix} \begin{pmatrix} a \\ b \\ \rho_3^\bullet \end{pmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{bmatrix},$$

and

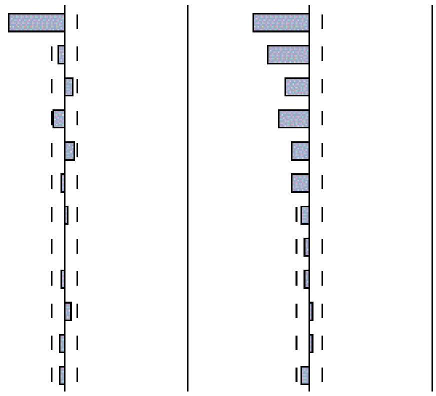
$$\begin{aligned} \rho_3^\bullet &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & 0 \\ 0 & \rho_1 & 0 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & 0 \\ \rho_1 & 1 & \rho_1 \\ 0 & \rho_1 & 1 \end{vmatrix}} = \frac{\rho_1 \begin{vmatrix} \rho_1 & 1 \\ 0 & \rho_1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix} - \rho_1 \begin{vmatrix} \rho_1 & \rho_1 \\ 0 & 1 \end{vmatrix}} = \frac{\rho_1^3}{1-2\rho_1^2} \\ &= \frac{\left(-\frac{\theta}{1+\theta^2}\right)^3}{1-2\left(-\frac{\theta}{1+\theta^2}\right)^2} = -\frac{\theta^3}{\left(1+\theta^2\right)\left[\left(1+\theta^2\right)^2 - 2\theta^2\right]} = -\theta^3 \frac{1-\theta^2}{1-\theta^8}. \end{aligned}$$

In general we have that

$$\rho_k^{\bullet} = -\theta^k \frac{1-\theta^2}{1-\theta^{2(k+1)}},$$

consequently we have that $\rho_k^{\bullet}$ declines with a power law if $|\theta| < 1$.

Hence for the $MA(1)$ model ρ_1 is non zero and ρ_k is zero for $k \geq 2$, whereas $\rho_k^{\bullet}$ decreases with a power law (compare with the $AR(1)$ model). Notice that for any value of θ the $MA(1)$ process is stationary.



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.469	-0.469	88.311	0.000
		2	-0.055	-0.353	89.553	0.000
		3	0.066	-0.197	91.302	0.000
		4	-0.096	-0.257	95.061	0.000
		5	0.084	-0.148	97.931	0.000
		6	-0.030	-0.143	98.302	0.000
		7	0.028	-0.068	98.615	0.000
		8	-0.003	-0.042	98.619	0.000
		9	-0.027	-0.044	98.929	0.000
		10	0.050	0.024	99.938	0.000
		11	-0.035	0.018	100.45	0.000
		12	-0.045	-0.063	101.28	0.000

Figure 3.1: ACF and Partial ACF for the $MA(1)$ process with $\theta = 0.8$, $\sigma^2 = 1.4$, $\mu = 1.0$.

In terms of **invertibility**, notice that equation (3) can be written as

$$y_t^* = \mu + (1-\theta L)\, u_t \Rightarrow \frac{1}{1-\theta L} y_t^* = \frac{\mu}{1-\theta} + u_t$$

and if $|\theta| < 1$ we have that

$$\Rightarrow \quad \left(1 + \theta L + \theta^2 L^2 + \theta^3 L^3 + \ldots\right) y_t^* = \frac{\mu}{1-\theta} + u_t$$

$$\Rightarrow \quad y_t^* = \frac{\mu}{1-\theta} - \theta y_{t-1}^* - \theta^2 y_{t-2}^* - \theta^3 y_{t-3}^* - \ldots + u_t$$

which is an $AR(\infty)$.

Prediction (with known parameters)

The unconditional distribution for an MA(1) model is Normal, as it is a linear combination of independent Normal random variables, with

$$E\left(y_t^*\right) = \mu, \quad Var\left(y_t^*\right) = \sigma^2\left(1+\theta^2\right).$$

Hence

$$y_t^* \sim N\left(\mu, \sigma^2\left(1+\theta^2\right)\right).$$

In this set up the best prediction for y_{t+1} is its unconditional mean, i.e.

$$\widehat{y_{t+1}^*} = \mu,$$

and the error variance is given by

$$Var\left(y_{t+1}^* - \widehat{y_{t+1}^*}\right) = Var\left(y_{t+1}^* - \mu\right) = Var\left(u_{t+1} - \theta u_t\right) = \sigma^2\left(1+\theta^2\right).$$

For the conditional distribution we need to assume a value for u_0, say $u_0 = 0$. Then

$$u_1 = y_1^* - \mu + \theta u_0$$

which is known, provided that the parameters μ and θ are known. Consequently

$$u_2 = y_2^* - \mu + \theta u_1$$

is known and so forth. Hence

$$u_t = y_t^* - \mu + \theta u_{t-1}$$

is known. Hence for $t = t+1$ we have that y_{t+1}^*, conditional on the values of y_t^*, y_{t-1}^*, y_{t-2}^*,..., y_2^*, y_1^*, and u_0, is distributed as Normal, because the only stochastic element is u_{t+1} which is Normally distributed, with mean $\mu - \theta u_t$ and variance σ^2, i.e.

$$\begin{aligned} E\left(y_{t+1}^*|y_t^*, y_{t-1}^*, ..., y_1^*, u_0\right) &= \mu - \theta u_t \\ & \quad and \\ Var\left(y_{t+1}^*|y_t^*, y_{t-1}^*, ..., y_1^*, u_0\right) &= E\left(u_{t+1}\right)^2 = \sigma^2. \end{aligned}$$

and it follows that

$$y_{t+1}^*|y_t^*, y_{t-1}^*, ..., y_1^*, u_0 \sim N\left(\mu - \theta u_t, \sigma^2\right).$$

Notice that the distribution is not the same for all t (as the $y_t{}'s$ have different values for different $t{}'s$ with probability 1). Hence the conditional predictor of y_{t+1}^* is now $\mu - \theta u_t$ with an error variance of σ^2 (the variance of u_{t+1}). Consequently, the conditional predictor is better than the unconditional one as it has smaller error variance. Furthermore, notice that we need to condition not only on y_t^*, as in the case of the $AR(1)$, but on the whole series of the previous $y_t^{*\prime}s$, i.e. on $y_t^*, y_{t-1}^*, ..., y_1^*$ as well as on u_0. To understand this difference recall that the $MA(1)$ model, if invertible, is an $AR(\infty)$, and consequently the last observation depends on all previous ones.

Estimation

To estimate the $MA(1)$ model in (3) we can employ either the Method of Moments or the maximum likelihood.

To apply the **Method of Moments** we equate the theoretical mean and first order autocorrelation with their sample counterparts we could estimate μ and θ, i.e.

$$\widehat{\mu} = \frac{1}{T}\sum_{t=1}^{T} y_t^* = \overline{y^*}, \quad \widehat{\theta} \quad \text{solve}\, s \quad -\frac{\widehat{\theta}}{1+\widehat{\theta}^2} = \frac{\sum_{t=2}^{T}\left(y_t^* - \overline{y^*}\right)\left(y_{t-1}^* - \overline{y^*}\right)}{\sqrt{\sum_{t=2}^{T}\left(y_t^* - \overline{y^*}\right)^2 \sum_{t=2}^{T}\left(y_{t-1}^* - \overline{y^*}\right)^2}}.$$

Notice that the equation in $\widehat{\theta}$ is a quadratic with two roots such that $\widehat{\theta_1} = \frac{1}{\widehat{\theta_2}}$ and consequently if they are real one one is less than 1, in absolute value, the other will be greater than 1. To raise this indeterminacy we adopt the convention that θ is less than 1 in absolute value, and consequently we choose the solution which has absolute value less than 1. This also is a condition for the invertibility of the $MA(1)$ process.

For the **Maximum Likelihood**, let $L(\mu, \alpha, \sigma^2) = L\left(y_1^*, y_2^*, ..., y_T^*; \mu, \alpha, \sigma^2\right)$ be the likelihood function for the random variables then, as in the $AR(1)$ case, it can be written as

$$L(\mu, \alpha, \sigma^2) = L\left(y_T^* | y_{T-1}^*, ..., y_2^*, y_1^*\right) L\left(y_{T-1}^* | y_{T-2}^* ..., y_2^*, y_1^*\right) ... L\left(y_2^* | y_1^*\right) L\left(y_1^*\right).$$

Now the conditional distribution of y_t^* is (see Prediction section):

$$y_t^* | y_{t-1}^*, y_{t-2}^*, ... y_1^*, u_0 \sim N\left(\mu - \theta u_{t-1}, \sigma^2\right).$$

Hence the Likelihood can be written as:

$$L(\mu, \alpha, \sigma^2) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_t^* - \mu + \theta u_{t-1})}{\sigma^2}\right)$$

and the log-Likelihood is

$$\ell(\mu, \alpha, \sigma^2) = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln \sigma^2 - \sum_{t=1}^{T} \frac{\left(y_t^* - \mu + \theta u_{t-1}\right)^2}{2\sigma^2}.$$

Recall that

$$u_t = y_t^* - \mu + \theta u_{t-1}$$

for $t = 1, ..., T$. Hence, the $u_t\prime s$ depend on both μ and θ. This is important for the first order conditions which are given by:

$$\frac{\partial \ell}{\partial \mu} = \sum_{t=1}^{T} (y_t^* - \mu + \theta u_{t-1}) \left(1 - \theta \frac{\partial u_{t-1}}{\partial \mu}\right) = 0,$$

where

$$\frac{u_t}{\partial \mu} = -1 + \theta \frac{\partial u_{t-1}}{\partial \mu} \quad for \quad t = 1, 2, ..., T \quad and \quad \frac{\partial u_0}{\partial \mu} = 0$$

$$\frac{\partial \ell}{\partial \theta} = -\sum_{t=1}^{T} \left(y_t^* - \mu + \theta u_{t-1}\right) \left(\theta \frac{\partial u_{t-1}}{\partial \theta} + u_{t-1}\right) = 0$$

where

$$\frac{u_t}{\partial \theta} = u_{t-1} + \theta \frac{\partial u_{t-1}}{\partial \theta} \quad for \quad t = 1, 2, ..., T \quad and \quad \frac{\partial u_0}{\partial \theta} = 0,$$

and

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \sum_{t=2}^{T} \frac{\left(y_t^* - \mu + \theta u_{t-1}\right)^2}{2\sigma^4} = 0,$$

Notice that the equations do not have an explicit solution. The dependence of the $u_t\prime s$ on the parameters is the main reason for not being able to estimate the parameters by a simple regression.

4 Mixed Models

Mixed models have both parts, i.e. an autoregressive and a moving average part. The order of the $ARMA$ models is defined from the order of the AR and MA part, i.e. $ARMA\left(3, 2\right)$ means a process which depends on 3 lags of its own value and 2 lagged errors, i.e.

$$y_t^* = \mu + \alpha_1 y_{t-1}^* + \alpha_2 y_{t-2}^* + \alpha_3 y_{t-3}^* + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2}.$$

In general an $ARMA\left(p, q\right)$ is written as

$$y_t^* = \mu + \alpha_1 y_{t-1}^* + \alpha_2 y_{t-2}^* + ... + \alpha_p y_{t-p}^* + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} ... - \theta_q u_{t-q}.$$

Notice that the first number in the parenthesis refer to the AR part and the second to the MA part.

4.1 ARMA of Order 1,1

These models are given by

$$y_t^* = \mu + \alpha y_{t-1}^* + u_t - \theta u_{t-1} \tag{4}$$

where y_t^* is the observed process and u_t are $iid\left(0, \sigma^2\right)$.

Properties

Notice that equation (4) is written as:

$$y_t^* = \frac{\mu}{1-\alpha} + \frac{1-\theta L}{1-\alpha L} u_t$$

and provided that $|\alpha| < 1$ we have that

$$y_t^* = \frac{\mu}{1-\alpha} + (1-\theta L)\left(1 + \alpha L + \alpha^2 L^2 + \alpha^3 L^3 + ...\right) u_t.$$

Multiplying the Lag polynomials and collecting terms we get:

$$y_t^* = \frac{\mu}{1-\alpha} + \left[1 + (\alpha-\theta) L + (\alpha-\theta)\alpha L^2 + (\alpha-\theta)\alpha^2 L^3 + ...\right] u_t$$

and multiplying through we get

$$y_t^* = \frac{\mu}{1-\alpha} + u_t + (\alpha-\theta) u_{t-1} + (\alpha-\theta)\alpha u_{t-2} + (\alpha-\theta)\alpha^2 u_{t-3} + ... \tag{5}$$

which is an $MA(\infty)$ process. Hence

$$E\left(y_t^*\right) = \frac{\mu}{1-\alpha},$$

as $E\left(u_t\right) = 0$ for all $t\prime s$. Furthermore,

$$\begin{aligned} Var\left(y_t^*\right) &= \sigma^2\left[1 + (\alpha-\theta)^2 + (\alpha-\theta)^2\alpha^2 + (\alpha-\theta)^2\alpha^4 + ...\right] \\ &= \sigma^2 + \sigma^2(\alpha-\theta)^2\left[1 + \alpha^2 + \alpha^4 + ...\right] \end{aligned}$$

and provided that $|\alpha| < 1$ we have that

$$Var\left(y_t^*\right) = \sigma^2 + \frac{\sigma^2\left(\alpha - \theta\right)^2}{1 - \alpha^2} = \sigma^2 \frac{1 - 2\alpha\theta + \theta^2}{1 - \alpha^2}.$$

To find the **autocorrelation** function, notice that in deviation, from the mean $\frac{\mu}{1-\alpha}$, form the model is written:

$$y_t = \alpha y_{t-1} + u_t - \theta u_{t-1}.$$

Multiply both sides by y_{t-k} and taking expectations we get:

$$E\left(y_t y_{t-k}\right) = \alpha E\left(y_{t-1} y_{t-k}\right) + E\left(u_t y_{t-k}\right) - \theta E\left(u_{t-1} y_{t-k}\right).$$

Clearly, $E\left(u_t y_{t-k}\right) = 0$ for any k, as y_{t-k} depends on u_{t-k} and the previous $u_t's$, see equation (5). Now for $k = 1$ we have

$$E\left(y_t y_{t-1}\right) = \alpha Var\left(y_{t-1}\right) - \theta E\left(u_{t-1} y_{t-1}\right)$$

where $E\left(u_{t-1} y_{t-1}\right)$

$$\begin{aligned} E\left(u_{t-1} y_{t-1}\right) &= \alpha E\left(u_{t-1} y_{t-2}\right) + E\left(u_{t-1}^2\right) - \theta E\left(u_{t-1} u_{t-2}\right) \Rightarrow \\ E\left(u_{t-1} y_{t-1}\right) &= 0 + \sigma^2 - 0 = \sigma^2. \end{aligned}$$

Hence

$$\gamma_1 = \alpha\gamma_0 - \theta\sigma^2 = \alpha\sigma^2 + \alpha\frac{\sigma^2\left(\alpha - \theta\right)^2}{1 - \alpha^2} - \theta\sigma^2 = \sigma^2\frac{\left(\alpha - \theta\right)\left(1 - \alpha\theta\right)}{1 - \alpha^2}$$

and

$$\rho_1 = \frac{\left(\alpha - \theta\right)\left(1 - \alpha\theta\right)}{1 - 2\alpha\theta + \theta^2}$$

For $k \geq 2$, we have that

$$E\left(y_t y_{t-k}\right) = \alpha E\left(y_{t-1} y_{t-k+1}\right) \Rightarrow \gamma_k = \alpha\gamma_{k-1}$$

and consequently,

$$\rho_k = \alpha\rho_{k-1}.$$

Hence

$$\rho_k = \begin{cases} \frac{(\alpha-\theta)(1-\alpha\theta)}{1-2\alpha\theta+\theta^2} & for \quad k=1 \\ \alpha^{k-1}\rho_1 & for \quad k \geq 2 \end{cases},$$

i.e. for $k \geq 2$ the autocorrelation function of an $ARMA(1,1)$ is the same as the equivalent of an $AR(1)$ model, i.e. declines as a power law (for $|\alpha| < 1$). Hence we can conclude that the ARMA(1,1) model is stationary if and only if $|\alpha| < 1$.

In terms of the **partial correlation** we know that

$$\rho_1 = \rho_1^\bullet.$$

Now to find $\rho_2^\bullet$ we have to find the theoretical coefficient of y_{t-2} of the regression (projection) of y_t on y_{t-1} and y_{t-2}, i.e.

$$y_t = \alpha y_{t-1} + \rho_2^\bullet y_{t-2},$$

where $y_t = y_t^* - \frac{\mu}{1-\alpha}$ for all t. Hence

$$y_t = \begin{pmatrix} y_{t-1} & y_{t-2} \end{pmatrix} \begin{pmatrix} \alpha \\ \rho_2^\bullet \end{pmatrix}.$$

Now from the normal equations we have

$$E\left[\begin{pmatrix} y_{t-1} & y_{t-2} \end{pmatrix}\begin{pmatrix} y_{t-1} \\ y_{t-1} \end{pmatrix}\right]\begin{pmatrix} \alpha \\ \rho_2^\bullet \end{pmatrix} = E\left[\begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix} y_t\right]$$

and it follows

$$\begin{aligned} E\begin{bmatrix} (y_{t-1})^2 & y_{t-1}y_{t-2} \\ y_{t-1}y_{t-2} & (y_{t-2})^2 \end{bmatrix}\begin{pmatrix} \alpha \\ \rho_2^\bullet \end{pmatrix} &= E\begin{bmatrix} y_{t-1}y_t \\ y_{t-2}y_t \end{bmatrix} \\ \begin{bmatrix} \gamma_0 & \gamma_1 \\ \gamma_1 & \gamma_0 \end{bmatrix}\begin{pmatrix} \alpha \\ \rho_2^\bullet \end{pmatrix} &= \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} \end{aligned}$$

or dividing by γ_0

$$\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}\begin{pmatrix} \alpha \\ \rho_2^\bullet \end{pmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix}.$$

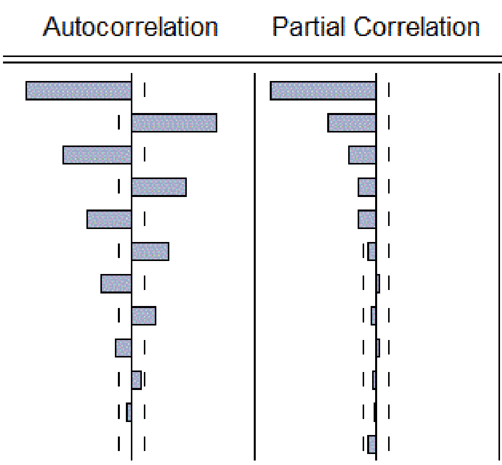
Hence

$$\rho_2^{\bullet} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \rho_1 \frac{\alpha - \rho_1}{1 - \rho_1^2},$$

as $\rho_2 = \alpha\rho_1$. Substituting for $\rho_1 = \frac{(\alpha-\theta)(1-\alpha\theta)}{1-2\alpha\theta+\theta^2}$ we get

$$\rho_2^{\bullet} = \frac{(\alpha - \theta)(1 - \alpha\theta)\theta}{[1 - \theta(\alpha - \theta)]^2 - \theta^2}$$

It is easy but tedious to prove that $\rho_k^{\bullet}$ will decline with a power law, depending on $\alpha - \theta$. Hence we can conclude that $\rho_k^{\bullet}$ behaves like in the $MA(1)$ case.



	AC	PAC	Q-Stat	Prob
1	-0.893	-0.893	320.58	0.000
2	0.715	-0.406	526.76	0.000
3	-0.574	-0.226	659.78	0.000
4	0.459	-0.143	745.02	0.000
5	-0.375	-0.147	802.24	0.000
6	0.316	-0.068	842.93	0.000
7	-0.258	0.019	870.11	0.000
8	0.197	-0.041	886.00	0.000
9	-0.137	0.028	893.74	0.000
10	0.082	-0.021	896.53	0.000
11	-0.040	-0.013	897.21	0.000
12	0.003	-0.072	897.21	0.000

Figure 4.1: ACF and Partial ACF for the $ARMA(1,1)$ process with $\alpha = -0.8$, $\theta = 0.8$, $\sigma^2 = 1.4$, $\mu = 1.0$.

In terms of **invertibility** the process is invertible if $|\alpha| < 1$, so that it can be written as an $MA(\infty)$, and $|\theta| < 1$ so that it can be written as an $AR(\infty)$.

Prediction (with known parameters)

The unconditional distribution for an $ARMA(1,1)$ model is Normal, as it is a linear combination of independent Normal random variables, with

$$E\left(y_t^*\right)=\frac{\mu}{1-\alpha},\quad Var\left(y_t^*\right)=\frac{1-2\alpha\theta+\theta^2}{1-\alpha^2}.$$

Hence

$$y_t^*\sim N\left(\frac{\mu}{1-\alpha},\frac{1-2\alpha\theta+\theta^2}{1-\alpha^2}\right).$$

In this set up the best prediction for y_{t+1} is its unconditional mean, i.e.

$$\widehat{y_{t+1}^*}=\frac{\mu}{1-\alpha},$$

and the error variance is given by

$$\begin{aligned}Var\left(y_{t+1}^*-\widehat{y_{t+1}^*}\right) &= Var\left(y_{t+1}^*-\frac{\mu}{1-\alpha}\right)\\ &= Var\left(u_t+(\alpha-\theta)u_{t-1}+(\alpha-\theta)\alpha u_{t-2}+(\alpha-\theta)\alpha^2u_{t-3}+...\right)\\ &= \frac{1-2\alpha\theta+\theta^2}{1-\alpha^2}.\end{aligned}$$

For the conditional distribution we need to assume a value for u_1, say $u_1=0$, and the value of y_1 is constant in repeating sampling. Then

$$u_2=y_2^*-\mu-\alpha y_1+\theta u_1$$

which is known, provided that the parameters μ, α,and θ are known. Consequently

$$u_3=y_3^*-\mu-\alpha y_2+\theta u_2$$

is known and so forth. Hence

$$u_t=y_t^*-\mu-\alpha y_{t-1}+\theta u_{t-1}$$

is known. Hence for $t=t+1$ we have that y_{t+1}^*, conditional on the values of y_t^*, y_{t-1}^*, y_{t-2}^*,..., y_2^*, y_1^*, and u_1, is distributed as Normal, because the only stochastic element is u_{t+1} which is Normally distributed, with mean $\mu+\alpha y_t-\theta u_t$ and variance σ^2, i.e.

$$\begin{aligned} E\left(y^*_{t+1}|y^*_t, y^*_{t-1}, ..., y^*_1, u_1\right) &= \mu+\alpha y_t-\theta u_t \\ & \quad and \\ Var\left(y^*_{t+1}|y^*_t, y^*_{t-1}, ..., y^*_1, u_1\right) &= E\left(u_{t+1}\right)^2 = \sigma^2. \end{aligned}$$

and it follows that

$$y^*_{t+1}|y^*_t, y^*_{t-1}, ..., y^*_1, u_0 \sim N\left(\mu+\alpha y_t-\theta u_t, \sigma^2\right).$$

Notice that the distribution is not the same for all t (as the $y_t\prime s$ have different values for different $t\prime s$ with probability 1). Hence the conditional predictor of y^*_{t+1} is now $\mu+\alpha y_t-\theta u_t$ with an error variance of σ^2 (the variance of u_{t+1}). Consequently, the conditional predictor is better than the unconditional one as it has smaller error variance. Furthermore, notice that we need to condition not only on y^*_t, as in the case of the $AR(1)$, but on the whole series of the previous $y^{*\prime}_t s$, i.e. on $y^*_t, y^*_{t-1}, ..., y^*_1$ as well as on u_1, as in the case of the $MA(1)$ model. To understand this difference recall that the $ARMA(1,1)$ model, if invertible, is an $AR(\infty)$, and consequently the last observation depends on all previous ones.

Estimation

To estimate the $ARMA(1,1)$ model in (2) we can employ the maximum likelihood.

For the **Maximum Likelihood**, let $L(\mu, \alpha, \sigma^2) = L\left(y^*_1, y^*_2, ..., y^*_T; \mu, \alpha, \sigma^2\right)$ be the likelihood function for the random variables. Now assuming that y_1 is constant, and consequently it does not contribute to the likelihood, and that $u_1 = 0$ we can write:

$$L(\mu, \alpha, \sigma^2|y_1, u_1) = L\left(y^*_T|y^*_{T-1}, .., y^*_2, y^*_1, u_1\right) L\left(y^*_{T-1}|y^*_{T-2}, .., y^*_2, y^*_1, u_1\right) ..L\left(y^*_2|y^*_1, u_1\right).$$

Now the conditional distribution of y_t^* is (see Prediction section):

$$y_t^*|y_{t-1}^*, y_{t-2}^*, ...y_1^*, u_0 \sim N\left(\mu + \alpha y_{t-1} - \theta u_{t-1}, \sigma^2\right).$$

Hence the Likelihood can be written as:

$$L(\mu, \alpha, \sigma^2) = \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_t^* - \mu - \alpha y_{t-1} + \theta u_{t-1})}{\sigma^2}\right)$$

and the log-Likelihood is

$$\ell(\mu, \alpha, \sigma^2) = -\frac{T-1}{2} \ln(2\pi) - \frac{T-1}{2} \ln \sigma^2 - \sum_{t=2}^{T} \frac{(y_t^* - \mu - \alpha y_{t-1} + \theta u_{t-1})^2}{2\sigma^2}.$$

Recall that

$$u_t = y_t^* - \mu - \alpha y_{t-1} + \theta u_{t-1}$$

for $t = 2, ..., T$. Hence, the $u_t\prime s$ depend on both μ, α,and θ. This is important for the first order conditions which are given by:

$$\frac{\partial \ell}{\partial \mu} = \sum_{t=2}^{T} (y_t^* - \mu + \theta u_{t-1}) \left(1 - \theta \frac{\partial u_{t-1}}{\partial \mu}\right) = 0,$$

where

$$\frac{u_t}{\partial \mu} = -1 + \theta \frac{\partial u_{t-1}}{\partial \mu} \quad for \quad t = 2, ..., T \quad and \quad \frac{\partial u_1}{\partial \mu} = 0.$$

$$\frac{\partial \ell}{\partial \alpha} = \sum_{t=2}^{T} (y_t^* - \mu - \alpha y_{t-1} + \theta u_{t-1}) \left(y_{t-1} - \theta \frac{\partial u_{t-1}}{\partial \alpha}\right)$$

where

$$\frac{\partial u_t}{\partial \alpha} = -y_{t-1} + \theta \frac{\partial u_{t-1}}{\partial \alpha} \quad for \quad t = 2, ..., T \quad and \quad \frac{\partial u_1}{\partial \alpha} = 0.$$

$$\frac{\partial \ell}{\partial \theta} = -\sum_{t=2}^{T} (y_t^* - \mu - \alpha y_{t-1} + \theta u_{t-1}) \left(u_{t-1} + \theta \frac{\partial u_{t-1}}{\partial \theta}\right)$$

where

$$\frac{\partial u_t}{\partial \theta} = u_{t-1} + \theta \frac{\partial u_{t-1}}{\partial \theta} \quad for \quad t = 2, ..., T \quad and \quad \frac{\partial u_1}{\partial \theta} = 0.$$

Finally,

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{T-1}{2}\frac{1}{\sigma^2} + \sum_{t=2}^{T} \frac{\left(y_t^* - \mu - \alpha y_{t-1} + \theta u_{t-1}\right)^2}{2\sigma^4}.$$

Notice that the equations do not have an explicit solution. Again, the dependence of the $u_t \prime s$ on the parameters is the main reason for not being able to estimate the parameters by a simple regression.

5 Testing for Autocorrelation

We shall mainly employ the Portmantau Q test.

5.1 Autocorrelation Coefficients

Given a stationary (second order) time series y_t we have already defined the k^{th} order autocovariance and autocorrelation coeficients, γ_k and ρ_k, as:

$$\gamma_k = Cov\left(y_t, y_{t-k}\right), \quad \rho_k = \frac{\gamma_k}{\gamma_0}.$$

For a given sample $\{y_t\}_{t=1}^{T}$ autocovariance and autocorrelation coefficients can be estimated in the natural way by replacing population moments with the sample counterparts:

$$\begin{aligned} \overset{\wedge}{\gamma}_k &= \frac{1}{T}\sum_{t=k+1}^{T}(r_t - \bar{r_T})(r_{t-k} - \bar{r_T}) \quad for\ 0 \leq k < T \quad and \\ \overset{\wedge}{\rho}_k &= \frac{\overset{\wedge}{\gamma}_k}{\overset{\wedge}{\gamma}_0}, \quad where \quad \bar{r_T} = \frac{1}{T}\sum_{t=1}^{T} r_t \end{aligned}$$

Depending on the assumed process for y_t we can derive the distributions of $\overset{\wedge}{\gamma}_k$ and $\overset{\wedge}{\rho}_k$. If y_t is white noise, has variance σ^2, its distribution is symmetric and the sixth moment is proportional to σ^6, then:

$$\begin{aligned} E\left(\overset{\wedge}{\rho}_k\right) &= -\frac{T-k}{T(T-1)} + \bigcirc(T^{-2}) \qquad and \\ Cov\left(\overset{\wedge}{\rho}_k, \overset{\wedge}{\rho}_l\right) &= \begin{cases} \frac{T-k}{T^2} + \bigcirc(T^{-2}) & if \quad k = l \neq 0 \\ \bigcirc(T^{-2}) & otherwise \end{cases}. \end{aligned}$$

Notice that although, under the white noise assumption, the true $\rho_k = 0$ for all $k\prime s$, the sample autocorrelations are negatively biased. This negative bias comes from the fact that by estimating the sample mean and subtracting it from the data results in deviations that sum to zero. Hence on average positive deviations are followed by negative ones and vice versa and consequently result in negative sum of the product.

In large samples, under the assumption that the true $\rho_k = 0$, we have that

$$\sqrt{T}\overset{\wedge}{\rho}_k \overset{A}{\sim} N(0,1).$$

5.2 Portmanteau Statistics

Since the white noise assumption implies that all autocorrelations are zero we can use the Box-Pierce $Q - statistic$. Under the null $H_0 : \rho_1 = \rho_2 = ...\rho_m = 0$ it easy to see that:

$$Q_m = T\sum_{k=1}^{m}\left(\overset{\wedge}{\rho}_k\right)^2 \overset{A}{\sim} \chi^2_m.$$

The Ljung-Box small sample correction is:

$$Q^*_m = T(T+2)\sum_{k=1}^{m}\frac{\left(\overset{\wedge}{\rho}_k\right)^2}{T-k} \overset{A}{\sim} \chi^2_m.$$

Notice that for unnecessarily big m the tests has low power, whereas for too small m it does not pick up higher possible correlation. This is one of the reasons that most econometric packages print the Q^*_m for various values of m (see Figures 2.1, 3.1, and 4.1).

Remark R.1 *When the residuals of an ARMA(p,q) are employed for the* Q_m *test then we have that, under the null of no extra autocorrelations,*

$$Q_m \overset{A}{\sim} \chi^2_{m-p-q}.$$

Part II

Conditional Heteroskedasticity

6 Conditional Heteroskedastic Models

These are models which have time varying conditional variance, i.e. the conditional variance is a function of all available information at time $t-1$.

6.1 ARCH(1)

The Autoregressive Conditional Heteroskedasticity models of order (1) are given by (Engle 1982)

$$y_t = \mu + u_t \quad where \quad u_t|I_{t-1} \sim N\left(0, \sigma_t^2\right) \quad and \quad \sigma_t^2 = c + \alpha u_{t-1}^2 \tag{6}$$

where y_t^* is the observed process and $I_{t-1} = \{y_{t-1}, y_{t-2}, ...\}$. Notice that given I_{t-1} σ_t^2 (the conditional variance) is known, i.e. non-stochastic. Of course unconditionally, i.e. when the conditioning set is the empty set, σ_t^2 is stochastic.

Properties

To capitalise on the properties of the $ARMA$ models, notice that the conditional variance can be written as:

$$u_t^2 = c + \alpha u_{t-1}^2 + u_t^2 - \sigma_t^2 \Rightarrow u_t^2 = c + \alpha u_{t-1}^2 + v_t \tag{7}$$

where $v_t = u_t^2 - \sigma_t^2$ (this parameterisation was suggested by Pantula). Now v_t has the following properties:

$$\begin{aligned} E\left(v_t\right) &= E\left(u_t^2 - \sigma_t^2\right) = E\left[E\left(u_t^2 - \sigma_t^2\right)|I_{t-1}\right] = E\left[E\left(u_t^2|I_{t-1}\right) - \sigma_t^2\right] \\ &= E\left[\sigma_t^2 - \sigma_t^2\right] = 0 \end{aligned}$$

where the second equality follows from the Law of Iterated Expectations, and the forth from the definition of σ_t^2. Furthermore,

$$\begin{aligned} E\left(v_t v_{t-k}\right) &= E\left[\left(u_t^2-\sigma_t^2\right)\left(u_{t-k}^2-\sigma_{t-k}^2\right)\right] = E\left[E\left(u_t^2-\sigma_t^2\right)\left(u_{t-k}^2-\sigma_{t-k}^2\right)|I_{t-1}\right] \\ &= E\left[E\left(u_t^2|I_{t-1}\right)\left(u_{t-k}^2-\sigma_{t-k}^2\right)-\sigma_t^2\left(u_{t-k}^2-\sigma_{t-k}^2\right)\right] \\ &= E\left[\sigma_t^2\left(u_{t-k}^2-\sigma_{t-k}^2\right)-\sigma_t^2\left(u_{t-k}^2-\sigma_{t-k}^2\right)\right]=0. \end{aligned}$$

Hence v_t is a martingale sequence, i.e. a zero mean and uncorrelated stochastic sequence. Additionally,

$$\begin{aligned} E\left(u_{t-1}^2 v_t\right) &= E\left[u_{t-1}^2\left(u_t^2-\sigma_t^2\right)\right] = E\left[u_{t-1}^2 E\left(u_t^2|I_{t-1}\right)-u_{t-1}^2\sigma_t^2\right] \\ &= E\left[u_{t-1}^2\sigma_t^2-u_{t-1}^2\sigma_t^2\right]=0. \end{aligned}$$

Consequently, equation (7) describes an $AR(1)$ model. This in turn means that the process u_t^2 is **1st order stationary** iff $|\alpha|<1$, consequently y_t is **2nd order stationary** under the same condition.

Furthermore,

$$E\left(y_t\right)=E\left(\mu+u_t\right)=\mu+E\left(u_t\right)=\mu+E\left[E\left(u_t|I_{t-1}\right)\right]=\mu+E\left[0\right]=\mu,$$

and

$$Var\left(y_t\right)=Var\left(\mu+u_t\right)=Var\left(u_t\right)=E\left(u_t^2\right)=\frac{c}{1-\alpha}.$$

The **autocorrelation function** of the process u_t^2 (say $\rho_k^{(u^2)}$) is given by the autocorrelation function of the $AR(1)$ model, i.e.

$$corr\left(u_t^2,u_{t-k}^2\right)=\rho_k^{(u^2}=\alpha^k.$$

Notice that as σ_t^2 is a conditional variance then it must be positive (with probability 1). This is achieved by imposing the so-called **positivity constraints**, i.e. σ_t^2 is positive with probability 1 $(P[\sigma_t^2>0]=1)$ if and only if

$$c>0 \quad and \quad \alpha\geq 0.$$

Estimation

To estimate the parameters of the model in (6), we employ the maximum likelihood. From (6) we get that

$$y_t|I_{t-1} \sim N\left(\mu, \sigma_t^2\right) \quad and \quad \sigma_t^2 = c + \alpha u_{t-1}^2.$$

Assuming that $u_0 = 0$ we have that the likelihood is given by

$$L(\mu, c, \alpha|u_0) = \prod_{t=1}^{T} L\left(y_t|I_{t-1}, u_0\right),$$

and the log-likelihood is:

$$\ell(\mu, c, \alpha|u_0) = -\frac{T}{2}\ln\left(2\pi\right) - \frac{1}{2}\sum_{t=1}^{T}\ln\sigma_t^2 - \sum_{t=1}^{T}\frac{\left(y_t-\mu\right)^2}{2\sigma_t^2}.$$

The first order conditions are:

$$\frac{\partial\ell(\mu, c, \alpha|u_0)}{\partial\mu} = \sum_{t=1}^{T}\frac{1}{\sigma_t^2}\alpha u_{t-1} + \sum_{t=1}^{T}\frac{\left(y_t-\mu\right)\sigma_t^2 - \alpha u_{t-1}\left(y_t-\mu\right)^2}{\sigma_t^4} \quad where \quad u_0 = 0.$$

$$\frac{\partial\ell(\mu, c, \alpha|u_0)}{\partial c} = -\frac{1}{2}\sum_{t=1}^{T}\frac{1}{\sigma_t^2} + \sum_{t=1}^{T}\frac{\left(y_t-\mu\right)^2}{2\sigma_t^4}.$$

$$\frac{\partial\ell(\mu, c, \alpha|u_0)}{\partial\alpha} = -\frac{1}{2}\sum_{t=1}^{T}\frac{1}{\sigma_t^2}u_{t-1}^2 + \sum_{t=1}^{T}\frac{\left(y_t-\mu\right)^2}{2\sigma_t^4}u_{t-1}^2 \qquad where \qquad u_0 = 0.$$

It is obvious that the conditions do not have explicit solutions.

6.2 GARCH(1,1)

The Generalised ARCH models of order $(1, 1)$ are given by

$$y_t = \mu + u_t \quad where \quad u_t|I_{t-1} \sim N\left(0, \sigma_t^2\right) \quad and \quad \sigma_t^2 = c + \alpha u_{t-1}^2 + \beta\sigma_{t-1}^2 \quad (8)$$

where y_t is the observed process and $I_{t-1} = \{y_{t-1}, y_{t-2}, ...\}$. Notice that given I_{t-1} σ_t^2 is known, i.e. non-stochastic. Of course unconditionally, i.e. when the conditioning set is the empty set, σ_t^2 is stochastic. Notice that we can also write

$$y_t = \mu + u_t \quad where \quad \frac{u_t}{\sqrt{\sigma_t^2}} = z_t \sim iid \quad N\left(0, 1\right) \quad and \quad \sigma_t^2 = c + \alpha u_{t-1}^2 + \beta\sigma_{t-1}^2.$$

Properties

Again, to capitalise on the properties of the $ARMA$ models, we employ the Pantula parameterisation, i.e.

$$u_t^2 = c+(\alpha+\beta)\,u_{t-1}^2+\beta\left(\sigma_{t-1}^2-u_{t-1}^2\right)+u_t^2-\sigma_t^2 \Rightarrow u_t^2 = c+(\alpha+\beta)\,u_{t-1}^2+v_t-\beta v_{t-1} \quad (9)$$

where $v_t = u_t^2-\sigma_t^2$, which is a martingale sequence as:

$$E\left(v_t\right) = E\left[E\left(u_t^2-\sigma_t^2\right)|I_{t-1}\right] = E\left[E\left(u_t^2|I_{t-1}\right)-\sigma_t^2\right] = E\left[\sigma_t^2-\sigma_t^2\right] = 0,$$

and

$$\begin{aligned} E\left(v_t v_{t-k}\right) &= E\left[\left(u_t^2-\sigma_t^2\right)\left(u_{t-k}^2-\sigma_{t-k}^2\right)\right] = E\left[E\left(u_t^2-\sigma_t^2\right)\left(u_{t-k}^2-\sigma_{t-k}^2\right)|I_{t-1}\right] \\ &= E\left[E\left(u_t^2|I_{t-1}\right)\left(u_{t-k}^2-\sigma_{t-k}^2\right)-\sigma_t^2\left(u_{t-k}^2-\sigma_{t-k}^2\right)\right] \\ &= E\left[\sigma_t^2\left(u_{t-k}^2-\sigma_{t-k}^2\right)-\sigma_t^2\left(u_{t-k}^2-\sigma_{t-k}^2\right)\right] = 0. \end{aligned}$$

Furthermore,

$$E\left(y_t\right) = E\left(\mu+u_t\right) = \mu+E\left(u_t\right) = \mu+E\left[E\left(u_t|I_{t-1}\right)\right] = \mu+E\left[0\right] = \mu,$$

and

$$Var\left(y_t\right) = Var\left(\mu+u_t\right) = Var\left(u_t\right) = E\left(u_t^2\right) = \frac{c}{1-\alpha-\beta},$$

given that $|\alpha+\beta| < 1$, for stationarity reasons.

The **autocorrelation function** of the process u_t^2 is given by the autocorrelation function of the $ARMA(1,1)$ model, i.e.

$$corr\left(u_t^2, u_{t-k}^2\right) = \rho_k^{(u^2} = \begin{cases} \frac{\alpha[1-(\alpha+\beta)\beta]}{1-2\alpha\beta-\beta^2} & for \quad k=1 \\ (\alpha+\beta)^{k-1}\rho_1 & for \quad k \geq 2 \end{cases}.$$

Notice that as σ_t^2 is a conditional variance then it must be positive (with probability 1). This is achieved by imposing the so-called **positivity constraints**, i.e. σ_t^2 is positive with probability 1 if and only if

$$c > 0, \quad \beta \geq 0 \quad and \quad \alpha \geq 0.$$

Estimation

To estimate the parameters of the model in (8), we employ the maximum likelihood. From (8) we get that

$$y_t|I_{t-1} \sim N\left(\mu, \sigma_t^2\right) \quad and \quad \sigma_t^2 = c + \alpha u_{t-1}^2 + \beta \sigma_{t-1}^2.$$

Assuming that $u_0 = 0$ and $\sigma_0^2 = \frac{c}{1-\alpha-\beta}$, the unconditional variance, we have that the likelihood is given by

$$L(\mu, c, \alpha, \beta | u_0, \sigma_0^2) = \prod_{t=1}^{T} L\left(y_t | I_{t-1}, u_0, \sigma_0^2\right),$$

and the log-likelihood is:

$$\ell(\mu, c, \alpha, \beta | u_0, \sigma_0^2) = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln\sigma_t^2 - \sum_{t=1}^{T}\frac{(y_t-\mu)^2}{2\sigma_t^2}.$$

The first order conditions are:

$$\begin{aligned}
\frac{\partial \ell(\mu, c, \alpha, \beta | u_0, \sigma_0^2)}{\partial \mu} &= -\frac{1}{2}\sum_{t=1}^{T}\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \mu} + \sum_{t=1}^{T}\frac{2\left(y_t-\mu\right)\sigma_t^2 + \frac{\partial \sigma_t^2}{\partial \mu}\left(y_t-\mu\right)^2}{2\sigma_t^4} \\
where \quad \frac{\partial \sigma_t^2}{\partial \mu} &= -2\alpha u_{t-1} + \beta\frac{\partial \sigma_{t-1}^2}{\partial \mu}, \qquad u_0 = 0, \qquad \frac{\partial \sigma_0^2}{\partial \mu} = 0.
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \ell(\mu, c, \alpha, \beta | u_0, \sigma_0^2)}{\partial c} &= -\frac{1}{2}\sum_{t=1}^{T}\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial c} + \sum_{t=1}^{T}\frac{\left(y_t-\mu\right)^2}{2\sigma_t^4}\frac{\partial \sigma_t^2}{\partial c} \\
where \quad \frac{\partial \sigma_t^2}{\partial \mu} &= 1 + \beta\frac{\partial \sigma_{t-1}^2}{\partial \mu}, \qquad \frac{\partial \sigma_0^2}{\partial c} = \frac{1}{1-\alpha-\beta}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \ell(\mu, c, \alpha, \beta | u_0, \sigma_0^2)}{\partial \alpha} &= -\frac{1}{2}\sum_{t=1}^{T}\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \alpha} + \sum_{t=1}^{T}\frac{\left(y_t-\mu\right)^2}{2\sigma_t^4}\frac{\partial \sigma_t^2}{\partial \alpha} \\
where \quad \frac{\partial \sigma_t^2}{\partial \alpha} &= u_{t-1}^2 + \beta\frac{\partial \sigma_{t-1}^2}{\partial \alpha}, \qquad u_0 = 0, \qquad \frac{\partial \sigma_0^2}{\partial \alpha} = -\frac{c}{\left(1-\alpha-\beta\right)^2}.
\end{aligned}$$

$$\begin{aligned}\frac{\partial \ell(\mu,c,\alpha,\beta|u_0,\sigma_0^2)}{\partial \beta} &= -\frac{1}{2}\sum_{t=1}^{T}\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \beta}+\sum_{t=1}^{T}\frac{(y_t-\mu)^2}{2\sigma_t^4}\frac{\partial \sigma_t^2}{\partial \beta}\\ where\quad \frac{\partial \sigma_t^2}{\partial \beta} &= \sigma_{t-1}^2+\beta\frac{\partial \sigma_{t-1}^2}{\partial \beta},\quad \sigma_0^2=\frac{c}{1-\alpha-\beta},\quad u_0=0,\\ \frac{\partial \sigma_0^2}{\partial \alpha} &= -\frac{c}{(1-\alpha-\beta)^2}.\end{aligned}$$

It is obvious that the conditions do not have explicit solutions.

6.3 GQARCH(1,1)

The Generalised Quadratic ARCH models of order $(1,1)$ (Sentana. 1995) are given by

$$y_t=\mu+u_t \quad where \quad u_t|I_{t-1}\sim N\left(0,\sigma_t^2\right) \quad and \quad \sigma_t^2=c+\alpha\left(u_{t-1}-\gamma\right)^2+\beta\sigma_{t-1}^2 \tag{10}$$

where y_t^* is the observed process and $I_{t-1}=\{y_{t-1},y_{t-2},...\}$. Notice that given I_{t-1} σ_t^2 is known, i.e. non-stochastic. Of course unconditionally, i.e. when the conditioning set is the empty set, σ_t^2 is stochastic. Notice that the conditional variance equation can be reparameterized as:

$$\sigma_t^2=\omega+\alpha u_{t-1}^2-\delta u_{t-1}+\beta\sigma_{t-1}^2 \quad where \quad \omega=c+\gamma^2 \quad and \quad \delta=2\alpha\gamma.$$

To explore the properties of the model the last parameterisation is very useful.

Furthermore, recall that the equation in (10) can be also written as:

$$y_t=\mu+u_t \quad where \quad \frac{u_t}{\sqrt{\sigma_t^2}}=z_t \quad iid\sim N\left(0,1\right) \quad and \quad \sigma_t^2=c+\alpha\left(u_{t-1}-\gamma\right)^2+\beta\sigma_{t-1}^2$$

Properties

Again, to capitalise on the properties of the $ARMA$ models, we employ the Pantula parameterisation, i.e.

$$u_t^2=\omega+\left(\alpha+\beta\right)u_{t-1}^2-\delta u_{t-1}+v_t-\beta v_{t-1}$$

where $v_t = u_t^2 - \sigma_t^2$, which is a martingale sequence as:

$$E(v_t) = E\left[E\left(u_t^2 - \sigma_t^2\right)|I_{t-1}\right] = E\left[E\left(u_t^2|I_{t-1}\right) - \sigma_t^2\right] = E\left[\sigma_t^2 - \sigma_t^2\right] = 0,$$

and

$$\begin{aligned} E(v_t v_{t-k}) &= E\left[\left(u_t^2 - \sigma_t^2\right)\left(u_{t-k}^2 - \sigma_{t-k}^2\right)\right] = E\left[E\left(u_t^2 - \sigma_t^2\right)\left(u_{t-k}^2 - \sigma_{t-k}^2\right)|I_{t-1}\right] \\ &= E\left[E\left(u_t^2|I_{t-1}\right)\left(u_{t-k}^2 - \sigma_{t-k}^2\right) - \sigma_t^2\left(u_{t-k}^2 - \sigma_{t-k}^2\right)\right] \\ &= E\left[\sigma_t^2\left(u_{t-k}^2 - \sigma_{t-k}^2\right) - \sigma_t^2\left(u_{t-k}^2 - \sigma_{t-k}^2\right)\right] = 0. \end{aligned}$$

Furthermore,

$$E(y_t) = E(\mu + u_t) = \mu + E(u_t) = \mu + E\left[E\left(u_t|I_{t-1}\right)\right] = \mu + E[0] = \mu,$$

and

$$\begin{aligned} Var(y_t) &= Var(\mu + u_t) = Var(u_t) = E\left(u_t^2\right) = E\left(z_t^2\sigma_t^2\right) \\ &= E\left(z_t^2\right)E\left(\sigma_t^2\right) = E\left(\sigma_t^2\right) = \frac{\omega}{1-\alpha-\beta}, \end{aligned}$$

given that $|\alpha + \beta| < 1$, for stationarity reasons.

It is not very difficult to prove that the **autocorrelation function** of the process u_t^2 is given by

$$Cov(u_t^2, u_{t-k}^2) = Cov(\sigma_t^2, \sigma_{t-k}^2) = (\alpha+\beta)^k \left[\frac{2\omega^2\alpha^2 + 4\alpha^2\delta^2\omega(1-\alpha-\beta)}{\left(1-3\alpha^2-\beta^2-2\alpha\beta\right)(1-\alpha-\beta)^2}\right],$$

provided that $3\alpha^2 + \beta^2 + 2\alpha\beta < 1$, a stronger condition than $\alpha + \beta < 1$, and

$$Cov(u_t^2, u_{t-k}) = E(u_t^2 u_{t-k}) = E(\sigma_t^2 u_{t-k}) = -2(\alpha+\beta)^{k-1}\frac{\delta\alpha\omega}{1-\alpha-\beta}$$

Notice that as σ_t^2 is a conditional variance then it must be positive (with probability 1). This is achieved by imposing the so-called **positivity constraints**, i.e. σ_t^2 is positive with probability 1 if and only if

$$c > 0, \quad \beta \geq 0 \quad and \quad \alpha \geq 0.$$

Estimation

To estimate the parameters of the model in (10), we employ the maximum likelihood. From (10) we get that

$$y_t|I_{t-1} \sim N\left(\mu, \sigma_t^2\right) \quad and \quad \sigma_t^2 = \omega + \alpha u_{t-1}^2 - \delta u_{t-1} + \beta\sigma_{t-1}^2.$$

Assuming that $u_0 = 0$ and $\sigma_0^2 = \frac{c}{1-\alpha-\beta}$, the unconditional variance, we have that the likelihood is given by

$$L(\mu, \omega, \alpha, \beta, \delta|u_0, \sigma_0^2) = \prod_{t=1}^{T} L\left(y_t|I_{t-1}, u_0, \sigma_0^2\right),$$

and the log-likelihood is:

$$\ell(\mu, \omega, \alpha, \beta, \delta|u_0, \sigma_0^2) = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln\sigma_t^2 - \sum_{t=1}^{T}\frac{(y_t-\mu)^2}{2\sigma_t^2}.$$

The first order conditions are similar to $GARCH(1,1)$ process they are recursive and do not have explicit solutions.

6.4 EGARCH(1,1)

The Exponential $GARCH$ models of order $(1,1)$ (Nelson 1991) are given by

$$\begin{aligned} y_t &= \mu + u_t \quad where \quad u_t|I_{t-1} \sim N\left(0, \sigma_t^2\right) \\ and \quad \ln\sigma_t^2 &= c + \alpha z_{t-1} + \gamma\left(|z_{t-1}| - E\,|z_{t-1}|\right) + \beta\ln\sigma_{t-1}^2 \\ where \quad z_t &= \frac{u_t}{\sqrt{\sigma_t^2}} \quad iid \sim N(0,1) \end{aligned} \tag{11}$$

where y_t^* is the observed process and $I_{t-1} = \{y_{t-1}, y_{t-2}, ...\}$. Notice that given I_{t-1} σ_t^2 is known, i.e. non-stochastic. Of course unconditionally, i.e. when the conditioning set is the empty set, σ_t^2 is stochastic. Furthermore under the assumed normality $E\,|z_t| = \sqrt{\frac{2}{\pi}}$ for all $t\prime s$.

Compared to $GARCH(1,1)$ models, the $EGARCH(1,1)$ models **do not** require any **positivity constraints** on the parameters so that the conditional variance is positive. Furthermore, the 2^{nd} order stationarity of $EGARCH\,(1,1)$ is satisfied if $|\beta| < 1$ whereas the 2^{nd} order stationarity of $GARCH\,(1,1)$ requires $|\alpha + \beta| < 1$ (see equation (10)), i.e. the sum of two coefficients to be less than 1. This constrain is difficult to impose.

Properties

The process y_t is **2nd order Stationarity** if and only if $|\beta| < 1$. The unconditional mean and variance of y_t^* is:

$$E\,(y_t) = E\,(\mu + u_t) = \mu + E\,(u_t) = \mu + E\,[E\,(u_t|I_{t-1})] = \mu + E\,[0] = \mu,$$

and

$$\begin{aligned} Var\,(y_t) &= Var\,(\mu + u_t) = Var\,(u_t) = E\left(u_t^2\right) = E\left(z_t^2\sigma_t^2\right) = E\left(z_t^2\right)E\left(\sigma_t^2\right) = E\left(\sigma_t^2\right) \\ &= \exp\left(\frac{c - \gamma\sqrt{\frac{2}{\pi}}}{1-\beta}\right)\prod_{i=0}^{\infty}\left[\Phi\left(\beta^i\gamma^*\right)\exp\left(\frac{\beta^{2i}\left(\gamma^*\right)^2}{2}\right) + \exp\left(\frac{\beta^{2i}\delta^2}{2}\right)\Phi\left(\beta^i\delta\right)\right], \end{aligned}$$

where $\gamma^* = \gamma + \alpha$, $\delta = \gamma - \alpha$ and $\Phi\,(k)$ is the value of the cumulative standard Normal evaluated at k, i.e. $\Phi\,(k) = \int_{-\infty}^{k}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)dx$.

The **autocovariance function** of the process u_t^2 is rather complicated and given by

$$\begin{aligned} Cov[u_t^2, u_{t-k}^2] &= Cov[\sigma_t^2, \sigma_{t-k}^2] = \exp\left(2\frac{c - \gamma\sqrt{\frac{2}{\pi}}}{1-\beta}\right) \times \\ &\left\{\begin{array}{c} \varpi_k^{**}\prod_{i=0}^{k-1}\left[\exp\left(\frac{\beta^{2i}(\gamma^*)^2}{2}\right)\Phi(\beta^i\gamma^*) + \exp\left(\frac{\beta^{2i}\delta^2}{2}\right)\Phi(\beta^i\delta)\right] \\ -E\left(\sigma_t^2\right) \end{array}\right\}, \end{aligned}$$

where γ^* and δ as before and

$$\varpi_k^{**} = \prod_{i=0}^{\infty} \left\{ \begin{array}{c} \exp\left(\frac{[(1+\beta^k)\beta^i\gamma^*]^2}{2}\right)\Phi\left[\left(1+\beta^k\right)\beta^i\gamma^*\right] \\ +\exp\left(\frac{[(1+\beta^k)\beta^i\delta]^2}{2}\right)\Phi\left[\left(1+\beta^k\right)\beta^i\delta\right] \end{array} \right\}.$$

The **dynamic asymmetry (leverage)** is given by

$$\begin{aligned} Cov[u_t^2, u_{t-k}] &= E[u_t^2 u_{t-k}] = E[\sigma_t^2 u_{t-k}] = \beta^{k-1}\exp\left(\frac{3}{2}\frac{c-\gamma\sqrt{\frac{2}{\pi}}}{1-\beta}\right)\varpi_k^* \\ &\prod_{i=0}^{k-2}\left[\exp\left(\frac{\beta^{2i}\left(\gamma^*\right)^2}{2}\right)\Phi(\beta^i\gamma^*) + \exp\left(\frac{\beta^{2i}\delta^2}{2}\right)\Phi(\beta^i\delta)\right] \\ &\left[\gamma^*\Phi(\beta^{k-1}\gamma^*)\exp\left(\frac{\beta^{2k-2}\left(\gamma^*\right)^2}{2}\right) - \delta\Phi(\beta^{k-1}\delta)\exp\left(\frac{\beta^{2k-2}\delta^2}{2}\right)\right] \end{aligned}$$

where

$$\varpi_k^* = \prod_{i=0}^{\infty} \left[\begin{array}{c} \exp\left(\frac{[(\frac{1}{2}+\beta^k)\beta^i\gamma^*]^2}{2}\right)\Phi\left[(\tfrac{1}{2}+\beta^k)\beta^i\gamma^*\right] \\ +\exp\left(\frac{[(\frac{1}{2}+\beta^k)\beta^i\delta]^2}{2}\right)\Phi\left[(\tfrac{1}{2}+\beta^k)\beta^i\delta\right] \end{array} \right].$$

Estimation

To estimate the parameters of the model in (11), we employ the maximum likelihood. Assuming that $z_0 = 0$ and $\ln\sigma_0^2 = \frac{c}{1-\beta}$, the unconditional variance, we have that the likelihood is given by

$$L(\mu, c, \alpha, \beta, \gamma | z_0, \sigma_0^2) = \prod_{t=1}^{T} L\left(y_t | I_{t-1}, z_0, \ln\sigma_0^2\right),$$

and the log-likelihood is:

$$\ell(\mu, c, \alpha, \beta, \gamma | z_0, \sigma_0^2) = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln\sigma_t^2 - \sum_{t=1}^{T}\frac{(y_t-\mu)^2}{2\sigma_t^2}.$$

The first order conditions are again recursive and consequently do not have explicit solutions.

6.5 Stochastic Volatility of Order 1

Let us consider the following SV(1) model:

$$\begin{aligned} y_t &= \mu + u_t \quad where \quad u_t = z_t\sqrt{\sigma_t^2}, \qquad (12) \\ \ln(\sigma_t^2) &= \alpha_0 + \beta\ln(\sigma_{t-1}^2) + \eta_{t-1} \\ and \quad \begin{pmatrix} z_t \\ \eta_t \end{pmatrix} \quad iid &\sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma_\eta \\ \rho\sigma_\eta & \sigma_\eta^2 \end{pmatrix}\right]. \end{aligned}$$

Notice that given the information set $J_{t-1} = \{y_{t-1}, y_{t-2}, ...\eta_{t-1}, \eta_{t-2}, ...\}$. the conditional variance σ_t^2 is known, i.e. non-stochastic. However, our information set is $I_{t-1} = \{y_{t-1}, y_{t-2}, ...\} \subset J_{t-1}$, as the only observed process is the y_t, and consequently σ_t^2 is stochastic (this is the origin of the name of this model). This is the main difference between this model and various heteroskedastic models where the conditional variance is non-stochastic, given I_{t-1}.

Properties

Under the normality assumption and for $|\beta| < 1$ the process $\{y_t$ is **covariance (2nd order)** and **strictly stationary**, and we can also invert the $AR(1)$ representation of the conditional variance and write the MA representation as $\ln(h_t) = \frac{\alpha_0}{1-\beta} + \sum_{i=0}^{\infty}\psi_i\eta_{t-1-i}$ where $\psi_i = \beta^i$. The unconditional mean and variance of y_t is:

$$E(y_t) = E(\mu + u_t) = \mu + E(u_t) = \mu + E[E(u_t|I_{t-1})] = \mu + E[0] = \mu,$$

and

$$\begin{aligned} Var(y_t) &= Var(\mu + u_t) = Var(u_t) = E\left(u_t^2\right) = E\left(z_t^2\sigma_t^2\right) = E\left(z_t^2\right)E\left(\sigma_t^2\right) = E\left(\sigma_t^2\right) \\ &= \exp\left(\frac{\alpha_0}{1-\beta} + \frac{\sigma_\eta^2}{2(1-\beta^2)}\right). \end{aligned}$$

The **autocorrelation function** of the process u_t^2 is less complicated, as compared to $EGARCH(1,1)$, and given by

$$\rho(u_t^2, u_{t-k}^2) = \frac{(1+\beta^{2k-2}\rho^2\sigma_\eta^2)\exp\left(\beta^k \frac{\sigma_\eta^2}{1-\beta^2}\right) - 1}{3\exp\left(\frac{\sigma_\eta^2}{1-\beta^2}\right) - 1}$$

and

$$Cov(\sigma_t^2, \sigma_{t-k}^2) = \exp\left(2\frac{\alpha_0}{1-\beta} + \frac{\sigma_\eta^2}{1-\beta^2}\right)\left[\exp\left(\beta^k \frac{\sigma_\eta^2}{1-\beta^2}\right) - 1\right].$$

The **dynamic asymmetry (leverage)** is given by

$$E\left(\sigma_t^2 u_{t-k}\right) = \rho\sigma_\eta\beta^{k-1}\exp\left(\frac{3\alpha_0}{2(1-\beta)} + \sigma_\eta^2\frac{\beta^k + \frac{5}{4}}{2(1-\beta^2)}\right).$$

Notice that for $\rho = 0$, as in Harvey et. al. (1994), the correlation of the squared errors is the same as in Shephard (1996) and Taylor (1984). Furthermore, in this case we have that $E\left(h_t\varepsilon_{t-k}\right) = 0$, i.e. the leverage effect is 0. Moreover, it is easy to prove that $\rho(u_t^2, u_{t-k}^2) \le \frac{1}{3}\rho(\sigma_t^2, \sigma_{t-k}^2) < \frac{1}{3}$.

Estimation

The main difficulty of employing the SV(1) model is that it is not obvious how to evaluate the likelihood, i.e. the distribution of y_t given $I_{t-1} = \{y_{t-1}, y_{t-2}, ...\}$. One solution is to employ the Generalised Method of Moments. Another possibility is to apply the Kalman Filter and maximise the likelihood of one-step prediction error, as if this distribution was normal (quasi-likelihood). Other possibilities include MCMC and Bayesian methods.

7 Conditional Heteroskedastic in Mean Models

These are models which have the time varying conditional variance in the mean specification as well, i.e.

$$r_t = \delta h_t + \varepsilon_t, \qquad \varepsilon_t = z_t\sqrt{h_t} \tag{13}$$

where

$$h_t^\lambda = \omega + \alpha\lambda f^\nu(z_{t-1})h_{t-1}^\lambda + \beta h_{t-1}^\lambda + \phi_\eta \lambda \eta_{t-1}, \tag{14}$$

$$f(z_{t-1}) = |z_{t-1} - b| - \gamma(z_{t-1} - b), \tag{15}$$

and

$$\begin{pmatrix} z_t \\ \eta_t \end{pmatrix} \sim iid \quad N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma_\eta \\ \rho\sigma_\eta & \sigma_\eta^2 \end{pmatrix}\right]. \tag{16}$$

A variety of conditionally heteroskedastic in mean models are nested in the above parametrization. However, here we consider only three possible values of the parameter λ and two possible values of ν, i.e. $\lambda \in \{0, 1/2, 1\}$and $\nu \in \{1, 2\}$. Among others, the following models are nested in our parameterization:

- Standard GARCH (GARCH(1,1); Bollerslev (1986) and GARCH(1,1)-M; Engle et. al. (1987)) ($\lambda = 1$, $\nu = 2$, $\phi_\eta = b = \gamma = 0$).

- Nonlinear Asymmetric GARCH (NLGARCH(1,1); Engle and Ng (1993)) ($\lambda = 1$, $\nu = 2$, $\phi_\eta = \gamma = 0$).

- Glosten-Jagannathan-Runkle GARCH (GJR-GARCH(1,1) and GJR-GARCH(1,1)-M; Glosten et al. (1993)) ($\lambda = 1$, $\nu = 2$, $\phi_\eta = b = 0$).

- Threshold GARCH (TGARCH(1,1); Zakoian (1994)) ($\lambda = \frac{1}{2}$, $\nu = 1$, $\phi_\eta = b = 0$).

- Absolute Value GARCH (AVGARCH(1,1); Taylor (1986) and Schwert (1989) ($\lambda = \frac{1}{2}$, $\nu = 1$, $\phi_\eta = b = \gamma = 0$).

- Exponential GARCH (EGARCH(1,1) and EGARCH(1,1)-M; Nelson (1991)) ($\lambda = 0$, $\nu = 1$, $\phi_\eta = 0$, $b = 0$).

- Stochastic Volatility (SV(1); Harvey and Shephard (1996)) ($\lambda = 0$, $\alpha = 0$, $\phi_\eta = 1$).

Furthermore, we consider the Quadratic GARCH (GQARCH(1,1) and GQARCH(1,1)-M; Sentana (1995)) which does not fall within the modelling of equations 14 and 15.

In principle, we can impose restrictions of the parameters of the model described in equations 13-16, so that the stochastic process $\{r_t\}$ is second order and strictly stationary. We further assume that the process $\{h_t\}$ started from some finite value in the distant past and the conditional variance is positive with probability one. Now we have that under the appropriate conditions of second order and strict stationarity we have that $\gamma_k = Cov(r_t, r_{t-k})$ is a quadratic function in δ given by (Arvanitis and Demos 2004 and 2004b):

$$\begin{aligned}
\gamma_k &= Cov(r_t, r_{t-k}) = E\left(r_t r_{t-k}\right) - E(r_t)E(r_{t-k}) \\
&= E\left[\left(\delta h_t + \varepsilon_t\right)\left(\delta h_{t-k} + \varepsilon_{t-k}\right)\right] - E(\delta h_t + \varepsilon_t)E(\delta h_{t-k} + \varepsilon_{t-k}) \\
&= E\left(\delta h_t \delta h_{t-k} + \varepsilon_t \delta h_{t-k} + \delta h_t \varepsilon_{t-k} + \varepsilon_t \varepsilon_{t-k}\right) - \delta^2 E(h_t)E(h_{t-k}) \\
&= \delta^2\left[E\left(h_t h_{t-k}\right) - E(h_t)E(h_{t-k})\right] + \delta E\left(h_t \varepsilon_{t-k}\right) \\
&= \delta^2 Cov(h_t, h_{t-k}) + \delta E\left(h_t \varepsilon_{t-k}\right)
\end{aligned}$$

It is important to mention that the assumption of normality of the errors is by no means a necessary condition for the above formula. It can be weakened and substituted for higher moment conditions. From formula above it is immediately obvious that if $E\left(h_t \varepsilon_{t-k}\right)$ is zero, as in the GARCH-M model of Engle et al. (1987), then the k-order autocovariance of the series has the sign of the autocovariance of the conditional variance irrespective of the value of δ; this could explain the poor empirical performance of GARCH-M models (see Fiorentini and Sentana (1998)). Ideally, one would like to employ a model which can be compatible with either negative or positive mean autocorrelations, and potentially different from of the sign of the autocorrelation of the conditional variance. As an example consider the volatility clustering observed in financial data. This implies that the autocorrelation of the returns' conditional variance is positive. However, short horizon returns are positively autocorrelated, whereas long horizon ones are negatively autocorrelated (see Poterba and Summers (1988)).

It is a well documented fact that in all applied work with financial data, the estimated values of β are positive, indicating the observed volatility clustering, and the estimated values of ρ are negative due to the asymmetry effect (see Harvey and Shephard (1996)). These two facts imply that $E\left(h_t \varepsilon_{t-k}\right)$ is negative and $Cov(h_t, h_{t-k})$ is positive, for any k. Consequently, γ_k is negative for $\delta \in \left(0, -E\left(h_t \varepsilon_{t-k}\right)/Cov(h_t, h_{t-k})\right)$ and is positive for $\delta \in \left(-E\left(h_t \varepsilon_{t-k}\right)/Cov(h_t, h_{t-k}), \infty\right)$, i.e. there are positive values of δ, the price of risk, that can incorporate both negative and positive autocorrelations of the observed series. Furthermore, notice that for positive β the autocorrelation of the squared errors, of any order, is positive for any ρ.

It turns out that not all first-order dynamic heteroskedasticity in mean models are compatible with negative autocorrelations of observed series. In fact, the GARCH(1,1)-M, AVGARCH(1,1)-M and SV(1)-M with uncorrelated mean and variance errors models are not compatible with data sets that exhibit negative autocorrelations. This statement rests on the assumption of symmetric distributions of the errors, which are mostly employed in applied work. On the other hand, all the other models considered above are compatible with either positive or negative autocorrelations of the observed data. In fact, under the assumption of symmetric distribution of the errors, models which incorporate the leverage effect can also accommodate the observed series autocorrelation of either sign.

7.1 The Log Conditional Variance Models

Let us consider the model in equations 13-16 and reparameterize equation 14 so that now it reads

$$\frac{h_t^\lambda - 1}{\lambda} = \omega^* + \alpha f^\nu(z_{t-1}) h_{t-1}^\lambda + \beta \frac{h_{t-1}^\lambda - 1}{\lambda} + \phi_\eta \eta_{t-1}, \quad where \quad \omega^* = \frac{\omega - 1 - \beta}{\lambda}.$$

Taking the limit now as $\lambda \to 0$ we have a family of first-order processes where the natural log of the conditional variance is modeled in the variance equation (Demos 2002), i.e.

$$\ln h_t = \omega^* + \alpha f^\nu(z_{t-1}) + \beta \ln h_{t-1} + \phi_\eta \eta_{t-1}$$

Now for $\alpha = 0$ and $\phi_\eta = 1$ we have the first-order Stochastic Volatility of Harvey and Shephard (1996), whereas for $\nu = 1$ and $\phi_\eta = 0$ we have the first-order Exponential GARCH of Nelson (1991). In fact we can consider a generalization of the above equation, i.e.

$$\ln h_t = \alpha_0 + \alpha \sum_{i=0}^{\infty} \theta_i f(z_{t-1-i}) + \phi_\eta \sum_{i=0}^{\infty} \psi_i \eta_{t-1-i} \tag{17}$$

where $\theta_0 = \psi_0 = 1$. Under the normality assumption, $\{\delta_t\}$, $\{\ln h_t\}$ and $\{h_t\}$ are covariance and strictly stationary if $|\varphi| < 1$, $\sum_{i=0}^{\infty} \theta_i^2$ and $\sum_{i=0}^{\infty} \psi_i^2$ are finite. In such a case, the observed process $\{r_t\}$ is covariance and strictly stationary as well. Furthermore, an advantage of using $\lambda = 0$, over $\lambda = 1/2$ or $\lambda = 1$, is that in this case we do not need to constrain further the parameter space to get the positivity of the conditional variance.

Let us now state the following Lemmas which will be needed in the sequel (see Demos 2002 for a proof).

Lemma 7.1 *For the following Gaussian AR(1) process* $y_t = \alpha_0 + \beta y_{t-1} + \eta_t$, *where* $|\beta| < 1$, $\eta_t \sim i.i.d.N(0, \sigma_\eta^2)$ *and* τ *any finite number we have that*

$$Cov\left(\exp(\tau y_t), \exp(\tau y_{t-k})\right) = \exp\left(2\tau\mu_y + \tau^2\sigma_y^2\right)\left[\exp\left(\frac{\tau^2\beta^k\sigma_\eta^2}{1-\beta^2}\right) - 1\right]$$

where $\mu_y = \frac{\alpha_0}{1-\beta}$ *and* $\sigma_y^2 = \frac{\sigma_\eta^2}{1-\beta^2}$.

A special case of the above Lemma for $\tau = 1$ can be found in Granger and Newbold (1976).

Lemma 7.2 *If* $\begin{pmatrix} z \\ \eta \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma_\eta \\ \rho\sigma_\eta & \sigma_\eta^2 \end{pmatrix}\right]$ *then for any* κ, τ *finite real numbers we have that*

$$E\left\{\exp\left[\tau f(z)\right]\exp(\kappa\eta)\right\} = \Phi(A-b)\exp\left(\Gamma\right) + \exp\left(\Delta\right)\Phi(b-B)$$

$$E\left\{z\exp\left[\tau f(z)\right]\exp(\kappa\eta)\right\} = A\Phi(A-b)\exp\left(\Gamma\right) + B\Phi(b-B)\exp\left(\Delta\right)$$

$$\begin{aligned} E\left\{z^2\exp\left[\tau f(z)\right]\exp(\kappa\eta)\right\} &= \frac{2\tau}{\sqrt{2\pi}}\exp\left(\frac{\kappa^2\sigma_\eta^2 - (b-\kappa\sigma_\eta\rho)^2}{2}\right) \\ &\quad +(1+A^2)\Phi(A-b)\exp\left(\Gamma\right) + (1+B^2)\Phi(b-B)\exp\left(\Delta\right) \end{aligned}$$

where

$$\Delta = \frac{\left[\kappa\sigma_\eta - \tau(1+\gamma)\right]^2}{2} + \tau(1+\gamma)[b+\kappa\sigma_\eta(1-\rho)],$$

$$\Gamma = \frac{\left[\kappa\sigma_\eta + \tau(1-\gamma)\right]^2}{2} - \tau(1-\gamma)[b+\kappa\sigma_\eta(1-\rho)],$$

$$B = \kappa\rho\sigma_\eta - \tau(1+\gamma) \quad and \quad A = \kappa\rho\sigma_\eta + \tau(1-\gamma).$$

Theorem 7.3 *For the models described in 13, 15, 16 and 17 above we have that:*

$$\begin{aligned} E\left(h_t\varepsilon_{t-k}\right) &= \exp\left(\frac{3\alpha_0}{2}\right)\varpi_k^{(1/2)}\prod_{i=0}^{k-2}\left[\exp\left(\Gamma_i\right)\Phi(A_{0,i}^{(0)}-b) + \exp\left(\Delta_i\right)\Phi(b-B_i)\right] \\ &\quad \left[A_{0,k-1}^{(0)}\Phi(A_{0,k-1}^{(0)}-b)\exp\left(\Gamma_{0,k-1}^{(0)}\right) + B_{0,k-1}^{(0)}\Phi(b-B_{0,k-1}^{(0)})\exp\left(\Delta_{0,k-1}^{(0)}\right)\right], \end{aligned}$$

$$E\left(h_th_{t-k}\right) = \exp(2\alpha_0)\varpi_k^{(1)}\prod_{i=0}^{k-1}\Phi(A_{0,i}^{(0)}-b)\exp\left(\Gamma_{0,i}^{(0)}\right) + \exp\left(\Delta_{0,i}^{(0)}\right)\Phi(b-B_{0,i}^{(0)}),$$

$$E(\varepsilon_t^2\varepsilon_{t-k}^2) = E(h_t h_{t-k})D\left[\Phi(A_{0,k-1}^{(0)} - b)\exp\left(\Gamma_{0,k-1}^{(0)}\right) + \exp\left(\Delta_{0,k-1}^{(0)}\right)\Phi(b - B_{0,k-1}^{(0)})\right]^{-1},$$

and

$$E\left(h_t\right) = \exp(2\alpha_0)\varpi_0^{(0)}$$

where

$$\varpi_k^{(j)} = \prod_{i=0}^{\infty}\left[\exp\left(\Gamma_{k,i}^{(j)}\right)\Phi(A_{k,i}^{(j)} - b) + \exp\left(\Delta_{k,i}^{(j)}\right)\Phi(b - B_{k,i}^{(j)})\right],$$

$$A_{k,i}^{(j)} = \phi_\eta(j\psi_i + \psi_{i+k})\rho\sigma_\eta + \alpha(j\theta_i + \theta_{i+k})(1-\gamma),$$

$$B_{k,i}^{(j)} = \phi_\eta(j\psi_i + \psi_{i+k})\rho\sigma_\eta - \alpha(j\theta_i + \theta_{i+k})(1+\gamma),$$

$$\begin{aligned}\Delta_{k,i}^{(j)} &= \frac{[\phi_\eta(j\psi_i + \psi_{i+k})\sigma_\eta - \alpha(j\theta_i + \theta_{i+k})(1+\gamma)]^2}{2} \\ &\quad + \alpha(j\theta_i + \theta_{i+k})(1+\gamma)[b + \phi_\eta(j\psi_i + \psi_{i+k})\sigma_\eta(1-\rho)],\end{aligned}$$

$$\begin{aligned}\Gamma_{k,i}^{(j)} &= \frac{[\phi_\eta(j\psi_i + \psi_{i+k})\sigma_\eta + \alpha(j\theta_i + \theta_{i+k})(1-\gamma)]^2}{2} \\ &\quad - \alpha(j\theta_i + \theta_{i+k})(1-\gamma)[\phi_\eta(j\psi_i + \psi_{i+k})\sigma_\eta(1-\rho) + b],\end{aligned}$$

and

$$\begin{aligned}D &= \frac{2\alpha\theta_{k-1}}{\sqrt{2\pi}}\exp\left(\frac{(\phi_\eta\psi_{k-1})^2\sigma_\eta^2 - (b - \phi_\eta\psi_{k-1}\sigma_\eta\rho)^2}{2}\right) \\ &\quad + \left(1 + (A_{0,k-1}^{(0)})^2\right)\Phi(A_{0,k-1}^{(0)} - b)\exp\left(\Gamma_{0,k-1}^{(0)}\right) \\ &\quad + \left(1 + (B_{k-1}^{(0)})^2\right)\Phi(b - B_{0,k-1}^{(0)})\exp\left(\Delta_{0,k-1}^{(0)}\right).\end{aligned}$$

(Proof Demos 2002)

Let us now turn our attention to individual models where the log of the conditional variance is modeled.

Stochastic Volatility in Mean

Consider the model in equations 13, 15-17 and impose the constraints $\phi_\eta = 1$, and $\alpha = 0$ to get the $SV - M$ model, i.e.

$$r_t = \delta h_t + \varepsilon_t, \qquad \varepsilon_t = z_t \sqrt{h_t} \tag{18}$$

where

$$\ln(h_t) = \alpha_0 + \sum_{i=0}^{\infty} \psi_i \eta_{t-1-i} \tag{19}$$

and $(z_t, \eta_t)^{/}$ are as in equations 16.

Notice that this model nests all invertible ARMA representations of the conditional variance process, i.e. $\ln(h_t) = \alpha_0 + \sum_{i=1}^{p} \beta_i \ln(h_{t-i}) + \sum_{i=0}^{q} \xi_i \eta_{t-1-i}$ with the assumption that the roots of $\left[1 - \sum_{i=1}^{p} \beta_i x^i\right]$ lie inside the unit circle we can invert the ARMA representation of the conditional variance equation to get: $\ln(h_t) = \alpha_0^* + \sum_{i=0}^{\infty} \psi_i \eta_{t-1-i}$ where $\alpha_0^* = \frac{\alpha_0}{1-\beta_1-\beta_2-\ldots-\beta_p}$, $\psi_0 = 1$, and ψ_i, for $i = 1, 2, \ldots$, are the usual coefficients for the MA representation of an ARMA model (see e.g. Anderson (1994)). In terms of stationarity, and under the normality assumption, the square summability of the $\psi_i^{\prime} s$ and $|\varphi| < 1$ guarantee the second order and strict stationarity of the $\{h_t\}$ process, and consequently the covariance and strict stationarity of the $\{r_t\}$ process.

Now we can state the following Lemma:

Lemma 7.4 *For the SV-M above and under the assumptions of stationarity we have that:*

$$E\left(h_t \varepsilon_{t-k}\right) = \rho \sigma_\eta \psi_{k-1} \exp\left(\frac{3\alpha_0}{2} + \frac{\sigma_\eta^2}{2} \sum_{i=0}^{k-1} \psi_i^2 + \frac{\sigma_\eta^2}{2} \sum_{i=0}^{\infty} (\psi_{k+i} + \frac{1}{2}\psi_i)^2\right),$$

$$Cov(h_t, h_{t-k}) = \exp\left(2\alpha_0 + \sigma_\eta^2 \sum_{i=0}^{\infty} \psi_i^2\right) \left[\exp\left(\sigma_\eta^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}\right) - 1\right]$$

and

$$\rho(\varepsilon_t^2, \varepsilon_{t-k}^2) = \frac{(1+\psi_{k-1}^2\rho^2\sigma_\eta^2)\exp\left(\sigma_\eta^2\sum_{i=0}^{\infty}\psi_{i+k}\psi_i\right)-1}{3\exp\left(\sigma_\eta^2\sum_{i=0}^{\infty}\psi_i^2\right)-1}$$

Assume for the moment that δ is positive, i.e. the price of risk is positive. Then in light of the above lemma and theorem 7.3 there are positive values of δ associated with either negative or positive values of the autocovariance, γ_k, (autocorrelation) of the process $\{r_t\}$.

If there is no correlation between the mean and variance innovations, i.e. $\rho = 0$ as in Harvey et al. (1994), we have that $E\left(h_t\varepsilon_{t-k}\right) = 0$ and consequently γ_k is positive when $Cov(h_t, h_{t-k})$ is independent of the value of δ. If, on the other hand, the $Cov(h_t, h_{t-k})$ is negative, then there are positive values of δ that are compatible with negative values of γ_k, i.e. for $\delta \in (0, -(1-\varphi^2)Cov(h_t, h_{t-k})/\varphi^{k-1}\sigma_u^2 E\left(h_t h_{t-k}\right))$ γ_k is negative and for $\delta \in (-(1-\varphi^2)Cov(h_t, h_{t-k})/\varphi^{k-1}\sigma_u^2 E\left(h_t h_{t-k}\right), \infty)$ γ_k is positive.

Since in most applications a first order SV-M model is considered with time invariant coefficient for the conditional variance in the mean equation, let us now turn our analysis to this model.

First-Order Stochastic Volatility in Mean with Constant Coefficients

Let us consider the following SV(1)-M model:

$$\begin{aligned} r_t &= \delta_t h_t + \varepsilon_t \quad where \quad \varepsilon_t = z_t\sqrt{h_t}, \quad \ln(h_t) = \alpha_0 + \beta\ln(h_{t-1}) + \eta_{t-1} \\ and \quad \begin{pmatrix} z_t \\ \eta_t \end{pmatrix} &\sim iidN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma_\eta \\ \rho\sigma_\eta & \sigma_\eta^2 \end{pmatrix}\right]. \end{aligned}$$

Under the normality assumption and for $|\beta| < 1$ the process $\{r_t\}$ is covariance and strictly stationary, and we can also invert the AR(1) representation of the conditional variance and write the MA representation as $\ln(h_t) = \alpha_0^* + \sum_{i=0}^{\infty} \psi_i \eta_{t-1-i}$ where $\alpha_0^* = \alpha_0/(1-\beta)$ and $\psi_i = \beta^i$. Now we can state the following Lemma

Corollary 1 *For the above SV-M we have that:*

$$E\left(h_t \varepsilon_{t-k}\right) = \rho \sigma_\eta \beta^{k-1} \exp\left(\frac{3\alpha_0}{2(1-\beta)} + \sigma_\eta^2 \frac{\beta^k + \frac{5}{4}}{2(1-\beta^2)}\right),$$

$$Cov(h_t, h_{t-k}) = \exp\left(2\frac{\alpha_0}{1-\beta} + \frac{\sigma_\eta^2}{1-\beta^2}\right)\left[\exp\left(\beta^k \frac{\sigma_\eta^2}{1-\beta^2}\right) - 1\right]$$

and

$$\rho(\varepsilon_t^2, \varepsilon_{t-k}^2) = \frac{(1 + \beta^{2k-2}\rho^2\sigma_\eta^2)\exp\left(\beta^k \frac{\sigma_\eta^2}{1-\beta^2}\right) - 1}{3\exp\left(\frac{\sigma_\eta^2}{1-\beta^2}\right) - 1}$$

To our knowledge, in all applied work with financial data, the estimated values of β are positive, indicating the observed volatility clustering, and the estimated values of ρ are negative due to the asymmetry effect (see Harvey and Shephard (1996)). These two facts imply that $E\left(h_t \varepsilon_{t-k}\right)$ is negative and $Cov(h_t, h_{t-k})$ is positive, for any k. Consequently, γ_k is negative for $\delta \in (0, -E\left(h_t \varepsilon_{t-k}\right)/Cov(h_t, h_{t-k}))$ and is positive for $\delta \in (-E\left(h_t \varepsilon_{t-k}\right)/Cov(h_t, h_{t-k}), \infty)$, i.e. there are positive values of δ, the price of risk, that can incorporate both negative and positive autocorrelations of the observed series. Furthermore, notice that for positive β the autocorrelation of the squared errors, of any order, is positive for any ρ.

Notice that for $\rho = 0$, as in Harvey et al. (1994), the correlation of the squared errors is same as in Shephard (1996) and Taylor (1984). Furthermore, in this case we have that $E\left(h_t \varepsilon_{t-k}\right) = 0$ and according to Theorem 1 γ_k is positive for any value of δ. Moreover, it is easy to prove that $\rho(\varepsilon_t^2, \varepsilon_{t-k}^2) \leq \frac{1}{3}\rho(h_t, h_{t-k}) < \frac{1}{3}$.

Exponential ARCH in Mean

Let us consider again the model in equations 13, 15-17. To derive the Time Varying (Mean) Parameter EGARCH(1,1)-M model set $\nu = 1$, $\phi_\eta = 0$ and $b = 0$, i.e.

$$r_t = \delta h_t + \varepsilon_t, \qquad \varepsilon_t = z_t \sqrt{h_t}$$
$$\ln(h_t) = \alpha_0^* + \sum_{i=0}^{\infty} \theta_i g\left(z_{t-1-i}\right), \qquad g\left(z_t\right) = \alpha_1 \left(|z_t| - \sqrt{\frac{2}{\pi}}\right) + \alpha_2 z_t,$$

where $z_t \sim iidN(0,1)$, $u_t \sim iidN(0,\sigma_u^2)$, z_t and u_t independent, δ_t as in (**??**) and $\theta_0 = 1$. Throughout this section we assume that $\alpha_1 \neq 0$. In case that $\alpha_1 = 0$ the above model is the same as the model in section 2.1 with $\rho = 1$. Notice that as it is the case for the SV-M model, this model nests all invertible ARMA representations of the conditional variance. Under the assumption of normality of $z_t\prime s$ the processes $\{\varepsilon_t\}$ and $\{h_t\}$ are covariance and strictly stationary if $\sum\limits_{i=0}^{\infty} \theta_i^2$ is finite (see Nelson (1991)). Consequently, under these conditions, the process $\{r_t\}$ is covariance and strictly stationary as well. We can state the following Lemma:

Lemma 7.5 *For the above EARCH-M model we have*

$$\begin{aligned} Cov[h_t, h_{t-k}] &= \exp[2(\alpha_0^* - \alpha_1 \sqrt{\frac{2}{\pi}} \sum_{i=0}^{\infty} \theta_i)] \\ &\left(\varpi_k^{**} \prod_{i=0}^{k-1} \exp\left(\frac{\theta_i^2 \alpha_1^{*2}}{2}\right) \Phi(\theta_i \alpha_1^*) + \exp\left(\frac{\theta_i^2 \alpha_2^{*2}}{2}\right) \Phi(\theta_i \alpha_2^*) - \varpi^2\right), \end{aligned}$$

where $\theta_0 = 1$, $\alpha_1^* = \alpha_1 + \alpha_2$, $\alpha_2^* = \alpha_1 - \alpha_2$, $\varpi = \prod\limits_{i=0}^{\infty} \left(\Phi(\theta_i \alpha_1^*) \exp\left(\frac{\theta_i^2 \alpha_1^{*2}}{2}\right) + \exp\left(\frac{\theta_i^2 \alpha_2^{*2}}{2}\right) \Phi(\theta_i \alpha_2^*)\right)$

and $\varpi_k^{**} = \prod_{i=0}^{\infty} \left[\begin{array}{c} \exp\left(\frac{[(\theta_i+\theta_{i+k})\alpha_1^*]^2}{2}\right)\Phi((\theta_i+\theta_{i+k})\alpha_1^*) \\ +\exp\left(\frac{[(\theta_i+\theta_{i+k})\alpha_2^*]^2}{2}\right)\Phi((\theta_i+\theta_{i+k})\alpha_2^*) \end{array} \right],$

$$
\begin{aligned}
E(h_t\varepsilon_{t-k}) &= \theta_{k-1}\exp\left(\frac{3}{2}(\alpha_0^* - \alpha_1\sqrt{\frac{2}{\pi}}\sum_{i=0}^{\infty}\theta_i)\right)\varpi_k^* \\
&\quad \prod_{i=0}^{k-2}\left[\exp\left(\frac{\theta_i^2\alpha_1^{*2}}{2}\right)\Phi(\theta_i\alpha_1^*) + \exp\left(\frac{\theta_i^2\alpha_2^{*2}}{2}\right)\Phi(\theta_i\alpha_2^*)\right] \\
&\quad \left[\alpha_1^*\Phi(\theta_{k-1}\alpha_1^*)\exp\left(\frac{\theta_{k-1}^2\alpha_1^{*2}}{2}\right) - \alpha_2^*\Phi(\theta_{k-1}\alpha_2^*)\exp\left(\frac{\theta_{k-1}^2\alpha_2^{*2}}{2}\right)\right]
\end{aligned}
$$

where $\varpi_k^* = \prod_{i=0}^{\infty} \left[\begin{array}{c} \exp\left(\frac{[(\frac{1}{2}\theta_i+\theta_{i+k})\alpha_1^*]^2}{2}\right)\Phi((\frac{1}{2}\theta_i+\theta_{i+k})\alpha_1^*) \\ +\exp\left(\frac{[(\frac{1}{2}\theta_i+\theta_{i+k})\alpha_2^*]^2}{2}\right)\Phi((\frac{1}{2}\theta_i+\theta_{i+k})\alpha_2^*) \end{array} \right].$

Also

$$
\begin{aligned}
Cov[\varepsilon_t^2, \varepsilon_{t-k}^2] &= Cov[h_t, h_{t-k}] + \theta_{k-1}E\left(h_t h_{t-k}\right) \left[\begin{array}{c} \exp\left(\frac{\theta_{k-1}^2(\alpha_1^*)^2}{2}\right)\mathbf{\Phi}(\theta_{k-1}\alpha_1^*) \\ +\exp\left(\frac{\theta_{k-1}^2(\alpha_2^*)^2}{2}\right)\mathbf{\Phi}(\theta_{k-1}\alpha_2^*) \end{array} \right]^{-1} \\
&\quad \left[\begin{array}{c} \alpha_1\sqrt{\frac{2}{\pi}} + \theta_{k-1}\alpha_1^{*2}\Phi(\theta_{k-1}\alpha_1^*)\exp\left(\frac{\theta_{k-1}^2\alpha_1^{*2}}{2}\right) \\ +\theta_{k-1}\alpha_2^{*2}\Phi(\theta_{k-1}\alpha_2^*)\exp\left(\frac{\theta_{k-1}^2\alpha_2^{*2}}{2}\right) \end{array} \right]
\end{aligned}
$$

Clearly the sign of $E(h_t\varepsilon_{t-k})$ depends on the sign of θ_{k-1} and the relative values of α_1^* and α_2^*. Let us now turn our attention to the first-order EGARCH in Mean model with time invariant δ_t.

First-Order EGARCH in Mean with Constant Coefficients

Consider now the following EGARCH(1)-M model:

$r_t = \delta h_t + \varepsilon_t$

$\varepsilon_t = z_t\sqrt{h_t}$ where $z_t \sim iid\ N(0,1)$ and

$\ln(h_t) = \alpha_0^* + \beta\ln(h_{t-1}) + g(z_{t-1})$ where $g(z_t) = \alpha_1\left(|z_t| - \sqrt{\frac{2}{\pi}}\right) + \alpha_2 z_t$.
Provided that $|\beta_1| < 1$ we can invert the above equation as:

$$\ln(h_t) = \alpha_0 + \sum_{i=0}^{\infty}\beta^i g(z_{t-1-i}) \tag{20}$$

where $\alpha_0 = \frac{\alpha_0^*}{1-\beta}$

Assuming normality and $|\beta| < 1$ we get the second order and strict stationarity of $\{r_t\}$ and we can state the following Corollary:

Corollary 2 *For the EGARCH(1)-M and under the assumptions that* $z_t \backsim i.i.d.N(0,1)$ *and* $|\beta| < 1$ *we have that*

$$Cov[h_t, h_{t-k}] = \exp[2(\frac{\alpha_0^*}{1-\beta} - \frac{\alpha_1}{1-\beta}\sqrt{\frac{2}{\pi}})]$$
$$\left(\varpi_k^{**}\prod_{i=0}^{k-1}\exp\left(\frac{\beta^{2i}\alpha_1^{*2}}{2}\right)\Phi(\beta^i\alpha_1^*) + \exp\left(\frac{\beta^{2i}\alpha_2^{*2}}{2}\right)\Phi(\beta^i\alpha_2^*) - \varpi^2\right),$$

where $\alpha_1^* = \alpha_1+\alpha_2$, $\alpha_2^* = \alpha_1-\alpha_2$, $\varpi = \prod_{i=0}^{\infty}\left(\Phi(\beta^i\alpha_1^*)\exp\left(\frac{\beta^{2i}\alpha_1^{*2}}{2}\right) + \exp\left(\frac{\beta^{2i}\alpha_2^{*2}}{2}\right)\Phi(\beta^i\alpha_2^*)\right)$ *and* $\varpi_k^{**} = \prod_{i=0}^{\infty}\left[\begin{array}{c}\exp\left(\frac{[(1+\beta^k)\beta^i\alpha_1^*]^2}{2}\right)\Phi((1+\beta^k)\beta^i\alpha_1^*) \\ +\exp\left(\frac{[(1+\beta^k)\beta^i\alpha_2^*]^2}{2}\right)\Phi((1+\beta^k)\beta^i\alpha_2^*)\end{array}\right]$,

$$E(h_t\varepsilon_{t-k}) = \beta^{k-1}\exp\left(\frac{3}{2}(\frac{\alpha_0^*}{1-\beta} - \frac{\alpha_1}{1-\beta}\sqrt{\frac{2}{\pi}})\right)\varpi_k^*$$
$$\prod_{i=0}^{k-2}\left[\exp\left(\frac{\beta^{2i}\alpha_1^{*2}}{2}\right)\Phi(\beta^i\alpha_1^*) + \exp\left(\frac{\beta^{2i}\alpha_2^{*2}}{2}\right)\Phi(\beta^i\alpha_2^*)\right]$$
$$\left[\alpha_1^*\Phi(\beta^{k-1}\alpha_1^*)\exp\left(\frac{\beta^{2k-2}\alpha_1^{*2}}{2}\right) - \alpha_2^*\Phi(\beta^{k-1}\alpha_2^*)\exp\left(\frac{\beta^{2k-2}\alpha_2^{*2}}{2}\right)\right]$$

where $\varpi_k^* = \prod_{i=0}^{\infty}\left[\begin{array}{c}\exp\left(\frac{[(\frac{1}{2}+\beta^k)\beta^i\alpha_1^*]^2}{2}\right)\Phi((\frac{1}{2}+\beta^k)\beta^i\alpha_1^*) \\ +\exp\left(\frac{[(\frac{1}{2}+\beta^k)\beta^i\alpha_2^*]^2}{2}\right)\Phi((\frac{1}{2}+\beta^k)\beta^i\alpha_2^*)\end{array}\right]$.

Also

$$Cov[\varepsilon_t^2, \varepsilon_{t-k}^2] = Cov[h_t, h_{t-k}] + \beta^{k-1} E\left(h_t h_{t-k}\right) \begin{bmatrix} \exp\left(\frac{\beta^{2k-2}(\alpha_1^*)^2}{2}\right) \mathbf{\Phi}(\beta^{k-1}\alpha_1^*) \\ +\exp\left(\frac{\beta^{2k-2}(\alpha_2^*)^2}{2}\right) \mathbf{\Phi}(\beta^{k-1}\alpha_2^*) \end{bmatrix}^{-1} \begin{bmatrix} \alpha_1\sqrt{\frac{2}{\pi}} + \beta^{k-1}\alpha_1^{*2}\Phi(\beta^{k-1}\alpha_1^*)\exp\left(\frac{\beta^{2k-2}\alpha_1^{*2}}{2}\right) \\ +\beta^{k-1}\alpha_2^{*2}\Phi(\beta^{k-1}\alpha_2^*)\exp\left(\frac{\beta^{2k-2}\alpha_2^{*2}}{2}\right) \end{bmatrix}$$

In Figure 1 we present the autocorrelation function of the conditional variance for an EGARCH(1)-M process when the parameter values are as in Nelson (1991).[1] Furthermore, we graph two approximations. The first one is employing the formula $\rho(h_t, h_{t-k}) = \beta^{k-1}\rho(h_t, h_{t-1})$, whereas the second is given by $\rho(h_t, h_{t-k}) = \left(\frac{\rho(h_t,h_{t-2})}{\rho(h_t,h_{t-1})}\right)^{k-2} \rho(h_t, h_{t-2})$, for $k \geq 2$. Notice that the exact ACF is always between these two approximations, closer to the second. Furthermore, for high positive values of β the first approximation is quite inaccurate, being more so as k increases.

It is clear that the sign of $E[h_t\varepsilon_{t-k}]$ depends on the relative values of β, α_1^*, and α_2^*. However, notice that under the assumptions of volatility clustering, leverage and asymmetry effects, i.e. $\beta > 0$, $\alpha_1 > 0$ and $\alpha_2 < 0$, we have that $\alpha_2^* > \alpha_1^*$ and $\alpha_2^* > 0$. Hence $\alpha_1^* \exp\left(\frac{(\beta^{k-1}\alpha_1^*)^2}{2}\right) \mathbf{\Phi}(\beta^{k-1}\alpha_1^*) - \alpha_2^* \exp\left(\frac{(\beta^{k-1}\alpha_2^*)^2}{2}\right) \mathbf{\Phi}(\beta^{k-1}\alpha_2^*) < 0$ as the exponential and the cumulative distribution functions are non-decreasing. Consequently, $E[h_t\varepsilon_{t-k}]$ is negative for any k.

Now if the random sequence $\{g\left(z_t\right)\}$ had a normal distribution, then we would be able to apply Lemma 1 and conclude that the autocovariance of the conditional variance is positive for positive β. However, although $\{g\left(z_t\right)\}$ is an independent

[1] Since Nelson (1991) estimated an EGARCH(2,1)-M, we use the largest estimated root as our β parameter.

sequence of random variables, its distribution is clearly non-normal. Nevertheless, under the assumptions of volatility clustering, leverage and asymmetry effects it is possible to prove that at least the first-order autocovariance is positive up to forth-order approximation, i.e. expanding the autocovariance with respect to β around zero. Under these facts we can say that either positive or negative autocorrelations of the observed series are compatible with positive expected value of the price of risk (see Theorem 1).

7.2 Comparing the Ln Models

Comparing the two models we can see that both models can incorporate positive and negative autocorrelations of the observed series with positive expected value of risk price. Furthermore, both models are compatible with cyclical effects in the ACF of the squared errors, for negative β. The SV-M models have a relative advantage over the EGARCH-M models, in that the autocorrelation functions of the conditional variance and the squared errors are easier to evaluate. Any other comparison between the two models is difficult due to the complexity of the formulae for the EGARCH model. Nevertheless, notice that the symmetric first-order EGARCH model, i.e. when $\alpha_1 = 0$, has the same ACFs for the conditional variance and the squared errors as a SV model with $\rho = 1$ and $\sigma_\eta^2 = \alpha_1^2$ (see Demos (2002) and He et al. (1999)).

Harvey and Shephard (1996) estimated a SV(1) model for the data in Nelson (1991). The estimates for β, ρ, and σ_η^2 were 0.9877, -0.66, and 0.016, respectively. In Figure 2 we graph the ACF of the conditional variance and the squared errors for the two models and the above mentioned parameter values. Notice that although the ACF of the two conditional variances are very close this is not the case for the ACF of the squared errors. In fact, we know that the ACF of the squared errors for first-order SV(-M) models is bound from above by 1/3 (see end of section 2.1). It seems that this bound is higher for the EGARCH(-M) models. However, the complexity of the formulae for these models makes the verification of this conjecture very difficult. Nevertheless it is possible to compare the symmetric first-order EGARCH(-M), i.e. when $\alpha_2 = 0$, with the SV(-M) one. For this we shall need the following Proposition.

Proposition 3 *For the symmetric Gaussian first-order EGARCH(-M) model we have that*

$$\rho(h_t, h_{t-k}) = \frac{\exp\left(\frac{\beta^k \alpha_1^2}{1-\beta^2}\right) \prod_{i=0}^{\infty} \frac{2\Phi((1+\beta^k)\beta^i \alpha_1)}{\prod_{j=ki}^{k(i+2)-1} 2\Phi(\beta^j \alpha_1)} - 1}{\exp\left(\frac{\alpha_1^2}{1-\beta^2}\right) \prod_{i=0}^{\infty} \frac{\Phi(2\beta^i \alpha_1)}{2\Phi^2(\beta^i \alpha_1)} - 1}$$

and

$$\rho(\varepsilon_t^2, \varepsilon_{t-k}^2) = \frac{\exp\left(\frac{\beta^k \alpha_1^2}{1-\beta^2}\right) \prod_{i=0}^{\infty} \frac{\left[2\Phi((1+\beta^k)\beta^i \alpha_1)\right]}{\prod_{j=ki}^{k(i+2)-1} 2\Phi(\beta^j \alpha_1)}}{3\exp\left(\frac{\alpha_1^2}{1-\beta^2}\right) \prod_{i=0}^{\infty} \frac{\Phi(2\beta^i \alpha_1)}{2\Phi^2(\beta^i \alpha_1)} - 1} \left[1 + \frac{\beta^{k-1}\alpha_1 \sqrt{\frac{1}{2\pi}}}{\exp\left(\frac{\beta^{2k-2}(\alpha_1)^2}{2}\right) \mathbf{\Phi}(\beta^{k-1}\alpha_1)} + \beta^{2k-2}\alpha_1^2\right] - \frac{1}{3\exp\left(\frac{\alpha_1^2}{1-\beta^2}\right) \prod_{i=0}^{\infty} \frac{\Phi(2\beta^i \alpha_1)}{2\Phi^2(\beta^i \alpha_1)} - 1}.$$

Specifically, for $k = 1$ and β close to 1 we have:

$$\rho(\varepsilon_t^2, \varepsilon_{t-1}^2) \approx \frac{1}{3} \exp\left(-\frac{\alpha_1^2}{2}\right) \left[1 + \frac{\exp\left(-\frac{\alpha_1^2}{2}\right)\alpha_1}{\sqrt{2\pi}\mathbf{\Phi}(\alpha_1)} + \alpha_1^2\right].$$

In light of the above Proposition, it is immediately obvious that the upper bound for the ACF of the squared errors is higher than 1/3. For high values of $|\alpha_1|$ the first-order ACF is dominated by the $\frac{1}{3}\exp\left(-\frac{\alpha_1^2}{2}\right)$ term. In fact, solving numerically the first order condition for $-2 < \alpha_1 < 2$, we find that the maximum is around 0.468 ($\alpha_1 \approx 0.84$).

Recall that the ACF for the conditional variance and the squared errors of a symmetric SV(1)(-M) model are given by $\rho(h_t, h_{t-k}) = \frac{\exp\left(\beta^k \frac{\sigma_\eta^2}{1-\beta^2}\right) - 1}{\exp\left(\frac{\sigma_\eta^2}{1-\beta^2}\right) - 1}$ and

$\rho(\varepsilon_t^2, \varepsilon_{t-k}^2) = \frac{\exp\left(\beta^k \frac{\sigma_\eta^2}{1-\beta^2}\right)-1}{3\exp\left(\frac{\sigma_\eta^2}{1-\beta^2}\right)-1}$, respectively. Let us assume that the estimated parameters of such a model and an EGARCH(1)(-M) process are such, so that they result in $\rho^{EGARCH}(h_t, h_{t-k}) \approx \rho^{SV}(h_t, h_{t-k})$, for small k. Then, provided that α_1 is positive, we get that $\rho^{EGARCH}(\varepsilon_t^2, \varepsilon_{t-k}^2) > \rho^{SV}(\varepsilon_t^2, \varepsilon_{t-k}^2)$. However, any other comparison of the two models is extremely difficult. Nevertheless, as the applicability of symmetric conditional heteroskedasticity models is rather limited (see e.g. Gallant et al. (1997)), and a comparison of the properties of general order asymmetric SV-M with those of EGARCH-M is extremely difficult, we do not pursue this topic any further.

7.3 Modelling the Standard Deviation

All proofs of this section can be found in Arvanitis and Demos (2004) and (2004b). Let us now consider the following alternative set of constraints on the model in equations 13-15: $\phi_\eta = 0$, i.e. there is no stochastic error in the conditional variance equation, $\lambda = \frac{1}{2}$, i.e. the standard deviation is modeled. Let us also relax the normality assumption and instead assume that $z_t's$ are $iid\ (0,1)$ random variables. This yields the following model:

$$r_t = \delta_{t-1} h_t + \varepsilon_t, \qquad \varepsilon_t = z_t \sqrt{h_t} \tag{21}$$

where

$$h_t^{1/2} = \omega + \alpha f^\nu(z_{t-1}) h_{t-1}^{1/2} + \beta h_{t-1}^{1/2}, \tag{22}$$

$$f(z_{t-1}) = |z_{t-1} - b| - \gamma(z_{t-1} - b), \tag{23}$$

and

$$z_t \sim iid(0,1). \tag{24}$$

When $\nu = 1$ and $\phi_u = \varphi = b = 0$ we get the TGARCH(1,1)-M of Zakoian (1994) whereas if $\gamma = 0$ as well we get the AVGARCH(1,1)-M of Taylor (1986) and Schwert (1989). Under the assumption $E\left(\alpha f^{\nu}(z) + \beta\right)^4 < 1$ both the conditional variance process, $\{h_t\}$, and the observed process, $\{r_t\}$, are covariance and strictly stationary (see He and Terasvirta (1999)). Furthermore, for the positivity of the conditional variance we need $\beta > 0$, $\alpha > 0$ and $-1 \leq \gamma \leq 1$. In this setup we can state the following Lemma.

Lemma 7.6 *If the $z_t's$ are $i.i.d.$ $(0,1)$ random variables with $E\left(\alpha f^{\nu}(z_t) + \beta\right)^4 < 1$ then for $k > 0$ we have that:*

$$E\left(h_t \varepsilon_{t-k}\right) = B^{k-1}\widetilde{B}E(h^{3/2}) + 2\omega\widetilde{A}E(h)\frac{B^k - A^k}{B - A}$$

and

$$Cov(h_t, h_{t-k}) = B^k V(h) + 2\omega AE\left(h^{3/2}\right)\frac{B^k - A^k}{B - A} - 2\omega^2 E(h)\frac{A(B^k - A^k)}{(1 - A)(B - A)},$$

which is positive for any k, where

$$E(h) = \omega^2\frac{1 + A}{(1 - A)(1 - B)}, \quad E(h^{3/2}) = \omega^3\frac{1 + AB + 2A + 2B}{(1 - A)(1 - B)(1 - \Gamma)},$$

$$V(h) = E(h^2) - E^2(h), \quad E(h_t^2) = \omega^4\frac{(1 + B)\left[3(A + \Gamma) + 1 + A\Gamma\right] + 4(B + A\Gamma)}{(1 - A)(1 - B)(1 - \Gamma)(1 - \Delta)},$$

$$A = E\left(\alpha f^{\nu}(z) + \beta\right), \quad \widetilde{A} = E\left[z\left(\alpha f^{\nu}(z) + \beta\right)\right] = \alpha E\left(z f^{\nu}(z)\right),$$

$$B = E\left(\alpha f^{\nu}(z) + \beta\right)^2, \quad \widetilde{B} = E\left[z\left(\alpha f^{\nu}(z) + \beta\right)^2\right] = \alpha^2 E\left(z f^{2\nu}(z)\right) + 2\alpha\beta\widetilde{A},$$

$$\Gamma = E\left(\alpha f^{\nu}(z) + \beta\right)^3 \quad and \quad \Delta = E\left(\alpha f^{\nu}(z) + \beta\right)^4.$$

Unfortunately the expressions derived are too complicated to draw any conclusions. In fact, in order to draw any conclusions in terms of the signs of various expectations we need, at least, to specify the distribution of the $z_t's$. Hence, let us assume that they are normally distributed and moreover that $b = 0$.

7.4 The First-Order TGARCH-M under Normality

The Gaussian the TGARCH(1,1)-M model (see Zakoian (1994)) is defined as follows..

$$\begin{aligned} r_t &= \delta h_t + \varepsilon_t, \quad where \quad \varepsilon_t = z_t\sqrt{h_t}, \\ h_t^{1/2} &= \omega + \alpha f(z_{t-1})h_{t-1}^{1/2} + \beta h_{t-1}^{1/2}, f(z_{t-1}) = |z_{t-1}| - \gamma z_{t-1}, \quad and \quad z_t \sim iidN(0,1). \end{aligned}$$

For stationarity assume that $E\left(\alpha f^{\nu}(z) + \beta\right)^4 < 1$ whereas for positivity of the conditional variance assume that $\beta > 0$, $\alpha > 0$ and $-1 < \gamma < 1$. We can now state the following Corollary.

Corollary 4 *For the above first-order TGARCH-M model and for $k > 0$ we have that:*

$$E\left(h_t\varepsilon_{t-k}\right) = -B^{k-1}\alpha^2\gamma(1+2\beta)E(h^{3/2}) - 2\omega\alpha\gamma E(h)\frac{B^k - A^k}{B - A}$$

and

$$Cov(h_t, h_{t-k}) = B^k V(h) + 2\omega A E\left(h^{3/2}\right)\frac{B^k - A^k}{B - A} - 2\omega^2 E(h)\frac{A(B^k - A^k)}{(1-A)(B-A)}$$

where

$$E(h) = \omega^2\frac{1+A}{(1-A)(1-B)}, \quad E(h^{3/2}) = \omega^3\frac{1+AB+2A+2B}{(1-A)(1-B)(1-\Gamma)},$$

$$\begin{aligned} E(h^2) &= \omega^4\frac{(1+B)\left[3(A+\Gamma)+1+A\Gamma\right]+4(B+A\Gamma)}{(1-A)(1-B)(1-\Gamma)(1-\Delta)}, \\ V(h) &= E(h^2) - E^2(h), A = \alpha\sqrt{\frac{2}{\pi}} + \beta, \end{aligned}$$

$B = \alpha^2(1+\gamma^2) + 2\alpha\beta\sqrt{\frac{2}{\pi}} + \beta^2$, $\Gamma = 2\sqrt{\frac{2}{\pi}}\alpha^3(1+3\gamma^2) + 3\alpha\beta\left(\alpha(1+\gamma^2) + \beta\sqrt{\frac{2}{\pi}}\right) + \beta^3$ *and* $\Delta = 3\alpha^4(1+6\gamma^2+\gamma^4) + 2\alpha^2\beta(1+\gamma^2)\left(4\sqrt{\frac{2}{\pi}}\alpha + 3\beta\right) + 4\sqrt{\frac{2}{\pi}}\alpha\beta^3 + \beta^4$.
(for proof see Arvanitis Demos (2004)

It is now apparent that the covariance, of any order, between the conditional variance and the lagged error is negative, then there are positive values of δ which are compatible with either negative or positive values of γ_k.

If we impose one additional restriction that the leverage parameter, γ, is zero, we get the Absolute Value GARCH(1,1)-M of Taylor (1986) and Schwert (1989). Notice that for this model the covariance (correlation) of the conditional variance with the lagged errors is zero (see above Corollary). Consequently, the autocovariance of the observed process can have only the sign of the autocovariance of the conditional variance which is positive.

7.5 Modelling the Conditional Variance

For $\lambda = 1$, $\phi_\eta = 0$ we get the following GARCH-M model:

$$r_t = \delta h_t + \varepsilon_t, \qquad \varepsilon_t = z_t\sqrt{h_t} \tag{25}$$

where

$$h_t = \omega + \alpha f^\nu(z_{t-1})h_{t-1} + \beta h_{t-1}, \tag{26}$$

$$f(z_{t-1}) = |z_{t-1} - b| - \gamma(z_{t-1} - b), \tag{27}$$

and

$$z_t \sim iid(0, 1). \tag{28}$$

When $\nu = 2$, $\phi_u = \phi = \gamma = 0$ we get the NLGARCH(1,1)-M of Engle and Ng (1993). Furthermore, if $b = 0$, but $\gamma \neq 0$, we get the GJR-GARCH(1,1)-M of Glosten et al. (1993) and if in addition $\gamma = 0$ we get the GARCH(1,1) of Bollerslev (1986) and GARCH(1,1)-M of Engle et al. (1987).

Under the assumption that $|\varphi| < 1$ and $E\left(f^\nu(z_{t-1}) + \beta\right)^2 < 1$ the process $\{r_t\}$ is second order and strictly stationary (see He and Terasvirta (1999)), and if $\alpha > 0$ and $\beta > 0$ the positivity constraint for the conditional variance is satisfied, as well. Consequently, we can state the following Lemma:

Lemma 7.7 *For the first-order GARCH model in equations 25-28 we have that*

$$E(h_t \varepsilon_{t-k}) = \alpha E\left[z f^{\nu}(z)\right] A^{k-1} E(h^{3/2}),$$

and

$$Cov(h_t, h_{t-k}) = \omega^2 A^k \frac{B - A^2}{(1-A)^2 (1-B)}$$

where $A = E\left(f^{\nu}(z_{t-1}) + \beta\right)$ *and* $B = E\left(f^{\nu}(z_{t-1}) + \beta\right)^2$.

It is obvious that if $E\left[z f^{\nu}(z)\right] = 0$ then the sign of the autocovariance of the observed series, $\{r_t\}$, is positive for any value of δ.

In most models considered in applied work $\nu = 2$. In such a case $E\left[z f^{\nu}(z)\right] = E(z^3)(1+\gamma^2) - 2b(1+\gamma^2) - 2\gamma E\left[|z-b|(z^2 - bz)\right]$. Consequently, for the GARCH(1,1)-M models we get that $E\left[z f^{\nu}(z)\right] = E(z^3)$, for the GJR-GARCH(1,1)-M we get that $E\left[z f^{\nu}(z)\right] = E(z^3)(1+\gamma^2) - 2\gamma E(z^2|z|)$ and for the NLGARCH(1,1)-M model we get that $E\left[z f^{\nu}(z)\right] = E(z^3) - 2b$. Consequently, for the GARCH(1,1)-M model we get that for symmetric distributions of $z_t's$ the autocovariance of the observed series can only be positive independent of the value of δ.

7.6 First-Order GQARCH in Mean

Consider the following GQARCH(1,1) in Mean model:

$$r_t = \delta h_t + \varepsilon_t, \quad \varepsilon_t = z_t \sqrt{h_t} \quad where \quad z_t \sim iid(0,1)$$

sequence of random variables and

$$h_t = \omega + \alpha(\varepsilon_{t-1} - b)^2 + \beta h_{t-1}.$$

In terms of stationarity we need $\alpha^2 E(z^4) + \beta^2 + 2\alpha\beta < 1$, whereas we need both α and β to be positive for positivity of the conditional variance (see Sentana (1995)).

The Sastry Pantula parameterization (see Bollerslev (1986)) of the conditional variance of the GARCH model applies also to GQARCH-M, i.e.

$$\varepsilon_t^2 = \omega^* + (\alpha+\beta)\varepsilon_{t-1}^2 - 2\alpha b\varepsilon_{t-1} + \nu_t - \beta\nu_{t-1}$$

where $\omega^* = \omega + b^2$ and $\nu_t = \varepsilon_t^2 - h_t$ is a martingale difference sequence, as

$$E(\nu_t) = E(\varepsilon_t^2 - h_t) = E\left(E_{t-1}(\varepsilon_t^2) - h_t\right) = 0,$$

and for $k > 0$

$$\begin{aligned} E(\nu_t\nu_{t-k}) &= E\left((\varepsilon_t^2 - h_t)(\varepsilon_{t-k}^2 - h_{t-k})\right) \\ &= E(h_t\varepsilon_{t-k}^2) - E(h_t\varepsilon_{t-k}^2) - E(h_th_{t-k}) + E(h_th_{t-k}) = 0. \end{aligned}$$

We can state the following Lemma.

Lemma 7.8 *For a GQARCH(1,1) in Mean model we have:*

$$E(h_t\varepsilon_{t-k}) = E(\varepsilon_t^2\varepsilon_{t-k}) = (\alpha+\beta)^{k-1}\left[\alpha E(z^3)E(h^{3/2}) - 2\alpha bE\left(h\right)\right]$$

and

$$Cov(h_t, h_{t-k}) = (\alpha+\beta)^k\left[E(h^2) - E^2(h)\right] = (\alpha+\beta)^kV(h)$$

where $E(h^2) = \frac{\omega^{*2} + \left(4\alpha^2b^2 + 2\omega^*\alpha + 2\omega^*\beta\right)E(h) - 4\alpha^2bE(z^3)E(h^{3/2})}{1-\alpha^2E(z^4)-\beta^2-2\alpha\beta}$, $E(h) = \frac{\omega^*}{1-\alpha-\beta}$, *and* $E(h^{\frac{3}{2}})$ *a finite number.*

From the above lemma it is clear that if the z_t*'s* are symmetrically distributed, i.e. $E(z^3) = 0$, and the asymmetry parameter b is positive then, $E(h_t\varepsilon_{t-k})$ is less than zero for any value of k. Consequently, the autocovariance of the process is negative for $\delta \in (0, -2\alpha b(\alpha+\beta)^{-1}E\left(h\right)/V(h))$ and positive for $\delta \in (-2\alpha b(\alpha+\beta)^{-1}E\left(h\right)/V(h), \infty)$. Furthermore, notice that if in the GQARCH(1,1)-M we set $b = 0$ we get the GARCH(1,1)-M. Hence, if z_t has a symmetric distribution, the autocovariance of any order of the GARCH(1,1)-M can have only non-negative values independent of the value δ (see also Hong (1991) and Fiorentini and Sentana (1998)).

7.7 Conclusions

It turns out that not all first-order dynamic heteroskedasticity in mean models, with time invariant conditional variance coefficient in the mean, δ, are compatible with negative autocorrelations of observed series. In fact, the GARCH(1,1)-M, AVGARCH(1,1)-M and SV(1)-M with uncorrelated mean and variance errors models are not compatible with data sets that exhibit negative autocorrelations. This statement rests on the assumption of symmetric distributions of the errors, which are mostly employed in applied work. On the other hand, all the other models considered here are compatible with either positive or negative autocorrelations of the observed data. In fact, under the assumption of symmetric distribution of the errors, models which incorporate the leverage effect can also accommodate the observed series autocorrelation of either sign.

Finally, the expressions for the autocovariances of the square residuals for the SV-M and EGARCH-M models make it possible to see how well the theoretical properties of the models fit the observed data, something which is important for the identification of the order of the conditional variance process along the lines of Bollerslev (1984) and Nelson (1991). A relative advantage of the SV-M model over the EGARCH-M one is that for the former it is simpler to evaluate the autocovariance function of the conditional variance and the squared errors.

In fact, in terms of formulae simplicity the SV-M and GARCH with $\lambda = 1$ and $\nu = 2$ models have an advantage over the rest. Comparing the two models, in terms of the autocorrelation function of the squared errors we have the following Proposition:

Proposition 5 *If the decline from the first to the second order autocorrelation in the conditional variance or squared errors is the same for a SV(1)-M and a G(Q)ARCH(1,1)-M model then the autoregressive parameter in the log conditional variance equation of the SV model, β, is bigger or equal to the persistence* *of the G(Q)ARCH model, i.e. the sum of GARCH and ARCH coefficients.*

In other words, if the β in the SV model is equal to the variance persistence parameter, as defined by the sum of the ARCH and GARCH parameters, the autocorrelation function of the conditional variance and the squared errors for the SV model decline at a rate which is at most equal and generally lower to those of the G(Q)ARCH model. The β and persistence parameters play an important part in the stationarity of the SV(1)-M and G(Q)ARCH(1,1)-M models. Consequently, if the objective function of a maximization procedure were the matching of the ACFs of the conditional variance or the squared errors the G(Q)ARCH(1,1)-M models would produce estimated values that are within the stationary region of the parameter space with higher probability.

However, in the majority of the applications of these models, a Quasi Maximum Likelihood method is employed for parameter estimation, leaving the structure in the squared errors as a diagnostic tool. Nevertheless, intuition suggests that considering an in Mean model would make these results relevant to estimation as well. This is something that we investigate now. Another interesting topic would be to investigate the autocorrelation function of the squared observed variables and the so called Taylor effect. Again this is a complete project on its own that is under consideration.

Of course our analysis is by no means exhaustive. Many members of the dynamic heteroskedasticity in mean family are not analyzed, e.g. models where λ and ν take any non-integer values in the range $[0, 2]$ as in Ding et al. (1993), or models where a fourth-order power of the errors is added in the conditional variance equation, as in Yang and Bewley (1995).

Part III

Non-stationarity

8 Non-stationary processes

For this chapter stationary process means second order weakly stationary, i.e. the mean, variance and autocovariances of the process are independent of t. There are several reasons to pay attention on stationarity.

1) For a Radmon Walk (see below), a shock at time t will have the same effect on the process at time t, $t+1$, $t+2$,..... The same is not true for a stationary process, e.g. for the stationary $AR(1)$ the effect on the process at $t+j$ is α^j, where α is the AR coefficient. In other words, the Impulse-Response function of a Random Walk is 1 at all horizons.

2) The "spurious regression" problems, i.e. when two non-stationary variables seem to be connected when they are not related to one another (see below on this).

3) The forecast variance of a simple Random Walk grows linearly with the forecast horizon.

4) Heuristically, the autocorrelation function of a Random Walk is 1 for any lag. In effect this means that the correlogram would die out very slowly.

5) Dealing with non-stationary variables in a regression framework would invalidate most of the asymptotic analysis, i.e. the usual 't-ratios' could not follow a t-distribution, and the F-statistic would not follow an F-distribution etc.

There are two models that are usually employed to describe non-stationarity. First, the $Trend\ Stationary\ (TS)$ process:

$$y_t = a + bt + u_t, \quad where \quad u_t\ is\ WN, \tag{29}$$

and WN stands for White Noise. In fact the term $a + bt$ is the deterministic term, as it is completely predictable. It can be any polynomial function in t, e.g. it can be of the following form: a, $a + bt$, $a + bt + ct^2$, etc.

Second, the $random\ walk\ (RW)$ with $drift$, i.e.

$$x_t = b + x_{t-1} + z_t, \quad where \quad z_t\ is\ WN, \tag{30}$$

where b is called the drift. Assuming that $x_0 = 0$, substituting backwards we get that

$$x_t = 2b + x_{t-2} + z_t + z_{t-1} = ... = bt + z_t + z_{t-1} + ... + z_1$$

justifying in this way the name drift.

8.1 Trend Stationary Process

Let us consider the TS process in equation (29). Assuming normality, i.e. $u_t \sim N(0, \sigma^2)$, all classical assumptions apply and consequently, the usual t-statistic and F-statistic have the standard Students-t and F distributions. However, as soon as we drop the assumption of normality things change.

Let us rewrite equation (29) as:

$$\begin{aligned} y_t &= (1, t)\begin{pmatrix} a \\ b \end{pmatrix} + u_t = x_t'\gamma + u_t, \quad where \quad x_t = \begin{pmatrix} 1 \\ t \end{pmatrix} \qquad (31) \\ \gamma &= \begin{pmatrix} a \\ b \end{pmatrix} \quad and \quad u_t\ iid\ with\ E(u_t) = 0, E(u_t^2) = \sigma^2 \quad and \quad E(u_t^4) < \infty. \end{aligned}$$

Let $\widehat{\gamma}$ denote the OLS estimator of γ based on n observations, i.e.

$$\widehat{\gamma} = \begin{pmatrix} \widehat{a} \\ \widehat{b} \end{pmatrix} = \left(\sum_{t=1}^{n} x_t x_t'\right)^{-1} \sum_{t=1}^{n} x_t y_t.$$

Substituting out y_t, employing equation (31), we get

$$\widehat{\gamma} = \left(\sum_{t=1}^{n} x_t x_t'\right)^{-1} \sum_{t=1}^{n} x_t \left(x_t'\gamma + u_t\right) = \gamma + \left(\sum_{t=1}^{n} x_t x_t'\right)^{-1} \sum_{t=1}^{n} x_t u_t$$

and it follows that

$$\begin{aligned} \widehat{\gamma} - \gamma &= \left(\sum_{t=1}^{n} x_t x_t'\right)^{-1} \sum_{t=1}^{n} x_t u_t \Rightarrow \\ \sqrt{n}\left(\widehat{\gamma} - \gamma\right) &= \left(\frac{1}{n}\sum_{t=1}^{n} x_t x_t'\right)^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t u_t. \end{aligned}$$

Under standard assumptions we have that $\frac{1}{n}\sum_{t=1}^{n} x_t x_t' \xrightarrow{P} M$, where M a non-singular matrix and $\xrightarrow{P}$ denotes convergence in probability. Further, $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_t u_t \xrightarrow{D} N\left(0, \sigma^2 M\right)$. It follows that $\sqrt{n}\left(\widehat{\gamma} - \gamma\right) \xrightarrow{D} N\left(0, \sigma^2 M^{-1}\right)$. However, for the TS case we have that:

$$\widehat{\gamma} - \gamma = \begin{pmatrix} n & \sum_{t=1}^{n} t \\ \sum_{t=1}^{n} t & \sum_{t=1}^{n} t^2 \end{pmatrix}^{-1} \sum_{t=1}^{n} \begin{pmatrix} u_t \\ t u_t \end{pmatrix}. \qquad (32)$$

Now notice that

$$\sum_{t=1}^{n} t = \frac{n\left(n+1\right)}{2} = O\left(n^2\right), \qquad \sum_{t=1}^{n} t^2 = \frac{n\left(n+1\right)\left(2n+1\right)}{6} = O\left(n^3\right)$$

and it follows that $\begin{pmatrix} n & \sum_{t=1}^{n} t \\ \sum_{t=1}^{n} t & \sum_{t=1}^{n} t^2 \end{pmatrix} = \begin{pmatrix} n & \frac{n(n+1)}{2} \\ \frac{n(n+1)}{2} & \frac{n(n+1)(2n+1)}{6} \end{pmatrix}$. Notice also, for future reference that

$$\sum_{t=1}^{n} t^3 = \left[\frac{n\left(n+1\right)}{6}\right]^2 = O\left(n^4\right).$$

Hence, scaling this matrix by $\frac{1}{n}$ would lead to non-convergent ellements. On the other hand scaling by $\frac{1}{n^3}$ would result to a singular limiting matrix. Consequently, $\widehat{a}$ and $\widehat{b}$ require different rates for asymptotic convergence. In fact we have to scale $\widehat{a}$ by $\sqrt{n}$ and $\widehat{b}$ by $n^{3/2}$. Hence, employing equation (32), we get

$$\begin{aligned}
\begin{pmatrix} \sqrt{n}\left(\widehat{a}-a\right) \\ n^{3/2}\left(\widehat{b}-b\right) \end{pmatrix} &= \begin{pmatrix} \sqrt{n} & 0 \\ 0 & n^{3/2} \end{pmatrix} \begin{pmatrix} \widehat{a}-a \\ \widehat{b}-b \end{pmatrix} = N_n \left(\sum_{t=1}^{n} x_t x_t' \right)^{-1} \sum_{t=1}^{n} x_t u_t \\
&= N_n \left(\sum_{t=1}^{n} x_t x_t' \right)^{-1} N_n N_n^{-1} \sum_{t=1}^{n} x_t u_t \\
&= \left(N_n^{-1} \sum_{t=1}^{n} x_t x_t' N_n^{-1} \right)^{-1} N_n^{-1} \sum_{t=1}^{n} x_t u_t,
\end{aligned}$$

where $N_n = \begin{pmatrix} \sqrt{n} & 0 \\ 0 & n^{3/2} \end{pmatrix}$. Now

$$N_n^{-1} \sum_{t=1}^{n} x_t x_t' N_n^{-1} = \begin{pmatrix} 1 & n^{-1}\frac{n+1}{2} \\ n^{-1}\frac{n+1}{2} & n^{-2}\frac{(n+1)(2n+1)}{6} \end{pmatrix} \to \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix} = M.$$

For the next term, notice that

$$N_n^{-1} \sum_{t=1}^{n} x_t u_t = \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_t \\ \frac{1}{n^{3/2}} \sum_{t=1}^{n} t u_t \end{pmatrix}$$

and as u_t is iid with finite fourth moment we get that $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_t \xrightarrow{D} N\left(0, \sigma^2\right)$. Further we can prove that $\frac{1}{n^{3/2}} \sum_{t=1}^{n} t u_t \xrightarrow{D} N\left(0, \sigma^2/3\right)$ (see Hamilton 1994), and the linear combination of the two elements converges to normality we conclude that

$$\begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_t \\ \frac{1}{n^{3/2}} \sum_{t=1}^{n} t u_t \end{pmatrix} \xrightarrow{D} N\left(0, \sigma^2 M\right),$$

and it follows that

$$\begin{pmatrix} \sqrt{n}\left(\widehat{a}-a\right) \\ n^{3/2}\left(\widehat{b}-b\right) \end{pmatrix} \xrightarrow{D} N\left(0, \sigma^2 M^{-1}\right) \tag{33}$$

establishing the $superconsistency$ of $\widehat{b}$.

Further, it is an easy exercise to prove the asymptotic normality of the $t-statistics$ for $\widehat{a}$ and $\widehat{b}$. Here we consider only the $t-statistic$ for the hypothesis $H_0 : b = b_0$ (see Hamilton 1994 for details).

$$\begin{aligned} t_{\widehat{b}} &= \frac{\widehat{b}-b_0}{\sqrt{s^2 m^{22}}}, \quad where \ \ m^{22} \ is \ the \ (2,2) \ element \ of \ \left(\sum_{t=1}^{n} x_t x_t'\right)^{-1} \ and \\ s^2 &= \frac{1}{n-2}\sum_{t=1}^{n}\left(y_t - \widehat{a} - t\widehat{b}\right)^2. \end{aligned}$$

Hence

$$\begin{aligned} t_{\widehat{b}} &= \frac{\widehat{b}-b_0}{\sqrt{s^2 \begin{pmatrix} 0 & 1 \end{pmatrix} \left(\sum_{t=1}^{n} x_t x_t'\right)^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}}} = \frac{n^{3/2}\left(\widehat{b}-b_0\right)}{\sqrt{s^2 \begin{pmatrix} 0 & n^{3/2} \end{pmatrix} \left(\sum_{t=1}^{n} x_t x_t'\right)^{-1} \begin{pmatrix} 0 \\ n^{3/2} \end{pmatrix}}} \\ &= \frac{n^{3/2}\left(\widehat{b}-b_0\right)}{\sqrt{s^2 \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{n} & 0 \\ 0 & n^{3/2} \end{pmatrix} \left(\sum_{t=1}^{n} x_t x_t'\right)^{-1} \begin{pmatrix} \sqrt{n} & 0 \\ 0 & n^{3/2} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix}}} \\ &= \frac{n^{3/2}\left(\widehat{b}-b_0\right)}{\sqrt{s^2 \begin{pmatrix} 0 & 1 \end{pmatrix} \left(\begin{pmatrix} \sqrt{n} & 0 \\ 0 & n^{3/2} \end{pmatrix}^{-1} \sum_{t=1}^{n} x_t x_t' \begin{pmatrix} \sqrt{n} & 0 \\ 0 & n^{3/2} \end{pmatrix}^{-1}\right)^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}}}. \end{aligned}$$

Now from above we know that $\begin{pmatrix} \sqrt{n} & 0 \\ 0 & n^{3/2} \end{pmatrix}^{-1} \sum_{t=1}^{n} x_t x_t' \begin{pmatrix} \sqrt{n} & 0 \\ 0 & n^{3/2} \end{pmatrix}^{-1} \rightarrow M$. Further, $s^2 \xrightarrow{P} \sigma^2$ and $n^{3/2}\left(\widehat{b}-b\right) \xrightarrow{D} N\left(0, \sigma^2 m^{22}\right)$ (by equation (33)).

Hence we get:

$$t_{\widehat{b}} \xrightarrow{P} \frac{n^{3/2}\left(\widehat{b}-b_0\right)}{\sqrt{\sigma^2 \begin{pmatrix} 0 & 1 \end{pmatrix} M^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}}} \xrightarrow{D} N(0,1).$$

8.2 Over and Under Differencing

Let us now consider the cases where the true model is RW and we treat it as TS $(under-differencing)$ as well as the reverse case, i.e. the true model is TS and we treat it as RW $(over-differencing)$.

Under-Differencing

Assume that we estimate the following equation

$$y_t = \widehat{a} + \widehat{b}t + \widehat{e_t}$$

where $\widehat{a}$ and $\widehat{b}$ denote the LS coefficients, $\widehat{e_t}$ denotes the LS residual, and $t = 1, ...n$. However, the true model is

$$y_t = \mu + y_{t-1} + u_t,$$

where u_t is such that $S_t = \sum_{j=1}^{t} u_j$ sutisfy an appropriate functional Central Limit Theorem (see Durlauf and Phillips 1988 for details). In this setup it is posible to prove that (see Theorem 3.1 in Durlauf and Phillips 1988), as $n \to \infty$:

1) $n^{-1/2}\widehat{a} \xrightarrow{D} N(0, 2\sigma^2/15)$, where $\sigma^2 = \lim_{n\to\infty} n^{-1} \sum E\left(S_n^2\right)$

2) $n^{1/2}\left(\widehat{b}-\mu\right) \xrightarrow{D} N(0, 6\sigma^2/5)$

3) The $t-statistics$, $t_{a=0}$ and $t_{b=0}$, diverge

4) The DW (Durbin-Watson) statistic goes in probability to 0

5) The R^2 has a nondegenerate distribution.

The results do not change if $\mu = 0$ (see Theorems 2.1 and 2.2 in Durlauf and Phillips 1988), just set $\mu = 0$ in 2). Notice that the asymtotic distribution of $\widehat{a}$ is non-standard, whereas the one of $\widehat{b}$ is. However, the $t_{b=0}$ diverges, invalidating any standard inference. Notice that the DW is going to 0 indicating that, for large data sets the probability of mistaken a RW process as a TS one is very low. However, a low DW statistic does not necessarily imply a RW process. The R^2 converges in distribution to a nondegenerate random variable with an expected value approximately 0.44 (see Nelson and Kang 1981).

Another effect of under-differencing is the spurious periodicity of the autocorrelations of the fitted values, $\widehat{y_t} = \widehat{a} + \widehat{b}t$ (see Nelson and Kang 1981 and Patterson 2011).

Over-Differencing

Consider the following process

$$y_t = a + bt + \varepsilon_t,$$

where ε_t is iid with $\mathsf{E}(\varepsilon_t^4) < \infty$. It follows that

$$\Delta y_t = b + \varepsilon_t - \varepsilon_{t-1},$$

i.e. over-differencing induces a nonivertible $MA\,(1)$ process. It seems that over-diffrencing is less of problem, as compared to under-differencing, provided that the autocorrelation of the errors are taken into acount. In fact, estimating the following model

$$\Delta y_t = b + \varepsilon_t - \theta\varepsilon_{t-1},$$

by ML, provided that ε_t is $iid\ N\,(0, \sigma^2)$, it is possible to prove that $\widehat{\theta}$ is n consistent (see Shephard 1993 and Sargan and Bhargava 1983). However, although

$\widehat{\theta}$ converges to 1 faster than the standard $\sqrt{n}$ rate, its asymptotic distribution is not normal (for an indirect estimator with $non-iid$ and/or nonnormal ε_t see Arvanitis 2013). Finally, for implications of over-differencing for the spectral density function see Patterson (2011).

8.3 Spurious Regressions

Assume two independent RWs, i.e.

$$y_t = y_{t-1} + u_t, \quad and \quad x_t = x_{t-1} + v_t,$$

where u_t and v_t are independent and satisfy some heterogeneity and weak dependence conditions (see Phillips 1986). Now according to Durlauf and Phillips (1988) (Theorem 5.1) the coefficients in the least squares regression

$$y_t = \widehat{a} + \widehat{b}t + \widehat{c}x_t + \widehat{u}_t, \quad t = 1, ..., n$$

have the following asymptotic behaviour:

(1) $\widehat{c}$ converges weakly to a nondegenerate random variable;

(2) $\widehat{a}$ diverges;

(3) $\widehat{b} \overset{P}{\to} 0$;

(4) s^2, the estimate of the error variance, diverges;

(5) $t_{c=0}$, the $t-statistic$ for $c=0$, diverges; and

(6) the Durbin-Watson statistic goes in probability to zero, i.e. $DW \overset{P}{\to} 0$.

Notice that the only consistent estimator is $\widehat{b}$. $\widehat{c}$ has a nondegenarate asymptotic distribution, which explains the simulation results in Granger and Newbold (1974), althought their results concern a regression without a time trend. Even in this case, i.e. when a time trend is not included, $\widehat{c}$ has a nondegenarate asymptotic distribution which differs by the presence of terms which express the interaction between the time trend and the nonstationary series (see Phillips

1986 and Durlauf and Phillips 1988). The $t_{c=0}$ test, diverges as in the non-detrended case (Phillips 1988). Hence, regardless of the inclusion of the time trend, the $t - statistic$ test will diverge and consequently, the nonstationarity of the underlying series is the critical issue, rather than inappropriate detrending (see Durlauf and Phillips 1988). Again the DW statistic converges in probability to zero. To quote Durlauf and Phillips (1988) "The inappropriateness of the conventional $t - statistic$ test should thus become apparent to the investigator from the inspection of residual DW diagnostics. Once again the results indicate that the employment of conventional significance tests must be suspect until the stationarity of the dependent variable is resolved."

Consider now the regression of the two independent RWs without the drift, i.e.

$$y_t = \widehat{a} + \widehat{c}x_t + \widehat{u_t}, \quad t = 1, ..., n.$$

Further, assume that the two independent RWs have a drift, i.e.

$$y_t = \gamma_y + y_{t-1} + u_t, \quad and \quad x_t = \gamma_x + x_{t-1} + v_t,$$

where u_t is $iid\left(0, \sigma_y^2\right)$, v_t is $iid\left(0, \sigma_x^2\right)$ are u_t and v_t are independent for all $t's$. Under regularity assumptions made explicit in Phillips (1986), Entorf (1992) was able to get the following:

$1^*)$ $\widehat{c} \xrightarrow{P} \frac{\gamma_y}{\gamma_x}$;

$2^*)$ $\widehat{a}$ diverges; and

$3^*)$ $DW \xrightarrow{P} 0$.

It is evident that is cases that the RWs have a drift the regression coeficient, $\widehat{c}$, converges to a constant, unlike in cases where the are no drifts (see (1) above) where the coefficient converges to a random variable. In both case, the consant of the regression $\widehat{a}$ diverges. Finally, notice that in both cases the DW statistic goes to 0 in probability, hence DW can be an important tool to detect spurious regressions, in large samples.

9 Unit Roots

In order to derive unit root tests, against the stationary alternatives, we need first to derive the distribution of the OLS estimator of the autoregressive parameter under the unit root null hypothesis. This distribution turns out to be a function of Wiener processes and hence a brief introduction to the Wiener processes is needed.

9.1 The Wiener Process

Let ΔW be the change in the Wiener process, or standard Brownian motion, say W, during a small time interval Δt. Then have that:

$$\Delta W = z\sqrt{\Delta t}$$

where $z \sim N(0,1)$. It follows that $E(\Delta W) = 0$, $V(\Delta W) = \Delta t$. Further, for any disjoint small time intervals, the values of ΔW are independent, as the zs are independent.

Now consider the value of W during a long period T, say $W(T)$. Then break the long period T in to n non-overlapping equal small intervals Δt, i.e. $T = n\Delta t$. Then we have that

$$W(T) - W(0) = \sum_{i=1}^{n} z_i\sqrt{\Delta t},$$

and since $z_i \sim iidN(0,1)$, assuming that $W(0) = 0$ we get that

$$E(W(T)) = 0, \quad V(W(T)) = n\Delta t = T.$$

Now consider $S_T = \sum_{t=1}^{T} z_t$, where $z_t \sim iidN(0,1)$. Then S_t is a random walk as $S_t = S_{t-1} + z_t$. Assuming that $S_0 = 0$, we have that

$$E(S_T) = 0, \quad V(S_T) = T,$$

as in the case of $W(T)$.

Now divide the interval $[0,1]$ into T intervals of length $1/T$, i.e. the points of the interval are $0, 1/T, 2/T, ..., (T-1)/T, 1$. Further define a new index r in $[0,1]$ corresponding to the time t, in $\{0, 1, ..., T\}$, by the relation

$$\frac{i-1}{T} \leq r < \frac{i}{T} \quad for \quad i = 1, ..., T.$$

Let $[rT]$ denote the integer part of rT, i.e. for $T = 50$ and $r = 0.45$, $[rT] = [0.45 * 50] = [22.5] = 22$. Finally define the step function

$$X_T(r) = \frac{1}{\sqrt{T}} S_{[rT]}.$$

According to the so called Donsker's Functional Central Limit Theorem we get that

$$X_T(r) \xrightarrow{D} W(r).$$

Furthermore, from the Continuous Mapping Theorem we have that if $g(.)$ is a continuous function on $[0,1]$ we have that

$$g(X_T(r)) \xrightarrow{D} g(W(r)),$$

see Billingsley (1968) section 5 for proofs.

Results on Wiener Process

Some basic results, employing the Donsker's theorem, are provided in the sequel.

Suppose that y_t is a random walk, i.e.

$$y_t = y_{t-1} + z_t$$

where $z_t \sim iidN(0,1)$, for $t = 1, 2, ..., T$. Assume that $y_0 = 0$ for simplicity. Let $\overline{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$, then we have the following lemma.

Lemma 9.1

$$(i)\ \frac{1}{\sqrt{T}}\overline{y} \xrightarrow{D} \int_0^1 W(r)\,dr, \quad (ii)\ \frac{1}{T^2}\sum_{t=1}^{T} y_t^2 \xrightarrow{D} \int_0^1 [W(r)]^2\,dr,$$

$$(iii)\ \frac{1}{T^{\frac{5}{2}}}\sum_{t=1}^{T} ty_t \xrightarrow{D} \int_0^1 rW(r)\,dr, \quad (iv)\ \frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T} tz_t \xrightarrow{D} \int_0^1 rdW(r),$$

$$and \quad (v)\ \frac{1}{T}\sum_{t=1}^{T} y_{t-1}z_t \xrightarrow{D} \int_0^1 W(r)\,dW(r).$$

Proof of Lemma 9.1. (i) Consider the step function

$$X_T(r) = \frac{1}{\sqrt{T}}y_{[rT]} = \frac{1}{\sqrt{T}}y_{i-1} \quad for\ \frac{i-1}{T} \leq r < \frac{i}{T} \quad and \quad X_T(1) = \frac{1}{\sqrt{T}}y_T$$

where $X_T(r)$ is a step function with steps $\frac{y_i}{\sqrt{T}}$ at $\frac{i}{T}$ and is constant between steps. Hence it follows that

$$\int_0^1 X_T(r)\,dr = \sum_{i=1}^{T}\int_{\frac{i-1}{T}}^{\frac{i}{T}} X_T(r)\,dr = \sum_{i=1}^{T}\frac{1}{\sqrt{T}}y_{i-1}\int_{\frac{i-1}{T}}^{\frac{i}{T}} dr = \frac{1}{T}\sum_{i=1}^{T}\frac{1}{\sqrt{T}}y_{i-1}$$

as

$$\int_{\frac{i-1}{T}}^{\frac{i}{T}} dr = r|_{\frac{i-1}{T}}^{\frac{i}{T}} = \frac{i}{T} - \frac{i-1}{T} = \frac{1}{T}.$$

Now, at least asymptotically, $\frac{1}{T}\sum_{i=1}^{T} y_{i-1} \simeq \frac{1}{T}\sum_{i=1}^{T} y_i$ we have that

$$\int_0^1 X_T(r)\,dr = \frac{1}{\sqrt{T}}\overline{y}.$$

Now by Donsker's Theorem we have that

$$X_T(r) \xrightarrow{D} W(r)$$

and it follows that

$$\frac{1}{\sqrt{T}}\overline{y} \xrightarrow{D} \int_0^1 W(r)\,dr.$$

(ii) With the same logic we have that

$$\int_0^1 [X_T(r)]^2\, dr = \sum_{i=1}^{T} \int_{\frac{i-1}{T}}^{\frac{i}{T}} [X_T(r)]^2\, dr = \sum_{i=1}^{T} \frac{1}{T} y_{i-1}^2 \int_{\frac{i-1}{T}}^{\frac{i}{T}} dr = \frac{1}{T^2} \sum_{i=1}^{T} y_{i-1}^2.$$

Now, at least asymptotically, $\frac{1}{T}\sum_{i=1}^{T} y_{i-1}^2 \simeq \frac{1}{T}\sum_{i=1}^{T} y_i^2$ we have that

$$\int_0^1 [X_T(r)]^2\, dr = \frac{1}{T^2}\sum_{i=1}^{T} y_i^2.$$

Now by Donsker's Theorem and the Continuous Mapping Theorem we have that

$$[X_T(r)]^2 \xrightarrow{D} [W(r)]^2$$

and it follows that

$$\frac{1}{T^2}\sum_{t=1}^{T} y_t^2 \xrightarrow{D} \int_0^1 [W(r)]^2\, dr.$$

(iii) Now

$$\begin{aligned} \int_0^1 r X_T(r)\, dr &= \sum_{i=1}^{T} \int_{\frac{i-1}{T}}^{\frac{i}{T}} r X_T(r)\, dr = \sum_{i=1}^{T} \frac{1}{\sqrt{T}} y_{i-1} \int_{\frac{i-1}{T}}^{\frac{i}{T}} r dr \\ &= \frac{1}{\sqrt{T}} \sum_{i=1}^{T} y_{i-1}\left(\frac{i-1}{T^2} - \frac{1}{2T^2}\right) = \frac{1}{T^{5/2}}\sum_{i=1}^{T}(i-1)\, y_{i-1} - \frac{1}{T^{5/2}}\sum_{i=1}^{T} y_{i-1} \end{aligned}$$

as

$$\int_{\frac{i-1}{T}}^{\frac{i}{T}} r dr = \frac{1}{2}\, r^2\Big|_{\frac{i-1}{T}}^{\frac{i}{T}} = \frac{1}{2}\frac{i^2-(i-1)^2}{T^2} = \frac{1}{2}\frac{2i-1}{T^2} = \frac{i-1}{T^2} - \frac{1}{2T^2}.$$

Further, as $\frac{1}{T^{5/2}}\sum_{i=1}^{T} y_{i-1} \xrightarrow{P} 0$ and $\frac{1}{T^{5/2}}\sum_{i=1}^{T}(i-1)\, y_{i-1} = \frac{1}{T^{5/2}}\sum_{i=1}^{T} t y_t - \frac{1}{T^{3/2}} y_T \xrightarrow{P} \frac{1}{T^{5/2}}\sum_{i=1}^{T} t y_t$ we get that

$$\frac{1}{T^{5/2}}\sum_{i=1}^{T}(i-1)\, y_{i-1} - \frac{1}{T^{5/2}}\sum_{i=1}^{T} y_{i-1} \xrightarrow{P} \frac{1}{T^{5/2}}\sum_{i=1}^{T} t y_t.$$

Now by Donsker's Theorem and the Continuous Mapping Theorem we have that

$$r X_T(r) \xrightarrow{D} r W(r)$$

it follows that

$$\frac{1}{T^{5/2}}\sum_{i=1}^{T}ty_t \xrightarrow{D} \int_0^1 rW(r)\,dr.$$

(iv) Notice

$$\begin{aligned}\frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T}tz_t &= \frac{1}{T^{\frac{2}{2}}}\sum_{i=1}^{T}i\frac{1}{\sqrt{T}}z_i=\sum_{i=1}^{T}\frac{i}{T}\int_{\frac{i-1}{T}}^{\frac{i}{T}}dX_T(r)=\sum_{i=1}^{T}\int_{\frac{i-1}{T}}^{\frac{i}{T}}rdX_T(r)\\ &= \int_0^1 rdX_T(r)\xrightarrow{D}\int_0^1 W(r)\,dW(r).\end{aligned}$$

(v) Last

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}y_{t-1}z_t &= \sum_{i=1}^{T}\frac{1}{\sqrt{T}}y_{i-1}\frac{1}{\sqrt{T}}z_i=\sum_{i=1}^{T}\int_{\frac{i-1}{T}}^{\frac{i}{T}}X_T(r)\,dX_T(r)\\ &= \int_0^1 X_T(r)\,dX_T(r)\xrightarrow{D}\int_0^1 W(r)\,dW(r).\end{aligned}$$

■

Notice the following mapping (see Maddala and Kim 1998):

$$\sum \Rightarrow \int, \quad t\Rightarrow r, \quad z_t \Rightarrow dW(r) \quad and \quad y_t \Rightarrow W(r).$$

Now as $z_t \sim iidN(0,1)$ and $y_T=\sum_{t=1}^{T}z_t$ it follows that $\frac{y_T}{\sqrt{T}}\sim N(0,1)=W(1)$. The following lemma presents the relation between $W(r)$ and the normal distribution.

Lemma 9.2 *(i)* $\int_0^1 W(r)\,dr \sim N\left(0,\frac{1}{3}\right)$*, (ii)* $\int_0^1 rdW(r) \sim N\left(0,\frac{1}{3}\right)$*, (iii)* $\int_0^1 (r-a)\,W(r)\,dr \sim N\left(0,\frac{8-25a+20a^2}{60}\right)$*, (iv) If* $W(r)$ *and* $W(r)$ *are independent Wiener processes then* $\left(\int_0^1 (W(r))^2\,dr\right)^{-\frac{1}{2}}\int_0^1 W(r)\,dV(r)\sim N(0,1)$*, (v) Now if* $z_t \sim N(0,\sigma_z^2)$ *but not independent and* $lim_{T\to\infty}E(y_t^2)=\sigma_y^2>0$ *then* $\int_0^1 W(r)\,dW(r)\sim\frac{\sigma_y^2}{2}(\chi_1^2-1)+\left(\frac{\sigma_y^2-\sigma_z^2}{2}\right)$.

Proof of Lemma 9.2. (i) Notice that

$$\begin{aligned}
\frac{1}{\sqrt{T}}\overline{y} &= \frac{1}{T^{3/2}}\sum_{t=1}^{T} y_t = \frac{1}{T^{3/2}}\sum_{t=1}^{T}\sum_{i=1}^{t} z_i = \frac{1}{T^{3/2}}\left(Tz_1 + (T-1)\,z_2 + ... + z_T\right) \\
&= \frac{1}{T^{3/2}}\sum_{t=1}^{T} tz_{T-t+1}.
\end{aligned}$$

It follows that $Var\left(\frac{1}{\sqrt{T}}\overline{y}\right) = \frac{1}{T^3}Var\left(\sum_{t=1}^{T} tz_{T-t+1}\right) = \frac{1}{T^3}\sum_{t=1}^{T} t^2 = \frac{1}{T^3}\frac{T(T+1)(2T+1)}{6} \rightarrow \frac{1}{3}$ for $T \rightarrow \infty$. Further, the distribution of $\frac{1}{\sqrt{T}}\overline{y}$ is normal, as it is a sum of independent normal variates. But from the previous lemma, 9.1(i), we have that $\frac{1}{\sqrt{T}}\overline{y} \xrightarrow{D} \int_0^1 W(r)\,dr$ and it follows that

$$\int_0^1 W(r)\,dr \sim N\left(0, \frac{1}{3}\right).$$

(ii) Again notice that $Var\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T} tz_t\right) = \frac{1}{T^3}\frac{T(T+1)(2T+1)}{6}$, $E\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T} tz_t\right) = 0$ and as the $z_t's$ are normally distributed we get that $\frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T} tz_t \xrightarrow{D} N\left(0, \frac{1}{3}\right)$. From lemma 9.1(iv) we have that $\frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T} tz_t \xrightarrow{D} \int_0^1 rdW(r)$ and it follows that

$$\int_0^1 rdW(r) \sim N\left(0, \frac{1}{3}\right).$$

(iii) Notice first that

$$\begin{aligned}
\sum_{t=1}^{T} y_t &= Tz_1 + (T-1)\,z_2 + (T-2)\,z_3 + ... + 2z_{t-1} + z_T \\
&= T\sum_{t=1}^{T} z_t - \sum_{t=2}^{T}(t-1)\,z_t = T\sum_{t=1}^{T} z_t - \sum_{t=2}^{T} tz_t - \sum_{t=2}^{T} z_t
\end{aligned}$$

and

$$\begin{aligned}
\sum_{t=1}^{T} ty_t &= z_1 + 2\,(z_1+z_2) + 3\,(z_1+z_2+z_3) + ... + T\,(z1 + z2 + z3 + ... + z_{T-1} + z_T) \\
&= \frac{T\,(T+1)}{2}z_1 + \left[\frac{T\,(T+1)}{2} - 1\right] z_2 + \left[\frac{T\,(T+1)}{2} - 1 - 2\right] z_3 + ... \\
&\quad + \left[\frac{T\,(T+1)}{2} - 1 - 2 - ... - (T-1)\right] z_T \\
&= \frac{T\,(T+1)}{2}\sum_{t=1}^{T} z_t - \sum_{t=2}^{T}\frac{t\,(t-1)}{2}z_t = \frac{T\,(T+1)}{2}\sum_{t=1}^{T} z_t - \frac{1}{2}\sum_{t=2}^{T} t^2 z_t + \frac{1}{2}\sum_{t=2}^{T} tz_t.
\end{aligned}$$

Hence

$$Var\left(\frac{1}{T^{\frac{5}{2}}}\sum_{t=1}^{T}ty_t - a\frac{1}{T^{3/2}}\sum_{t=1}^{T}y_t\right) = \frac{1}{T^5}Var\left(\sum_{t=1}^{T}ty_t\right) + a^2\frac{1}{T^3}Var\left(\sum_{t=1}^{T}y_t\right) - 2a\frac{1}{T^4}Cov\left(\sum_{t=1}^{T}ty_t, \sum_{t=1}^{T}y_t\right). \quad (34)$$

Now for the first term of the above equation we have:

$$\begin{aligned}
\frac{1}{T^5}Var\left(\sum_{t=1}^{T}ty_t\right) &= \frac{1}{T^5}Var\left[\frac{T(T+1)}{2}\sum_{t=1}^{T}z_t - \frac{1}{2}\sum_{t=2}^{T}t^2z_t + \frac{1}{2}\sum_{t=2}^{T}tz_t\right] \\
&= \frac{1}{T^5}\frac{T^2(T+1)^2}{4}Var\left(\sum_{t=1}^{T}z_t\right) + \frac{1}{4T^5}Var\left[\sum_{t=2}^{T}t^2z_t\right] + \frac{1}{4T^5}Var\left[\sum_{t=2}^{T}tz_t\right] \\
&\quad - \frac{T(T+1)}{2T^5}Cov\left[\sum_{t=1}^{T}z_t, \sum_{t=2}^{T}t^2z_t\right] + \frac{T(T+1)}{2T^5}Cov\left[\sum_{t=1}^{T}z_t, \sum_{t=2}^{T}tz_t\right] \\
&\quad - \frac{1}{2}\frac{1}{T^5}Cov\left[\sum_{t=2}^{T}t^2z_t, \sum_{t=2}^{T}tz_t\right] \\
&= \frac{T^3(T+1)^2}{4T^5} + \frac{1}{4T^5}\sum_{t=2}^{T}t^4 + \frac{1}{4T^5}\sum_{t=2}^{T}t^2 - \frac{T(T+1)}{2T^5}\sum_{t=2}^{T}t^2 \\
&\quad + \frac{T(T+1)}{2T^5}\sum_{t=2}^{T}t - \frac{1}{2T^5}\sum_{t=2}^{T}t^3 \to \frac{1}{4} + \frac{1}{20} - \frac{1}{6} = \frac{8}{60}
\end{aligned}$$

as $\frac{T(T+1)}{2T^5}\sum_{t=2}^{T}t \to 0$ and $\frac{1}{2T^5}\sum_{t=2}^{T}t^3 \to 0$ for $T \to \infty$. For the second term in equation (34)

$$\begin{aligned}
\frac{1}{T^3}Var\left(\sum_{t=1}^{T}y_t\right) &= \frac{1}{T^3}Var\left(T\sum_{t=1}^{T}z_t - \sum_{t=2}^{T}tz_t - \sum_{t=2}^{T}z_t\right) \\
&= \frac{1}{T^3}Var\left(T\sum_{t=1}^{T}z_t\right) + \frac{1}{T^3}Var\left(\sum_{t=2}^{T}tz_t\right) + \frac{1}{T^3}Var\left(\sum_{t=2}^{T}z_t\right) \\
&\quad - \frac{2}{T^2}Cov\left(\sum_{t=1}^{T}z_t, \sum_{t=2}^{T}tz_t\right) - \frac{2}{T^2}Var\left(\sum_{t=1}^{T}z_t\right) + \frac{2}{T^3}Cov\left(\sum_{t=2}^{T}tz_t, \sum_{t=2}^{T}z_t\right) \\
&= 1 + \frac{1}{T^3}\sum_{t=2}^{T}t^2 - \frac{2}{T^2}\sum_{t=2}^{T}t \to 1 + \frac{1}{3} - 1 = \frac{1}{3},
\end{aligned}$$

as $\frac{1}{T^3}Var\left(\sum_{t=2}^{T} z_t\right) \to 0$, $\frac{2}{T^3}Cov\left(\sum_{t=2}^{T} tz_t, \sum_{t=2}^{T} z_t\right) \to 0$ and $\frac{2}{T^2}Var\left(\sum_{t=1}^{T} z_t\right) \to 0$ for $T \to \infty$.

$$\begin{aligned}
\frac{1}{T^4}Cov\left(\sum_{t=1}^{T} ty_t, \sum_{t=1}^{T} y_t\right) &= \frac{1}{T^4}Cov\left(\begin{array}{r} \frac{T(T+1)}{2}\sum_{t=1}^{T} z_t - \frac{1}{2}\sum_{t=2}^{T} t^2 z_t + \frac{1}{2}\sum_{t=2}^{T} tz_t, T\sum_{t=1}^{T} z_t \\ -\sum_{t=2}^{T} tz_t - \sum_{t=2}^{T} z_t \end{array}\right) \\
&= \frac{T^3(T+1)}{2T^4} - \frac{T(T+1)}{2T^4}\sum_{t=2}^{T} t - \frac{1}{2T^3}\sum_{t=2}^{T} t^2 + \frac{1}{2T^4}\sum_{t=2}^{T} t^3 \\
&\to \frac{1}{2} - \frac{1}{4} - \frac{1}{6} + \frac{1}{8} = \frac{5}{24}
\end{aligned}$$

as $\frac{T(T+1)}{T^4}Var\left(\sum_{t=2}^{T} z_t\right) \to 0$, $\frac{1}{T^4}Cov\left(\sum_{t=2}^{T} t^2 z_t, \sum_{t=2}^{T} z_t\right) \to 0$, $\frac{1}{T^4}Var\left(\sum_{t=2}^{T} tz_t\right) \to 0$, $\frac{1}{T^3}Cov\left(\sum_{t=2}^{T} tz_t, \sum_{t=1}^{T} z_t\right) \to 0$ and $\frac{1}{T^4}Cov\left(\sum_{t=2}^{T} tz_t, \sum_{t=2}^{T} z_t\right) \to 0$ for $T \to \infty$. Consequently, substituting in equation (34) we get:

$$Var\left(\frac{1}{T^{\frac{5}{2}}}\sum_{t=1}^{T} ty_t - a\frac{1}{T^{3/2}}\sum_{t=1}^{T} y_t\right) = \frac{8}{60} + a^2\frac{1}{3} - 2a\frac{5}{24}.$$

But from the previous lemma, 9.1(i) and (v), below, we get that $\frac{1}{T^{\frac{5}{2}}}\sum_{t=1}^{T} ty_t - a\frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T} y_t \xrightarrow{D} \int_0^1 (r-a)W(r)\,dr$ and it follows that

$$\int_0^1 (r-a)W(r)\,dr \sim N\left(0, \frac{8-25a+20a^2}{60}\right).$$

(iv) Let $\mathcal{F}_t^1 = \sigma(W(r), r \le t)$, $\mathcal{F}_t^2 = \sigma(dV(r)\, r \le t)$ and $\mathcal{F}_t = \sigma(\mathcal{F}_t^1 \cup \mathcal{F}_t^2)$ be the natural filtration generated by the two independent Wiener processes. First, by the conditional Ito Isometry (see e.g. Steele 2001) we have that

$$E\left[\left(\int_s^t W(r)\,dV(r)\right)^2 |\mathcal{F}_t\right] = E\left[\int_s^t W^2(r)\,dr|\mathcal{F}_t\right],$$

and as

$$E\left|\int_0^t W(r)\,dV(r)\right| < \infty, \qquad \sum W_{t_n}\left(V_{t_{n+1}} - V_{t_n}\right) \xrightarrow{L_2} \int_0^t W(r)\,dV(r)$$

and due to independence and normality, which is preserved in the limit we get that

$$\int_0^t W(r)\,dV(r)|\mathcal{F}_t \sim N\left(0, \int_0^t W^2(r)\,dr\right).$$

(v) Notice that

$$\sum_{t=1}^{T} y_t^2 = \sum_{t=1}^{T} y_{t-1}^2 + 2\sum_{t=1}^{T} y_{t-1}z_t + \sum_{t=1}^{T} z_t^2.$$

Hence

$$2\sum_{t=1}^{T} y_{t-1}z_t = \sum_{t=1}^{T} y_t^2 - \sum_{t=1}^{T} y_{t-1}^2 - \sum_{t=1}^{T} z_t^2 = y_T^2 - \sum_{t=1}^{T} z_t^2.$$

and it follows that

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}z_t = \frac{1}{2}\left(\frac{1}{T}y_T^2 - \frac{1}{T}\sum_{t=1}^{T} z_t^2\right).$$

Further, $y_T = \sum_{t=1}^{T} z_t$ and from normality of the $z_t\prime s$ we get $\frac{1}{\sqrt{T}}y_T \sim N\left(0, \sigma_y^2\right)$ and $\frac{1}{T}y_T^2 \sim \sigma_y^2\chi_1^2$. From the Law of Large Numbers we have that $\frac{1}{T}\sum_{t=1}^{T} z_t^2 \to Var\left(z_t\right) = \sigma_z^2$. Hence

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}z_t \xrightarrow{D} \frac{\sigma_y^2}{2}\left(\chi_1^2 - \frac{\sigma_z^2}{\sigma_y^2}\right) = \frac{\sigma_y^2}{2}\left(\chi_1^2 - 1\right) + \frac{\sigma_y^2 - \sigma_z^2}{2},$$

and from lemma 9.1(v) we have that $\frac{1}{T}\sum_{t=1}^{T} y_{t-1}z_t \xrightarrow{D} \int_0^1 W\left(r\right)dW\left(r\right)$. Now if $z_t \sim iidN\left(0,1\right)$, then $\frac{1}{\sqrt{T}}y_T \sim N\left(0,1\right)$ and $\frac{1}{T}y_T^2 \sim \chi_1^2$ and $\frac{1}{T}\sum_{t=1}^{T} z_t^2 \to Var\left(z_t\right) = 1$, and it follows that

$$\int_0^1 W\left(r\right)dW\left(r\right) \sim \frac{1}{2}\left(\chi_1^2 - 1\right).$$

■

9.2 Unit Root Tests without Deterministic Trend

To visualize the difference between a stationary and a random walk process we depict in the following figures an $AR\left(1\right)$ process with an autoregressive coefficient of 0.89 and a random walk, where the errors are drawn from a normal distribution. The same random errors are employed for the simulations of the processes.

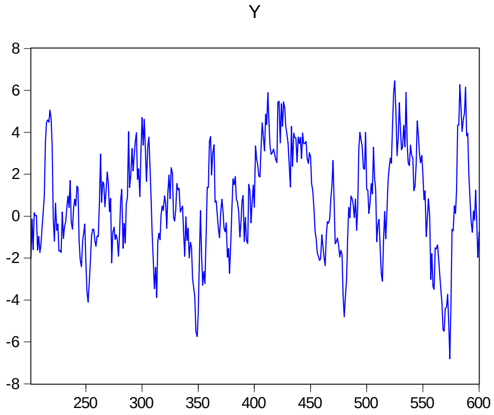


Fig. 9.2: Graph of the $AR(1)$ process $y_t = 0.89y_{t-1} + u_t$, $u_t \sim N(0, 1.4)$.

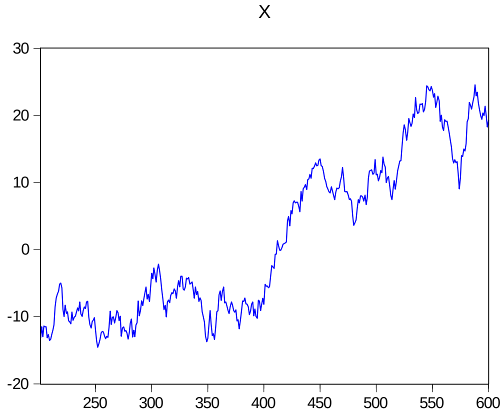


Fig. 9.2: Graph of the random walk $x_t = x_{t-1} + u_t$, $u_t \sim N(0, 1.4)$.

Furthermore, the following figures present the correlogram of the two processes:

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.875	0.875	308.20	0.000
		2	0.771	0.027	548.38	0.000
		3	0.677	-0.010	734.15	0.000
		4	0.594	-0.003	877.62	0.000
		5	0.527	0.022	990.61	0.000
		6	0.464	-0.011	1078.5	0.000
		7	0.364	-0.193	1132.6	0.000
		8	0.272	-0.049	1163.0	0.000
		9	0.218	0.094	1182.5	0.000
		10	0.140	-0.140	1190.5	0.000
		11	0.089	0.034	1193.8	0.000
		12	0.049	0.016	1194.8	0.000
		13	0.008	-0.008	1194.8	0.000
		14	-0.015	0.038	1194.9	0.000
		15	-0.034	-0.037	1195.4	0.000
		16	-0.044	0.051	1196.2	0.000
		17	-0.038	0.058	1196.8	0.000
		18	-0.031	-0.036	1197.2	0.000
		19	-0.014	0.077	1197.3	0.000
		20	0.001	-0.015	1197.3	0.000

Fig. 9.2: Correlogram of $y_t = 0.89y_{t-1} + u_t$, $u_t \sim N(0, 1.4)$.

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.990	0.990	395.08	0.000
		2	0.981	0.048	784.10	0.000
		3	0.971	-0.052	1166.3	0.000
		4	0.962	-0.015	1541.8	0.000
		5	0.952	0.032	1910.9	0.000
		6	0.943	-0.014	2273.7	0.000
		7	0.931	-0.103	2628.7	0.000
		8	0.920	-0.013	2976.1	0.000
		9	0.910	0.044	3316.4	0.000
		10	0.898	-0.073	3648.7	0.000
		11	0.886	0.005	3973.5	0.000
		12	0.876	0.039	4291.4	0.000
		13	0.865	-0.018	4601.9	0.000
		14	0.855	0.078	4906.6	0.000
		15	0.847	0.038	5205.9	0.000
		16	0.839	0.044	5500.4	0.000
		17	0.831	0.021	5790.5	0.000
		18	0.824	-0.008	6076.3	0.000
		19	0.818	0.067	6358.7	0.000
		20	0.812	-0.015	6637.7	0.000

Fig. 9.2: Correlogram of $x_t = x_{t-1} + u_t$, $u_t \sim N(0, 1.4)$

There are distinct differences between the graphs and the correlograms of the two processes. However, in practice things are not so obvious. Let us now return to the statistical inference for a unit root.

Suppose that y_t is a normal random walk, i.e.

$$y_t = \alpha y_{t-1} + z_t$$

where $z_t \sim iidN(0,1)$, for $t = 1, 2, ..., T$. Assume that $y_0 = 0$ for simplicity (for more general assumptions on the distribution of z_t's and the initial y_0 see Phillips 1987). Now the maximum likelihood estimator is the same as the OLS one and is given by

$$\widehat{\alpha} = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2} = \alpha + \frac{\sum_{t=1}^{T} y_{t-1} z_t}{\sum_{t=1}^{T} y_{t-1}^2}$$

For $|\alpha| < 1$ we have that

$$\frac{\sqrt{T}\left(\widehat{\alpha} - \alpha\right)}{\sqrt{1-\alpha^2}} \xrightarrow{D} N(0,1)$$

(see Mann and Wald 1943). White (1958) showed that

$$\frac{|\alpha|^T \left(\widehat{\alpha} - \alpha\right)}{\alpha^2 - 1} \xrightarrow{D} Cauchy,$$

for $|\alpha| > 1$.

Now for $\alpha = 1$, we have, from Phillips (1987), that

$$T\left(\widehat{\alpha} - 1\right) = \frac{\frac{1}{T}\sum_{t=1}^{T} y_{t-1} z_t}{\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2} \xrightarrow{D} \frac{\int_0^1 W(r)\, dW(r)}{\int_0^1 \left[W(r)\right]^2 dr} = \frac{\frac{1}{2}\left[\left(W(1)\right)^2 - 1\right]}{\int_0^1 \left[W(r)\right]^2 dr} \qquad (35)$$

by the Continuous Mapping Theorem and lemma 9.1 *(ii)* and *(v)*, and the last equality follows from lemma 9.2 *(v)*. In this case, i.e. when $\alpha = 1$, we have that the t-statistic is given by:

$$t_{\widehat{\alpha}} = \frac{\widehat{\alpha} - 1}{s}\left(\sum_{t=1}^{T} y_{t-1}^2\right)^{1/2} \quad where \;\; s^2 = \frac{1}{T}\sum_{t=1}^{T} \widehat{z}_t^{\,2} \;\; and \; \widehat{z}_t = y_t - \widehat{\alpha} y_{t-1}.$$

Notice, first, that $\widehat{\alpha}$ is consistent, by equation (35), i.e. $p\lim \widehat{\alpha} = 1$. Consequently, s^2 is a consistent estimator of the variance of z_t, i.e. $p\lim s^2 = 1$.

Hence we have that

$$\begin{aligned} t_{\widehat{\alpha}} &= \frac{T(\widehat{\alpha}-1)}{T\left(\frac{s^2}{\sum_{t=1}^{T} y_{t-1}^2}\right)^{1/2}} = \frac{T(\widehat{\alpha}-1)}{\left(\frac{s^2}{\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2}\right)^{1/2}} \xrightarrow{D} \frac{\frac{\int_0^1 W(r)dW(r)}{\int_0^1 [W(r)]^2 dr}}{\left(\frac{1}{\int_0^1 [W(r)]^2 dr}\right)^{1/2}} \\ &= \frac{\int_0^1 W(r)\, dW(r)}{\left(\int_0^1 [W(r)]^2 dr\right)^{1/2}} = \frac{1}{2}\frac{(W(1))^2-1}{\left(\int_0^1 [W(r)]^2 dr\right)^{1/2}}, \end{aligned}$$

where we employed, again, the Continuous Mapping Theorem and lemma 9.1 *(ii)*.

Tables 8.5.1 and 8.5.2 in Fuller (1976) provide critical values for both of these statistics, i.e. for $T(\widehat{\alpha}-1)$ and $t_{\widehat{\alpha}}$. These values are employed for testing the null of unit root versus the alternative of stationarity, under the maintained hypothesis that there is no deterministic trend.

9.3 Unit Root Tests with Drift

Suppose, as in the previous section, that y_t is a normal random walk, i.e.

$$y_t = y_{t-1} + z_t$$

where $z_t \sim iidN(0,1)$, for $t = 1, 2, ..., T$. Assume now that one estimates the following equation:

$$y_t = c + \alpha y_{t-1} + z_t. \tag{36}$$

In this case, the asymptotic distribution of $T(\widehat{\alpha}-1)$ and $t_{\widehat{\alpha}}$ are again functions of demeaned Wienner processes, i.e.

$$T(\widehat{\alpha}-1) \xrightarrow{D} \frac{\int_0^1 W^*(r)\, dW(r)}{\int_0^1 [W^*(r)]^2 dr} \quad and \quad t_{\widehat{\alpha}} \xrightarrow{D} \frac{\int_0^1 W^*(r)\, dW(r)}{\left(\int_0^1 [W^*(r)]^2 dr\right)^{1/2}},$$

where $W^*(r) = W(r) - \int W(r)\, dr$. The critical values of $T(\widehat{\alpha}-1)$ and $t_{\widehat{\alpha}}$ are presented in Tables 8.5.1 and 8.5.2 in Fuller (1976). Naturally, one would like to test if $c = 0$, as well. Consequently, the critical values of the F-statistic for the joint hypothesis $c = 0$ and $a = 1$ are provided in Dickey and Fuller (1981) and are evaluated via Monte Carlo, as the distribution of this statistic is not standard.

Now suppose that y_t is a normal random walk with drift, i.e.

$$y_t = \mu + y_{t-1} + z_t$$

where $z_t \sim iidN(0,1)$, for $t = 1, 2, ..., T$, and for simplicity $y_0 = 0$, the estimating equation remains the one in equation (36). Notice that in this case

$$y_t = \mu t + \sum_{i=1}^{t} z_i = \mu t + S_t.$$

Now it is clear (see Trend Stationary Process section) that the coefficients $\widehat{\alpha}$ and $\widehat{c}$ need appropriate scaling. Hence

$$\begin{pmatrix} \sqrt{T}(\widehat{c}-\mu) \\ T^{3/2}(\widehat{a}-1) \end{pmatrix} = \begin{pmatrix} 1 & T^{-2}\sum y_{t-1} \\ T^{-2}\sum y_{t-1} & T^{-3}\sum y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} T^{-1/2}\sum z_t \\ T^{-3/2}\sum y_{t-1}z_t \end{pmatrix}.$$

Now

$$\sum y_{t-1}z_t = \sum \mu(t-1)z_t + \sum S_{t-1}z_t = \mu \sum t z_t - \mu \sum z_t + \sum S_{t-1}z_t.$$

Notice that from lemmata 9.1(iv) and 9.2(ii) $T^{-3/2}\mu \sum t z_t \xrightarrow{D} N\left(0, \frac{\mu^2}{3}\right)$. $T^{-3/2}\sum z_t \xrightarrow{P} 0$ (as under our assumptions and by the Strong Law of Large Number $T^{-1}\sum z_t \to 0$ almost surely), and from lemma 9.2*(v)* we have that $\frac{1}{T}\sum S_{t-1}z_t \xrightarrow{D} \frac{1}{2}(\chi_1^2 - 1)$ and it follows that $T^{-3/2}\sum S_{t-1}z_t \xrightarrow{P} 0$. It follows that

$$T^{-3/2}\sum y_{t-1}z_t \xrightarrow{D} N\left(0, \frac{\mu^2}{3}\right).$$

Furthermore, $T^{-1/2}\sum z_t \sim N(0,1)$ and

$$Cov\left(T^{-1/2}\sum z_t, T^{-3/2}\sum y_{t-1}z_t\right) = E\left(T^{-2}\sum z_t \sum y_{t-1}z_t\right) \to \frac{\mu}{2},$$

as $E\left(T^{-2}\sum z_t \sum t z_t\right) = T^{-2}\frac{T(T+1)}{2}$, $E\left(T^{-2}\left(\sum z_t\right)^2\right) = \frac{1}{T}$, and $E\left(T^{-2}\sum z_t \sum S_{t-1} z_t\right) = 0$. Hence

$$\begin{pmatrix} T^{-1/2}\sum z_t \\ T^{-3/2}\sum y_{t-1} z_t \end{pmatrix} \xrightarrow{D} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, M\right),$$

where $M = \begin{pmatrix} 1 & \frac{\mu}{2} \\ \frac{\mu}{2} & \frac{\mu^2}{3} \end{pmatrix}$.

On the other hand,

$$\begin{aligned} T^{-3}\sum y_{t-1}^2 &= T^{-3}\sum \left[\mu\left(t-1\right) + S_{t-1}\right]^2 \\ &= T^{-3}\sum \mu^2\left(t-1\right)^2 + 2T^{-3}\mu\sum\left(t-1\right)S_{t-1} + T^{-3}\sum S_{t-1}^2. \end{aligned}$$

Notice that

$$T^{-(\nu+1)}\sum t^{\nu} \to \frac{1}{\nu+1}, \qquad for \;\; \nu = 0, 1, 2, \ldots$$

and it follows that $T^{-3}\sum \mu^2\left(t-1\right)^2 \to \frac{\mu^2}{3}$. Taking into account lemma 9.2*(iii)* we get that $T^{-3}\sum t S_{t-1} \xrightarrow{P} 0$ and form lemma 9.1(ii) we get that $T^{-3}\sum S_{t-1}^2 \xrightarrow{P} 0$. Hence

$$T^{-3}\sum y_{t-1}^2 \xrightarrow{P} \frac{\mu^2}{3}.$$

Further,

$$T^{-2}\sum y_{t-1} = T^{-2}\sum \mu\left(t-1\right) + T^{-2}\sum S_{t-1} \xrightarrow{P} \frac{\mu}{2},$$

and it follows that

$$\begin{pmatrix} 1 & T^{-2}\sum y_{t-1} \\ T^{-2}\sum y_{t-1} & T^{-3}\sum y_{t-1}^2 \end{pmatrix}^{-1} \xrightarrow{P} \begin{pmatrix} 1 & \frac{\mu}{2} \\ \frac{\mu}{2} & \frac{\mu^2}{3} \end{pmatrix}^{-1} = M^{-1}.$$

Hence,

$$\begin{pmatrix} \sqrt{T}\left(\widehat{c} - \mu\right) \\ T^{3/2}\left(\widehat{a} - 1\right) \end{pmatrix} \xrightarrow{D} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, M^{-1}\right) = N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 & -\frac{6}{\mu} \\ -\frac{6}{\mu} & \frac{12}{\mu^2} \end{pmatrix}\right),$$

proving that when there is a drift in the random walk specification the time trend terms dominates (West 1988), leading to normal asymptotic distribution, i.e.

$$T^{3/2}\left(\widehat{a}-1\right)\xrightarrow{D}N\left(0,\frac{12}{\mu^{2}}\right).$$

The asymptotic normality of $\widehat{a}$ applies, as well, in case that y_t is a normal random walk with drift and trend, i.e.

$$y_t=\mu+\beta t+y_{t-1}+z_t$$

where $z_t\sim iidN\left(0,1\right)$, for $t=1,2,...,T$, and the estimating equation is as in equation (36). In fact it is possible to prove that

$$T^{5/2}\left(\widehat{a}-1\right)\xrightarrow{D}N\left(0,\frac{180}{\mu^{2}}\right)$$

see West (1988) or Maddala and Kim (1998).

Now notice that in equation (36) when the hypothesis $H_0:\alpha=1$ is not rejected, y_t is a random walk with drift c, i.e. it has a trend for $c\neq 0$. When on the other hand H_0 is rejected then y_t is stationary around a constant mean, but has no trend. However, this rather asymmetric treatment does not do justice to the alternative, i.e. it could be the case that y_t does indeed have a trend (see Schmidt and Phillips 1992). One of course could add a trend in equation (36). But in this case, under the null y_t has a quadratic trend. This problem does not arise in the approach of Bhargava (1986). Following the set up in Bhargava (1986), Schmidt and Phillips (1992) consider the following Data Generating Process (DGP):

$$y_t=\psi+\xi t+x_t,\quad x_t=\beta x_{t-1}+\varepsilon_t$$

and $\varepsilon_t\sim iidN\left(0,\sigma_{\varepsilon}^{2}\right)$ (the iid assumption is not crucial). Notice that in this set up the unit root corresponds to $\beta=1$, however, the trend and drift can be present under the null and under the alternative. From the equation above it follows that

$$\begin{aligned} y_t &= \psi + \xi t + \beta \left(y_{t-1} - \psi - \xi \left(t-1 \right) \right) + \varepsilon_t \Rightarrow \\ \Delta y_t &= c_0 + c_1 t + \varphi y_{t-1} + \varepsilon_t, \quad where \\ c_0 &= \left(1-\beta \right) \psi + \beta \xi, \ \varphi = \beta - 1 \ and \ c_1 = \xi \left(1-\beta \right). \end{aligned}$$

Now if $\widehat{S_{t-1}}$ is the residual from the regression of y_{t-1} on a constant and t, then the estimator of φ is the same as the estimator of φ from the following equation

$$\Delta y_t = intercept + \varphi \widehat{S_{t-1}} + \varepsilon_t, \qquad for \ t = 1, 2, ..., T. \tag{37}$$

Further, if $\widehat{\varphi}$ is the least squares estimator from the above equation then under the null that $\varphi = 0$, $T\widehat{\varphi}$ and the $t-statistic$ have the Dickey-Fuller distribution.

On the other hand, under the null that $\beta = 1$

$$\Delta y_t = \xi + \varepsilon_t,$$

and ξ can be estimated as the average of the $\Delta y_t' s$, i.e.

$$\widetilde{\xi} = \frac{1}{T-1} \sum_{t=2}^{T} \Delta y_t = \frac{y_T - y_1}{T-1},$$

and

$$\widetilde{\psi} = y_1 - \widetilde{\xi}.$$

Let

$$\widetilde{S}_t = y_t - \widetilde{\psi} - \widetilde{\xi} t.$$

Now the LM test, for the null that $\beta = 1$, in Schmidt and Phillips (1992) is the $t - statistic$, say $\widetilde{\tau}$, of the estimate of φ in the following regression

$$\Delta y_t = intercept + \varphi \widetilde{S_{t-1}} + error. \tag{38}$$

Apart from critical values for $\widetilde{\tau}$, Schmidt and Phillips (1992) present also critical values for $T\widetilde{\varphi}$, as well. Notice that the difference between the LM and the DF

test is that the LM one employs $\widetilde{S_{t-1}}$ whereas the DF one employs $\widehat{S_{t-1}}$. Now, the LM test is expected to be more powerful, as under the null, $\widehat{S_{t-1}}$ is evaluated from a spurious regression (see Schmidt and Phillips 1992). This is indeed the case, expect from cases where $\varepsilon_0/\sigma_\varepsilon$ is large in absolute value. Further, Schmidt and Phillips (1992) prove that the LM test is related to the Bhargava (1986) one. Finally, Schmidt and Phillips (1992) deal also with the case that the errors are not iid, by introducing a correction ala Phillips and Perron (1988) (see below).

9.4 Similar Tests

Many inference procedures on unit roots are not invariant with respect to the values of the nuisance parameters. In this subsection we discuss tests that do not suffer from this problem. Hence, we discuss similar tests, i.e. tests for which the distribution of the test statistic under the null hypothesis is independent of nuisance parameters in the DGP (see Kiviet and Phillips 1992). If a test is not similar, then the appropriate critical values may depend upon unknown nuisance parameters (e.g. a constant), which will invalidate standard inferences. Let us consider the following $DGPs$ (in all cases $\varepsilon_t \sim iidN\left(0, \sigma_\varepsilon^2\right)$):

$$y_t = \alpha y_{t-1} + \varepsilon_t, \quad y_0 = 0 \tag{39}$$

$$y_t = \alpha y_{t-1} + \varepsilon_t, \quad y_0 \ arbitrary \tag{40}$$

$$y_t = \mu + \alpha y_{t-1} + \varepsilon_t, \quad y_0 \ arbitrary \quad and \tag{41}$$

$$y_t = \mu + \beta t + \alpha y_{t-1} + \varepsilon_t, \quad y_0 \ arbitrary. \tag{42}$$

Let us consider possible regression model:

$$y_t = \alpha y_{t-1} + \varepsilon_t, \tag{43}$$

$$y_t = \mu + \alpha_\mu y_{t-1} + \varepsilon_t, \quad and \tag{44}$$

$$y_t = \mu + \beta t + \alpha_\tau y_{t-1} + \varepsilon_t, \quad y_0 \; arbitrary. \tag{45}$$

Now for the DGP in equation (39) if one employes the regression (45) the appropriate critical values for $\widehat{\alpha_\tau}$ and its associate $t-statistic$ are given in the 3^{rd} part of Tables 85.1 and 8.5.2 in Fuller (1976). The same Tables can be employed to make inference in DGP in equation (41), even the is a non zero μ in this DGP. This is because regression in (45) yields a similar test (see Banerjee, Dolado, Galbraith and Hendry 1993). Similarity implies that the distributions of α_τ and its associated $t-statistic$ are not affected by the value, under the null, of the nuisance parameter, and the critical values are the same as the ones that would apply for $\mu = 0$, namely, those in the 3^{rd} part of Tables 8.5.1 and 8.5.2 in Fuller (1976).

Notice that in DGP in equation (39) there are no nuisance parameters, so that similarity is a trivial property. In general, a similar test having a Dickey-Fuller distribution requires that the regression employed contain more parameters than the DGP. In order to have a similar test for the DGP in equation (42), one would then need a regression with a term such as t^2, necessitating another block of critical values, not included in Tables 8.5.1 and 85.2 in fuller (1976). On the other hand, for the DGP in equation (40) we need at least the regression in equation (44) (with a constant) to allow for the unknown starting value. For the DGP in equation (41) we need the trend term in regression in equation (45) to allow for its effect (see Kiviet and Phillips 1992 or Banerjee et al 1993). Finally, in the case of exact parameterizations, such as DGP in equation (41) with regression in equation (44), we do not have similar tests with the Dickey-Fuller distributions. However, as West (1988) showed, the $t-statistics$ in the exactly parameterized case are asymptotically normal (see previous subsection). In finite samples, however, the Dickey-Fuller distributions may be a better approximation than the normal distribution (see Banerjee et al 1993).

9.5 Dropping the Independence assumption

There are two strands in the literature on how one can deal with possible correlation in the errors. One is based on changing the estimating equation and the other is based on modifying the test statistics.

Changing the Estimating Equation-The ADF test

Let us consider the following $ARIMA(p,1,0)$ model, i.e. the first difference is an $AR\left(p\right)$ process:

$$\Delta y_t = a_1 \Delta y_{t-1} + a_2 \Delta y_{t-2} + ...a_p \Delta y_{t-p} + z_t \qquad z_t \sim iidN\left(0,1\right).$$

Then we can test the unit root hypothesis by regressing Δy_t on p lags of Δy_t and y_{t-1}, i.e.

$$\Delta y_t = \rho y_{t-1} + \sum_{i=1}^{p} a_i \Delta y_{t-i} + e_t$$

and test the null hypothesis $H_0 : \rho = 0$. Then $\widehat{\rho}$ and the $t-statistic$ of $\widehat{\rho}$ follow the Dickey-Fuller distribution and the Tables in Fuller (1976) apply. This is the augmented Dickey-Fuller (ADF) test. Hence, the asymptotic distribution of $\widehat{\rho}$, under the null, is not affected by the presence of the lagged $\Delta y_t's$.

The reason behind this result is that the asymptotic correlation of an $I\left(1\right)$ and an $I\left(0\right)$ stochastic process is zero (see Maddala and Kim 1998). To see this consider the following DGP:

$$y_t = a y_{t-1} + z_t \qquad z_t \sim iidN\left(0,1\right).$$

Now under the null that $a = 1$ we have that $\Delta y_t = z_t$.

Consider the regression

$$y_t = a y_{t-1} + \beta \Delta y_{t-1} + z_t.$$

We shall demonstrate that the estimators $\widehat{a}$ and $\widehat{\beta}$ are asymptotically independent. As y_t is, under the null, $I(1)$ and Δy_t is $I(0)$, we need different scaling for the two estimators. Hence,

$$y_t = Ta\frac{y_{t-1}}{T} + \sqrt{T}\beta\frac{\Delta y_{t-1}}{\sqrt{T}} + z_t = \left(\frac{y_{t-1}}{T}, \frac{\Delta y_{t-1}}{\sqrt{T}}\right)\begin{pmatrix} Ta \\ \sqrt{T}\beta \end{pmatrix} + z_t.$$

It follows that as, under the null, $\Delta y_t = z_t$ and it follows that

$$\begin{pmatrix} T\widehat{a} \\ \sqrt{T}\widehat{\beta} \end{pmatrix} = \begin{pmatrix} \frac{1}{T^2}\sum y_{t-1}^2 & \frac{1}{T\sqrt{T}}\sum y_{t-1}z_{t-1} \\ \frac{1}{T\sqrt{T}}\sum y_{t-1}z_{t-1} & \frac{1}{T}\sum z_{t-1}^2 \end{pmatrix}^{-1}\begin{pmatrix} \frac{1}{T}\sum y_{t-1}y_t \\ \frac{1}{\sqrt{T}}\sum y_t z_{t-1} \end{pmatrix}.$$

Now we have from lemma 9.1(ii) and (v) we have that

$$\frac{1}{T^2}\sum y_{t-1}^2 \xrightarrow{D} \int_0^1 [W(r)]^2 dr, \quad \frac{1}{T}\sum_{t=1}^{T} y_{t-1}z_{t-1} \xrightarrow{D} \int_0^1 W(r)\, dW(r)$$

and it follows that

$$\frac{1}{T\sqrt{T}}\sum y_{t-1}z_{t-1} \xrightarrow{P} 0$$

and it follows that the distributions of $\widehat{a}$ and $\widehat{\beta}$ are asymptotically independent.

In case that error term is a stationary invertible $ARMA(p,q)$ process Said and Dickey (1984) suggest to approximate the $ARMA$ structure by a high order AR one. This is based on the fact that any invertible MA process can by approximated by a high order AR one. Consequently, Said and Dickey (1984) suggest to employ an AR approximation with order that is controled by $T^{1/3}$. Further the order p in the ADF procedure can be chosen via various information criteria such as AIC and BIC.

Altering the test statistics-The Phillips-Perron test

Phillips (1987) provides an alternative procedure that allows to employ the critical values in the two tables in Fuller (1976) while allowing for quite general $DGPs$, without adding extra elements in the regression model. Phillips suggests a non-parametric correction to the standard statistics to account for the autocorrelation that will be present.

Now under the assumption of not iid errors we have that

$$\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{D} \sigma_y^2 \int_0^1 [W(r)]^2 \, dr$$

and by lemma 9.2*(v)* we have that

$$T(\widehat{\alpha}-1) \xrightarrow{D} \frac{\frac{1}{2}\left(\chi_1^2 - 1\right) + \frac{\sigma_y^2-\sigma_z^2}{2\sigma_y^2}}{\int_0^1 [W(r)]^2 \, dr}$$

and analogously is adjusted the $t_{\widehat{\alpha}}$ ("$t-statistic$").

Phillips (1987), and Phillips and P. Perron (1988) suggest to estimate σ_z^2 and σ_y^2 by s^2 and s_{Tl}^2, respectively, where

$$\begin{aligned} s^2 &= \frac{1}{T}\sum_{t=1}^{T} e_t^2 \quad and \\ s_{Tl}^2 &= \frac{1}{T}\sum_{t=1}^{T} e_t^2 + \frac{2}{T}\sum_{j=1}^{l} w_{jl} \sum_{t=j+1}^{T} e_t e_{t-j} \quad where \quad w_{jl} = 1 - \frac{j}{l+1}, \end{aligned}$$

and e_t are the residuals from any of the regressions in equations 43, 44 or 45. The weights w_{jl} are such that s_{Tl}^2 is not negative (see Newey and West 1987). The choice of l should increase as T increases. However for $l = O\left(T^{1/4}\right)$, s_{Tl}^2 provides a consistent estimator of σ_y^2.

In comparing the ADF procedure with the Phillips-Perron one, notice that it may be the case that many lags of Δy_t may be needed for regressors to correct for an MA error. This could distort the size of the ADF test (see Schwert

1989a). On the other hand, the same problem appears to the Phillips-Perron test if the errors have a strong negative first order autocorrelation (see e.g. Schwert 1989a, and Phillips and Perron 1988). However, notice that Perron and Ng (1996) suggested useful modifications of the Phillips-Perron tests that solve this problem. Finally, the power of the ADF test could affected if many lags of Δy_t are needed, due to the fact for each additional lag of Δy_t results in losing one more initial observation (see Maddala and Kim 1998).

9.6 Testing more than one parameter and Alternative Tests

In cases such the $DGPs$ in equations (41) and (42) it is possible to test jointly if the parameters satisfy the null hypotheses, i.e. $\mu = 0$ and $\alpha = 1$ for the DGP in (41), and $\mu = 0$, $\beta = 0$ and $\alpha = 1$ for the DGP and (42). Dickey, and Fuller (1981) provide, by Monte Carlo, critical values for Likelihood Ratio, $t - type$ and $F - type$ statistics for the parameters of the two $DGPs$.

The critical values of Dickey and Fuller (1981) are derived under the hypothesis that the errors are white noise processes. However, they the same distributions apply if the errors follow an AR process and the ADF regression is correctly specified. Of course, the non-parametric correction of Phillps-Perron type can be applied. However, the Phillips-Perron corrections to the standard Dickey-Fuller statistics must be employed cautiously. Schwert (1989a) demonstrate, via Monte Carlo, that the critical values of the ADF test statistics, given by the standard Dickey-Fuller tables, are much more robust to the presence of moving average terms in the errors of the random-walk process than are the corresponding non-parametrically adjusted Dickey- Fuller statistics (see Schwert 1989a, and Banerjee et al. 1993 for an example).

There have been several tests for stationarity as null, although these are not as numerous as tests using unit AR root as null, e.g. Tanaka (1990), Kwiatkowski, Phillips, Schmidt, and Shin (1992), Saikkonen and Luukkonen(1993), Choi (1994), Leybourne and McCabe (1994), and Arellano and Pantula (1995). For an extensive discussion see Maddala and Kim (1998).

10 Cointegration

In order to motivate the notion of cointegration, let us consider a simple example taken from Engle and Granger (1987) and reproduced in Banerjee et al. (1993). Two series $\{x_t\}$ and $\{y_t\}$ are each integrated of order 1 and evolve according to the following data-generation process:

$$x_t + ay_t = u_t, \qquad u_t = u_{t-1} + \varepsilon_t \tag{46}$$

$$x_t + by_t = e_t, \qquad e_t = \varphi e_{t-1} + v_t, \qquad |\varphi| < 1, \tag{47}$$

$$\begin{pmatrix} \varepsilon_t \\ v_t \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon,v} \\ \sigma_{\varepsilon,v} & \sigma_v^2 \end{pmatrix}\right).$$

Solving for x_t and y_t from the above system we get that

$$x_t = \frac{a}{a-b}e_t - \frac{b}{a-b}u_t \quad and \tag{48}$$

$$y_t = \frac{1}{a-b}u_t - \frac{1}{a-b}e_t \tag{49}$$

provided that $a \neq b$. Now x_t and y_t are $I(1)$ random variables, as they are linear combinations of a random walk, u_t, and a stationary random variable, e_t. However, $x_t + by_t$ is stationary. In this example the vector $(1, b)'$ is called the **cointegrating vector** and $x_t + by_t$ is the **long-run equilibrium** relationship and the regression in (47) is called **cointegrating regression**. The case $a = b$ is excluded as if $a = b$ then we have that

$$x_t + ay_t - x_t - by_t = u_t - e_t \Rightarrow u_t = e_t$$

which is impossible as u_t is a random walk and e_t is stationary. Notice that x_t and y_t are both driven by the same random walk process, i.e. the u_t. Provided now that x_t and y_t have a long-run equilibrium, although they are random walks, it is only natural to think that there must be a mechanism that ties them down, which is called the **Error Correction Mechanism** (ECM). To see this subtract x_{t-1} and y_{t-1} from equations (48) and (49), respectively, multiply (49) by b and add. We get

$$\Delta x_t = -b\Delta y_t + (\varphi - 1)\, e_{t-1} + v_t. \tag{50}$$

This describes the ECM of the processes x_t and y_t. Notice that, as by assumption $|\varphi| < 1$, we have that the coefficient of e_{t-1} is negative. In the following figure the cointegrated x_t and z_t are presented where the cointegrating vector is $(1, -0.8)'$.

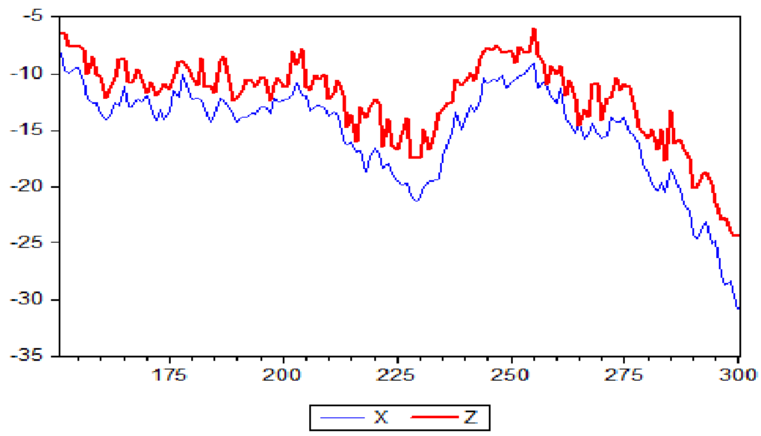


Fig. 10: $x_t = x_{t-1} + u_t$, $u_t \sim N(0, 1)$, $z_t = 0.8x_t + v_t$, $v_t \sim N(0, 1)$

The ECM is closely related to the 'general to specific modelling' notion emphasized by Hendry. In this context the ECM can be interpreted as a reparameterisation of the general 'auto-regressive distributed lag' (ADL) or 'dynamic linear regression' (DLR) models (see Alogoskoufis and Smith 1991, and Maddala and Kim 1998).

As an example, consider the first order ADL for z_t and x_t, i.e.

$$z_t = a_0 + b_0 x_t + b_1 x_{t-1} + a_1 z_{t-1} + \varepsilon_t$$

where ε_t is white noise. Now subtracting from both sides z_{t-1} and x_{t-1} we get

$$\Delta z_t = b_0 \Delta x_t - \lambda \left(z_{t-1} - \alpha_0 - \gamma x_{t-1} \right) + \varepsilon_t$$

where $\lambda = 1 - a_1$, $\alpha_0 = \frac{a_0}{1-a_1}$ and $\gamma = \frac{b_0+b_1}{1-a_1}$, which of course is the ECM of z_t and x_t.

In this ECM parameterisation b_0 is interpreted as impact effect, λ as a scalar adjustment coefficient, and γ as a long-run effect (see Alogoskoufis and Smith 1991). The ECM and DLR as well as other reparameterizations are all observationally equivalent and consequently, there are no statistical criteria that one can use to choose between them (see Alogoskoufis and Smith 1991). To quote Alogoskoufis and Smith (1991) "They are all observationally equivalent, thus there are no statistical criteria that we can use to choose between them. The questions that arise then, relate to the parameters of interest from the point of view of economic theory. These questions can only be answered by an explicit theory".

Let us turn our attention to testing and estimating the cointegrating regression. First, we shall consider the two variables case, the so called Engle-Granger method.

10.1 The Engle-Granger method

Assume that x_t and y_t are both $I\left(1\right)$ random variables and

$$x_t = a y_t + u_t \tag{51}$$

where u_t is $I(0)$. First of all, the cointegrating vector $(1, -a)'$ is unique. To see this consider that there exist a second cointegrating vector, say $(1, -a^*)'$ such that $a \neq a^*$, i.e.

$$x_t = a^* y_t + v_t$$

where v_t is $I(0)$. Then subtracting the two equations we get

$$(a^* - a) y_t = u_t - v_t.$$

But the right hand side variables are $I(0)$ and consequently the linear combination of $I(0)$ random variables is a $I(0)$ random variable. However, as y_t is by assumption $I(1)$ we have that a $I(1)$ random variable equals a $I(0)$ random variable, which is a contradiction.

Second, the OLS estimator of a, say $\widehat{a}$, is superconsistent (Stock 1987), i.e. it converges at a rate $n^{1-\delta}$, for any $\delta > 0$, instead of the usual $n^{1/2}$. This is because the term $\sum_{t=1}^{n} y_t^2$ is $O_p(n^2)$. However, there is strong Monte Carlo evidence that $\widehat{a}$ is substantially biased in small samples (see Banerjee et al. 1986 and Banerjee et al. 1993). In fact Banerjee et al. (1986) demonstrate that the bias of $\widehat{a}$ is related to $1 - R^2$, where R^2 is the one from the regression in (51). Hence, for high value of R^2 the bias is small.

Another consequence of the high convergence rate is fact that the OLS estimator of the cointegrated vector does not require the assumption that the regressors are uncorrelated withy the error term. Hence, any of the cointegrated variables can be employed as dependent variable in the regression, i.e. if instead (51) we estimate

$$y_t = b x_t + \varepsilon_t$$

the OLS estimator of b is still superconsistent. In fact, if the R^2 from the regression in (51) is very close to 1 then $\widehat{b} \approx \frac{1}{\widehat{a}}$ (see Maddala and Kim 1998 for more on this normalisation issue).

The Two-step Procedure

Engle and Granger (1987) suggested a two-step procedure to estimate the ECM regression, equation (50). In the first step estimate the long-run (cointegrating) relationship, by the following regression

$$x_t = ay_t + u_t.$$

and get the residuals, say $\widehat{u}_t$. Then, in the second step, estimate the Error Correction relationship by the following regression:

$$\Delta x_t = \gamma \Delta y_t + \theta \widehat{u}_{t-1} + v_t.$$

As an example consider the two time series depicted in Figure (10). Regressing z_t on x_t , i.e. cointegrating regression,

$$z_t = c + ax_t + u_t$$

we get the following results:

Dependent Variable: z_t				
Method: Least Squares				
Incl. observations: 300				
Variable	Coefficient	Std. Error	t-Statistic	Prob
c	-0.081777	0.128689	-0.635465	0.5256
a	0.790082	0.006993	112.9888	0.0000

The Error Correction regression is

$$\Delta z_t = c + \gamma \Delta x_t + \theta \widehat{u}_{t-1} + v_t,$$

where $\widehat{u}_{t-1}$ is the residuals from the cointegrating regression and we get the following results:

Dependent Variable: Δz_t				
Method: Least Squares				
Incl. observations: 299				
Variable	Coefficient	Std. Error	t-Statistic	Prob
c	-0.015233	0.058147	-0.261977	0.7935
γ	0.657618	0.057516	11.43358	0.0000
θ	-0.994096	0.057594	-17.26040	0.0000

Although $\widehat{a}$, the coefficient of x_t, is superconsistent however, in small samples it is biased (see Banerjee et al. 1986) and inefficient, as compared to Full Information MLE.

The Three-Step Procedure

Engle and Yoo (1991) proposed a 3-step procedure, the 2 step from the Engle and Granger (1987) one and a third one to correct for the small sample bias of a , the cointegrating regression coefficient, and provides a set of standard errors so that the t-statistics are valid.

In the third step regress the residuals for the ECM on the right hand variables of the cointegrating regression multiplied by the minus the error correction parameter, i.e. by $-\widehat{\theta}$. For our example, regress $\widehat{v}_t$ on a constant and $-\widehat{\theta}x_t$. In doing this we get the following results (where $HEL = -\widehat{\theta}x_t$):

Dependent Variable: RESECM				
Method: Least Squares				
Incl. observations: 299				
Variable	Coefficient	Std. Error	t-Statistic	Prob
C	0.024588	0.128273	0.191682	0.8481
HEL	0.001512	0.007044	0.214661	0.8302

Then the bias corrected estimator of a, say $\widetilde{a}$, is given by: $\widetilde{a} = \widehat{a} + HEL$, i.e. $\widetilde{a} = 0.790082 + 0.001512 = 0.791594$ and its standard error is given by the standard error of HEL, i.e. $s.e.\left(\widetilde{a}\right) = 0.007044$, as compared with 0.006993 of the cointegrating regression.

Notice that in case that x_t has a drift, i.e. $x_t = \mu + x_{t-1} + u_t$, then the estimator of a, $\widehat{a}$, is normally distributed, as in section Unit Root with Drift, above (for details see Maddala and Kim 1998).

Let us turn our attention on the multivariate case.

10.2 The Johansen method

Let us assume that X_t is an $I\left(1\right)$ $\left(k \times 1\right)$ vector that obeys the following VAR equation:

$$X_t = A_1 X_{t-1} + A_2 X_{t-2} + ... + A_p X_{t-p} + V_t$$

where $V_t \sim iidN\left(0, \Omega\right)$. It follows that

$$\Delta X_t = B_1 X_{t-1} + B_2 \Delta X_{t-1} + B_3 \Delta X_{t-2} ... + B_p \Delta X_{t-p+1} + V_t$$

where $B_1 = \sum_{i=1}^{p} A_i - I$ and $B_i = -\sum_{j=i}^{p} A_j$ for $i = 2, ..., p$.

As an example consider the case where $p = 3$. Then

$$X_t = A_1 X_{t-1} + A_2 X_{t-2} + A_3 X_{t-3} + V_t$$

and subtracting from both sides X_{t-1} and adding and subtracting $A_3 X_{t-2}$ in the right hand side we get:

$$X_t - X_{t-1} = -X_{t-1} + A_1 X_{t-1} + \left(A_2 + A_3\right) X_{t-2} - A_3 \left(X_{t-2} - X_{t-3}\right) + V_t.$$

Adding and subtracting $\left(A_2 + A_3\right) X_{t-1}$ we get

$$X_t - X_{t-1} = \left(A_1 + A_2 + A_3 - I\right) X_{t-1} - \left(A_2 + A_3\right)\left(X_{t-1} - X_{t-2}\right) - A_3 \left(X_{t-2} - X_{t-3}\right) + V_t$$

or

$$\Delta X_t = B_1 X_{t-1} + B_2 \Delta X_{t-1} + B_3 \Delta X_{t-2} + V_t \tag{52}$$

where $B_1 = A_1 + A_2 + A_3 - I$, $B_2 = -A_2 - A_3$ and $B_3 = -A_3$, as required.

As now ΔX_{t-1} and ΔX_{t-2} are stationary, i.e. $I(0)$, but X_{t-1} is $I(1)$ for the above equation to be meaningful the B_1 matrix must be of reduced rank, say r. Then $B_1 = ab'$, where a is an $n \times r$ matrix and b' is an $r \times n$ one. This in fact means that $b' X_{t-1}$ are the r cointegrating equations. Notice that in this set-up, a has the interpretation of Error Correction terms.

Since our interest is in a and b' we first eliminate B_2 and B_3 (consecrate out B_2 and B_3), by regressing ΔX_t on ΔX_{t-1} and ΔX_{t-2} and call the residuals R_{0t}. Further regress X_{t-1} on the same variables and call the residuals R_{1t}. Hence, the regression in (52) has become

$$R_{0t} = ab' R_{1t} + V_t.$$

To understand the logic, consider the following regression (see Johansen 1995):

$$x_t = \alpha y_t + \beta z_t + u_t. \tag{53}$$

where x_t, y_t and z_t are zero mean univariate random variables. Then, the first order condition for β, either from a normal likelihood or from a regression, is given by

$$\sum_t \left(x_t - \alpha y_t - \widehat{\beta} z_t \right) z_t = 0$$

and it follows that

$$\widehat{\beta} = \frac{\sum_t x_t z_t}{\sum_t z_t^2} - \alpha \frac{\sum_t y_t z_t}{\sum_t z_t^2}.$$

Now substituting out β from the regression equation (53) we get

$$x_t - \frac{\sum_t x_t z_t}{\sum_t z_t^2} z_t = \alpha \left(y_t - \frac{\sum_t y_t z_t}{\sum_t z_t^2} z_t \right) + u_t.$$

However notice that $x_t - \frac{\sum_t x_t z_t}{\sum_t z_t^2} z_t$ is the residual, at time t, from the regression of x_t on z_t, say R_{0t}, whereas $y_t - \frac{\sum_t y_t z_t}{\sum_t z_t^2} z_t$ is the residual, at time t, of the y_t on z_t, say R_{1t}. Hence, by concetnrating out β, α can be estimated from the regression of R_{0t} on R_{1t}, i.e.

$$R_{0t} = \alpha R_{1t} + u_t.$$

Hence concentrating out B_2 and B_3 the likelihood function, of a sample of T observations, is proportional to:

$$L(a, b, \Omega) = |\Omega|^{-\frac{T}{2}} \exp \left[-\frac{1}{2} \sum_{t=1}^{T} (R_{0t} - ab' R_{1t})' \, \Omega^{-1} (R_{0t} - ab' R_{1t}) \right].$$

Now if b were known, a and Ω can be estimated in the usual regression of R_{0t} on $b' R_{1t}$. Hence, $\widehat{a}$ and $\widehat{\Omega}$, as functions of b, are given by:

$$\begin{aligned} \widehat{a}(b) &= S_{01} b (b' S_{11} b)^{-1} \\ \widehat{\Omega}(b) &= S_{00} - S_{01} b (b' S_{11} b)^{-1} b' S_{10} \quad where \\ S_{00} &= \frac{1}{T} \sum_{t=1}^{T} R_{0t} R_{0t}', \quad S_{10} = \frac{1}{T} \sum_{t=1}^{T} R_{1t} R_{0t}' \;\; and \;\; S_{11} = \frac{1}{T} \sum_{t=1}^{T} R_{1t} R_{1t}', \end{aligned}$$

and they are the sample counterparts of the variance of R_{0t}, the covariance of R_{0t} and R_{1t}, and the variance of R_{1t}, respectively (see Johansen 1991). After concentrating out a and Ω from the likelihood function we get that the likelihood becomes proportional to:

$$L(b) = \left|\widehat{\Omega}(b)\right|^{-\frac{T}{2}} = \left|S_{00} - S_{01}b\left(b'S_{11}b\right)^{-1}b'S_{10}\right|^{-\frac{T}{2}}.$$

Hence maximizing the likelihood with respect to b is equivalent to minimize $\left|S_{00} - S_{01}b\left(bS_{11}b'\right)^{-1}b'S_{10}\right|$ with respect to b.

Now for appropriate order matrices we know that if R represent a nonsingular $n \times n$ matrix, S an $n \times m$ matrix, T a nonsingular $m \times m$ matrix, and U an $m \times n$ matrix, then (see Theorem 18.1.1 in Harville 1997)

$$|R + STU| = |R|\,|T|\,\left|T^{-1} + UR^{-1}S\right|. \tag{54}$$

Consequently, setting $R = S_{00}$, $T = \left(b'S_{11}b\right)^{-1}$, $S = -S_{01}b$ and $U = b'S_{10}$ we get that

$$\left|S_{00} - S_{01}b\left(b'S_{11}b\right)^{-1}b'S_{10}\right| = \frac{\left|b'S_{11}b - b'S_{10}S_{00}^{-1}S_{01}b\right|\,|S_{00}|}{|b'S_{11}b|}$$

and consequently, we should minimize (recall that b' is an $r \times n$)

$$\frac{\left|b'S_{11}b - b'S_{10}S_{00}^{-1}S_{01}b\right|\,|S_{00}|}{|b'S_{11}b|} = |S_{00}|\,\frac{\left|b'\left(S_{11} - S_{10}S_{00}^{-1}S_{01}\right)b\right|}{|b'S_{11}b|}.$$

This expression is minimized by solving the eigenvalue problem (see Lemma A.8 in Johansen 1995, or Banerjee et al. 1993, or Anderson 2003)

$$\begin{aligned}\left|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}\right| &= 0 \quad or \\ \left|S_{11}^{-1}S_{10}S_{00}^{-1}S_{01} - \lambda I\right| &= 0\end{aligned}$$

where $\lambda_i's$ are the r largest eigenvalues of $S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}$, which are the r canonical correlations of R_{1t} and R_{0t}.

Now it is known that if λ is an eigenvalue of a matrix, say A, then $1 - \lambda$ is the eigenvalue of $I - A$. Further, the determinant of a matrix equals the product of its eigenvalues. Hence we get that

$$\prod_{i=1}^{k}(1-\lambda_i)=\left|I-S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}\right|=\frac{\left|S_{11}-S_{10}S_{00}^{-1}S_{01}\right|}{\left|S_{11}\right|}.$$

It follows that finding the r largest eigenvalues we can find the eigenvectors that correspond to each of them and consequently to estimate the $n\times r$ matrix b, say $\widehat{b}$. Then $\widehat{b}'X_{t-1}=Z_{t-1}\sim I\left(0\right)$.

Consequently, employing equation (54) we get that

$$L_{\max}^{-\frac{2}{T}}=\left|S_{00}-S_{01}\widehat{b}\left(\widehat{b}'S_{11}\widehat{b}\right)^{-1}\widehat{b}'S_{10}\right|=C\prod_{i=1}^{r}\left(1-\widehat{\lambda}_i\right)$$

and the Likelihood Ratio test, for the null that there are at most r cointegrating rations $(0\leq r<k)$, is given

$$LR=-T\sum_{i=r+1}^{k}\log\left(1-\widehat{\lambda}_i\right),\qquad r=0,1,...,k-1.$$

Johansen (1988) suggests a χ^2 approximation to the non-standard distribution of the LR statistic, i.e.

$$LR\simeq\left[0.85-\frac{0.58}{2\left(k-r\right)^2}\right]\chi^2_{2(k-r)^2},\qquad r=0,1,...,k-1.$$

Hence, having estimated the r cointegrating relations, equation (52) is balanced and consequently a , the Error Correction coefficients, can be estimated by the following regression

$$\Delta X_t=aZ_{t-1}+B_2\Delta X_{t-1}+B_3\Delta X_{t-2}+V_t.$$

The main critique to the Johansen procedure is that it is extremely sensitive to the normality assumptions of the errors. Departure of this assumption could lead to spurious cointegration, high variance and high probability of producing outliers. There are alternative to Johansen procedures in a multivariate setup. Interested reads are referred to Maddala and Kim (1998).

References

Alogoskoufis, G. and R. Smith (1991) "On error correction models: Specification, interpretation, estimation" *Journal of Economic Surveys* 5, 97–128.

Anderson T. W. (2003) " *The Statistical Analysis of Time Series*" Wiley Classics, J. Wiley & Sons, N. York.

Arellano, C. and S. G. Pantula (1995) "Testing for Trend Stationarity versus Difference Stationarity" *Journal of Time Series Analysis* 16, 147-164.

Arvanitis, S. (2013) "A Simple Example of An Indirect Estimator With Discontinuous Limit Theory in the MA(1) Model" DP

Arvanitis, S. and A. Demos (2004) "Time Dependence and Moments of a family of Time-Varying parameter GARCH in Mean Models" *Journal of Time Series Analysis* 25, 1-25.

Arvanitis, S. and A. Demos (2004b) "Conditional Heteroskedasticity in Mean Models" in *Quantitative Methods in Finance in Honor of Prof. A. Kintis*, Editor A. Refenes, Typothito, 2004.

Banerjee, A., J.J. Dolado, D.F. Hendry and G.W. Smith (1986) "Exploring Equilibrium Relationships in Econometrics through Static Models: Some Monte Carlo Evidence" *Oxford Bulletin of Economics and Statistics*, 48, 253-277.

Banerjee, A., J.J. Dolado, J.W. Galbraith and D.F. Hendry (1993) *"Co-integration, error correction, and the econometric analysis of non-stationary data"* Oxford University Press, N. York.

Bhargava, A. (1986) "On the Theory of Testing for Unit Roots in Observed Time Series" *The Review of Economic Studies*, 53, 369-384.

Billingsley, P. (1968) "Convergence of Probability Measures", New York, J. Wiley.

Bollerslev, T. (1986) "Generalized Autoregressive Conditional Heteroskedasticity" *Journal of Econometrics* 31, 307-327.

Bollerslev, T. (1988) "On the Correlation Structure for the Generalized Autoregressive Conditional Heteroskedastic Process" *Journal of Time Series Analysis* 9, 121-131.

Bollerslev, T., R. Y. Chou, and K. F. Kroner (1992) "ARCH Modelling in Finance: A Review of the Theory and Empirical Evidence" *Journal of Econometrics* 52, 5-59.

Brooks, C. (2008) "Introductory Econometrics for Finance", second edition, Cambr. Univ. Press.

Choi, I. (1994) "Residual-Based Tests for the Null of Stationary with Application to US Macroeconomic Time Series" *Econometric Theory* 10, 720-746.

Chou, R., R. F. Engle and A. Kane (1992) "Measuring Risk Aversion from Excess Returns on a Stock Index" *Journal of Econometrics* 52, 201-224.

Demos, A. (2002) "Moments and Dynamic Structure of a Time-Varying-Parameter Stochastic Volatility in Mean Model", *The Econometrics Journal* 5.2, 345-357.

Cuthbertson, K., S.G. Hall and M.P. Taylor (1992), "*Applied Econometric Techniques*" APhilip Allan.

Dhrymes, P. (1998) "*Time series, unit roots, and cointegration*" Academic Press.

Dickey, D.A. and W.A Fuller (1979) "Distribution of the Estimator for Autoregressive Series with a Unit Root" *Journal of the American Statistical Association* 74, 427-431.

Dickey, D.A. and W.A Fuller (1981) "Likelihood ratio statistics for autoregressive time series with a unit root" *Econometrica* 49, 1057-1072.

Ding Z., C. W. J. Granger and R. F. Engle (1993) "A Long Memory Property of Stock Market Returns and a New Model" *Journal of Empirical Finance* 1, 83-106.

Durlauf, S.N. and P.C.B. Phillips (1988) "Trends versus random walks in time series analysis" *Econometrica* 56, 1333–54.

Engle, R. F. (1982) "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation" *Econometrica* 50, 987-1007.

Engle, R.F. and C.W.J. Granger (1987) "Co-Integration and error Correction: Representation, Estimation, and Testing" *Econometrica* 55, 251-276.

Engle, R. F., D. M. Lilien and R. P. Robins (1987) "estimating Time-Varying Risk Premia in the Term Structure: The ARCH-M Model" *Econometrica* 55, 391-407.

Engle, R. F. and V. K. Ng (1993) "Measuring and Testing the Impact of News on Volatility" *Journal of Finance* 48, 1749-1778.

Engle, R.F. and B.S. Yoo (1991) "Cointegrated economic time series: An overview with new results" *Long-Run Economic Relationships. Readings in Cointegration*, R.F. Engle and C.W.J. Granger editors, Oxford University Press.

Entorf, H. (1992) "Random walks with drift, simultaneous equation errors, and small samples: Simulating the bird's-eye view" Discussion Paper no. 9222 CREST.

Fiorentini G. and Sentana, E. (1998) "Conditional Means of Time Series Processes and Time Series Processes for Conditional Means" *International Economic Review* 39, 1101-1118.

Francq, C. and J.M. Zakoian (2010), *Garch Models; Structure, Statistical Inference and Fianancial Applications*, J. Wiley & Sons, UK.

French, K. R., W. G. Schwert and R. F. Stambaugh (1987) "Expected Stock Returns and Volatility" *Journal of Financial Economics* 19, 3-29.

Fuller, W.A. (1976), *Introduction to Statistical Time Series*, J. Wiley & Sons, New York.

Gallant A.R., D. Hsieh and G. Tauchen (1997) "Estimation of Stochastic

Volatility Models with Diagnostics" *Journal of Econometrics* 81, 159-192.

Glosten, L. R., R. Jagannathan and D. E. Runkle (1993) "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks" *Journal of Finance* 48, 1779-1801.

Gourieroux C. and J. Jasiak (2001) *Financial Econometrics*, Princeton

Gourieroux C. (1997) *"ARCH Models and Financial Applications"*, Springer

Gradshteyn, I.S. and I.M. Ryzhik (1994) "*Table of Integrals, Series, and Products*" A. Jeffrey editor, Academic Press.

Granger, W.J. and P. Newbold (1974) "Spurious Regressions in Econometrics," *Journal of Econometrics* 2, 111-120.

Granger, C.W.J. and P. Newbold (1976) "Forecasting Transformed Series" *Journal of the Royal Statistical Society* 38B, 189-203.

Hamilton J. D. (1994) "*Time Series Analysis*", Princeton University Press.

Harvey, A. (1993) *"Time Series Models"* Harvester Wheatsheaf.

Harvey, A., E. Ruiz and N. Shephard (1994) "Multivariate Stochastic Variance Models" *Review of Economic Studies* 61, 247-264.

Harvey, A. and N. Shephard (1996) "Estimation of an Asymmetric Stochastic Volatility Model for asset Returns" *Journal of Business and Economic Statistics* 14, 429-434.

Harville, D.A. (1997) "Matrix Algebra from a Statisticians Perspective", Springer.

He C. (2000) "Moments and the Autocorrelation Structure of the Exponential GARCH(p,q) Process" Working Paper 359, Stockholm School of Economics.

He C. and T. Terasvirta (1999a) "Properties of Moments of a Family of GARCH Processes" *Journal of Econometrics* 92, 173-192.

He C. and T. Terasvirta (1999b) "Fourth Moment Structure of the GARCH(p,q) Process" *Econometric Theory* 15, 824-846.

He C., T. Terasvirta and H. Malmsten (1999) "Fourth Moment Structure of a Family of First-Order Exponential GARCH Models" Working Paper 345, Stockholm School of Economics.

Hentschel, L. (1995) "All in the Family. Nesting symmetric and asymmetric GARCH models" *Journal of Financial Economics* 39, 71-104.

Hong E.P. (1991) "The Autocorrelation Structure for the GARCH-M Process" *Economics Letters 37*, 129-132.

Johansen, S. (1988) "Statistical Analysis of Cointegration Vectors", *Journal of Economic Dynamics and Control* 12, 231-254.

Johansen, S. (1991) "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models", *Econometrica* 59, 1551-1580.

Johansen, S. (1995) "*Likelihood Based Inference in Cointegrated Vector Autoregression Models*" Oxford Univ. Press.

Karanasos M. "The Second Moment and Autocovariance Function of the Squared Errors of the GARCH Model" *Journal of Econometrics* 90, 63-76.

Kiviet, J.F. and G.D.A. Phillips (1992) "Exact similar tests for unit roots and cointegration" *Oxford Bulletin of Economics and Statistics* 54, 349-367.

Kwiatkowski D., P.C.B. Phillips, P. Schmidt, and Y. Shin (1992) "Testing the Null Hypothesis of Stationary against the Alternative of a Unit Root" *Journal of Econometrics* 54, 159-178.

Leybourne S.J. and B.P.M. McCabe (1994) "A Consistent Test for a Unit Root" *Journal of Business and Economic Statistics*, 12, 157-166.

Maddala, G.S. and I.M. Kim (1998) "*Unit roots cointegration and structural change*" sixth print 2004, Cambr. Univ. Press.

Mann, H.B. and A. Wald (1943), "On Stochastic Limit and Order Relationships", *Annals of Mathematical Statistics* 14, 217-277.

Merton, R. C. (1980) "On Estimating the Expected Return on the Market: An Exploratory Investigation" *Journal of Financial Economics* 8, 323-361.

Mills T. (1997) *"The Econometric Modelling of Financial Time Series"*, Cambridge Univ. Press

Nelson, C.R. and H. Kang (1981) "Spurious Periodicity in Inappropriately Detrended Time Series", *Econometrica*, 49, 741-751.

Nelson, D.B. (1991) "Conditional Heteroskedasticity in Asset Returns: A New Approach" *Econometrica* 59, 347-370.

Newey, W.K. and K.D. West (1987), "A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation-Consistent Covariance Matrix" *Econometrica* 55, 703-708.

Patterson, K.D. (2011) *"Unit Root Tests in Time Series, Volume 1: Key Concepts and Problems"* Palgrave and Macmillan.

Perron, P. and S. Ng (1996), "Useful Modifications to Some Unit Root Tests with Dependent Errors and their Local Asymptotic Properties" *Review of Economic Studies*, 63, 435-465.

Phillips, P.C.B. (1986) "Understanding Spurious Regressions in Econometlics," *Journal of Econometrics* 33, 311-340.

Phillips, P.C.B. (1987) "Time Series Regression with a Unit Root" *Econometrica*, 55, 277-301.

Phillips, P.C.B. and P. Perron (1988) "Testing for a Unit Root in Time Series Regression" *Biometrika*, 75, 335-346.

Poterba, J.M. and L.H. Summers (1988) "Mean Reversion in Stock Prices: Evidence and Implications" *Journal of Financial Economics* 22, 27-59.

Said, S.E and D.A. Dickey (1984) "Testing for unit roots in autoregressive-moving average models of unknown order" *Biometrika* 71, 599-607.

Saikkonen, P. and R. Luukkonen (1993) "Testing for a Moving Average Unit

Root in ARIMA Models" *Journal of the American Statistical Association* 88, 596-601.

Sargan, J.D. and A. Bhargava (1983), "Maximum Likelihood Estimation of Regression Models with First Order Moving Average Errors when the Root Lies on the Unit Circle.", *Econometrica* 51, pp. 799-820.

Schmidt, P. and P.C.B. Phillips (1992) "LM tests for a unit root in the presence of deterministic trends" Oxford Bulletin of Economics and Statistics 54, 257-287.

Schwert, G.W. (1989) "Why does Stock Market Volatility change over Time?" *Journal of Finance* 45, 1129-1155.

Schwert, G.W. (1989a), "Tests for Unit Roots: A Monte Carlo Investigation" *Journal of Business and Economic Statistics,* 7, 147-159.

Sentana, E. (1995) "Quadratic ARCH Models", *Review of Economic Studies* 62, 639-661.

Shephard N. (1993), "Distribution of the ML Estimator of an MA(1) and a Local Level Model", *Econometric Theory* 9, 377-401.

Shephard, N. (1996) "Statistical Aspects of ARCH and Stochastic Volatility" in "*Time Series Models in Econometrics, Finance and other Fields*" edit: Cox, D.R. D.V. Hinkley and O.E. Barndorff-Nielsen, Chapman and Hall.

Steele, J.M. (2001) "*Stochastic Calculus and Financial Applications*" Springer-Verlag.

Stock J.H. (1987) "Asymptotic properties of least squares estimators of cointegrating vectors", *Econometrica* 55, 1035-1056.

Tanaka K. (1990), "Testing for a Moving Average Unit Root" *Econometric Theory* 6, 433-444.

Taylor, S.J. (1995) "*Modelling Financial Time Series*" J. Wiley & Sons.

West, K.D. (1988) "Asymptotic normality, when regressors have a unit root", *Econometrica* 56, 1397-1417.

White, J.S. (1958) "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case", *The Annals of Mathematical Statistics* 29, 1188-1197.

Yang, M. and R. Bewley (1995) "Moving Average Conditional Heteroskedastic Models" *Economics Letters* 49, 367-372.

Zakoian, J. M. (1994) "Threshold Heteroskedastic Models" *Journal of Economic Dynamics and Control* 18, 931-955.