FORECASTING

Herman J. Bierens

Pennsylvania State University

April 30, 2012

1. *Conditional expectations as best forecasting schemes*

Consider a pair of random variables, X and Y, for which you know the joint distribution. Suppose that Y is not yet observed, and that you want to forecast Y, given that you have observed X. The question is: what is the best forecasting scheme?

Any forecasting scheme for Y is a function of X, for example

$$\hat{Y} = \varphi(X). \tag{1}$$

Using (1) as a forecast of Y, the mean-square forecast error involved is defined as

$$E\left[\left(Y - \hat{Y}\right)^2\right] = E\left[\left(Y - \varphi(X)\right)^2\right]. \tag{2}$$

Now the best forecasting scheme is the function (1) that minimizes (2). So now the question arises: for which function φ is the mean-square forecast error (2) minimal?

The answer is: the conditional expectation function: $\varphi(X) = E[Y|X]$, as will be shown in the Appendix, section A.1. Then the forecast error is $U = Y - \varphi(X) = Y - E[Y|X]$, which satisfies $E[U|X] = E[Y|X] - E[\varphi(X)|X] = \varphi(X) - \varphi(X) = 0.$ Thus, we can always write

$$Y = \varphi(X) + U, \text{ where } E[U|X] = 0. \tag{3}$$

which is a general form of a regression model, with U the error term and $\varphi(X)$ the regression function. The property $E[U|X] = 0$ is the usual assumption about the error term of a regression model. Thus, this assumption implies that the regression function $\varphi(X)$ is equal to the conditional expectation of Y given X: $\varphi(X) = E[Y|X]$. Therefore, for a regression model to be correctly specified, the regression function has to represent the conditional expectation of the dependent variable given the regressors.

2. *Out-of-sample forecasting with a linear regression model*

Consider the linear regression model

$$Y_j = \alpha + \beta . X_j + U_j, \quad j = 1,2,.....,n, \tag{4}$$

where the unobserved error terms are assumed to be independent normally distributed:

$$U_j \sim i.i.d.\ N(0,\sigma^2), \tag{5}$$

and each U_j is independent of all the explanatory variables $X_1, X_2, ..., X_n, ...$ Let $\hat{\beta}$ and $\hat{\alpha}$ be the OLS estimators of β and α in model (4) on the basis of the observations (Y_j, X_j) for $j = 1,...,n$.

Next, suppose we observe X_{n+1}. Then the one-step ahead out-of-sample forecast of Y_{n+1} is

$$\hat{Y}_{n+1} = \hat{\alpha} + \hat{\beta} . X_{n+1}, \tag{6}$$

where the OLS estimators $\hat{\alpha}$ and $\hat{\beta}$ are computed on the basis of the observations for $j = 1,2,...,n$. The actual but unknown value of Y_{n+1} is $Y_{n+1} = \alpha + \beta . X_{n+1} + U_{n+1}$, so that the forecast error is $Y_{n+1} - \hat{Y}_{n+1}$. In the Appendix it will be shown that

$$Y_{n+1} - \hat{Y}_{n+1} \sim N[0, \sigma^2_{Y_{n+1}-\hat{Y}_{n+1}}], \text{ where } \sigma^2_{Y_{n+1}-\hat{Y}_{n+1}} = \sigma^2 \left(\frac{n+1}{n} + \frac{(X_{n+1}-\bar{X})^2}{\sum_{j=1}^n (X_j - \bar{X})^2} \right). \tag{7}$$

Moreover, denoting,

$$\hat{\sigma}^2_{Y_{n+1}-\hat{Y}_{n+1}} = \hat{\sigma}^2 \left(\frac{n+1}{n} + \frac{(X_{n+1}-\bar{X})^2}{\sum_{j=1}^n (X_j - \bar{X})^2} \right), \tag{8}$$

where $\hat{\sigma}^2$ is the (unbiased) OLS estimator of σ^2:

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{j=1}^{n} \left(Y_j - \hat{\alpha} - \hat{\beta} . X_j \right)^2, \tag{9}$$

it can be shown that

Proposition 1. *In the case of model* (4), $(Y_{n+1} - \hat{Y}_{n+1}) / \hat{\sigma}_{Y_{n+1}-\hat{Y}_{n+1}} \sim t_{n-2}$.

This result can be used to construct a 95% confidence interval, say, of Y_{n+1}. Look up in the table of the t distribution the critical value t_* of the two-sided t-test at the 5% significance level with $n-2$ degrees of freedom. Note that if $n-2 > 30$ then t_* is approximately equal to the critical value of the two-sided standard normal test at the 5% significance level: $t_* \approx 1.96$. Then it follows from Proposition 1 that

$$\begin{aligned} P[-t_* \leq (Y_{n+1}-\hat{Y}_{n+1})/\hat{\sigma}_{Y_{n+1}-\hat{Y}_{n+1}} \leq t_*] &= P[-t_*\hat{\sigma}_{Y_{n+1}-\hat{Y}_{n+1}} \leq Y_{n+1}-\hat{Y}_{n+1} \leq t_*\hat{\sigma}_{Y_{n+1}-\hat{Y}_{n+1}}] \\ &= P[\hat{Y}_{n+1} - t_*\hat{\sigma}_{Y_{n+1}-\hat{Y}_{n+1}} \leq Y_{n+1} \leq \hat{Y}_{n+1} + t_*\hat{\sigma}_{Y_{n+1}-\hat{Y}_{n+1}}] = 0.95 \end{aligned} \tag{10}$$

Thus, the 95% confidence interval of Y_{n+1} is $[\hat{Y}_{n+1} - t_*\hat{\sigma}_{Y_{n+1}-\hat{Y}_{n+1}}, \hat{Y}_{n+1} + t_*\hat{\sigma}_{Y_{n+1}-\hat{Y}_{n+1}}]$.

Observe from (8) that $\hat{\sigma}^2_{Y_{n+1}-\hat{Y}_{n+1}}$ increases with $(X_{n+1}-\bar{X})^2$, and so does the width of the confidence interval. Thus, the father X_{n+1} is away from $\bar{X}$, the more unreliable the forecast $\hat{Y}_{n+1}$ of Y_{n+1} becomes. Also observe from (8) that $\hat{\sigma}^2_{Y_{n+1}-\hat{Y}_{n+1}} \geq \hat{\sigma}^2$, and that $\hat{\sigma}^2_{Y_{n+1}-\hat{Y}_{n+1}}$ gets close to $\hat{\sigma}^2$ if n is large because $\lim_{n\to\infty}\sum_{j=1}^n(X_j-\bar{X})^2 = \infty$.

These results apply to general regression models as well, with only a minor modification. Consider the general linear regression model

$$Y_j = \beta_0 + \beta_1.X_{1,j} + ... + \beta_k.X_{k,j} + U_j, \quad j = 1,2,.....,n, \tag{11}$$

where again the error terms U_j satisfy the normality condition (5) and are independent of all the regressors $X_{i,t}$ for $i = 1,...,k$ and all observation indices t. If the regressors $X_{i,n+1}$ are observed then we can forecast Y_{n+1} by

$$\hat{Y}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1.X_{1,n+1} + ... + \hat{\beta}_k.X_{k,n+1}, \tag{12}$$

where the $\hat{\beta}_i$'s are the OLS estimators of the corresponding parameters β_i on the basis of the observations for $j = 1,....,n$. It is no longer possible to give an explicit expression for the forecast standard error $\hat{\sigma}_{Y_{n+1}-\hat{Y}_{n+1}}$ without using matrix calculus, but your econometrics software will compute $\hat{\sigma}_{Y_{n+1}-\hat{Y}_{n+1}}$ or $\hat{\sigma}^2_{Y_{n+1}-\hat{Y}_{n+1}}$. Then

Proposition 2. *In the case of model* (12), $(Y_{n+1}-\hat{Y}_{n+1})/\hat{\sigma}_{Y_{n+1}-\hat{Y}_{n+1}} \sim t_{n-k-1}$.

The difference with the result in Proposition 1 are the degrees of freedom of the t distributions involved. In Proposition 1 the degrees of freedom is the sample size n minus the number of parameters (= 2) in model (4), whereas in Proposition 2 the degrees of freedom is the sample size

n minus the number of parameters (= k+1) in model (12).

Also in this case $\hat{\sigma}^2_{Y_{n+1}-\hat{Y}_{n+1}}$ increases with $(X_{i,n+1} - \bar{X}_i)^2$, where $\bar{X}_i = (1/n)\sum_{j=1}^{n} X_{i,j}$.

The construction of the 95% confidence interval of Y_{n+1} can be done in the same way as before. Look up in the table of the t distribution the critical value t_* of the two-sided t-test at the 5% significance level with $\boldsymbol{n - k - 1}$ degrees of freedom (instead of $n-2$ degrees of freedom!). Note again that if $n-k-1 > 30$ then t_* is approximately equal to the critical value of the two-sided standard normal test at the 5% significance level: $t_* \approx 1.96$. Then it follows from Proposition 2 that $[\hat{Y}_{n+1} - t_*\hat{\sigma}_{Y_{n+1}-\hat{Y}_{n+1}}, \hat{Y}_{n+1} + t_*\hat{\sigma}_{Y_{n+1}-\hat{Y}_{n+1}}]$ is the 95% confidence interval of Y_{n+1}.

3. *Forecasting stationary time series*

Let Y_t be a time series process, and suppose that we have observed Y_t for all time periods $t \le n$. The problem is how to predict the unknown future value Y_{n+1}, given that we have observed Y_t for all time periods $t \le n$. In practice of course we only observe Y_t from some initial time period onwards, say $t = 1$, but I will deal with that problem later.

The answer is similar to the simple case considered in section 1, namely the best predictor $\hat{Y}_{n+1}$ of Y_{n+1} is the conditional expectation of Y_{n+1} given the whole past of the time series up to the last period $t = n$:

$$\hat{Y}_{n+1} = E[Y_{n+1} | Y_n, Y_{n-1}, Y_{n-2}, Y_{n-3}, ...]. \tag{13}$$

Also in this case the right-hand side of (13) if a function of the conditioning variables, but this is in general a function with an infinite number of arguments, and its shape may depend on n as well:

$$E[Y_{n+1} | Y_n, Y_{n-1}, Y_{n-2}, Y_{n-3}, ...] = g_n(Y_n, Y_{n-1}, Y_{n-2}, Y_{n-3}, ...) \tag{14}$$

If indeed the shape of $g_n(.)$ changes with n, there is no way to determine its shape, so that (14) is not a feasible predictor. For the shape of $g_n(.)$ to be constant, $g_n(.) = g(.)$, we need to require that the time series Y_t is **stationary**:

Definition 1: *A time series process Y_t is said to be strictly stationary if for arbitrary integers $m_1 < m_2 < .... < m_n$ the joint distribution of $Y_{t-m_1}, ..., Y_{t-m_n}$ does not depend on the time index t.*

A weaker version of stationarity is **covariance stationarity**, which requires that the expectations and covariances are well-defined and do not depend on the time index *t*.

Definition 2: *A time series process* Y_t *is said to be covariance stationary* (*or weakly stationary*) *if for all integers t and m, the expectations* $E[Y_t]$ *and the covariances* $E[(Y_t - \mu)(Y_{t-m} - \mu)]$ *are finite and do not depend on the time index t*: $E[Y_t] = \mu$ *and* $E[(Y_t - \mu)(Y_{t-m} - \mu)] = \gamma(m)$. *The function* $\gamma(m)$ *is called the covariance function.*

Note that a strictly stationary time series process Y_t is covariance stationary if $E[Y_t^2] < \infty$, because then $E[Y_t]$ and $E[(Y_t - \mu)(Y_{t-m} - \mu)]$ are finite and do not depend on *t*.

Definition 3: *A covariance stationary process* Y_t *is said to be Gaussian if for any finite sequence* α_i, $i = 0,1,....,m$, *of coefficients,* $X_t = \sum_{i=0}^{m}\alpha_i Y_{t-i}$ *is normally distributed.*

It can now be shown:

Proposition 3. *If* Y_t *strictly stationary and* $E[Y_t^2] < \infty$, *then the shape of the function* $g_n(.)$ *in* (14) *does not change with n*: *There exists a function g*(.) *such that for all t*,

$$E[Y_t | Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4}, ...] = g(Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4}, ...). \tag{15}$$

If Y_t *is covariance stationary and Gaussian then* Y_t *is strictly stationary, and the conditional expectation function g is linear*:

$$g(Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4}, ...) = \beta_0 + \sum_{j=1}^{\infty}\beta_j Y_{t-j}. \tag{16}$$

Moreover, in that case the random variables

$$U_t = Y_t - E[Y_t | Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4}, ...] \tag{17}$$

are independent normally distributed with zero expectation and constant variance, *and for each t*, U_t *is independent of* $Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4}, ...$

Thus, a stationary Gaussian process Y_t can be written as

$$
\begin{aligned}
Y_t &= \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \beta_4 Y_{t-4} + \ldots\ldots. + U_t \\
&= \beta_0 + \textstyle\sum_{j=1}^{\infty} \beta_j Y_{t-j} + U_t, \ \text{where} \ U_t \sim i.i.d. \ N(0,\sigma^2).
\end{aligned} \tag{18}
$$

Model (18) is the general linear autoregressive model. Almost all time series models for univariate stationary time series[1] are special cases of (18).

Note that similar to (3),

$$E[U_t | Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4}, \ldots\ldots.] = 0, \tag{19}$$

regardless whether Y_t is Gaussian or not.

In order for (18) to be usable for forecasting, we need to estimate the parameters β_j for j = 0,1,2,..... In general this will not be possible without imposing restrictions on the β_j's, because it is impossible to estimate an infinite number of parameters.

If we assume that $\beta_j = 0$ for all $j > p$, where p is a given natural number, we get an Auto-Regression of order p, shortly an AR(p) model:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots.. + \beta_p Y_{t-p} + U_t. \tag{20}$$

If the order p in (20) is known and not too large, this model can be estimated by OLS. In practice we do not know p, but there are various ways to determine p, as I will show below.

4. *The AR*(1) *model*

The simplest AR(p) model is the one for the case $p = 1$:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + U_t. \tag{21}$$

In the Appendix it will be show, under the assumption that the errors U_t are independent $N(0,\sigma^2)$ distributed, that a necessary condition for the stationarity of the process (21) is that

$$|\beta_1| < 1. \tag{22}$$

and that then, by repeated backwards substitution of (21), we can write Y_t as

$$Y_t = \frac{\beta_0}{1-\beta_1} + \sum_{k=0}^{\infty} \beta_1^k U_{t-k}. \tag{23}$$

---

[1] The exceptions are ARCH and GARCH models, which will be discussed in the last section.

This is called the **infinite-order moving average** [MA(∞)] representation of the AR(1) process (21).

From this expression we can derive the covariance function of the AR(1) process, as follows. Denoting $\mu = \beta_0/(1-\beta_1)$, it follows from (23) that

$$\begin{aligned}
\gamma(m) &= E[(Y_t - \mu)(Y_{t-m} - \mu)] = E\left[\left(\Sigma_{k=0}^{\infty}\beta_1^k U_{t-k}\right)\left(\Sigma_{k=0}^{\infty}\beta_1^k U_{t-m-k}\right)\right] \\
&= E\left[\left(\Sigma_{k=0}^{m-1}\beta_1^k U_{t-k} + \Sigma_{k=m}^{\infty}\beta_1^k U_{t-k}\right)\left(\Sigma_{k=0}^{\infty}\beta_1^k U_{t-m-k}\right)\right] \\
&= E\left[\left(\Sigma_{k=0}^{m-1}\beta_1^k U_{t-k} + \beta_1^m\Sigma_{k=0}^{\infty}\beta_1^k U_{t-m-k}\right)\left(\Sigma_{k=0}^{\infty}\beta_1^k U_{t-m-k}\right)\right] \\
&= E\left[\left(\Sigma_{k=0}^{m-1}\beta_1^k U_{t-k}\right)\left(\Sigma_{k=0}^{\infty}\beta_1^k U_{t-m-k}\right)\right] + \beta_1^m E\left[\left(\Sigma_{k=0}^{\infty}\beta_1^k U_{t-m-k}\right)^2\right] \\
&= \beta_1^m E\left[\left(\Sigma_{k=0}^{\infty}\beta_1^k U_{t-m-k}\right)^2\right] = \beta_1^m\sigma^2\Sigma_{k=0}^{\infty}\beta_1^{2k} = \sigma^2\beta_1^m/(1-\beta_1^2).
\end{aligned} \tag{24}$$

6. *Lag operators*

A lag operator L is the instruction to shift the time back with one period: $LY_t = Y_{t-1}$. If we apply the lag operator again we get $L^2Y_t = L(LY_t) = LY_{t-1} = Y_{t-2}$, and more generally

$$L^mY_t \overset{def.}{=} Y_{t-m},\ m = 0,1,2,..... \tag{25}$$

Using the lag operator, the AR(1) model (21) can be written as

$$(1 - \beta_1 L)Y_t = \beta_0 + U_t. \tag{26}$$

In the previous section we have in several places used the equality

$$\sum_{k=0}^{\infty} z^k = \frac{1}{1-z},\ \text{provided that}\ |z| < 1, \tag{27}$$

which follows from the equalities

$$\Sigma_{k=0}^{\infty}z^k = 1 + \Sigma_{k=1}^{\infty}z^k = 1 + \Sigma_{k=0}^{\infty}z^{k+1} = 1 + z.\Sigma_{k=0}^{\infty}z^k.$$

Now suppose that we may treat $\beta_1 L$ as the variable z in (27). If so, it follows from (27) that

$$\frac{1}{1-\beta_1 L} = \sum_{k=0}^{\infty}(\beta_1 L)^k = \sum_{k=0}^{\infty}\beta_1^k L^k. \tag{28}$$

Applying this lag polynomial to both sides of (26) then yields

$$\begin{aligned}
Y_t &= \frac{1}{1 - \beta_1 L}(1 - \beta_1 L)Y_t = \sum_{k=0}^{\infty}\beta_1^k L^k\beta_0 + \sum_{k=0}^{\infty}\beta_1^k L^k U_t = \sum_{k=0}^{\infty}\beta_1^k\beta_0 + \sum_{k=0}^{\infty}\beta_1^k U_{t-k} \\
&= \frac{\beta_0}{1-\beta_1} + \sum_{k=0}^{\infty}\beta_1^k U_{t-k},
\end{aligned} \tag{29}$$

which is exactly the moving average representation (23). Note that in the second equality in (29) I have used the fact that the lag operator has no effect on a constant: $L\beta_0 = \beta_0$, hence $L^k\beta_0 = \beta_0$. Thus, the equality (28) holds if $|\beta_1| < 1$:

Proposition 4. *The lag polynomial* $\sum_{k=0}^{\infty}\beta^k L^k$ *may be treated as* $1/(1-\beta L)$, *in the sense that* $(1-\beta L)\sum_{k=0}^{\infty}\beta^k L^k = 1$, *provided that* $|\beta| < 1$.

7. *The AR*(2) *model*

Consider the AR(2) process

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + U_t, \tag{30}$$

where the errors U_t have the same properties as before. Similar to (26) we can write this model in lag-polynomial form as

$$(1 - \beta_1 L - \beta_2 L^2)Y_t = \beta_0 + U_t. \tag{31}$$

We can always write

$$1 - \beta_1 L - \beta_2 L^2 = (1 - \alpha_1 L)(1 - \alpha_2 L), \tag{32}$$

by solving the equations $\alpha_1 + \alpha_2 = \beta_1$, $\alpha_1\alpha_2 = -\beta_2$,[2] so that (31) can be written as

$$(1 - \alpha_1 L)(1 - \alpha_2 L)Y_t = \beta_0 + U_t. \tag{33}$$

Now if $|\alpha_1| < 1$ and $|\alpha_2| < 1$ then it follows from Proposition 4 that

[2] Although the solutions involved may be complex valued.

$$Y_t = \frac{1}{(1-\alpha_1 L)(1-\alpha_2 L)}\left(\beta_0 + U_t\right) = \left(\sum_{k=0}^{\infty}\alpha_1^k L^k\right)\left(\sum_{m=0}^{\infty}\alpha_2^m L^m\right)\beta_0 + \left(\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\alpha_1^k\alpha_2^m L^{k+m}\right)U_t$$

$$= \left(\sum_{k=0}^{\infty}\alpha_1^k\right)\left(\sum_{m=0}^{\infty}\alpha_2^m\right)\beta_0 + \sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\alpha_1^k\alpha_2^m U_{t-k-m} \tag{34}$$

$$= \frac{\beta_0}{(1-\alpha_1)(1-\alpha_2)} + \sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\alpha_1^k\alpha_2^m U_{t-k-m} = \frac{\beta_0}{1-\beta_1-\beta_2} + \sum_{j=0}^{\infty}\left(\sum_{m=0}^{j}\alpha_1^{j-m}\alpha_2^m\right)U_{t-j}$$

Note that under the assumption that the errors U_t are independent $N(0,\sigma^2)$ distributed, Y_t in (34) is normally distributed, with expectation $\mu = \beta_0/(1-\beta_1-\beta_2)$ and variance $\sigma^2\sum_{j=0}^{\infty}\left(\sum_{m=0}^{j}\alpha_1^{j-m}\alpha_2^m\right)^2$.

Consequently, the necessary conditions for the covariance stationarity of the AR(2) process (30) is that the errors U_t are covariance stationary and that the solutions $1/\alpha_1$ and $1/\alpha_2$ of the equation $0 = 1 - \beta_1.z - \beta_2.z^2 = (1 - \alpha_1.z)(1 - \alpha_2.z)$ are larger than one in absolute value.

Similar conditions apply to general AR(p) processes:

Proposition 5. *The necessary conditions for the covariance stationarity of the AR*(p) process (20) *are that the errors* U_t *are covariance stationary and the solutions* $z_1,\dots,z_p$ *of the equation* $0 = 1 - \beta_1 z - \beta_2 z^2 - \dots - \beta_p z^p$ *are all greater than one in absolute value*: $|z_j| > 1$ *for* $j = 1,\dots,p$.

8. *How to determine the order p of an AR*(p) *process*

8.1 *The partial autocorrelation function*.

If the correct order of an AR process is p_0 but you estimate the AR(p) model (20) with $p > p_0$ by OLS, then the OLS estimates of the coefficients $\beta_{p_0+1},\dots,\beta_p$ will be small and insignificant, because these coefficients are then all zero: $\beta_{p_0+1} = \beta_{p_0+2} = \dots = \beta_p = 0$. This suggests the following procedure for selecting p. Estimate the AR(p) model for $p = 1,2,\dots,\bar{p}$, where $\bar{p} > p_0$.

$$\hat{Y}_t = \hat{\beta}_{p,0} + \hat{\beta}_{p,1}Y_{t-1} + \hat{\beta}_{p,2}Y_{t-2} + \dots + \hat{\beta}_{p,p}Y_{t-p}, \tag{35}$$

where the $\hat{\beta}_{p,j}$'s are OLS estimates. Then the (estimated) partial autocorrelation function, PAC(p), is defined by

$$PAC(p) \overset{def.}{=} \hat{\beta}_{p,p},\ p = 1,2,3,...,\ PAC(0) = 1. \tag{36}$$

For example, suppose that an AR(p) model $Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + ... + \beta_p Y_{t-p} + U_t$ has been fitted for p = 1, 2, 3, 4, 5 to 500 observation of a time series Y_t, with the following estimation results:

p	β_0	β_1	β_2	β_3	β_4	β_5
1	0.04070	0.02760				
	(0.06481)	(0.04470)				
2	0.06902	0.05358	−0.71257			
	(0.04609)	(0.03189)	(0.03221)			
3	0.06068	0.10470	−0.72159	0.07156		
	(0.04607)	(0.04481)	(0.03231)	(0.04525)		
4	0.06264	0.10524	−0.76661	0.07215	−0.06511	
	(0.04612)	(0.04500)	(0.04548)	(0.04577)	(0.04534)	
5	0.06283	0.10032	−0.76805	0.04648	−0.06783	−0.03274
	(0.04624)	(0.04516)	(0.04575)	(0.05715)	(0.04592)	(0.04544)

The entries that are not enclosed in brackets are the OLS estimates of the AR parameters, and the entries in brackets are the standard errors of the corresponding OLS estimates. Then

p	$PAC(p)$	$(s.e.)$
1	0.02760	(0.04470)
2	−0.71257	(0.03221)
3	0.07156	(0.04525)
4	−0.06511	(0.04534)
5	−0.03274	(0.04544)

In EasyReg the PAC function can be computed automatically, via Menu > Data analysis > Auto/Cross correlation, and the results will then be displayed as a plot. For example, the PAC(p) for the AR(2) model

$$Y_t = 1.144123 Y_{t-1} - 0.5 Y_{t-2} + U_t,\ U_t\ i.i.d.\ N(0,1),\ t = 1,...,500, \tag{37}$$

is displayed in Figure 1 below. The dots are the lower and upper bound of the one and two times the standard error bands, which correspond to the 68% and 95% confidence intervals of $\hat{\beta}_{p,p}$, respectively. The value PAC(0) = 1 is arbitrary, and is chosen because PAC(p) < 1 for $p \geq 1$.

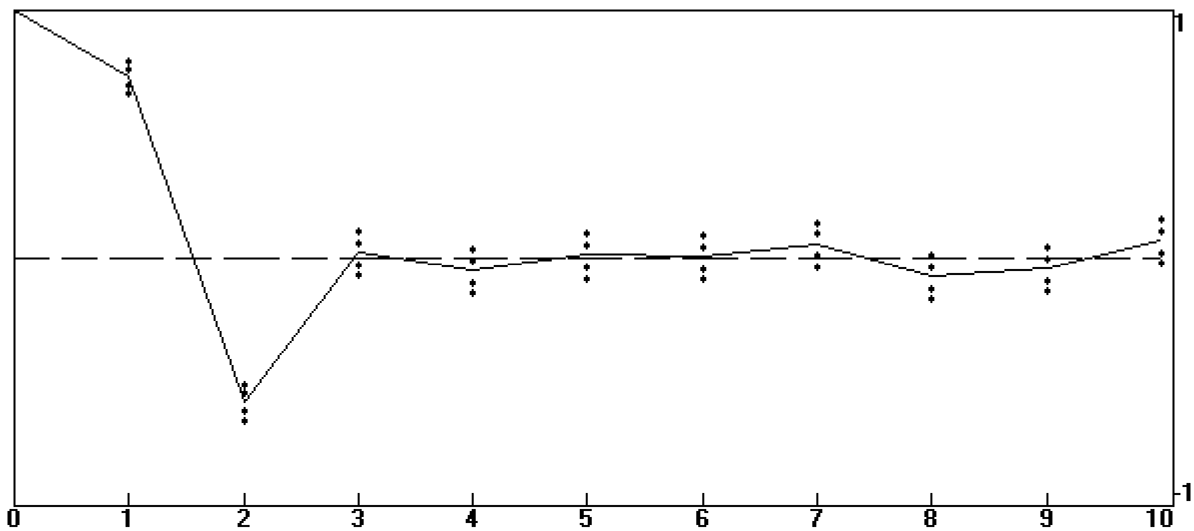

Figure 1: Partial autocorrelation function, PAC(m), of the AR(2) process (37)

In Figure 1, at $p = 3$, the zero level is contained in the smaller 68% confidence interval, and at $p = 4$ the zero level is contained in the larger 95% confidence interval. From $p = 3$ onwards the zero level is contained in either the 68% and/or 95% confidence intervals, which indicates that the true value of p is $p_0 = 2$.

8.2 *Information criteria*

An alternative approach to determine the order p of the AR(p) model (20) is to use the Akaike (1974, 1976), Hannan-Quinn (1979), or Schwarz (1978) information criteria:

Akaike: $c_n^{AR}(p) = \ln(\hat{\sigma}_p^2) + 2(1+p)/n,$

Hannan-Quinn: $c_n^{AR}(p) = \ln(\hat{\sigma}_p^2) + 2(1+p)\ln(\ln(n))/n,$

Schwarz:[3] $c_n^{AR}(p) = \ln(\hat{\sigma}_p^2) + (1+p)\ln(n)/n,$

[3] The Schwarz information criterion is also known as the Bayesian Information Criterion (BIC).

where n is the effective sample size of the regression (20) (so that Y_t is observed for $t = 1-p,...,n$), and $\hat{\sigma}_p^2$ is the OLS estimator of the error variance $\sigma^2 = E[U_t^2]$. Denoting by $\hat{p}$ the value of p for which $c_n^{AR}(p)$ is minimal:

$$c_n^{AR}(\hat{p}) = \min\{c_n^{AR}(1),....,c_n^{AR}(\bar{p})\},$$

where $\bar{p} > p_0$, with p_0 the true value of p, we have in the Hannan-Quinn and Schwarz cases: $\lim_{n\to\infty}P[\hat{p} = p_0] = 1$, and in the Akaike case $\lim_{n\to\infty}P[\hat{p} \geq p_0] = 1$ but $\lim_{n\to\infty}P[\hat{p} = p_0] < 1$. These results are explained in the Appendix. Thus, the Akaike criterion may "overshoot" the true value. Due to the latter, it is recommended to use the Hannan-Quinn or Schwarz criterion instead of the Akaike criterion. Note however that in small samples the Hannan-Quinn and Schwarz criteria may give different results for $\hat{p}$.

For example, for the same data on which Figure 1 was based, namely the AR(2) model $Y_t = 1.144123Y_{t-1} - 0.5Y_{t-2} + U_t$, t =1,...,500, with independent N(0,1) distributed errors U_t, and upper bound $\bar{p} = 4$, we get

p	*Akaike*	*Hannan-Quinn*	*Schwarz*
1	$5.14474E-01$	$5.21089E-01$	$5.31332E-01$
2	$1.08788E-01$	$1.18711E-01$	$1.34076E-01$
3	$1.12462E-01$	$1.25692E-01$	$1.46179E-01$
4	$1.13783E-01$	$1.30321E-01$	$1.55929E-01$

All three criteria are minimal for $p = 2$, hence $\hat{p} = 2$, which is equal to the true value $p_0 = 2$.

8.3 *The Wald test*

A third way to determine the correct order p_0 of the AR(p) model (20) is the following. Determine an upper bound $\bar{p} > p_0$ on the basis of the PAC function and the information criteria, estimate the model (20) for $p = \bar{p}$ and test whether p can be reduced, using the Wald test, via Options > Wald test of linear parameter restrictions > Test joint significance, in the "What to do next?" module of EasyReg. For example, for the same data as in the previous section, and $\bar{p} = 4$, we get the OLS results

Parameters	*OLS estimate*	*t-value*
β_0	0.06910	1.449
β_1	1.17283	26.033
β_2	−0.63395	−9.134
β_3	0.07841	1.130
β_4	−0.05090	−1.130

The t-value of β_4 is well within the range -1.96, +1.96, hence the null hypothesis that $\beta_4 = 0$ cannot be rejected at the 5% significance level. To test whether $\beta_3 = 0$ as well, you need to test the joint null hypothesis $\beta_3 = \beta_4 = 0$, using the Wald test. In this case the test result involved is:

```
Wald test:                                    1.45
Asymptotic null distribution:  Chi-square(2)
  p-value = 0.48398
  Significance levels:        10%          5%
  Critical values:           4.61        5.99
  Conclusions:             accept      accept
```

Thus, the null hypothesis $\beta_3 = \beta_4 = 0$ cannot be rejected, hence we may reduce p to 2.

Since β_2 is strongly significant, there is no need to test the null hypothesis $\beta_2 = \beta_3 = \beta_4 = 0$, but if we do the null hypothesis will be rejected:

```
Wald test:                                  253.39
Asymptotic null distribution:  Chi-square(3)
  p-value = 0.00000
  Significance levels:        10%          5%
  Critical values:           6.25        7.81
  Conclusions:             reject      reject
```

Thus, the test results involved lead to the same conclusion as the one on the basis of the PAC function and the information criteria, namely that $p_0 = 2$.

9. *Moving average processes*

A moving average process of order q, denoted by MA(q), takes the form

$$Y_t = \mu + U_t - \theta_1 U_{t-1} - .... - \theta_q U_{t-q}, \tag{38}$$

where $\mu = E[Y_t]$. Under regularity conditions an MA process has an infinite order AR representation, as I will demonstrate for the case $q = 1$.

Consider the MA(1) process

$$Y_t = \mu + U_t - \theta U_{t-1}. \tag{39}$$

Using the lag operator, we can write this MA(1) model as

$$Y_t = \mu + (1 - \theta L)U_t. \tag{40}$$

Now it follows from Proposition 4 and (40) that if $|\theta| < 1$ then

$$\sum_{j=0}^{\infty}\theta^j L^j Y_t = (1 - \theta L)^{-1}Y_t = (1 - \theta L)^{-1}\mu + U_t = \frac{\mu}{1-\theta} + U_t, \tag{41}$$

hence

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + ... + U_t = \beta_0 + \sum_{j=1}^{\infty}\beta_j Y_{t-j} + U_t,$$
$$\textit{where } \beta_0 = \mu/(1-\theta),\ \beta_i = -\theta^j \textit{ for } j = 1,2,3,..... \tag{42}$$

More generally we have:

Proposition 6. *If the solutions* $z_1, ..., z_q$ *of the equation* $0 = 1 - \theta_1 z - \theta_2 z^2 - ..... - \theta_q z^q$ *are all greater than one in absolute value*: $|z_j| > 1$ *for* $j = 1,...,q$, *then the MA*(q) *process* (38) *can be written as an infinite order AR process*: $Y_t = \beta_0 + \sum_{j=1}^{\infty}\beta_j Y_{t-j} + U_t$, *where* $\beta_0 = \mu/(1-\theta_1-\theta_2-....-\theta_q)$ *and* $1 - \sum_{j=1}^{\infty}\beta_j L^j = 1/(1-\theta_1 L-\theta_2 L^2-.....-\theta_q L^q)$.

This result is the main reasons for working with MA models, because the best one-step ahead linear forecast of Y_t takes the form $\beta_0 + \sum_{j=1}^{\infty}\beta_j Y_{t-j}$, which can be approximated quite well by an MA(q) model.

10. *How to determine the order q of a MA(q) process*

10.1 *The (regular) autocorrelation function*

The autocorrelation function of a time series process Y_t is defined by

$$\rho(m) = \frac{cov(Y_t, Y_{t-m})}{var(Y_t)}, \; m = 0,1,2,...... \tag{43}$$

Proposition 7. *For an MA(q)* process, $\rho(m) = 0$ *for* $m > q$, *and* $\rho(q) \neq 0$.

To see this, consider first the MA(1) process (39). For this process,

$$\begin{aligned} cov(Y_t, Y_{t-1}) &\overset{def.}{=} E[(Y_t - \mu)(Y_{t-1} - \mu)] = E[(U_t - \theta U_{t-1})(U_{t-1} - \theta U_{t-2})] \\ &= E[U_t U_{t-1}] - \theta.E[U_{t-1}^2] - \theta.E[U_t U_{t-2}] + \theta^2 E[U_{t-1} U_{t-2}] = -\theta.E[U_{t-1}^2], \end{aligned} \tag{44}$$

where the last equality follows from the conditions that the errors U_t are uncorrelated and have zero expectation: $E[U_t U_{t-k}] = E[U_t].E[U_{t-k}]$ for $k \neq 0$. Similarly we have for $m > 1$,

$$\begin{aligned} cov(Y_t, Y_{t-m}) &\overset{def.}{=} E[(Y_t - \mu)(Y_{t-m} - \mu)] = E[(U_t - \theta U_{t-1})(U_{t-m} - \theta U_{t-m-1})] \\ &= E[U_t U_{t-m}] - \theta.E[U_{t-1} U_{t-m}] - \theta.E[U_t U_{t-m-1}] + \theta^2 E[U_{t-1} U_{t-m-1}] = 0 \end{aligned} \tag{45}$$

and

$$\begin{aligned} var(Y_t) &\overset{def.}{=} E[(Y_t - \mu)^2] = E[(U_t - \theta U_{t-1})^2] \\ &= E[U_t^2] - 2\theta.E[U_t U_{t-1}] + \theta^2 E[U_{t-1}^2] = E[U_t^2] + \theta^2 E[U_{t-1}^2] = (1+\theta^2)E[U_t^2], \end{aligned} \tag{46}$$

where the last equality follows from the stationarity of U_t. Thus in the MA(1) case,

$$\rho(1) = -\frac{\theta}{1+\theta^2} \neq 0, \; \rho(m) = 0 \; for \; m = 2,3,..... \tag{47}$$

Along similar lines it can be shown that Proposition 7 holds.

The actual autocorrelation function cannot be calculated, but it can be estimated in various ways. EasyReg estimates $\rho(m)$ by

$$\hat{\rho}(m) = \frac{(1/(n-m))\sum_{t=m+1}^{n}(Y_t-\bar{Y})(Y_{t-m}-\bar{Y})}{\sqrt{(1/(n-m))\sum_{t=1}^{n-m}(Y_t-\bar{Y})^2}\sqrt{(1/(n-m))\sum_{t=m+1}^{n}(Y_{t-m}-\bar{Y})^2}}, \quad m = 0,1,2,...... \tag{48}$$

where $\bar{Y} = (1/n)\sum_{t=1}^{n}Y_t$.

For an AR(p) process the autocorrelation function does not provide information about p. To see this, consider again the AR(1) process (21) satisfying condition (22). Then it follows from (24) that $cov(Y_t, Y_{t-m}) = \gamma(m) = \sigma^2\beta_1^m/(1-\beta_1^2)$ and $var(Y_t) = \gamma(0) = \sigma^2/(1-\beta_1^2)$, hence in the AR(1) case, $\rho(m) = \beta_1^m$. Therefore, in this case the autocorrelation function will not drop sharply to zero for $m > 1$. The same applies to more general AR processes.

For example, for the same data on which Figure 1 was based we have:

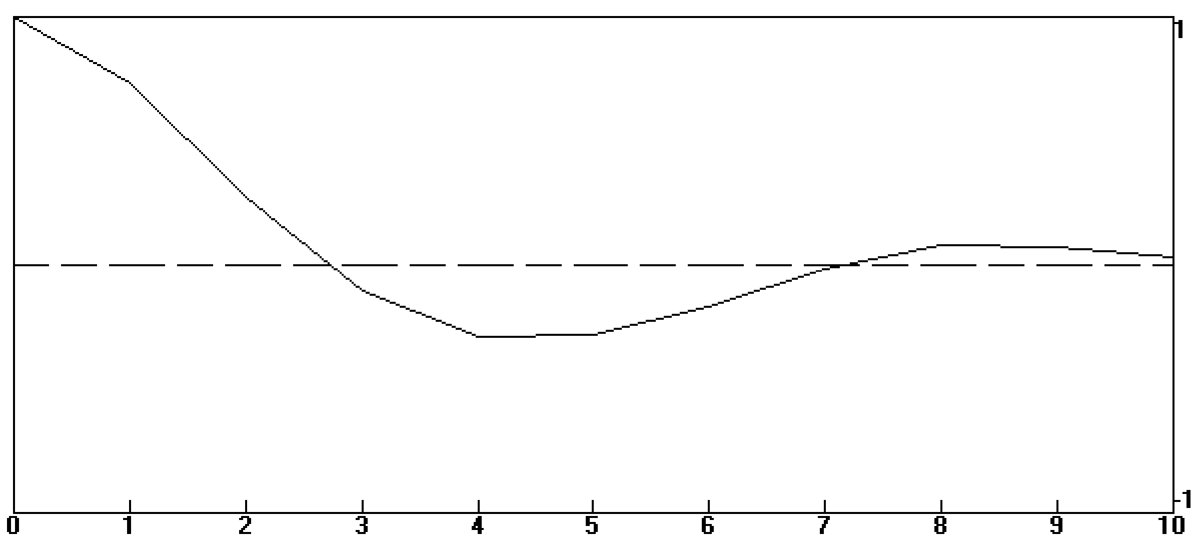

Figure 2: Estimated autocorrelation function $\hat{\rho}(m)$ of the AR(2) process (37).

To demonstrate how to use the estimated autocorrelation function to determine the order q of an MA(q) process, I have generated 500 observations according to the model

$$Y_t = U_t - 1.4U_{t-1} + 0.5U_{t-2}, \; U_t \sim i.i.d.\ N(0,1), \; t = 1,2,....,500 \tag{49}$$

The estimated autocorrelation function $\hat{\rho}(m)$ involved is displayed in Figure 3 below, for m = 0,1,...,10.

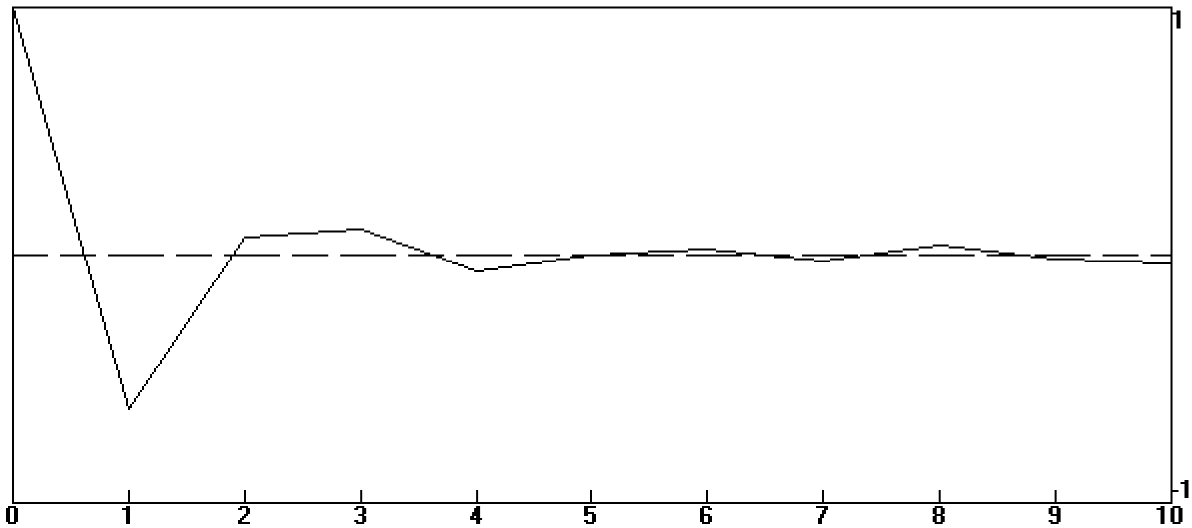

Figure 3: Estimated autocorrelation function $\hat{\rho}(m)$ of the MA(2) process (49)

Because $\hat{\rho}(m)$ is not endowed with standard error bands, it is not obvious at which value of m the true autocorrelation function $\rho(m)$ becomes zero. But at least we can determine an upper bound $\bar{q}$ of q from Figure 3: It seems that $\hat{\rho}(m)$ is approximately zero for $m \geq 5$, indicating that $q \leq 4$. Thus, let $\bar{q} = 4$.

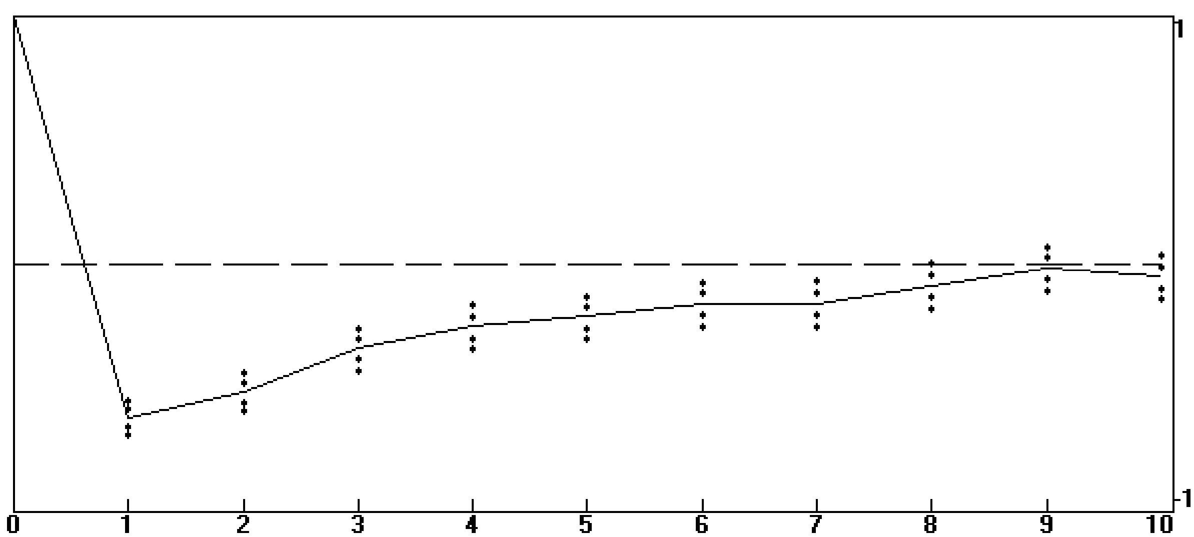

Figure 4: Partial autocorrelation function, PAC(m), of the MA(2) process (49)

The partial autocorrelation function of an MA(q) process is of no use for determining q or an upper bound of q, because of the AR(∞) representation of an MA(q) process. See Proposition 6. For example, the PAC(m) of the MA(2) process (49) does not drop sharply to zero for $m > 2$, as is demonstrated in Figure 4.

10.2 *Information criteria*

The three information criteria, Akaike, Hannan-Quinn and Schwarz also apply to MA processes. Therefore, estimate the MA(q) model (38) for $q = 1,2,3,4$ ($=\bar{q}$), and compare the information criteria:

q	*Akaike*	*Hannan−Quinn*	*Schwarz*
1	1.91941E−01	1.98556E−01	2.08799E−01
2	1.24771E−02	2.23999E−02	3.77647E−02
3	1.63628E−02	2.95933E−02	5.00797E−02
4	1.68145E−02	3.33526E−02	5.89606E−02

All three criteria are minimal for $q = 2$, which is the true value.

10.3 *Wald test*

As a double check, estimate the MA model (38) for $q = 4$ (in EasyReg via Menu > Single equation models > ARIMA estimation and forecasting), and test whether $\theta_3 = \theta_4 = 0$, using the Wald test:

Parameters	*Estimate*	*t−value*
μ	0.000234	0.050
θ_1	1.348470	29.979
θ_2	−0.427742	−5.637
θ_3	−0.074225	−0.978
θ_4	0.050943	1.127

```
Wald test:                                   1.29
Asymptotic null distribution:  Chi-square(2)
  p-value = 0.52588
  Significance levels:        10%          5%
  Critical values:           4.61        5.99
  Conclusions:             accept      accept
```

Thus, the null hypothesis $\theta_3 = \theta_4 = 0$ cannot be rejected, hence we may reduce q from 4 to q =2. Since θ_2 is strongly significant, we cannot reduce q further.

Re-estimating model (38) for q = 2 yields:

Parameters	*Estimate*	*t−value*
μ	0.000187	0.039
θ_1	1.348684	33.682
θ_2	−0.456211	−11.349

which is reasonably close to the true values of the parameters: $\mu = 0$, $\theta_1 = 1.4$, $\theta_2 = -0.5$.

11. *ARMA models*

11.1 *Introduction*

As I have shown before, both AR(p) models and MA(q) models are parsimonious[4] approximations of more general AR(∞) processes, $Y_t = \beta_0 + \sum_{j=1}^{\infty}\beta_j Y_{t-j} + U_t$, and the best one-step ahead linear forecast of Y_t given all past values of Y_t takes the form $\hat{Y}_t = \beta_0 + \sum_{j=1}^{\infty}\beta_j Y_{t-j}$. This suggests that even closer approximations can be achieved by combining the two types of models:

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + V_t, \\ V_t &= U_t - \theta_1 U_{t-1} - \dots - \theta_q U_{t-q}. \end{aligned} \quad (50)$$

This is called an ARMA(p,q) model. Denoting

$$\begin{aligned} \varphi_p(L) &= 1 - \beta_1 L - \beta_2 L^2 - \dots - \beta_p L^p, \\ \psi_q(L) &= 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q, \end{aligned} \quad (51)$$

---

4 In the sense that only a limited number of parameters are used.

an ARMA(p,q) model can be written more compactly as

$$\varphi_p(L)Y_t = \beta_0 + \psi_q(L)U_t. \tag{52}$$

Proposition 8. *If* $\varphi_p(z_1) = 0$ *implies* $|z_1| > 1$ *and* $\psi_q(z_2) = 0$ *implies* $|z_2| > 1$ *then the ARMA*(p,q) *process* (52) *is stationary, with AR*(∞) *representation* $\psi_q(L)^{-1}\varphi_p(L)Y_t = \beta_0/\psi_q(1) + U_t$ *and MA*(∞) *representation* $Y_t = \beta_0/\varphi_p(1) + \varphi_p(L)^{-1}\psi_q(L)U_t$.

The last result implies that $E[Y_t] = \beta_0/\varphi_p(1) + \varphi_p(L)^{-1}\psi_q(L)E[U_t] = \beta_0/\varphi_p(1)$.

I will demonstrate Proposition 8 for the case $p = q = 1$:

$$(1-\beta_1 L)Y_t = \beta_0 + (1-\theta_1 L)U_t. \tag{53}$$

The condition that $\varphi_1(z) = 0$ implies $|z| > 1$ is equivalent to $|\beta_1| < 1$, because $\varphi_1(z) = 1-\beta_1 z =$ 0 implies that $z = 1/\beta_1$. Similarly, the condition that $\psi_1(z) = 1-\theta_1 z = 0$ implies $|z| > 1$ is equivalent to $|\theta_1| < 1$. It follows now from Proposition 4 that $\psi_1(L)^{-1} = (1-\theta_1 L)^{-1} = \sum_{j=0}^{\infty}\theta_1^j L^j$, hence

$$\begin{aligned}\sum_{j=0}^{\infty}\theta_1^j L^j(1-\beta_1 L)Y_t &= \psi_1(L)^{-1}(1-\beta_1 L)Y_t = \sum_{j=0}^{\infty}\theta_1^j L^j\beta_0 + \sum_{j=0}^{\infty}\theta_1^j L^j(1-\theta_1 L)U_t \\ &= \sum_{j=0}^{\infty}\theta_1^j\beta_0 + U_t = \beta_0/(1-\theta_1) + U_t\end{aligned} \tag{54}$$

and

$$\begin{aligned}\sum_{j=0}^{\infty}\theta_1^j L^j(1-\beta_1 L)Y_t &= \sum_{j=0}^{\infty}\theta_1^j L^j Y_t - \beta_1\sum_{j=0}^{\infty}\theta_1^j L^{j+1}Y_t \\ = Y_t + \sum_{j=1}^{\infty}\theta_1^j Y_{t-j} - \beta_1\sum_{j=0}^{\infty}\theta_1^j Y_{t-j-1} &= Y_t + \theta_1\sum_{j=0}^{\infty}\theta_1^j Y_{t-1-j} - \beta_1\sum_{j=0}^{\infty}\theta_1^j Y_{t-1-j} \\ &= Y_t - (\beta_1-\theta_1)\sum_{j=0}^{\infty}\theta_1^j Y_{t-1-j}.\end{aligned} \tag{55}$$

Combining these results yields

$$Y_t = \beta_0/(1-\theta_1) + (\beta_1-\theta_1)\sum_{j=0}^{\infty}\theta_1^j Y_{t-1-j} + U_t, \tag{56}$$

which is the AR(∞) representation of the ARMA(1,1) process under review.

Similarly, it follows from Proposition 4 that $\varphi_1(L)^{-1} = (1-\beta_1 L)^{-1} = \sum_{j=0}^{\infty}\beta_1^j L^j$, hence

$$\begin{aligned} Y_t &= (1-\beta_1 L)^{-1}(1-\beta_1 L)Y_t = (1-\beta_1 L)^{-1}\beta_0 + (1-\beta_1 L)^{-1}(1-\theta_1 L)U_t \\ &= \beta_0/(1-\beta_1) + \Sigma_{j=0}^{\infty}\beta_1^j L^j(1-\theta_1 L)U_t \\ &= \beta_0/(1-\beta_1) + \Sigma_{j=0}^{\infty}\beta_1^j L^j U_t - \theta_1\Sigma_{j=0}^{\infty}\beta_1^j L^{j+1}U_t \\ &= \beta_0/(1-\beta_1) + U_t - (\theta_1-\beta_1)\Sigma_{j=0}^{\infty}\beta_1^j U_{t-1-j}, \end{aligned} \tag{57}$$

which is the MA(∞) representation of the ARMA(1,1) process under review.

11.2 *Common roots*

Observe from (56) and (57) that if $\beta_1 = \theta_1$ then $Y_t = \beta_0/(1-\beta_1) + U_t$, which is an ARMA(0,0) process (also called a *white noise* process). This is the **common roots** problem:

Proposition 9. *Let the conditions in Proposition 8 be satisfied. If there exists a* $\delta \neq 0$ *such that* $\varphi_p(1/\delta) = \psi_q(1/\delta) = 0$ *then we can write the lag polynomials in ARMA(p,q) model* (52) *as* $\varphi_p(L) = (1-\delta L)\varphi^*_{p-1}(L)$ *and* $\psi_q(L) = (1-\delta L)\psi^*_{q-1}(L)$, *where* $\varphi^*_{p-1}(L)$ *and* $\psi^*_{q-1}(L)$ *are lag polynomials or order* $p-1$ *and* $q-1$, *respectively, satisfying the conditions in Proposition 8. The ARMA(p,q) process* (52) *is then equivalent to the ARMA*$(p-1,q-1)$ *process* $\varphi^*_{p-1}(L)Y_t = \beta^*_0 + \psi^*_{q-1}(L)U_t$, *where* $\beta^*_0 = \varphi^*_{p-1}(1)E[Y_t]$.

Because the value of δ does not matter, the parameters in the lag polynomials $\varphi_p(L)$ and $\psi_q(L)$ are no longer identified. The same applies to the constant β_0 in model (52) because $\beta_0 = (1-\delta)\beta^*_0$ for arbitrary δ. For example, let for $p = q = 2$,

$$\begin{aligned} \varphi_2(L) &= (1-\delta L)(1-\beta L) = 1-2(\delta+\beta)L + \delta.\beta L^2 = 1-\beta_1 L-\beta_2 L^2 \\ \psi_2(L) &= (1-\delta L)(1-\theta L) = 1-2(\delta+\theta)L + \delta.\theta L^2 = 1-\theta_1 L-\theta_2 L^2 \end{aligned} \tag{58}$$

where $|\beta| < 1$, $|\theta| < 1$, and $|\delta| < 1$, and let $E[Y_t] = 0$. Then the ARMA(2,2) model $\varphi_2(L)Y_t = \psi_2(L)U_t$ is equivalent to the ARMA(1,1) model $(1-\beta L)Y_t = (1-\theta L)U_t$ for all values of δ. Hence, given β and θ , $\beta_1 = 2(\delta+\beta)$, $\beta_2 = -\delta.\beta$, $\theta_1 = 2(\delta+\theta)$, $\theta_2 = -\delta.\theta$ for arbitrary δ. As a consequence, the estimates of the parameters β_1, β_2, θ_1, θ_2 are no longer consistent, and the t-test and Wald test for testing the (joint) significance of the parameters are no longer valid. In particular, in the ARMA(2,2) case under review the Wald test of the null hypothesis $\beta_2 = \theta_2 = 0$

is no longer valid. Therefore, we should not use the Wald test to test whether the AR and MA orders p and q can be reduced to $p-1$ and $q-1$.

The problem of common roots in ARMA models is similar to the multicollinearity problem in linear regression. As in the latter case, the t values of the parameters will be deflated towards zero. Therefore, if all the t values of the ARMA parameters are insignificant this may indicate that the AR and MA lag polynomials have a common root.

Although we should not use the Wald test to test for common roots, we can still use the information criteria to determine whether the AR and MA orders p and q can be reduced to $p-1$ and $q-1$. In the case of a common root, the variance σ^2 of the errors U_t of the ARMA(p,q) model in Proposition 9 is the same as the variance of the errors U_t in the ARMA($p-1,q-1$) model $\varphi^*_{p-1}(L)Y_t = \beta^*_0 + \psi^*_{q-1}(L)U_t$. Therefore, the estimate $\hat{\sigma}^2_{p,q}$ of the errors U_t of the ARMA(p,q) model involved will be close to the estimate $\hat{\sigma}^2_{p-1,q-1}$ of the errors of the equivalent ARMA($p-1,q-1$) model, and asymptotically they will be equal:

$$\text{plim}_{n\to\infty}\hat{\sigma}^2_{p,q} = \text{plim}_{n\to\infty}\hat{\sigma}^2_{p-1,q-1} = \sigma^2. \qquad (59)$$

In the ARMA case the three information criteria take the form

Akaike:	$c_n^{ARMA}(p,q) = \ln(\hat{\sigma}^2_{p,q}) + 2(1+p+q)/n,$
Hannan-Quinn:	$c_n^{ARMA}(p,q) = \ln(\hat{\sigma}^2_{p,q}) + 2(1+p+q)\ln(\ln(n))/n,$
Schwarz:	$c_n^{ARMA}(p,q) = \ln(\hat{\sigma}^2_{p,q}) + (1+p+q)\ln(n)/n,$

Therefore, in the case of a common root, $c_n^{ARMA}(p-1,q-1) < c_n^{ARMA}(p,q)$ if n is large enough, due to (59).

To demonstrate the common roots phenomenon, I have generated a time series Y_t, t = 1,...,500, according to the ARMA(1,1) model

$$Y_t = 0.3 + 0.7Y_{t-1} + U_t + 0.5U_{t-1}, \quad U_t \sim i.i.d\ N(0,1), \qquad (60)$$

and estimated this model as an ARMA(2,2) model

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + U_t - \theta_1 U_{t-1} - \theta_2 U_{t-2}. \qquad (61)$$

The EasyReg estimation results involved are:

Parameters	*Estimate*	*t-value*
$\mu = \beta_0/(1-\beta_1-\beta_2)$	0.776272	3.273
β_1	1.087430	0.170
β_2	−0.267512	−0.058
θ_1	−0.166275	−0.026
θ_2	0.189439	0.055
σ	1.008897	

```
Information criteria:
Akaike:              2.76651E-02
Hannan-Quinn:        4.42031E-02
Schwarz:             6.98112E-02
```

Apart from the estimate of $\mu = E[Y_t]$, the AR and MA parameters are insignificant, due to a common root in the AR and MA lag polynomials.

Next, I have estimated the model as an ARMA(1,1) model:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + U_t - \theta_1 U_{t-1}. \tag{62}$$

The EasyReg estimation results are:

Parameters	*Estimate*	*t-value*
$\mu = \beta_0/(1-\beta_1)$	0.775967	3.249
β_1	0.720705	20.784
θ_1	−0.530183	−12.597
σ	1.006879	

```
Information criteria:
Akaike:              1.96937E-02
Hannan-Quinn:        2.96166E-02
Schwarz:             4.49814E-02
```

Indeed, the information criteria for the latter model are substantial lower (and thus better) than for the previous ARMA(2,2) model. Moreover, observe that in the latter case the estimates of β_1, θ_1 and σ are close to the true values $\beta_1 = 0.7$, $\theta_1 = -0.5$ and $\sigma = 1$, respectively, although at first sight the estimate $\hat{\mu} = 0.775967$ of $\mu = E[Y_t]$ seems quite different from the true value $\mu = 0.3/(1-0.7) = 1$. However, it can be shown that $\hat{\mu}$ is not significantly different from 1.

11.3 *How to distinguish an ARMA process from an AR process*

The AR(∞) representation (56) of the ARMA(1,1) process (60) is

$$\begin{aligned} Y_t &= \beta_0/(1-\theta_1) + (\beta_1-\theta_1)\Sigma_{j=0}^{\infty}\theta_1^j Y_{t-1-j} + U_t \\ &= 0.3/(1+0.5) + 1.2\Sigma_{j=0}^{\infty}(-0.5)^j Y_{t-1-j} + U_t \\ &= 0.2 + 1.2\Sigma_{j=0}^{\infty}(-0.5)^j Y_{t-1-j} + U_t \\ &= 0.2 + 1.2Y_{t-1} - 0.6Y_{t-2} + 0.3Y_{t-3} - 0.15Y_{t-4} + 0.075Y_{t-5} + \ldots\ldots + U_t \end{aligned} \tag{63}$$

which is close to an AR(4) process. Therefore, the partial autocorrelation function, PAC(m), of this process will look like the PAC(m) of an AR process:

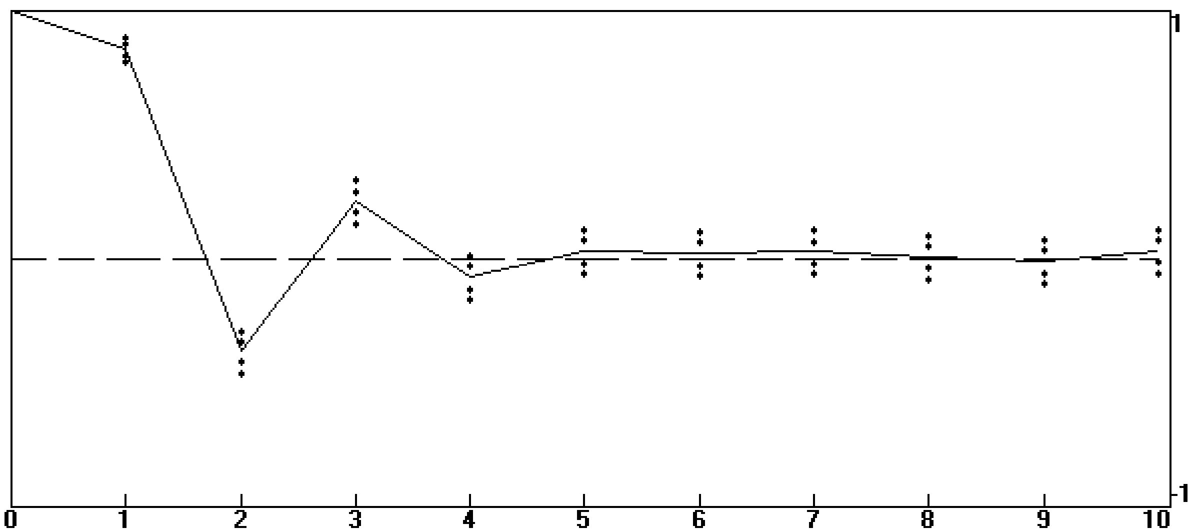

Figure 5: Partial autocorrelation function, PAC(m), of the ARMA(1,1) process (60)

Indeed, on the basis of this plot one may be tempted to conclude (erroneously) that the process is an AR(4) process, and the estimated autocorrelation function would actually corroborates this:

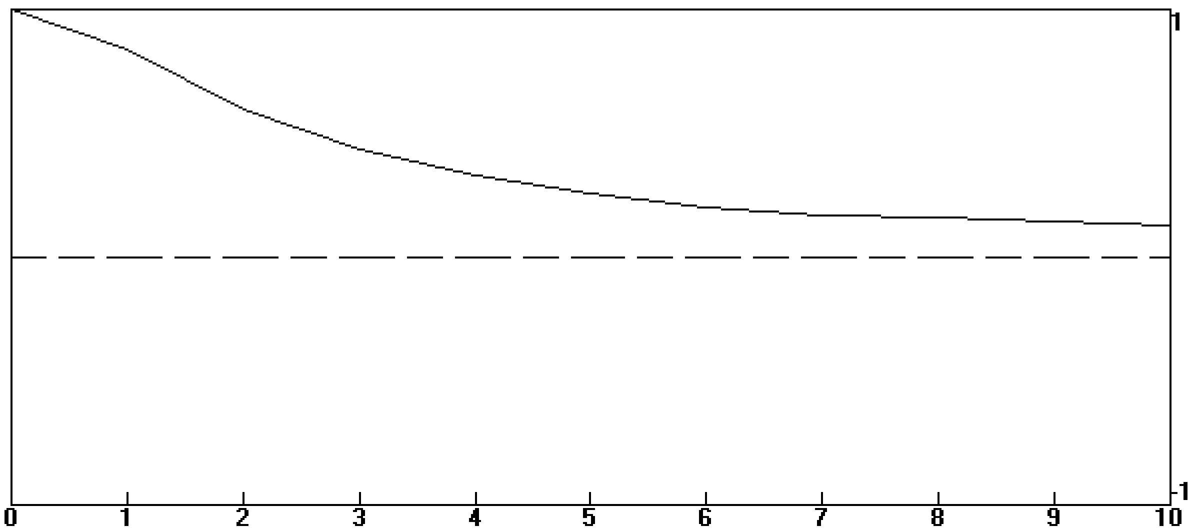

Figure 6: Estimated autocorrelation function $\hat{\rho}(m)$ of the ARMA(1,1) process (60)

Therefore, the partial and regular autocorrelation functions are of no help in distinguishing an ARMA model from an AR model.

So how to proceed? In this particular case I recommend the following: In first instance assume, on the basis of Figure 5, that the model is an AR(4) model, and estimate it by OLS or the ARIMA module. Then try all ARMA(p,q) models with $p + q \le 4$, and pick the model with the lowest value of one of the information criteria. For example, the Hannan-Quinn information criteria for the model (60) are:

p	q=0	q=1	q=2	q=3	q=4
4	$3.99778E-02$				
3	$3.79859E-02$	$4.15536E-02$			
2	$8.56998E-02$	$3.61047E-02$	$4.42031E-02$		
1	$22.9013E-02$	$2.96166E-02$	$3.57906E-02$	$3.92521E-02$	
0	$145.019E-02$	$54.6481E-02$	$19.5530E-02$	$10.7008E-02$	$7.73787E-02$

The smallest value of the Hannan-Quinn information criterion is 2.96166E-02 for $p = q = 1$, hence the conclusion is that the process involved is an ARMA(1,1) process.

This model selection procedure can be conducted automatically in EasyReg, via Menu > Single equation models > ARIMA model selection via information criteria. The only thing you have to do is to specify an initial, possibly over-parametrized, ARMA model. EasyReg will then compute the Akaike, Hannan-Quinn and Schwarz information criteria for this model and all sub-models, and indicate which model is optimal.

11.4 *Forecasting with an ARMA model*

In EasyReg the AR(∞) representation of an ARMA model is used as forecasting scheme, because it represents the conditional expectation function. For example, in the ARMA(1,1) case the forecasting scheme for Y_{n+1} given its past up to time n is

$$\hat{Y}_{n+1} = \beta_0/(1-\theta_1) + (\beta_1-\theta_1)\Sigma_{j=0}^{\infty}\theta_1^j Y_{n-j}, \tag{64}$$

where n is the last observed time period. Compare (56). In practice we have to replace the parameters involved by estimates. Moreover, usually we do not observe all values of Y_{n-j}, but only for $n-j \geq 1$, say. Therefore, replace Y_t for $t<1$ in (64) by its sample mean $\bar{Y} = (1/n)\Sigma_{t=1}^{n}Y_t$. Thus, the actual forecast of Y_{n+1} is:

$$\begin{aligned}\tilde{Y}_{n+1|n} &= \hat{\beta}_0/(1-\hat{\theta}_1) + (\hat{\beta}_1-\hat{\theta}_1)\Sigma_{j=0}^{n-1}\hat{\theta}_1^j Y_{n-j} + (\hat{\beta}_1-\hat{\theta}_1)\Sigma_{j=n}^{\infty}\hat{\theta}_1^j\bar{Y} \\ &= \frac{\hat{\beta}_0}{1-\hat{\theta}_1} + (\hat{\beta}_1-\hat{\theta}_1)\Sigma_{j=0}^{n-1}\hat{\theta}_1^j Y_{n-j} + \frac{(\hat{\beta}_1-\hat{\theta}_1)\hat{\theta}_1^n\bar{Y}}{1-\hat{\theta}_1},\end{aligned} \tag{65}$$

where $\hat{\beta}_1$ and $\hat{\theta}_1$ are the estimates of β_1 and θ_1, respectively, based on the data up to time n.

To forecast Y_{n+2} given its past up to time n, replace n in (65) by n+1, and the unobserved Y_{n+1} by its forecast:

$$\tilde{Y}_{n+2|n} = \hat{\beta}_0/(1-\hat{\theta}_1) + (\hat{\beta}_1-\hat{\theta}_1)\tilde{Y}_{n+1|n} + (\hat{\beta}_1-\hat{\theta}_1)\Sigma_{j=1}^{n}\hat{\theta}_1^j Y_{n+1-j} + \frac{(\hat{\beta}_1-\hat{\theta}_1)\hat{\theta}_1^{n+1}\bar{Y}}{1-\hat{\theta}_1}. \tag{66}$$

This procedure is called *recursive forecasting*. More generally, the h step ahead recursive forecast of Y_{n+h} given its past up to time n is

$$\tilde{Y}_{n+h|n} = \frac{\hat{\beta}_0}{1-\hat{\theta}_1} + (\hat{\beta}_1-\hat{\theta}_1)\Sigma_{j=0}^{h-2}\hat{\theta}_1^j\tilde{Y}_{n+h-1-j|n} + (\hat{\beta}_1-\hat{\theta}_1)\Sigma_{j=h+1}^{n+h-2}\hat{\theta}_1^j Y_{n+h-1-j} + \frac{(\hat{\beta}_1-\hat{\theta}_1)\hat{\theta}_1^{n+h-1}\bar{Y}}{1-\hat{\theta}_1}. \tag{67}$$

Note however, that in this case

$$\lim_{h\to\infty}\tilde{Y}_{n+h|n} = \hat{\beta}_0/(1-\hat{\theta}_1) + (\hat{\beta}_1-\hat{\theta}_1)\lim_{h\to\infty}\Sigma_{j=0}^{h-2}\hat{\theta}_1^j\tilde{Y}_{n+h-1-j|n}$$
$$= \hat{\beta}_0/(1-\hat{\theta}_1) + (\hat{\beta}_1-\hat{\theta}_1)\Sigma_{j=0}^{\infty}\hat{\theta}_1^j\lim_{h\to\infty}\tilde{Y}_{n+h|n} = \hat{\beta}_0/(1-\hat{\theta}_1) + \left((\hat{\beta}_1-\hat{\theta}_1)/(1-\hat{\theta}_1)\right)\lim_{h\to\infty}\tilde{Y}_{n+h|n}. \quad (68)$$

Solving this equality for $\lim_{h\to\infty}\tilde{Y}_{n+h|n}$ yields

$$\lim_{h\to\infty}\tilde{Y}_{n+h|n} = \hat{\beta}_0/(1-\hat{\beta}_1), \quad (69)$$

which is just the estimate of $\mu = E[Y_t]$. Compare (57). Thus, if we choose the forecast horizon h too large, the recursive forecast $\tilde{Y}_{n+h|n}$ will be close to the expectation $\mu = E[Y_t]$.

12. *ARMA models for seasonal time series*

12.1 *Seasonal dummy variables*

The effect of seasonality may manifest itself through seasonal varying expectations as well as seasonal patterns in the AR and/or MA lag polynomials. As to the former, time varying expectations can easily be modeled using seasonal dummy variables. For example, if Y_t is a quarterly time series, $E[Y_t]$ can be modeled either by

$$E[Y_t] = \mu_0 + \mu_1 Q_{1,t} + \mu_2 Q_{2,t} + \mu_3 Q_{3,t} \quad (70)$$

or

$$E[Y_t] = \mu_1^* Q_{1,t} + \mu_2^* Q_{2,t} + \mu_3^* Q_{3,t} + \mu_4^* Q_{4,t}, \quad (71)$$

where the $Q_{s,t}$'s are seasonal dummy variables:

$$Q_{s,t} = 1 \textit{ if the quarter of } t \textit{ is } s, \; Q_{s,t} = 0 \textit{ if not}. \quad (72)$$

The equivalence of (70) and (71) follows from the fact that $\Sigma_{s=1}^4 Q_{s,t} = 1$, so that

$$E[Y_t] = \mu_1^* Q_{1,t} + \mu_2^* Q_{2,t} + \mu_3^* Q_{3,t} + \mu_4^*\left(1 - Q_{1,t} - Q_{2,t} - Q_{3,t}\right)$$
$$= \mu_4^* + (\mu_1^*-\mu_4^*)Q_{1,t} + (\mu_2^*-\mu_4^*)Q_{2,t} + (\mu_3^*-\mu_4^*)Q_{3,t}, \quad (73)$$

hence $\mu_0 = \mu_4^*$, $\mu_1 = \mu_1^*-\mu_4^*$, $\mu_2 = \mu_2^*-\mu_4^*$, $\mu_3 = \mu_3^*-\mu_4^*$.

Note that if we had defined (71) as $E[Y_t] = \mu_0^* + \mu_1^* Q_{1,t} + \mu_2^* Q_{2,t} + \mu_3^* Q_{3,t} + \mu_4^* Q_{4,t}$ the parameters involved are no longer identified, because then (73) becomes

$$E[Y_t] = (\mu_0^* + \mu_4^*) + (\mu_1^*-\mu_4^*)Q_{1,t} + (\mu_2^*-\mu_4^*)Q_{2,t} + (\mu_3^*-\mu_4^*)Q_{3,t}, \quad (74)$$

which is also equivalent to (70). Hence

$$\mu_0 = \mu_0^* + \mu_4^*,\ \mu_1 = \mu_1^* - \mu_4^*,\ \mu_2 = \mu_2^* - \mu_4^*,\ \mu_3 = \mu_3^* - \mu_4^*, \tag{75}$$

which is a system of four equations in five unknowns.

The presence of seasonally varying expectations can be observed from the autocorrelation function. For example, let Y_t be a quarterly time series satisfying

$$Y_t = \mu_0 + \mu_1 Q_{1,t} + \mu_2 Q_{2,t} + \mu_3 Q_{3,t} + X_t \tag{76}$$

where X_t is zero-mean covariance stationary with covariance function $\gamma_x(m) = E(X_t.X_{t-m})$. The sample average of Y_t is

$$\begin{aligned}\bar{Y}_n &= \mu_0 + \mu_1(1/n)\Sigma_{t=1}^n Q_{1,t} + \mu_2(1/n)\Sigma_{t=1}^n Q_{2,t} + \mu_3(1/n)\Sigma_{t=1}^n Q_{3,t} + (1/n)\Sigma_{t=1}^n X_t \\ &\approx \mu_0 + 0.25\mu_1 + 0.25\mu_2 + 0.25\mu_3\end{aligned} \tag{77}$$

if n is large, because for each s = 1,2,3 the fraction of values of $Q_{s,t}$ for t = 1,...,n that are equal to 1 tends towards 0.25 if $n \to \infty$, and $\text{plim}_{n\to\infty}(1/n)\Sigma_{t=1}^n X_t = E[X_t] = 0$ by the law of large numbers. Then it can be shown[5] that there exists constants c_s, $s = 1,2,3,4,$ such that for $n \to \infty$,

$$\frac{1}{n-m}\sum_{t=m+1}^{n}(Y_t - \bar{Y}_n)(Y_{t-m} - \bar{Y}_n) \to \begin{cases}\gamma_x(m) + c_1 \text{ for } m = 0,4,8,12,.....\\ \gamma_x(m) + c_2 \text{ for } m = 1,5,9,13,.....\\ \gamma_x(m) + c_3 \text{ for } m = 2,6,10,14,.....\\ \gamma_x(m) + c_4 \text{ for } m = 3,7,11,15,.....\end{cases} \tag{78}$$

in probability. It follows now from (48) and (78) that the estimated autocorrelation function $\hat{\rho}(m)$ will have spikes at distances of four lags, and will not die out to zero:

$$\hat{\rho}(m) \to \rho(m),\ where\ \rho(m) = \begin{cases}(\gamma_x(m) + c_1)/(\gamma_x(0) + c_1) \text{ for } m = 0,4,8,12,.....\\ (\gamma_x(m) + c_2)/(\gamma_x(0) + c_1) \text{ for } m = 1,5,9,13,.....\\ (\gamma_x(m) + c_3)/(\gamma_x(0) + c_1) \text{ for } m = 2,6,10,14,.....\\ (\gamma_x(m) + c_4)/(\gamma_x(0) + c_1) \text{ for } m = 3,7,11,15,.....\end{cases} \tag{79}$$

in probability as $n \to \infty$.

For example consider the quarterly process

---

[5] But the derivation involved is too tedious and therefore omitted.

$$Y_t = 1 + 2Q_{1,t} - Q_{2,t} - 2Q_{3,t} + X_t, \text{ where } X_t \sim i.i.d.\ N(0,1), \tag{80}$$

for t = 1,2,.....,225. The estimated autocorrelation function $\hat{\rho}(m)$ of this process is displayed in Figure 7.

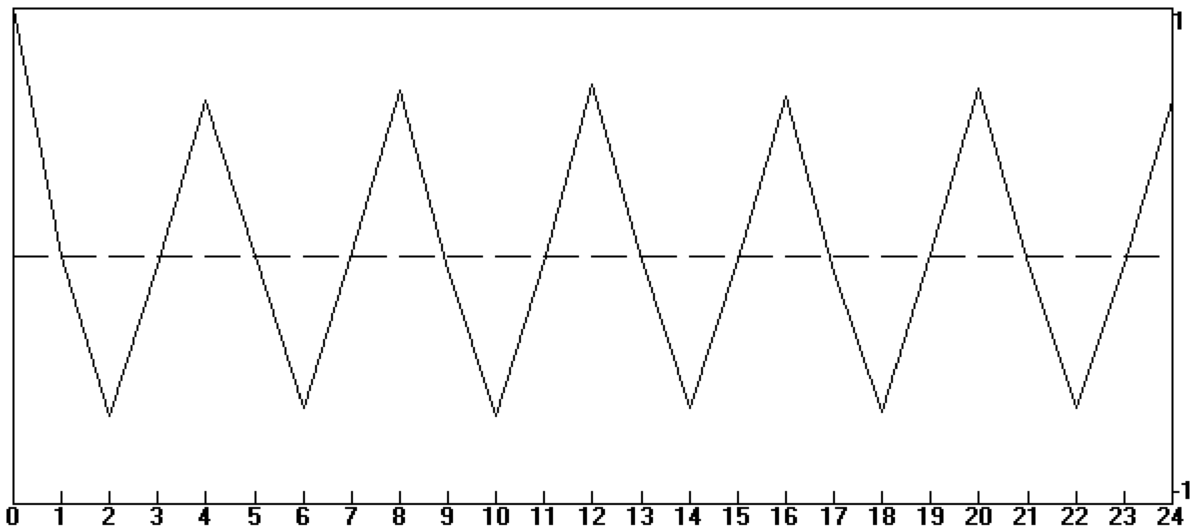

Figure 7: Estimated autocorrelation function $\hat{\rho}(m)$ of quarterly process (80)

12.2 *Seasonal lag polynomials*

Seasonality may also occur in the process X_t in (76) itself. For example, let X_t be a quarterly ARMA process

$$\varphi_p(L)\lambda_r(L^4)X_t = \beta_0 + \psi_q(L)\eta_s(L^4)U_t, \tag{81}$$

where $\varphi_p(L)$ and $\psi_q(L)$ are the non-seasonal AR and MA lag polynomials of orders p and q, respectively, defined before, and $\lambda_r(z)$ and $\eta_s(z)$ are the seasonal AR and MA polynomials of orders r and s, respectively.

In EasyReg these polynomials are specified via the window displayed in Figure 8 below. The coefficients a(1,i), i = 1,...,p, are the coefficients of the non-seasonal AR polynomial $\varphi_p(L)$, the coefficients a(2,i), i = 1,...,q, are the coefficients of the non-seasonal MA polynomial $\psi_q(L)$, the coefficients c(1,i), i = 1,...,r, are the coefficients of the seasonal AR polynomial $\lambda_r(L^4)$, and the coefficients c(2,i), i = 1,...,s, are the coefficients of the seasonal MA polynomial $\eta_s(L^4)$. The displayed specification is for $p = q = r = s = 2$.

The specification procedure for p, q, r and s is similar to the non-seasonal ARMA case: First, specify upper bounds of p, q, r and s, and then use the information criteria to select the correct p, q, r and s, via Menu > Single equation models > ARIMA model selection via information criteria.

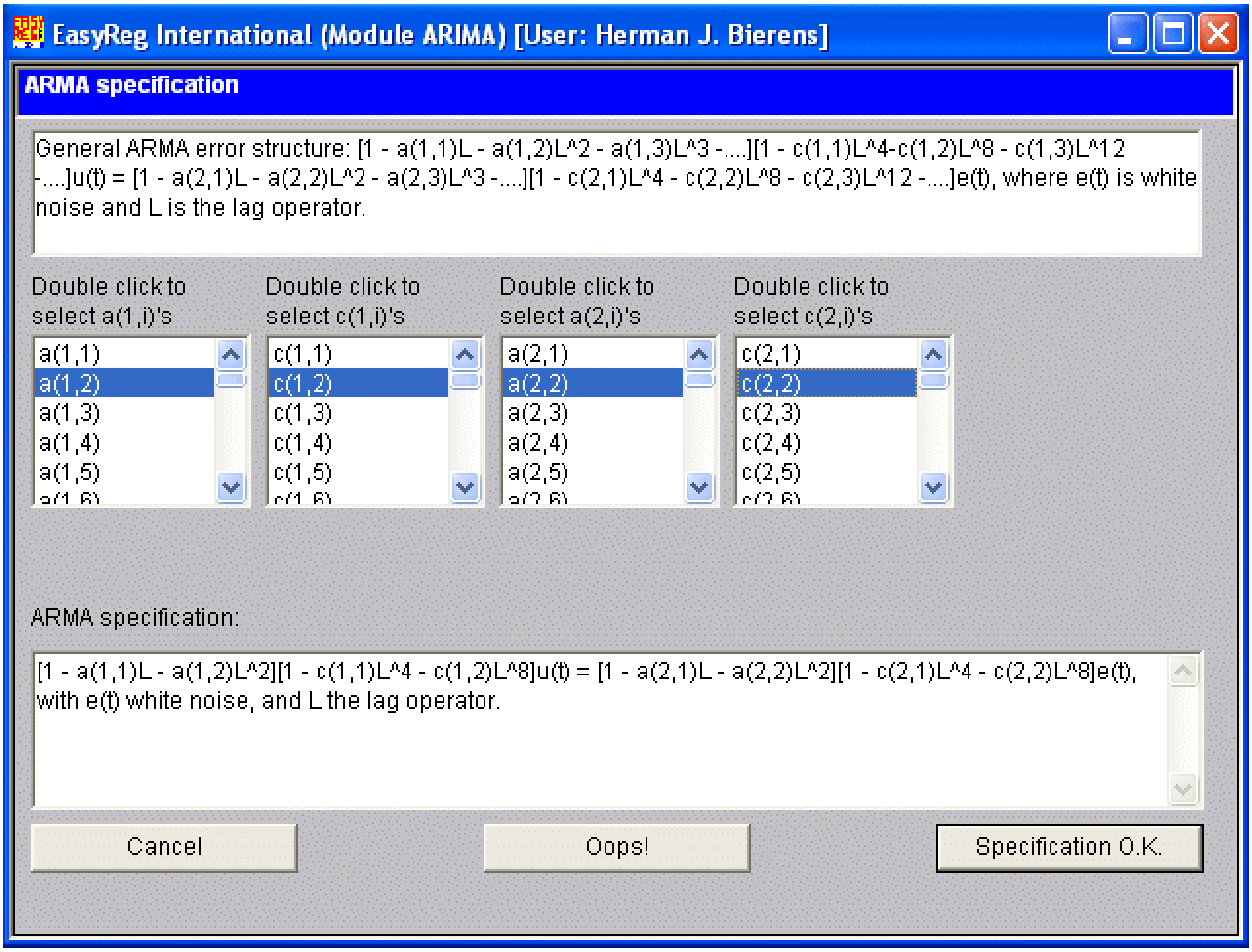

Figure 8: Specification of a seasonal ARMA model in EasyReg.

13. *Unit roots*

13.1 *What is a unit root process?*

Consider the AR(p) process (20), written as

$$\varphi_p(L)Y_t = \beta_0 + U_t, \text{ where } \varphi_p(L) = 1 - \beta_1 L - \beta_2 L^2 - .... - \beta_p L^p. \tag{82}$$

If $\varphi_p(1) = 0$, this process is called a unit root process. If so, $\varphi_p(L) = (1-L)\varphi^*_{p-1}(L)$, where

$\varphi^*_{p-1}(L)$ is a lag polynomial of order $p-1$. Moreover, the process (82) is then no longer stationary, because the condition in Proposition 5 is no longer satisfied. But if $\varphi^*_{p-1}(L)$ satisfies the condition in Proposition 5 then the first difference of Y_t, $\Delta Y_t = Y_t - Y_{t-1}$, is a stationary AR($p-1$) process: $\varphi^*_{p-1}(L)\Delta Y_t = \beta_0 + U_t$.

For example, consider the case $p = 1$, $\beta_0 = 0$:

$$Y_t = Y_{t-1} + U_t,\ where\ U_t \sim i.i.d.\ N(0,\sigma^2). \tag{83}$$

This process is called a random walk. By t times backwards substitution of (83) we get

$$Y_t = Y_0 + \Sigma_{j=1}^t U_j. \tag{84}$$

Because $var(\Sigma_{j=1}^t U_j) = \Sigma_{j=1}^t var(U_t) = \sigma^2 t$, it is clear that Y_t is not stationary.

13.2 *The Augmented Dickey-Fuller (ADF) tests*

Now the question arises how to distinguish the case (83) from a stationary AR(1) process $Y_t = \beta Y_{t-1} + U_t$, $|\beta| < 1$. The latter process can be written as

$$\Delta Y_t = Y_t - Y_{t-1} = (\beta-1)Y_{t-1} + U_t = \alpha Y_{t-1} + U_t, \tag{85}$$

say, where $\alpha = \beta - 1$. Now the non-stationary case (83) corresponds to $\alpha = 0$, and the stationary case ($|\beta| < 1$) corresponds to $-2 < \alpha < 0$. This suggests to estimate the model (85) by OLS, and to use the t value $\hat{t}_\alpha$ of α to test the null hypothesis $\alpha = 0$ against the alternative hypothesis $\alpha < 0$. The problem, however, is that under the null hypothesis $\alpha = 0$ the test statistic $\hat{t}_\alpha$ has no longer a standard normal distribution, as is demonstrated in Figure 9.

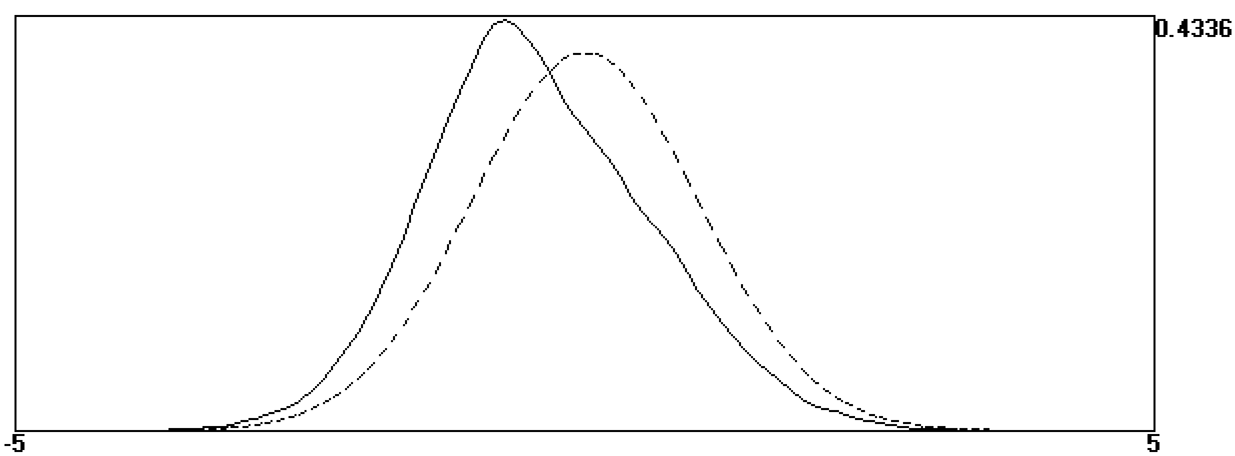

Figure 9: Density of $\hat{t}_\alpha$ under the null hypothesis $\alpha = 0$ (solid curve) compared with the standard normal density (dashed curve)

Therefore, the asymptotic critical values of the left-sided standard normal test are not valid. Instead we have under the null hypothesis $\alpha = 0$ and large sample size n,

$$P(\hat{t}_\alpha \leq -1.95) = 0.05, \quad P(\hat{t}_\alpha \leq -1.62) = 0.10. \tag{86}$$

Thus, the null hypothesis $\alpha = 0$ is rejected at the 5% significance level in favor of the alternative hypothesis $\alpha < 0$ if $\hat{t}_\alpha \leq -1.95$, and the null hypothesis $\alpha = 0$ is accepted at the 5% significance level if $\hat{t}_\alpha > -1.95$.

The assumption that the U_t's in (83) are independent is much too restrictive, though. To relax this assumption, consider an AR(2) process without intercept:

$$(1-\delta L)(1-\beta L)Y_t = U_t,\ \textit{where}\ |\beta| < 1,\ \textit{and either}\ \delta = 1\ \textit{or}\ |\delta| < 1. \tag{87}$$

The unit root hypothesis corresponds to the case $\delta = 1$, and the stationarity hypothesis corresponds to the case $|\delta| < 1$. Using the easy equality $(1-\delta L)(1-\beta L) = 1 - (\delta+\beta)L + \delta.\beta.L^2$ this model can be written as

$$\begin{aligned} \Delta Y_t &= Y_t - Y_{t-1} = (\delta+\beta)Y_{t-1} - Y_{t-1} - \delta.\beta.Y_{t-2} + U_t \\ &= (\delta+\beta)Y_{t-1} - Y_{t-1} - \delta.\beta.Y_{t-1} + \delta.\beta.Y_{t-1} - \delta.\beta.Y_{t-2} + U_t \\ &= (\delta+\beta-1-\delta.\beta)Y_{t-1} + \delta.\beta.(Y_{t-1} - Y_{t-2}) + U_t \\ &= \alpha Y_{t-1} + \gamma\Delta Y_{t-1} + U_t, \end{aligned} \tag{88}$$

where

$$\begin{aligned} \alpha &= \delta+\beta-1-\delta.\beta = -(1-\delta)(1-\beta), \\ \gamma &= \delta.\beta. \end{aligned} \tag{89}$$

Now the unit root hypothesis $\delta = 1$ corresponds to $\alpha = 0$, and the stationarity hypothesis $|\delta| < 1$ corresponds to $\alpha < 0$ (and $\alpha > -4$).

More generally we have:

Proposition 10. *An AR(p) process Y_t with intercept can always be written as*

$$\Delta Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \sum_{j=1}^{p-1}\gamma_j \Delta Y_{t-j} + U_t. \tag{90}$$

If the process Y_t is stationary then $\alpha_1 < 0$, and if Y_t is a unit root process then $\alpha_1 = 0$. If model (90) *is estimated without an intercept* (*thus $\alpha_0 = 0$*) *then under the unit root hypothesis $\alpha_1 = 0$ the t value $\hat{t}_{\alpha_1}$ of α_1 converges in distribution to a non-normal distribution with density displayed in Figure 9. If model* (90) *is estimated* ***with*** *an intercept then under the unit root hypothesis*

$\alpha_1 = 0$ *the t value* $\hat{t}_{\alpha_1}$ *of* α_1 *converges in distribution to a non-normal distribution with density displayed in Figure 10 below.*

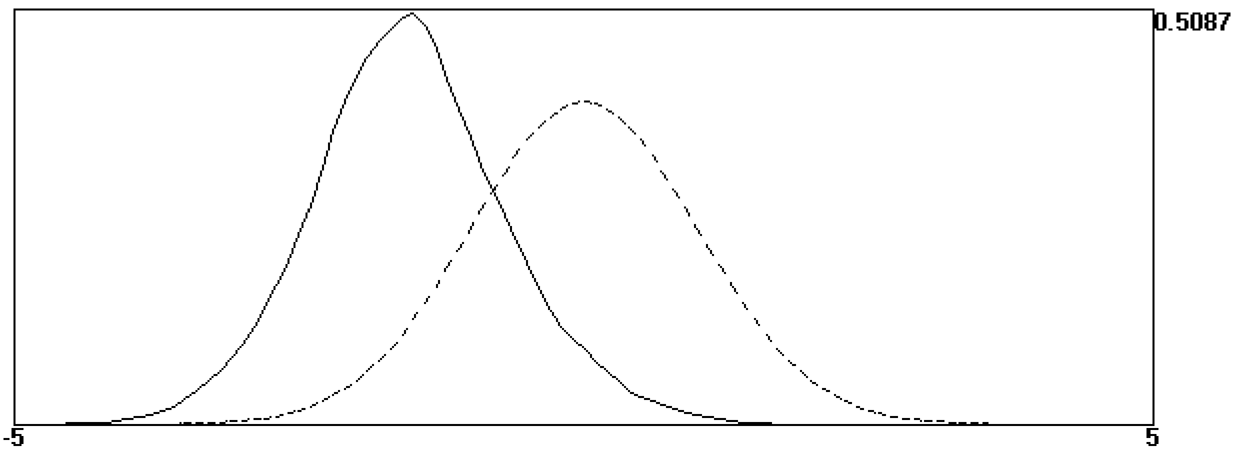

Figure 10: Density of $\hat{t}_{\alpha_1}$ for model (90) with intercept under the null hypothesis $\alpha_1 = 0$ (solid curve) compared with the standard normal density (dashed curve)

Observe that the density of $\hat{t}_{\alpha_1}$ in Figure 10 is shifted even more farther to the left of the standard normal density than in Figure 9, hence the left-sided standard normal test would result in a dramatically higher type 1 error than in the case without an intercept. In particular, we now have that under the null hypothesis $\alpha_1 = 0$ and for sufficiently long time series,

$$P(\hat{t}_{\alpha_1} \le -2.86) = 0.05, \quad P(\hat{t}_{\alpha_1} \le -2.57) = 0.1. \tag{91}$$

Thus, the null hypothesis $\alpha_1 = 0$ is rejected at the 5% significance level in favor of the alternative hypothesis $\alpha_1 < 0$ if $\hat{t}_{\alpha_1} \le -2.86$, and the null hypothesis $\alpha_1 = 0$ is accepted at the 5% significance level if $\hat{t}_{\alpha_1} > -2.86$.

The left-sided test of the unit root (null) hypothesis $\alpha_1 = 0$ against the (alternative) stationarity hypothesis $\alpha_1 < 0$ based on the t value $\hat{t}_{\alpha_1}$ of α_1 in model (90) is known as the Augmented Dickey-Fuller (ADF) test. If model (90) is estimated without an intercept the alternative hypothesis is zero-mean stationarity: $E[Y_t] = 0$. This is ADF test 1 in EasyReg. However, zero-mean stationarity is very rare for economic time series. Therefore, it is recommended to always include an intercept in model (90), so that the alternative hypothesis is stationarity about a constant. This is ADF test 2 in EasyReg.

13.3 *Unit root processes with drift*

Consider the random walk with a constant:

$$Y_t = Y_{t-1} + \tau + U_t. \tag{92}$$

This is a random walk with **drift**, with τ the drift parameter. By t times backwards substitution of (92) we get

$$Y_t = Y_0 + \tau.t + \sum_{j=1}^{t} U_j. \tag{93}$$

Due to the deterministic time trend $\tau.t$ this process moves upwards if $\tau > 0$ and downwards if $\tau < 0$. An example of a unit root process with drift is the log of the nominal GDP of the U.S., which is displayed in Figure 11 for the years 1909-1988.

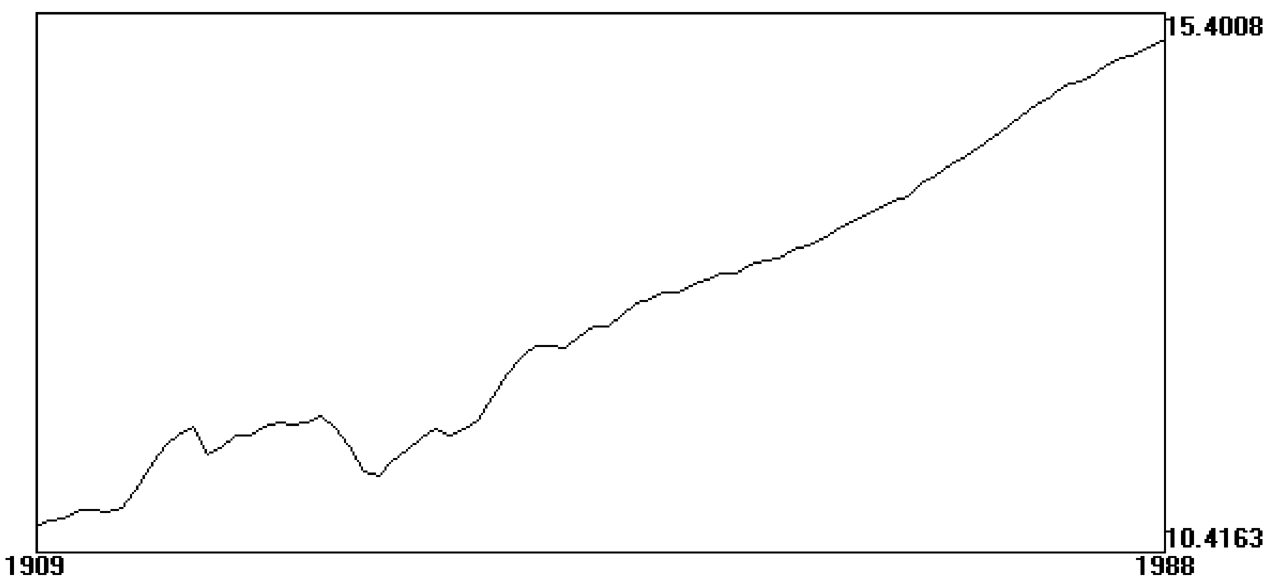

Figure 11: Log of nominal GDP of the U.S.

In this case the proper alternative to the unit root with drift hypothesis is linear trend stationarity:

$$Y_t = \tau_0 + \tau_1.t + X_t, \text{ where } X_t \text{ is zero-mean stationary.} \tag{94}$$

In particular, if X_t is an AR(p) process, then Y_t can be written as (90), but now including a time trend as well.

Proposition 11. *An AR(p) process Y_t with intercept and time trend can always be written as*

$$\Delta Y_t = \alpha_0 + \alpha_1 t + \alpha_2 Y_{t-1} + \sum_{j=1}^{p-1}\gamma_j \Delta Y_{t-j} + U_t. \tag{95}$$

If the process Y_t is trend stationary then $\alpha_2 < 0$, and if Y_t is a unit root process with drift process then $\alpha_2 = 0$. In the latter case the t value $\hat{t}_{\alpha_2}$ of α_2 converges in distribution to a non-normal distribution with density displayed in Figure 12.

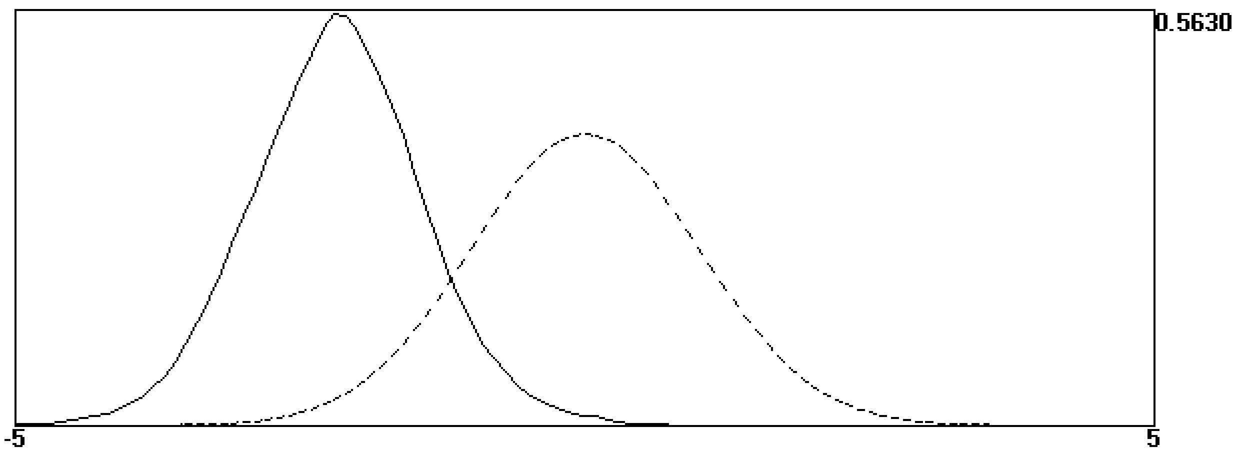

Figure 12: Density of $\hat{t}_{\alpha_2}$ for model (95) under the null hypothesis $\alpha_2 = 0$ (solid curve) compared with the standard normal density (dashed curve)

Note that in this case the asymptotic density of $\hat{t}_{\alpha_2}$ under the null hypothesis is even farther to the left of the standard normal density than in the previous two cases.

The asymptotic 5% and 10% quantiles are now

$$P(\hat{t}_{\alpha_2} \le -3.41) = 0.05, \quad P(\hat{t}_{\alpha_2} \le -3.13) = 0.1. \tag{96}$$

Thus, the null hypothesis $\alpha_2 = 0$ is rejected at the 5% significance level in favor of the alternative hypothesis $\alpha_2 < 0$ if $\hat{t}_{\alpha_2} \le -3.41$, and the null hypothesis $\alpha_2 = 0$ is accepted at the 5% significance level if $\hat{t}_{\alpha_1} > -3.41$.

13.4 *Choices you have to make*

In conducting the ADF test we have make three decisions: First, we have to determine whether to transform the time series to make it unbounded, because a bounded time series cannot be a unit root process. For example, the time series Y_t = “Percentage unemployment rate” takes only values between 0 and 100 (%). If Y_t were a unit root process without drift then $Y_t - Y_{t-1} = X_t$, where X_t is a zero-mean stationary process, and $Y_t = Y_0 + \sum_{j=1}^{t} X_j$. It follows from the central limit theorem that for $t \to \infty$, $(1/\sqrt{t})\sum_{j=1}^{t} X_j$ converges in distribution to the normal distribution with zero expectation, hence

$$\lim_{t \to \infty} P[Y_t \le 0] = \lim_{t \to \infty} P[Y_t \sqrt{t} \le 0] = \lim_{t \to \infty} P[Y_0/\sqrt{t} + (1/\sqrt{t})\sum_{j=1}^{t} X_j \le 0] = 1/2,$$

which is impossible. Therefore, we have to make Y_t unbounded. For the unemployment rate involved we can do that by transforming Y_t to $Y_t^* = \ln(Y_t/(100 - Y_t))$, and then test whether Y_t^* is a unit root process.

Now suppose that the time series Y_t is positive valued and not bounded from above, and that $Y_t - Y_{t-1} = X_t$, where X_t is a stationary process. Because $Y_t > 0$ we must have that for all t, $X_t > -Y_{t-1}$. However, this condition implies that X_t depends on the non-stationary process Y_{t-1}, and therefore X_t is non-stationary itself! Only if $X_t > 0$ for all t is it possible that $Y_t - Y_{t-1} = X_t$ with X_t is a stationary process, but then it is not possible that $Y_t < Y_{t-1}$, which is rare for economic time series. Therefore, also in the case where Y_t is positive valued one should make it unbounded, by taking the log transformation, before testing for a unit root.

Second, we have to determine what the appropriate alternative hypothesis is: Either stationarity about a constant (ADF test 2), or trend stationarity (ADF test 3). As to this choice, always plot the time series first and see whether there is a trend pattern in the time series, like in Figure 11. However, there are cases where the time series plot shows a trend but trend stationarity is impossible. For example, let us have a look at the time series $Y_t^* = \ln(Y_t/(100 - Y_t))$, where Y_t is the monthly percentage unemployment rate in the US from month 1948.01 to month 1995.09.

At first sight one would conclude from Figure 13 below that Y_t^* is either a unit root with drift process or a trend stationary process, with positive trend slope. However, in the latter case $\lim_{t\to\infty} Y_t^* = \infty$, hence $\lim_{t\to\infty} Y_t = 100$. Thus, if the trend stationarity hypothesis were true the unemployment rate in the far future would becomes 100%, which seems quite unlikely. Consequently, despite its appearance, Y_t^* cannot be trend stationary, hence the only plausible alternative hypothesis is that Y_t^* is stationary about a constant.

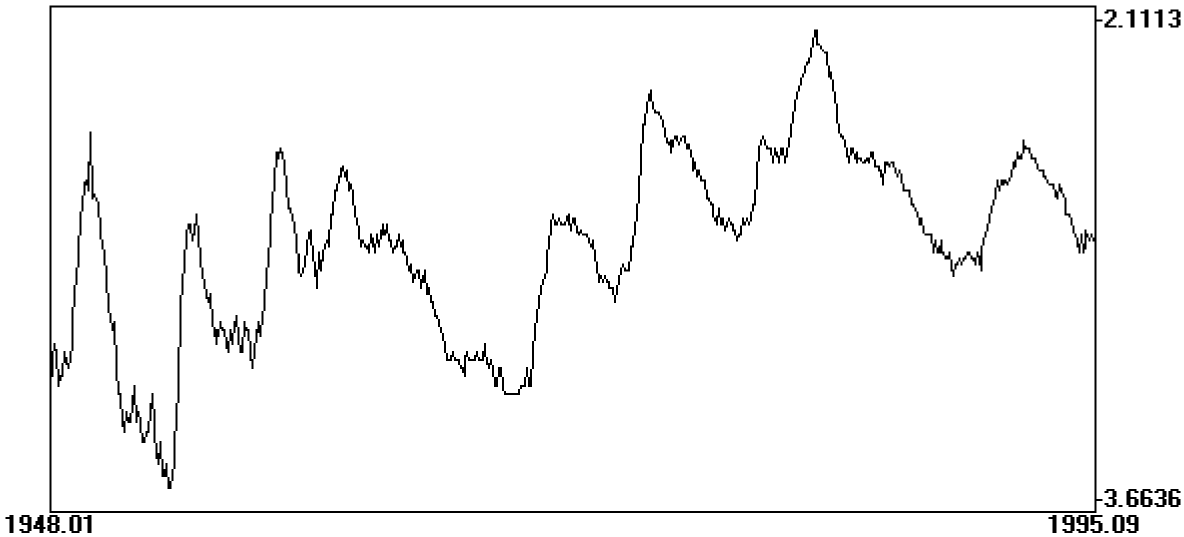

Figure 13: Plot of $\ln(Y_t/(100-Y_t))$, where Y_t is the monthly percentage unemployment rate in the US

Third, we have to choose *p*. There are various ways to do that. Given an initial choice of *p*, EasyReg will compute the Akaike, Hannan-Quinn and Schwarz information criteria, up to the initial choice of *p*. EasyReg will also automatically conduct a sequence of Wald tests to test whether the initial *p* can be reduced. Moreover, you can also fit an AR model to the first differences of the time series involved, and determine the appropriate order of the AR model involved as discussed earlier, because the choice of *p* is only critical under the null hypothesis.

13.5 *The Breitung tests*

The disadvantage of the ADF tests is that these tests only apply to AR(*p*) processes with a unit root, and that seasonal variation is excluded. An alternative test for a unit root that also applies to ARMA processes with a unit root in the AR lag polynomial is the Breitung test, which only requires to specify the alternative hypothesis.

Given a time series $Y_1, Y_2, ..., Y_n$, let

$$X_t = Y_t \text{ if the alternative hypothesis is zero-mean stationarity,} \tag{97}$$

$$X_t = Y_t - \bar{Y} \text{ if the alternative hypothesis is stationarity about a contstant,} \tag{98}$$

where $\bar{Y} = (1/n)\sum_{t=1}^{n} Y_t$, or

$$X_t = Y_t - \hat{\alpha} - \hat{\beta}.t \text{ if the alternative hypothesis is trend stationarity,} \tag{99}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the OLS estimates of the intercept and the slope parameter, respectively, of the linear regression of Y_t on the constant 1 and time *t*. Thus, in the latter case X_t is the OLS residual of the trend regression involved.

Next, let $S_t = \sum_{j=1}^{t} X_j$. Then the test statistic of the Breitung test is B_n/n, where

$$B_n = \frac{\sum_{t=1}^{n} S_t^2}{n.\sum_{t=1}^{n} X_t^2}. \tag{100}$$

Under the unit root hypothesis B_n/n converges in distribution, where the limiting distribution is non-normal and different for each of the three cases (97), (98) and (99). On the other hand, under one of the alternatives (97), (98) or (99), B_n itself converges in distribution, hence B_n/n converges in probability to zero. Thus, the Breitung test is a left-sided test: The unit root (with drift) hypothesis is rejected if B_n/n is smaller than a critical value. The critical values are different for the cases (97), (98) and (99).

In the case of seasonal data, the Breitung test should be conducted on seasonal moving averages in order to eliminate possible seasonal variation in the time series involved. For example, if Y_t is a quarterly time series satisfying (76), the Breitung test should be conducted with Y_t replaced by $\bar{Y}_t = (1/4)\sum_{i=0}^{3} Y_{t-i}$, because then the seasonal dummy variables in (76) are wiped out.

14. *ARIMA models*

A non-seasonal ARIMA(*p,r,q*) model for a time series Y_t takes the form

$$X_t = (1-L)^r Y_t, \quad \varphi_p(L)X_t = \beta_0 + \psi_q(L)U_t, \tag{101}$$

where X_t is a stationary ARMA(*p,q*) process. Compare (52) and Proposition 8. The parameter *r* indicates how many time we need to difference Y_t to get a stationary ARMA(*p,q*) process X_t. For economic time series *r* is usually either 0 or 1, and can be determined on the basis of the Breitung test. Once you have determined *r* and transformed Y_t to X_t, the ARMA model for X_t can be specified and estimated as before, and be used for forecasting out-of-sample values of X_t and Y_t.

15. *ARCH and GARCH models, and forecasting volatility*

15.1 *ARCH errors*

ARCH stands for Auto-Regressive Conditional Heteroskedasticity, and relates to the conditional variance of model errors. Let U_t be the error in a (conditional expectation) model for Y_t. As we have seen before, the errors of a correctly specified model for a time series Y_t should satisfy $E[U_t | Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4}, ...] = 0$. See (19). However, in general this condition does not imply that the conditional variance of U_t,

$$\sigma_t^2 = E[U_t^2 | Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4}, ...], \tag{102}$$

is constant. Only if Y_t is Gaussian it is guaranteed that σ_t^2 is constant.

In the case of ARCH(p) errors, U_t is specified as

$$\begin{aligned} &U_t = e_t\sqrt{\alpha_0 + \alpha_1 U_{t-1}^2 + \alpha_2 U_{t-2}^2 + ... + \alpha_p U_{t-p}^2} = e_t\sqrt{\alpha_0 + \textstyle\sum_{j=1}^p \alpha_j U_{t-j}^2}, \\ &\textit{where } e_t \sim \textit{i.i.d. } N(0,1),\ \alpha_0 > 0,\ \textit{and } \alpha_j \geq 0,\ j = 1,2,...,p. \end{aligned} \tag{103}$$

The restrictions on the parameters involved are needed to guarantee that the expression over which the square root is taken is always positive. It follows from (103), and the conditional expectation property (120) [with $V = e_t$ and $X = \sqrt{\alpha_0 + \sum_{j=1}^p \alpha_j U_{t-j}^2}$], that

$$\begin{aligned} E[U_t|Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-2}, ...] &= E[e_t|Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-2}, ...].\sqrt{\alpha_0 + \textstyle\sum_{j=1}^p \alpha_j U_{t-j}^2} \\ &= E[e_t].\sqrt{\alpha_0 + \textstyle\sum_{j=1}^p \alpha_j U_{t-j}^2} = 0. \end{aligned} \tag{104}$$

Moreover, similar to (104) we have

$$\begin{aligned} \sigma_t^2 &= E[U_t^2|Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-2}, ...] = E[e_t^2|Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-2}, ...].\left(\alpha_0 + \textstyle\sum_{j=1}^p \alpha_j U_{t-j}^2\right) \\ &= E[e_t^2].\left(\alpha_0 + \textstyle\sum_{j=1}^p \alpha_j U_{t-j}^2\right) = \alpha_0 + \textstyle\sum_{j=1}^p \alpha_j U_{t-j}^2. \end{aligned} \tag{105}$$

Denoting $V_t = U_t^2 - \sigma_t^2$, the result (105) can be written as an AR(p) model for U_t^2:

$$U_t^2 = \alpha_0 + \textstyle\sum_{j=1}^p \alpha_j U_{t-j}^2 + V_t, \tag{106}$$

where $E[V_t | Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4}, ...] = 0$. Denoting

$$\alpha(L) = 1 - \alpha_1 L - \alpha_2 L^2 - ... - \alpha_p L^p, \tag{107}$$

the process (106) is stationary if $\alpha(z) = 0$ implies $|z| > 1$. The latter condition has to be imposed as well.

15.2 *GARCH errors*

GARCH stands for Generalized Auto-Regressive Conditional Heteroskedasticity. In the case of GARCH(q,p) errors, U_t is specified as

$$U_t = e_t \sigma_t, \text{ where } e_t \sim i.i.d.\ N(0,1) \text{ and}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 U_{t-1}^2 + \alpha_2 U_{t-2}^2 + ... + \alpha_p U_{t-p}^2 + \theta_1 \sigma_{t-1}^2 + \theta_2 \sigma_{t-2}^2 + ... + \theta_q \sigma_{t-q}^2 \qquad (108)$$

$$\text{with } \alpha_0 > 0,\ \alpha_j \geq 0,\ j = 1,2,...,p,\ \theta_i \geq 0,\ i = 1,2,...,q.$$

Again, the restrictions on the parameters involved are needed to guarantee that $\sigma_t^2 > 0$. Denoting

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - .... - \theta_q L^q \qquad (109)$$

and using (107), the model for σ_t^2 in (108) can be written as

$$\theta(L)\sigma_t^2 = \alpha_0 + (1-\alpha(L))U_t^2. \qquad (110)$$

Assuming that $\theta(z) = 0$ implies $|z| > 1$, the lag polynomial $\theta(L)$ is invertible, hence it follows from (110) that

$$\sigma_t^2 = \alpha_0/\theta(1) + \theta(L)^{-1}(1-\alpha(L))U_t^2, \qquad (111)$$

and thus

$$U_t^2 = \alpha_0/\theta(1) + \theta(L)^{-1}(1-\alpha(L))U_t^2 + V_t, \qquad (112)$$

where $V_t = U_t^2 - \sigma_t^2$. Applying the lag polynomial $\theta(L)$ to both sides of equation (112) yields an ARMA(max(p,q),q) model for U_t^2:

$$\left(\theta(L) + \alpha(L) - 1\right)U_t^2 = \left(1 - \Sigma_{i=1}^q \theta_i L^i - \Sigma_{j=1}^p \alpha_j L^j\right)U_t^2 = \alpha_0 + \theta(L)V_t, \qquad (113)$$

Note that if we would choose $p = 0$, so that $\alpha(L) = 1$, then (113) becomes an ARMA model with common AR and MA lag polynomial $\theta(L)$ and thus with common roots: $\theta(L)U_t^2 = \alpha_0 + \theta(L)V_t$, which is equivalent to $U_t^2 = \alpha_0/\theta(1) + V_t$. Consequently, $\sigma_t^2 = \alpha_0/\theta(1)$ is then constant because $E[V_t | Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4}, ...] = 0$. This result follows also directly from (111): If $\alpha(L) = 1$ then (111) implies $\sigma_t^2 = \alpha_0/\theta(1)$. Therefore, in specifying a GARCH model for the errors U_t you have to choose $p \geq 1$.

15.3 *Estimating a model with* (*G*)*ARCH errors via EasyReg*

The option to estimate a model with ARCH or GARCH errors is only available in

EasyReg if you estimate a linear regression model first, via Menu > Single equation models > Linear regression models. The ARIMA module in EasyReg does not have this option. This is not a restriction, though. You can estimate an ARIMA model with GARCH errors as follows.

(1) First, determine whether you have to difference the time series in order to make it stationary, by testing whether the time series is a unit root process. If so, take the first differences of the time series, via Menu > Input > Transform variables > Time series transformations. Let Y_t be the stationary time series involved.

(2) Add missing values to the data, via Menu > Input > Prepare data for forecasting, in order to enable out of sample forecasting.

(3) Use the option "ARIMA section via information criteria" to specify an ARMA model. If Y_t is stationary about a constant, include the constant 1. If Y_t is trend stationary, include the constant 1 and the time *t,* and if Y_t is a seasonal time series you may include seasonal dummy variables as well.

(4) Regress Y_t on a constant, and eventually a time trend and/or seasonal dummy variables, via OLS.

(5) Once you have estimated this model, choose the option "Re-estimate the model with ARMA errors", under menu item "Options" in the "What to do next?" window.

(6) When done, choose the option "Re-estimate the model with GARCH errors". This yields an ARMA model with GARCH errors.

If after that you want to specify a different ARMA model for Y_t, or a different GARCH model for the errors, you have to redo the last three steps. The details of these steps are explained in the EasyReg guided tour on OLS estimation, which you can access via Tours > OLS.HTM.

To illustrate these procedure, I will choose Y_t = %DIF1[SP 500], which is the percentage change of the monthly SP 500 index in the US, starting at month 1 of 1950. The latter time series is in the EasyReg data base. Moreover, I have added missing values to the time series, to enable out-of-sample forecasting. The time series is stationary about a constant. The automatic ARMA model selection on the basis of the information criteria suggests that Y_t is an MA(1) process.

Regressing Y_t on a constant only:

```
Dependent variable:
Y = %DIF1[SP500 index]
  First available observation = 2(=1939.02)
  Last available observation  = 673(=1995.01)
  First chosen observation = 133(=1950.01)
  Last chosen observation  = 673(=1995.01)
  Number of usable chosen observations: 541
X variables:
X(1) = 1

Model:
Y = b(1)X(1)  + U,
where U is the error term, satisfying E[U|X(1)] = 0.

OLS estimation results
Parameters  Estimate    t-value    H.C. t-value
                         (S.E.)     (H.C. S.E.)
                      [p-value]  [H.C. p-value]
b(1)         0.67191      4.723           4.723
                      (0.14227)       (0.14227)
                      [0.00000]       [0.00000]

Notes:
1: S.E. = Standard error
2: H.C. = Heteroskedasticity Consistent. These t-values and
   standard errors are based on White's heteroskedasticity
   consistent variance matrix.
3: The two-sided p-values are based on the normal approximation.

Effective sample size (n):                              541
Variance of the residuals:                        10.949943
Standard error of the residuals (SER):              3.30907
Residual sum of squares (RSS):                  5912.969032
(Also called SSR = Sum of Squared Residuals)
Total sum of squares (TSS):                     5912.969032

Information criteria:
Akaike:              2.39518E+00
Hannan-Quinn:        2.39828E+00
Schwarz:             2.40312E+00
```

Next, choose the option “Re-estimate the model with ARMA errors”, and specify an ARMA(0,1) model:

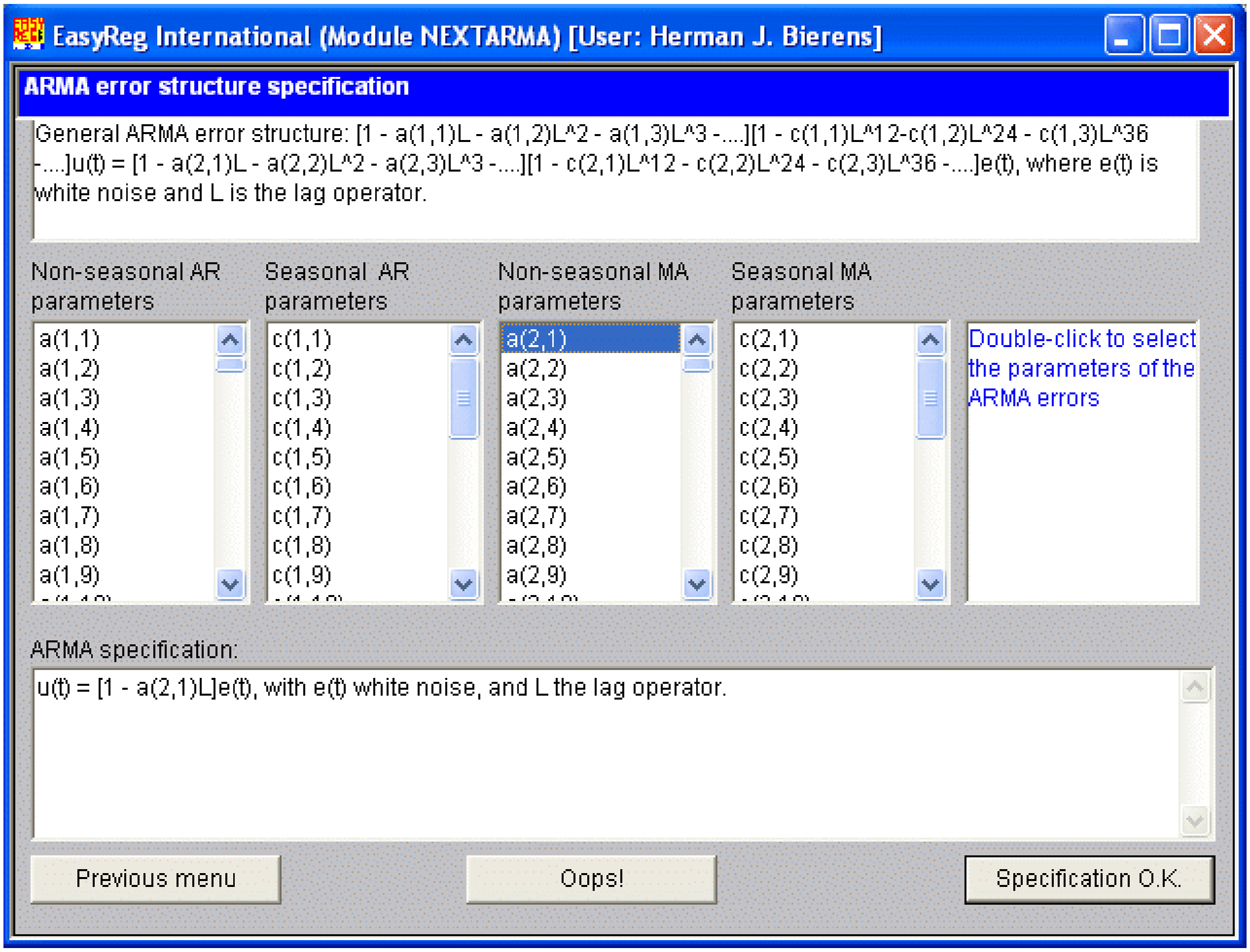

Figure 14: MA(1) error specification

The EasyReg results are:

```
Error specification: u(t) = [1 - a(2,1)L]e(t), with e(t) white
noise and L the lag operator.

Parameters    estimate t-value  [p-value] HC t-value) [HC p-value]
b(1)          0.672896   3.862 [0.00011]       3.872    [0.00011]
a(2,1)       -0.265969  -6.404 [0.00000]      -5.288    [0.00000]
RSS = 55.271441184E+02
s.e. = 32.022557328E-01
R-square = 0.0653
n = 541
```

```
Information criteria:
Akaike:              2.33140E+00
Hannan-Quinn:        2.33761E+00
Schwarz:             2.34727E+00
```

Next, re-estimate the MA(1) model with GARCH errors. I will specify an GARCH(1,1) model. See Figure 15.

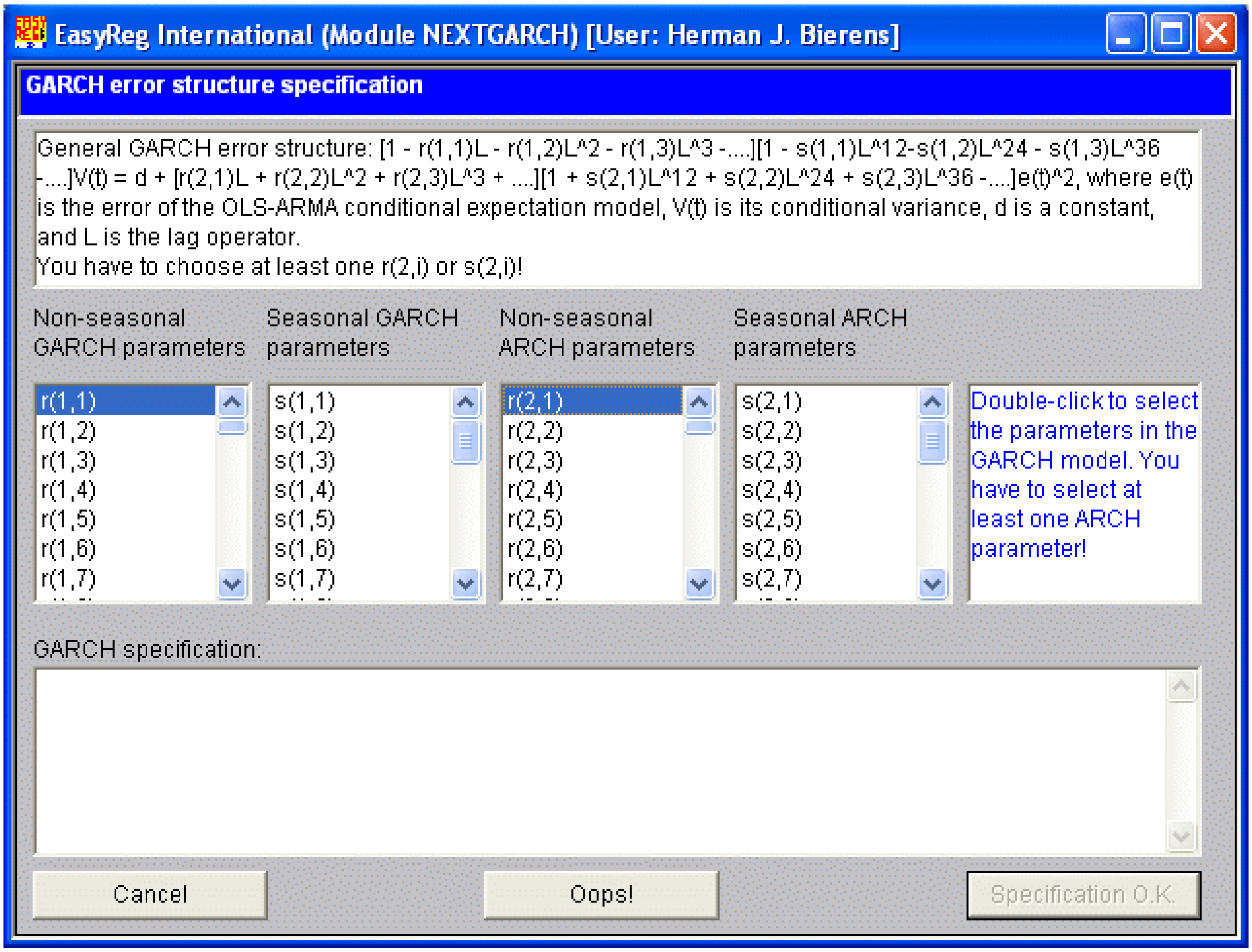

Figure 15

The EasyReg results are:

```
GARCH specification:
[1 - r(1,1)L]V(t) = d  + [r(2,1)L]e(t)^2,
where e(t) is the error of the OLS-ARMA conditional expectation
model, V(t) is its conditional variance, d is a constant, and L
```

```
is the lag operator.
Maximum likelihood estimation results:
Parameters  ML estimate t-value [p-value]
b(1)           0.764270   4.669 [0.00000]
a(2,1)        -0.121729  -4.916 [0.00000]
d              8.656218   3.076 [0.00210]
r(1,1)         0.001671   0.006 [0.99546]
r(2,1)         0.149719   2.983 [0.00286]
[The two-sided p-values are based on the normal approximation]

Log-Likelihood = -13.890927734E+02
RSS            = 56.461249666E+02
s.e.           = 32.455839592E-01
R-square       = 0.0451
n              = 541
Information criteria:
     Akaike:              2.315885497
     Hannan-Quinn:        2.331403141
     Schwarz:             2.355565897
```

The parameter `r(1,1)` is not significant, indicating the model is an MA(1)-ARCH(1) model. To re-estimate the model as an MA(1)-ARCH(1) model we have to start all over again. The results are:

```
Maximum likelihood estimation results:
Parameters  ML estimate t-value [p-value]
b(1)           0.768396   4.703 [0.00000]
a(2,1)        -0.119612  -4.851 [0.00000]
d              8.685394  15.866 [0.00000]
r(2,1)         0.147329   3.106 [0.00190]
[The two-sided p-values are based on the normal approximation]

Log-Likelihood = -13.890972900E+02
RSS            = 56.498197461E+02
s.e.           = 32.436213715E-01
R-square       = 0.0445
n              = 541
Information criteria:
     Akaike:              2.312205337
     Hannan-Quinn:        2.324619452
     Schwarz:             2.343949657
```

EasyReg also provides the option to plot the GARCH variances:

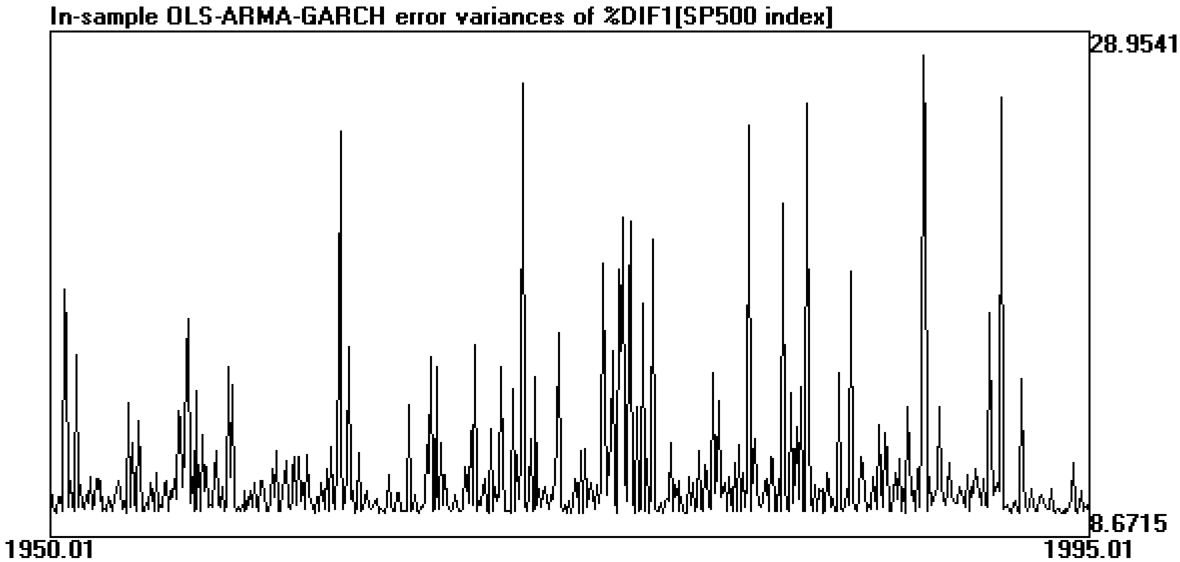

Figure 16

15.4 *Forecasting volatility*

In finance the (G)ARCH variance σ_t^2 is called the **volatility** of a financial time series. EasyReg does not allow you to recursively forecast the conditional variance σ_t^2 directly, so that you have to use the following indirect approach.

First, add at least one missing value to the data, via Menu > Input > Prepare time series for forecasting, in order to enable the forecast option.

After completion of step (6) above, open "Options > Write the residuals to the input file".

Next, take the square of the residuals $\hat{U}_t$, via Menu > Input > Transform variables > Multiplicative transformation, with power 2. Then $\hat{U}_t^2$ is added to the data file.

Now estimate an ARMA model for $\hat{U}_t^2$, similar to the ARMA(max(p,q),q) model (113) for GARCH(p,q) errors, or an AR(p) model for $\hat{U}_t^2$ in the ARCH(p) case.

Finally, choose the recursive forecast option. Then you get the recursive out-of-sample forecasts of σ_t^2.

Technical Appendix

This appendix contains some tedious derivations together with advanced material. None is part of the course: You may skip it if you wish. I will not aks questions on the exam about the contents of this appendix, nor is understanding of this material essential for doing well on the exam.

A.1 *Optimality of conditional expectations for forecasting*

To show that the conditional expectation function $E[Y|X]$ is the best forecast of Y given X, recall that the conditional expectation $E[Y|X]$ is a function of X, for example

$$E[Y|X] = g(X). \tag{114}$$

In particular, if $f(y,x)$ is the joint density of Y and X, then $f_x(x) = \int_{-\infty}^{\infty} f(y,x)dy$ is the marginal density of X, $f(y|x) = f(y,x)/f_x(x)$ is the conditional density of Y given $X = x$, and

$$E[Y|X=x] \overset{def.}{=} \int_{-\infty}^{\infty} yf(y|x)dy = g(x), \tag{115}$$

is the conditional expectation of Y given $X = x$. Plugging in X for x, we get (114).

Define

$$U = Y - E[Y|X] = Y - g(X). \tag{116}$$

Substituting (116) (in the form $Y = g(X) + U$) in (2) now yields:

$$\begin{aligned} E[(Y - \hat{Y})^2] &= E[(U + g(X) - \varphi(X))^2] \\ &= E[U^2 + 2.U(g(X) - \varphi(X)) + (g(X) - \varphi(X))^2] \\ &= E[U^2] + 2.E[U(g(X) - \varphi(X))] + E[(g(X) - \varphi(X))^2]. \end{aligned} \tag{117}$$

Next, recall the following properties of conditional expectations: For any function ψ of X, and random variable V and Z,

$$E[\psi(X)|X] = \psi(X), \tag{118}$$

$$E\big[E[Z|X]\big] = E[Z], \tag{119}$$

$$E[V.\psi(X)|X] = \psi(X).E[V|X]. \tag{120}$$

Property (119) is known as the law of iterated expectations.

It follows from (114), (116) and (118), with $\psi(x) = g(x)$ in (118), that

$$E[U|X] = E[Y|X] - E[g(X)|X] = E[Y|X] - g(X) = g(X) - g(X) = 0. \tag{121}$$

Moreover it follows from (120) and (121), with $V = U$ and $\psi(X) = g(X) - \varphi(X)$ in (120), that

$$E\left[U\big(g(X) - \varphi(X)\big)\,\middle|\,X\right] = \big(g(X) - \varphi(X)\big)E[U|X] = 0, \tag{122}$$

hence it follows from (119), with $Z = U\big(g(X) - \varphi(X)\big)$, and (122) that

$$E\left[U\big(g(X) - \varphi(X)\big)\right] = E\left[E\big(U\big(g(X) - \varphi(X)\big)|X\big)\right] = E\left[\big(g(X) - \varphi(X)\big)E(U|X)\right] = 0. \tag{123}$$

The latter result implies that (117) can be written as

$$E\left[\big(Y - \hat{Y}\big)^2\right] = E\left[U^2\right] + E\left[\big(g(X) - \varphi(X)\big)^2\right]. \tag{124}$$

Clearly, (124) is minimal for $\varphi(X) = g(X)$. Thus the conditional expectation $g(X) = E[Y|X]$ is the best forecasting scheme for Y given X. This argument can easily be generalized to the case with multiple explanatory variables.

A.2 *The distribution of the forecast error*

Recall that the OLS estimators of β and α in model (4) on the basis of the observations (Y_j, X_j) for $j = 1,...,n$ are

$$\hat{\beta} = \frac{\sum_{j=1}^n (X_j - \bar{X})(Y_j - \bar{Y})}{\sum_{j=1}^n (X_j - \bar{X})^2} = \frac{\sum_{j=1}^n (X_j - \bar{X})Y_j}{\sum_{j=1}^n (X_j - \bar{X})^2} = \beta + \frac{\sum_{j=1}^n (X_j - \bar{X})U_j}{\sum_{j=1}^n (X_j - \bar{X})^2} \tag{125}$$

and

$$\hat{\alpha} = \bar{Y} - \hat{\beta}.\bar{X} = \alpha + \sum_{j=1}^n \left(\frac{1}{n} - \frac{\bar{X}(X_j - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}\right).U_j, \tag{126}$$

respectively, where $\bar{X} = (1/n)\sum_{j=1}^n X_j$ and $\bar{Y} = (1/n)\sum_{j=1}^n Y_j$. Then the forecast error $Y_{n+1} - \hat{Y}_{n+1} = \alpha + \beta X_{n+1} + U_{n+1} - \hat{\alpha} - \hat{\beta}X_{n+1}$ can be written as

$$\begin{aligned} Y_{n+1} - \hat{Y}_{n-1} &= U_{n+1} - (\hat{\alpha} - \alpha) - (\hat{\beta} - \beta).X_{n-1} = U_{n+1} - \sum_{j=1}^n \left(\frac{1}{n} - \frac{\bar{X}(X_j - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}\right).U_j \\ &- \sum_{j=1}^n \left(\frac{X_{n+1}(X_j - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}\right)U_j = U_{n+1} - \sum_{j=1}^n \left(\frac{1}{n} + \frac{(X_{n+1} - \bar{X})(X_j - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}\right).U_j. \end{aligned} \tag{127}$$

where the second equality in (127) follows from (125) and (126). Moreover, due to the independence of the explanatory variables X_j and the model errors U_j the variance of the forecast error can be computed as follows.

$$\sigma^2_{Y_{n+1}-\hat{Y}_{n+1}} = \sigma^2 + \sum_{j=1}^{n}\left(\frac{1}{n} + \frac{(X_{n+1}-\bar{X})(X_j-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right)^2 .\sigma^2$$

$$= \sigma^2\left(1 + \frac{1}{n} + \frac{2}{n}.\frac{(X_{n+1}-\bar{X})\sum_{j=1}^{n}(X_j-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2} + \frac{(X_{n+1}-\bar{X})^2\sum_{j=1}^{n}(X_j-\bar{X})^2}{(\sum_{i=1}^{n}(X_i-\bar{X})^2)^2}\right) \quad (128)$$

$$= \sigma^2\left(\frac{n+1}{n} + \frac{(X_{n+1}-\bar{X})^2}{\sum_{j=1}^{n}(X_j-\bar{X})^2}\right).$$

The result (7) follows now from the fact that linear combinations of normally distributed random variables are normally distributed themselves.

A.3 *Stationarity condition and moving average representation for an AR*(1) *process*

By repeated backwards substitution of (21) we can write

$$\begin{aligned}
Y_t &= \beta_0 + \beta_1(\beta_0 + \beta_1 Y_{t-2} + U_{t-1}) + U_t.\\
&= \beta_0 + \beta_1\beta_0 + \beta_1^2 Y_{t-2} + U_t + \beta_1 U_{t-1}\\
&= \beta_0 + \beta_1\beta_0 + \beta_1^2(\beta_0 + \beta_1 Y_{t-3} + U_{t-2}) + U_t + \beta_1 U_{t-1}\\
&= \beta_0 + \beta_1\beta_0 + \beta_1^2\beta_0 + \beta_1^3 Y_{t-3} + U_t + \beta_1 U_{t-1} + \beta_1^2 U_{t-2}\\
&\quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&= \beta_0 + \beta_1\beta_0 + \beta_1^2\beta_0 + \ldots.. + \beta_1^{m-1}\beta_0 + \beta_1^m Y_{t-m}\\
&\quad + U_t + \beta_1 U_{t-1} + \beta_1^2 U_{t-2} + \ldots.. + \beta_1^{m-1} U_{t-m+1}\\
&= \beta_0\textstyle\sum_{k=0}^{m-1}\beta_1^k + \sum_{k=0}^{m-1}\beta_1^k U_{t-k} + \beta_1^m Y_{t-m}.
\end{aligned} \quad (129)$$

If Y_t is covariance stationary then $E[Y_t] = \mu$ and $E[(Y_t-\mu)(Y_{t-m}-\mu)] = \gamma(m)$ for all *t*. Then it follows from (129) that

$$\mu = \beta_0\textstyle\sum_{k=0}^{m-1}\beta_1^k + \sum_{k=0}^{m-1}\beta_1^k E[U_{t-k}] + \beta_1^m\mu = \beta_0\sum_{k=0}^{m-1}\beta_1^k + \beta_1^m\mu. \quad (130)$$

Next, subtract (130) from (129),

$$Y_t - \mu = \textstyle\sum_{k=0}^{m-1}\beta_1^k U_{t-k} + \beta_1^m(Y_{t-m} - \mu), \quad (131)$$

take the square of both sides,

$$(Y_t - \mu)^2 = \left(\sum_{k=0}^{m-1} \beta_1^k U_{t-k}\right)^2 + 2\beta_1^m \left(\sum_{k=0}^{m-1} \beta_1^k U_{t-k}\right)(Y_{t-m} - \mu) + \beta_1^{2m}(Y_{t-m} - \mu)^2 \tag{132}$$

and take expectations,

$$\begin{aligned} E[(Y_t - \mu)^2] &= E\left[\left(\sum_{k=0}^{m-1} \beta_1^k U_{t-k}\right)^2\right] + 2\beta_1^m \sum_{k=0}^{m-1} \beta_1^k E[U_{t-k}(Y_{t-m} - \mu)] + \beta_1^{2m} E(Y_{t-m} - \mu)^2] \\ &= \sigma^2 \sum_{k=0}^{m-1} \beta_1^{2k} + \beta_1^{2m} E(Y_{t-m} - \mu)^2]. \end{aligned} \tag{133}$$

The last equality in (133) follows from (19) and property (120), and the fact that under the assumption that the errors U_t are independent $N(0,\sigma^2)$ distributed,

$$E\left[\left(\sum_{k=0}^{m-1} \beta_1^k U_{t-k}\right)^2\right] = \sigma^2 \sum_{k=0}^{m-1} \beta_1^{2k}. \tag{134}$$

If Y_t is covariance stationary then $E[(Y_{t-m} - \mu)^2] = E[(Y_t - \mu)^2] = \gamma(0)$, so that (133) reads:

$$\gamma(0) = \sigma^2 \Sigma_{k=0}^{m-1} \beta_1^{2k} + \beta_1^{2m} \gamma(0). \tag{135}$$

However, if $|\beta_1| \geq 1$ then the right-hand side of (135) converges to ∞ if we let $m \to \infty$, which contradicts the condition that $\gamma(0) < \infty$. On the other hand, if $|\beta_1| < 1$ then $\beta_1^{2m} E[Y_{t-m}^2] = \beta_1^{2m} E[Y_1^2] \to 0$ as $m \to \infty$, hence it follows from (129) by letting $m \to \infty$ that

$$Y_t = \beta_0 \sum_{k=0}^{\infty} \beta_1^k + \sum_{k=0}^{\infty} \beta_1^k U_{t-k} = \frac{\beta_0}{1-\beta_1} + \sum_{k=0}^{\infty} \beta_1^k U_{t-k}. \tag{136}$$

Note that under the assumption that the errors U_t are independent $N(0,\sigma^2)$ distributed, Y_t in (136) is normally distributed, with expectation $\mu = \beta_0/(1-\beta_1)$ and variance $E[(\Sigma_{k=0}^{\infty} \beta_1^k U_{t-k})^2] = \sigma^2 \Sigma_{k=0}^{\infty} \beta_1^{2k} = \sigma^2/(1-\beta_1^2)$.

The expression at the right-hand side of (136) is called the Moving Average (MA) representation of a covariance stationary time series.

A.4 *AR order selection via information criteria*

Recall that for an AR(p) model the Akaike, Hannan-Quinn, and Schwarz information criteria take the form

Akaike: $c_n^{AR}(p) = \ln(\hat{\sigma}_p^2) + 2(1+p)/n,$

Hannan-Quinn: $c_n^{AR}(p) = \ln(\hat{\sigma}_p^2) + 2(1+p)\ln(\ln(n))/n,$

Schwarz: $c_n^{AR}(p) = \ln(\hat{\sigma}_p^2) + (1+p)\ln(n)/n,$

where n is the effective sample size and $\hat{\sigma}_p^2$ is the OLS estimator of the error variance $\sigma^2 = E[U_t^2]$. Denoting by $\hat{p}$ the value of p for which $c_n^{AR}(p)$ is minimal:

$$c_n^{AR}(\hat{p}) = \min\{c_n^{AR}(1),....,c_n^{AR}(\bar{p})\},$$

where $\bar{p} > p_0$, with p_0 the true value of p, we have in the Hannan-Quinn and Schwarz cases: $\lim_{n\to\infty}P[\hat{p} = p_0] = 1$, and in the Akaike case $\lim_{n\to\infty}P[\hat{p} \geq p_0] = 1$ but $\lim_{n\to\infty}P[\hat{p} = p_0] < 1$. Thus, the Akaike criterion may "overshoot" the true value.

These results are based on the following facts. If $p < p_0$ then $\text{plim}_{n\to\infty}\hat{\sigma}_p^2 > \text{plim}_{n\to\infty}\hat{\sigma}_{p_0}^2$, hence in all three cases, $\lim_{n\to\infty}P[c_n^{AR}(p_0) < c_n^{AR}(p)] = 1$, whereas for $p > p_0$,

$$n\left(\ln(\hat{\sigma}_{p_0}^2) - \ln(\hat{\sigma}_p^2)\right) \to \chi^2_{p-p_0} \tag{137}$$

in distribution if $n \to \infty$. The result (137) is due to the so-called likelihood-ratio test.[6] Then in the Akaike case,

$$n\left(c_n^{AR}(p_0) - c_n^{AR}(p)\right) = n\left(\ln(\hat{\sigma}_{p_0}^2) - \ln(\hat{\sigma}_p^2)\right) - 2(p-p_0) \to X_{p-p_0} - 2(p-p_0) \tag{138}$$

in distribution if $n \to \infty$, where $X_{p-p_0} \sim \chi^2_{p-p_0}$, hence

$$\lim_{n\to\infty}P[c_n^{AR}(p_0) > c_n^{AR}(p)] = P[X_{p-p_0} > 2(p-p_0)] > 0. \tag{139}$$

Consequently, in the Akaike case we have $\lim_{n\to\infty}P[\hat{p} \geq p_0] = 1$, but $\lim_{n\to\infty}P[\hat{p} > p_0] > 0$. Therefore, the Akaike criterion may asymptotically overshoot the correct number of parameters.

Since (137) implies $\text{plim}_{n\to\infty}n(\ln(\hat{\sigma}_{p_0}^2) - \ln(\hat{\sigma}_p^2))/\ln(\ln(n)) = 0$ and $\text{plim}_{n\to\infty}n(\ln(\hat{\sigma}_{p_0}^2) -$

---

6 Which, however, is beyond the undergraduate econometrics level, so you have to believe this.

$\ln(\hat{\sigma}_p^2))/\ln(n) = 0$ it follows that in the Hannan-Quinn case,

$$\text{plim}_{n\to\infty} n\left(c_n^{AR}(p_0) - c_n^{AR}(p)\right)/\ln(\ln(n)) = 2(p-p_0) \geq 2$$

and in the Schwarz case,

$$\text{plim}_{n\to\infty} n\left(c_n^{AR}(p_0) - c_n^{AR}(p)\right)/\ln(n) = p-p_0 \geq 1,$$

so that in both cases $\lim_{n\to\infty} P[c_n^{AR}(p_0) > c_n^{AR}(p)] = 0$. Hence, $\lim_{n\to\infty} P[\hat{p} = p_0] = 1$.

A.5 *Motivation for the Dickey-Fuller test*

Consider an AR(1) model without a constant

$$Y_t = \beta Y_{t-1} + U_t, \text{ where } U_t \sim i.i.d.\ N(0,1), \tag{140}$$

which can be written as

$$(1-\beta L)Y_t = U_t. \tag{141}$$

If β = 1 then the AR lag polynomial involved has a unit root: $1-z = 0$ implies $z = 1$. The model then becomes a random walk:

$$Y_t = Y_{t-1} + U_t. \tag{142}$$

This is a nonstationary process. To see this, assume for convenience that

$$Y_t = 0 \text{ for } t \leq 0. \tag{143}$$

Then if follows by backwards substitution of (142) that

$$Y_t = U_1 + U_2 + \dots U_t = \Sigma_{j=1}^t U_j, \tag{144}$$

so that under the normality assumption in (140), $Y_t \sim N(0,t)$. Thus, in this case the variance of Y_t blows up to infinity with *t*.

The model (140) can be rewritten as

$$\Delta Y_t = Y_t - Y_{t-1} = (\beta-1)Y_{t-1} + U_t = \alpha Y_{t-1} + U_t, \text{ where } \alpha = \beta - 1. \tag{145}$$

Then the unit root hypothesis β = 1 corresponds to α = 0, and the stationarity hypothesis |β| < 1 corresponds to $-2 < \alpha < 0$. This suggests to test the null hypothesis α = 0 by estimating model (145) by OLS and using the t-value $\hat{t}_\alpha$ of α for a left-sided t test. However, the problem is that under the null hypothesis α = 0 the t-value $\hat{t}_\alpha$ has no longer an asymptotic standard normal distribution.

To derive the distribution of $\hat{t}_\alpha$, let us first derive the asymptotic distribution of the OLS estimator $\hat{\alpha}$ of α in the case α = 0:

$$\hat{\alpha} = \frac{\sum_{t=1}^{n} Y_{t-1}\Delta Y_t}{\sum_{t=1}^{n} Y_{t-1}^2} = \frac{\sum_{t=1}^{n} U_t Y_{t-1}}{\sum_{t=1}^{n} Y_{t-1}^2}. \tag{146}$$

Note that by (142), $Y_t^2 - Y_{t-1}^2 = (U_t+Y_{t-1})^2 - Y_{t-1}^2 = U_t^2 + 2U_tY_{t-1}$, hence

$$2.\sum_{t=1}^{n} U_tY_{t-1} = \sum_{t=1}^{n}(Y_t^2-Y_{t-1}^2) - \sum_{t=1}^{n} U_t^2 = Y_n^2 - Y_0^2 - \sum_{t=1}^{n} U_t^2 = Y_n^2 - \sum_{t=1}^{n} U_t^2, \tag{147}$$

where the latter equality follows from (143). Therefore we can write

$$n.\hat{\alpha} = \frac{\frac{1}{2}\left(\left(Y_n/\sqrt{n}\right)^2 - (1/n)\sum_{t=1}^{n} U_t^2\right)}{(1/n)\sum_{t=1}^{n}\left(Y_{t-1}/\sqrt{n}\right)^2}. \tag{148}$$

Note that by (144) and the normality assumption in (140),

$$Y_n/\sqrt{n} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_t \sim N(0,1), \tag{149}$$

so that $\left(Y_n/\sqrt{n}\right)^2$ has a χ_1^2 distribution. Moreover, it follows from the law of large numbers that $(1/n)\sum_{t=1}^{n} U_t^2$ converges is probability to the expectation $E[U_t^2] = 1$, so that

$$(1/n)\sum_{t=1}^{n} U_t^2 = 1 + r_n, \text{ where } \operatorname{plim}_{n\to\infty} r_n = 0. \tag{150}$$

To determine the distribution of $(1/n)\sum_{t=1}^{n}\left(Y_{t-1}/\sqrt{n}\right)^2$, denote

$$\begin{aligned} W_n(x) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{[x.n]} U_t \ \ if \ 1/n \le x \le 1, \\ W_n(x) &= 0 \ \ if \ 0 \le x < 1/n, \end{aligned} \tag{151}$$

where $[x.n]$ means the largest natural number $\le x.n$. For example, [2.5] = 2, [5] = 5, [6.999] = 6. Then $Y_{t-1}/\sqrt{n} = W_n((t-1)/n)$, hence

$$\begin{aligned} (1/n)\sum_{t=1}^{n}\left(Y_{t-1}/\sqrt{n}\right)^2 &= (1/n)\sum_{t=1}^{n}\left(W_n((t-1)/n)\right)^2 = (1/n)\sum_{t=1}^{n}\left(W_n((t-1)/n)\right)^2\int_{t-1}^{t} dz \\ &= (1/n)\sum_{t=1}^{n}\int_{t-1}^{t}\left(W_n((t-1)/n)\right)^2 dz = (1/n)\sum_{t=1}^{n}\int_{t-1}^{t}\left(W_n(z/n)\right)^2 dz = (1/n)\int_0^n\left(W_n(z/n)\right)^2 dz \\ &= \int_0^1\left(W_n(x)\right)^2 dx, \end{aligned} \tag{152}$$

where the fourth equality follows from the fact that for $t-1 \leq z < t$, $[z] = t-1$, hence, $W_n(z/n) = W_n((t-1)/n)$, and the last equality follows by replacing z by $n.x$. Moreover, it follows from (150) and (151) that

$$\left(Y_n/\sqrt{n}\right)^2 - (1/n)\Sigma_{t=1}^{n}U_t^2 = \left(W_n(1)\right)^2 - 1 - r_n \tag{153}$$

Combining (148), (152) and (153) it follows now that

$$n.\hat{\alpha} = \frac{\frac{1}{2}\left(\left(W_n(1)\right)^2 - 1 - r_n\right)}{\int_0^1 \left(W_n(x)\right)^2 dx}. \tag{154}$$

Now let us have a closer look at the function $W_n(x)$. First, it follows from (151) and the normality assumption in (140) that for $0 < x \leq 1$,

$$W_n(x) = \frac{1}{\sqrt{n}}\sum_{t=1}^{[x.n]} U_t \sim N\left(0,\frac{[x.n]}{n}\right) \rightarrow N(0,x) \text{ in distribution if } n \rightarrow \infty \tag{155}$$

Moreover, for $0 < x < y \leq 1$,

$$\begin{aligned} &W_n(y) - W_n(x) = \frac{1}{\sqrt{n}}\sum_{t=[x.n]+1}^{[y.n]} U_t \sim N\left(0,\frac{[y.n]-[x.n]}{n}\right) \rightarrow N(0,y-x) \\ &\qquad\qquad \text{in distribution if } n \rightarrow \infty, \text{ and} \\ &W_n(y) - W_n(x) \text{ and } W_n(x) \text{ are independent.} \end{aligned} \tag{156}$$

This suggests the existence of a random function $W(x)$ on [0,1], called a Brownian motion or Wiener process, with the following properties:

$$\begin{aligned} &\text{for } 0 < x \leq 1,\ W(x) \sim N(0,x), \\ &\text{for } 0 < x < y \leq 1,\ W(y) - W(x) \sim N(0,y-x), \text{ and} \\ &\qquad\qquad W(x) \text{ and } W(y) - W(x) \text{ are independent,} \end{aligned} \tag{157}$$

such that $W_n(x) \rightarrow W(x)$ in distribution. This result, together with $\text{plim}_{n\rightarrow\infty} r_n = 0$, implies that

$$n.\hat{\alpha} = \frac{\frac{1}{2}\left(\left(W_n(1)\right)^2 - 1 - r_n\right)}{\int_0^1 \left(W_n(x)\right)^2 dx} \rightarrow \frac{\frac{1}{2}\left(\left(W(1)\right)^2 - 1\right)}{\int_0^1 \left(W(x)\right)^2 dx} \text{ in distribution if } n \rightarrow \infty. \tag{158}$$

The t value $\hat{t}_\alpha$ of $\hat{\alpha}$ is defined by

$$\hat{t}_\alpha = \frac{\sqrt{n}\,\hat{\alpha}.\sqrt{(1/n)\Sigma_{t=1}^{n} Y_{t-1}^2}}{\hat{\sigma}}, \tag{159}$$

which can be rewritten as

$$\hat{t}_\alpha = \frac{n\,\hat{\alpha}.\sqrt{(1/n^2)\Sigma_{t=1}^{n} Y_{t-1}^2}}{\hat{\sigma}} = \frac{n\,\hat{\alpha}.\sqrt{\int_0^1 \left(W_n(x)\right)^2 dx}}{\hat{\sigma}}, \tag{160}$$

where

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{t=1}^{n}\left(\Delta Y_t - \hat{\alpha} Y_{t-1}\right)^2. \tag{161}$$

Under the null hypothesis, and using the previous results, it follows that

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n-1}\sum_{t=1}^{n}\left(U_t - \hat{\alpha} Y_{t-1}\right)^2 = \frac{1}{n-1}\sum_{t=1}^{n} U_t^2 - \hat{\alpha}^2 \frac{1}{n-1}\sum_{t=1}^{n} Y_{t-1}^2 \\ &= \frac{n}{n-1}\left((1/n)\Sigma_{t=1}^{n} U_t^2\right) - \frac{1}{n-1}(n.\hat{\alpha})^2 (1/n^2)\Sigma_{t=1}^{n} Y_{t-1}^2 \\ &= \frac{n}{n-1}\left(1 + r_n\right) - \frac{1}{n-1}(n.\hat{\alpha})^2 \int_0^1 \left(W_n(x)\right)^2 dx \to 1 \text{ in probability if } n \to \infty. \end{aligned} \tag{162}$$

hence

$$\hat{t}_\alpha \to \frac{\frac{1}{2}\left(\left(W(1)\right)^2 - 1\right)}{\sqrt{\int_0^1 \left(W(x)\right)^2 dx}} \text{ in distribution if } n \to \infty. \tag{163}$$

The density function of the limiting random variable involved is displayed in Figure 9.

A.6 *Motivation for the Breitung test*

In order to explain the Breitung unit root test, suppose first that

$$Y_t = Y_{t-1} + U_t, \text{ where } U_t \sim i.i.d.\ N(0,1) \text{ and } Y_t = 0 \text{ for } t \leq 0, \tag{164}$$

so that

$$Y_t = U_1 + U_2 + \ldots U_t = \Sigma_{j=1}^{t} U_j = \sqrt{n}.W_n(t/n), \tag{165}$$

where again $W_n(x)$ is defined by (151). Then similar to (152),

$$(1/n^2)\Sigma_{t=1}^n Y_t^2 \;=\; (1/n^2)\Sigma_{t=0}^n Y_t^2 \;=\; (1/n).\Sigma_{t=0}^n \big(W_n(t/n)\big)^2 \;=\; W_n(1)/n \;+\; \int_0^1 \big(W_n(x)\big)^2$$
$$\approx\; \int_0^1 \big(W_n(x)\big)^2, \tag{166}$$

where the latter approximation follows from the fact that $W_n(1)/n \sim N(0,1/n^2) \rightarrow 0$ if $n \rightarrow \infty$. Moreover,

$$S_t \;=\; \Sigma_{j=1}^t Y_j \;=\; \sqrt{n}\,.\,\Sigma_{j=1}^t W_n(j/n) \;=\; \sqrt{n}\,.\,\Sigma_{j=1}^t \int_j^{j+1} W_n(z/n)dz$$
$$=\; \sqrt{n}\,.\int_1^{t+1} W_n(z/n)dz \;=\; n\sqrt{n}\,.\int_1^{t+1} W_n(z/n)d(z/n) \;=\; n\sqrt{n}\,.\int_0^{t+1} W_n(z/n)d(z/n) \tag{167}$$
$$=\; n\sqrt{n}\,.\int_0^{(t+1)/n} W_n(x)dx.$$

so that

$$(1/n^4)\Sigma_{t=1}^n S_t^2 \;=\; (1/n)\Sigma_{t=1}^n \left(\int_0^{t/n+1/n} W_n(x)dx\right)^2 \;\approx\; \int_0^n \left(\int_0^{z/n+1/n} W_n(x)dx\right)^2 d(z/n)$$
$$=\; \int_0^1 \left(\int_0^{y+1/n} W_n(x)dx\right)^2 dy \;\approx\; \int_0^1 \left(\int_0^{y} W_n(x)dx\right)^2 dy$$

Thus, with B_n defined by (100),

$$B_n/n \;=\; \frac{\Sigma_{t=1}^n S_t^2}{n^2.\Sigma_{t=1}^n Y_t^2} \;=\; \frac{(1/n^4)\,\Sigma_{t=1}^n S_t^2}{(1/n^2)\,\Sigma_{t=1}^n Y_t^2} \;\approx\; \frac{\int_0^1 \left(\int_0^y W_n(x)dx\right)^2 dy}{\int_0^1 \big(W_n(x)\big)^2} \;\rightarrow\; \frac{\int_0^1 \left(\int_0^y W(x)dx\right)^2 dy}{\int_0^1 \big(W(x)\big)^2} \tag{169}$$

in distribution if $n \rightarrow \infty$.

On the other hand, if

$$Y_t \;=\; U_t,\; where\;\; U_t \sim i.i.d.\; N(0,1), \tag{170}$$

then $S_t = \sqrt{n}\,W_n(t/n)$, hence

$$B_n = \frac{\sum_{t=1}^{n} S_t^2}{n.\sum_{t=1}^{n} Y_t^2} = \frac{n\sum_{t=1}^{n}\left(W_n(t/n)\right)^2}{n\sum_{t=1}^{n} U_t^2} = \frac{(1/n)\sum_{t=1}^{n}\left(W_n(t/n)\right)^2}{(1/n)\sum_{t=1}^{n} U_t^2} \approx \int_0^1 \left(W_n(x)\right)^2 dx$$
$$\to \int_0^1 \left(W(x)\right)^2 dx \qquad (171)$$

in distribution if $n \to \infty$. Consequently, $B_n/n \to 0$ in probability if $n \to \infty$.