ARMA MODELS

Herman J. Bierens

Pennsylvania State University

February 23, 2009

1. *Introduction*

Given a covariance stationary process Y_t with vanishing memory[1] and expectation $\mu = E[Y_t]$, the Wold decomposition states that

$$Y_t - \mu = \sum_{j=0}^{\infty}\alpha_j U_{t-j}, \text{ with } \alpha_0 = 1, \sum_{j=0}^{\infty}\alpha_j^2 < \infty, \tag{1}$$

where U_t is an uncorrelated zero-mean covariance stationary process. If in addition the process Y_t is Gaussian then the U_t's are i.i.d. $N(0,\sigma_u^2)$ [Why?]. If so then Y_t is strictly stationary [Why?]. In the sequel I will assume that (in the non-seasonal case) Y_t is covariance stationary and Gaussian, with a vanishing memory, so that the U_t's are i.i.d. $N(0,\sigma_u^2)$.

To approximate the process Y_t by a process that only involves a finite number of parameters, denote $X_t = Y_t - \mu$. Next, project X_t on $X_{t-1}, X_{t-2}, \ldots., X_{t-p}$ for some $p \geq 1$. This projection takes the form $\hat{X}_t = \sum_{j=1}^{p}\beta_j X_{t-j}$. Then we can write

$$X_t = \sum_{k=1}^{p}\beta_k X_{t-k} + V_t, \tag{2}$$

where

$$\begin{aligned} V_t &= X_t - \sum_{k=1}^{p}\beta_k X_{t-k} = U_t + \sum_{j=1}^{\infty}\alpha_j U_{t-j} - \sum_{k=1}^{p}\beta_k\left(U_{t-k} + \sum_{j=1}^{\infty}\alpha_j U_{t-k-j}\right) \\ &= U_t - \sum_{m=1}^{\infty}\theta_m U_{t-m}, \text{ say.} \end{aligned} \tag{3}$$

Since the θ_m's are functions of the α_j's for $j \geq m$ and $\sum_{j=0}^{\infty}\alpha_j^2 < \infty$, it follows that $\sum_{m=1}^{\infty}\theta_m^2 < \infty$. Consequently, for arbitrary small $\varepsilon > 0$ we can find a q such that

$$\mathrm{var}\left(V_t - \left(U_t - \sum_{m=1}^{q}\theta_m U_{t-m}\right)\right) = \sum_{m=q+1}^{\infty}\theta_m^2 < \varepsilon.$$

This motivates to specify V_t in (2) as $U_t - \sum_{m=1}^{q}\theta_m U_{t-m}$, which gives rise to the ARMA(p,q) model

$$X_t = \sum_{k=1}^{p}\beta_k X_{t-k} + U_t - \sum_{m=1}^{q}\theta_m U_{t-m}. \tag{4}$$

Substituting $X_t = Y_t - \mu$ in (4) then yields

---

[1] See Bierens (2004), Chapter 7.

$$Y_t = \beta_0 + \sum_{k=1}^{p}\beta_k Y_{t-k} + U_t - \sum_{m=1}^{q}\theta_m U_{t-m}, \quad (5)$$

where $\beta_0 = \left(1 - \sum_{k=1}^{p}\beta_k\right)\mu$. This is the general ARMA(p,q) model.

All stationary time series models are of the form (5) or are special cases of (5). In particular, the ARMA(p,0) case is known as the autoregressive model of order p, shortly an $AR(p)$ model:

$$Y_t = \beta_0 + \sum_{j=1}^{p}\beta_j Y_{t-j} + U_t, \quad (6)$$

and the ARMA(0,q) case is known as the moving average model of order q, shortly an MA(q) model:

$$Y_t = \beta_0 + U_t - \sum_{m=1}^{q}\theta_m U_{t-m}. \quad (7)$$

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

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

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + U_t. \quad (8)$$

I will first show that a necessary condition for the stationarity of the process (8) is that

$$|\beta_1| < 1. \quad (9)$$

By repeated backwards substitution of (8) we can write

$$Y_t = \beta_0\sum_{k=0}^{m-1}\beta_1^k + \sum_{k=0}^{m-1}\beta_1^k U_{t-k} + \beta_1^m Y_{t-m}. \quad (10)$$

Since Y_t is covariance stationary, $E[Y_t] = \mu$ and $E[(Y_t-\mu)(Y_{t-m}-\mu)] = \gamma(m)$ for all t. Then it follows from (10) that

$$\mu = \beta_0\sum_{k=0}^{m-1}\beta_1^k + \sum_{k=0}^{m-1}\beta_1^k E[U_{t-k}] + \beta_1^m\mu = \beta_0\sum_{k=0}^{m-1}\beta_1^k + \beta_1^m\mu. \quad (11)$$

Next, subtract (11) from (10),

$$Y_t - \mu = \sum_{k=0}^{m-1}\beta_1^k U_{t-k} + \beta_1^m(Y_{t-m} - \mu), \quad (12)$$

take the square of both sides,

$$(Y_t - \mu)^2 = \left(\sum_{k=0}^{m-1}\beta_1^k U_{t-k}\right)^2 + 2\beta_1^m\left(\sum_{k=0}^{m-1}\beta_1^k U_{t-k}\right)(Y_{t-m} - \mu) + \beta_1^{2m}(Y_{t-m} - \mu)^2 \quad (13)$$

and take expectations,

$$E[(Y_t - \mu)^2] = E\left[\left(\sum_{k=0}^{m-1}\beta_1^k U_{t-k}\right)^2\right] + 2\beta_1^m\sum_{k=0}^{m-1}\beta_1^k E[U_{t-k}(Y_{t-m} - \mu)] + \beta_1^{2m}E(Y_{t-m} - \mu)^2]$$
$$= \sigma_u^2\sum_{k=0}^{m-1}\beta_1^{2k} + \beta_1^{2m}E(Y_{t-m} - \mu)^2]. \tag{14}$$

The last equality in (14) follows from the Wold decomposition (1). Because $\gamma(0) = E[(Y_t - \mu)^2] = E[(Y_{t-m} - \mu)^2]$, (14) reads:

$$\gamma(0) = \sigma^2\Sigma_{k=0}^{m-1}\beta_1^{2k} + \beta_1^{2m}\gamma(0). \tag{15}$$

However, if $|\beta_1| \geq 1$ then the right-hand side of (15) converges to ∞ if $m \to \infty$, which contradicts the condition that $\gamma(0) < \infty$. On the other hand, if $|\beta_1| < 1$ then $\beta_1^{2m}\gamma(0) \to 0$ as $m \to \infty$, hence it follows from (10) by letting $m \to \infty$ that

$$Y_t = \beta_0\sum_{k=0}^{\infty}\beta_1^k + \sum_{k=0}^{\infty}\beta_1^k U_{t-k} = \frac{\beta_0}{1-\beta_1} + \sum_{k=0}^{\infty}\beta_1^k U_{t-k}, \tag{16}$$

which is the Wold decomposition (1) :

$$\mu = \frac{\beta_0}{1-\beta_1},\ \alpha_k = \beta_1^k,\ k = 0,1,2,3,...... \tag{17}$$

The expression at the right-hand side of (16) is also called the Moving Average (MA) representation of a covariance stationary time series. From this expression we can derive the covariance function of the AR(1) process, as follows:

$$\begin{aligned}
\gamma(m) &= E[(Y_t - \mu)(Y_{t-m} - \mu)] = E\left[\left(\Sigma_{k=0}^{\infty}\beta_1^k U_{t-k}\right)\left(\Sigma_{k=0}^{\infty}\beta_1^k U_{t-m-k}\right)\right] \\
&= E\left[\left(\Sigma_{k=0}^{m-1}\beta_1^k U_{t-k} + \Sigma_{k=m}^{\infty}\beta_1^k U_{t-k}\right)\left(\Sigma_{k=0}^{\infty}\beta_1^k U_{t-m-k}\right)\right] \\
&= E\left[\left(\Sigma_{k=0}^{m-1}\beta_1^k U_{t-k} + \beta_1^m\Sigma_{k=0}^{\infty}\beta_1^k U_{t-m-k}\right)\left(\Sigma_{k=0}^{\infty}\beta_1^k U_{t-m-k}\right)\right] \\
&= E\left[\left(\Sigma_{k=0}^{m-1}\beta_1^k U_{t-k}\right)\left(\Sigma_{k=0}^{\infty}\beta_1^k U_{t-m-k}\right)\right] + \beta_1^m E\left[\left(\Sigma_{k=0}^{\infty}\beta_1^k U_{t-m-k}\right)^2\right] \\
&= \beta_1^m E\left[\left(\Sigma_{k=0}^{\infty}\beta_1^k U_{t-m-k}\right)^2\right] = \beta_1^m\sigma_u^2\Sigma_{k=0}^{\infty}\beta_1^{2k} = \sigma_u^2\beta_1^m/(1-\beta_1^2),\ m = 0,1,2,3,......
\end{aligned} \tag{18}$$

3. *Lag operators*

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

$$L^m.Y_t \overset{def.}{=} Y_{t-m},\ m = 0,1,2,3,..... \tag{19}$$

Using the lag operator, the AR(1) model (8) can be written as

$$(1 - \beta_1 L)Y_t = \beta_0 + U_t. \tag{20}$$

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

$$\sum_{k=0}^{\infty} z^k = \frac{1}{1-z},\ \text{provided that}\ |z| < 1, \tag{21}$$

which follows from the equalities $\sum_{k=0}^{\infty} z^k = 1 + \sum_{k=1}^{\infty} z^k = 1 + \sum_{k=0}^{\infty} z^{k+1} = 1 + z.\sum_{k=0}^{\infty} z^k$. Now suppose that we may treat $\beta_1 L$ as the variable z in (21). If so, it follows from (21) that

$$\frac{1}{1-\beta_1 L} = \sum_{k=0}^{\infty} (\beta_1 L)^k = \sum_{k=0}^{\infty} \beta_1^k L^k. \tag{22}$$

Applying this lag function to both sides of (20) then yields

$$\begin{aligned} Y_t &= \frac{1}{1 - \beta_1 L}(1 - \beta_1 L)Y_t = \sum_{k=0}^{\infty} \beta_1^k L^k \beta_0 + \sum_{k=0}^{\infty} \beta_1^k L^k U_t = \sum_{k=0}^{\infty} \beta_1^k \beta_0 + \sum_{k=0}^{\infty} \beta_1^k U_{t-k} \\ &= \frac{\beta_0}{1-\beta_1} + \sum_{k=0}^{\infty} \beta_1^k U_{t-k}, \end{aligned} \tag{23}$$

which is exactly the moving average representation (16). Note that in the second equality in (23) 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 (22) holds if $|\beta_1| < 1$:

Proposition 1. *The lag function* $\sum_{k=0}^{\infty} \beta^k L^k$ *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* $|\beta| < 1$.

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

Consider the AR(2) process

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + U_t, \tag{24}$$

where the errors U_t have the same properties as before. Similar to (20) we can write this model in lag-polynomial form as

$$(1 - \beta_1 L - \beta_2 L^2) Y_t = \beta_0 + U_t. \tag{25}$$

We can always write

$$1 - \beta_1 L - \beta_2 L^2 = (1 - \alpha_1 L)(1 - \alpha_2 L), \tag{26}$$

by solving the equations $\alpha_1 + \alpha_2 = \beta_1$, $\alpha_1\alpha_2 = -\beta_2$,[2] so that (25) can be written as

$$(1 - \alpha_1 L)(1 - \alpha_2 L) Y_t = \beta_0 + U_t. \tag{27}$$

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

$$\begin{aligned} Y_t &= \frac{1}{(1-\alpha_1 L)(1-\alpha_2 L)}\left(\beta_0 + U_t\right) = \left(\sum_{k=0}^{\infty}\alpha_1^k L^k\right)\left(\sum_{m=0}^{\infty}\alpha_2^m L^m\right)\beta_0 + \left(\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\alpha_1^k\alpha_2^m L^{k+m}\right)U_t \\ &= \left(\sum_{k=0}^{\infty}\alpha_1^k\right)\left(\sum_{m=0}^{\infty}\alpha_2^m\right)\beta_0 + \sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\alpha_1^k\alpha_2^m U_{t-k-m} \qquad (28) \\ &= \frac{\beta_0}{(1-\alpha_1)(1-\alpha_2)} + \sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\alpha_1^k\alpha_2^m U_{t-k-m} = \frac{\beta_0}{1-\beta_1-\beta_2} + \sum_{j=0}^{\infty}\left(\sum_{m=0}^{j}\alpha_1^{j-m}\alpha_2^m\right)U_{t-j}. \end{aligned}$$

Note that Y_t in (28) has expectation $\mu = \beta_0/(1-\beta_1-\beta_2)$ and variance $\sigma_u^2\sum_{j=0}^{\infty}\left(\sum_{m=0}^{j}\alpha_1^{j-m}\alpha_2^m\right)^2$.

Consequently, the necessary conditions for the covariance stationarity of the AR(2) process (24) 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:

---

[2] Although the solutions involved may be complex valued. If $\beta_2 \neq 0$ then the solutions are $\alpha_1 = 0.5\beta_1 \pm 0.5\sqrt{\beta_1^2 + 4\beta_2}$, $\alpha_2 = -\beta_2/\alpha_1$. If $\beta_1^2 + 4\beta_2 < 0$ then α_1 and α_2 are complex conjugate: $\alpha_1 = 0.5\beta_1 + i.0.5\sqrt{-\beta_1^2-4\beta_2}$, $\alpha_2 = 0.5\beta_1 - i.0.5\sqrt{-\beta_1^2-4\beta_2}$.

Proposition 2. *The necessary conditions for the covariance stationarity of the AR*(*p*) *process* (6) *are that the errors* U_t *are covariance stationary and the solutions* $z_1,\ldots,z_p$ *of the equation* $0 = 1 - \beta_1 z - \beta_2 z^2 - \ldots.. - \beta_p z^p$ *are all greater than one in absolute value*: $|z_j| > 1$ *for j* = 1,...,*p*.

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

5.1 *The partial autocorrelation function*.

If the correct order of an AR process is p_0 but you estimate the AR(*p*) model (6) with $p > p_0$ by OLS, then the OLS estimates of the coefficients $\beta_{p_0+1},\ldots.,\beta_p$ will be small and insignificant, because these coefficients are then all zero: $\beta_{p_0+1} = \beta_{p_0+2} = \ldots. = \beta_p = 0$. This suggests the following procedure for selecting *p*. Estimate the AR(*p*) model for *p* = 1,2,....,$\bar{p}$, where $\bar{p} > p_0$.

$$\hat{Y}_t = \hat{\beta}_{p,0} + \hat{\beta}_{p,1}Y_{t-1} + \hat{\beta}_{p,2}Y_{t-2} + \ldots.. + \hat{\beta}_{p,p}Y_{t-p}, \tag{29}$$

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

$$PAC(p) \overset{def.}{=} \hat{\beta}_{p,p},\ p = 1,2,3,\ldots,\ PAC(0) = 1. \tag{30}$$

For example, suppose that an AR(*p*) model $Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots.. + \beta_p Y_{t-p} + U_t$ 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 (see Bierens 2008a) 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

$$Y_t = 1.144123Y_{t-1} - 0.5Y_{t-2} + U_t,\ U_t \ i.i.d.\ N(0,1),\ t = 1,...,500, \tag{31}$$

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 $p \geq 1$.

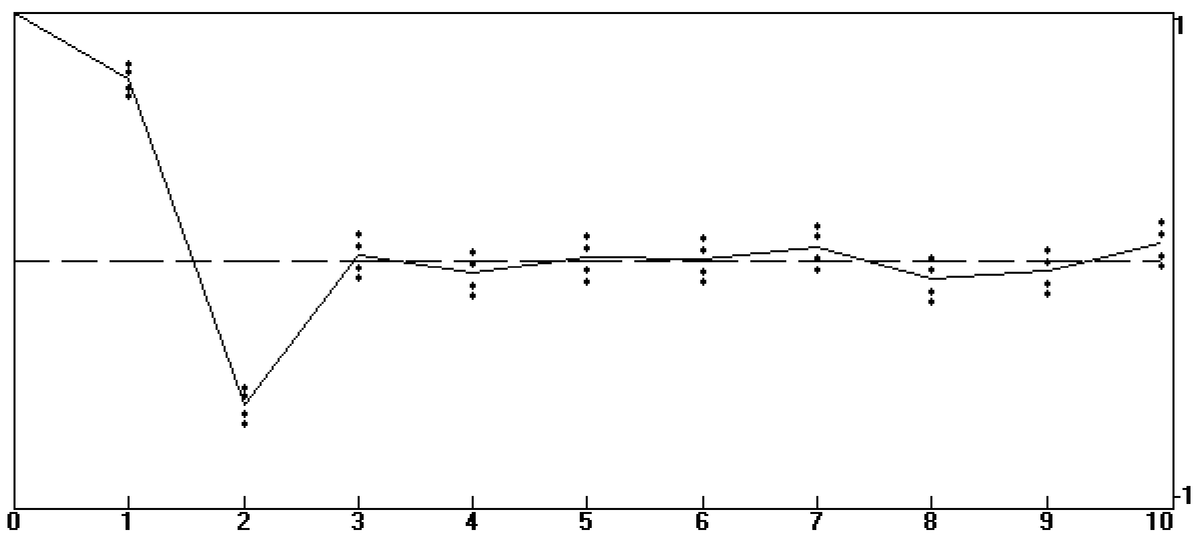

Figure 1: Partial autocorrelation function, PAC(m), of the AR(2) process (31)

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 $p_0 = 2$.

5.2 *Information criteria*

An alternative approach to determine the order p of the AR(p) model (6) 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] $c_n^{AR}(p) = \ln(\hat{\sigma}_p^2) + (1+p)\ln(n)/n,$

where n is the effective sample size of the regression (6) (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:

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

where $\overline{p} > p_0$, 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 $p > p_0$,

$$n\left(\ln(\hat{\sigma}_{p_0}^2) - \ln(\hat{\sigma}_p^2)\right) \to_d \chi^2_{p-p_0}, \tag{32}$$

where $\to_d$ indicates convergence in distribution. The result (32) is due to the likelihood-ratio test. Then in the Akaike case,

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

where $X_{p-p_0} \sim \chi^2_{p-p_0}$, hence $\lim_{n\to\infty}P[c_n^{AR}(p_0) > c_n^{AR}(p)] = P[X_{p-p_0} > 2(p-p_0)] > 0$.

---

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

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 (32) implies $\text{plim}_{n\to\infty}n(\ln(\hat{\sigma}^2_{p_0}) - \ln(\hat{\sigma}^2_p))/\ln(\ln(n)) = 0$ and $\text{plim}_{n\to\infty}n(\ln(\hat{\sigma}^2_{p_0}) - \ln(\hat{\sigma}^2_p))/\ln(n) = 0$ 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, $\lim_{n\to\infty}P[\hat{p} = p_0] = 1$. 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 $p_0 = 2$.

5.3 *The Wald test*

A third way to determine the correct order p_0 of the AR(p) model (6) 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 (6) 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 so 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 $p_0 = 2$.

6. *Moving average processes*

Recall that a moving average process of order q, denoted by MA(q), takes the form

$$Y_t = \mu + U_t - \theta_1 U_{t-1} - .... - \theta_q U_{t-q}, \tag{33}$$

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

$$Y_t = \mu + U_t - \theta U_{t-1}. \tag{34}$$

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

$$Y_t = \mu + (1 - \theta L)U_t. \tag{35}$$

Now it follows from Proposition 1 and (35) that if $|\theta| < 1$ then

$$\sum_{j=0}^{\infty}\theta^j L^j Y_t = (1 - \theta L)^{-1} Y_t = (1 - \theta L)^{-1}\mu + U_t = \frac{\mu}{1-\theta} + U_t, \tag{36}$$

hence

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + ... + U_t = \beta_0 + \sum_{j=1}^{\infty}\beta_j Y_{t-j} + U_t,$$
$$\textit{where } \beta_0 = \mu/(1-\theta),\ \beta_i = -\theta^j \textit{ for } j = 1,2,3,..... \tag{37}$$

More generally we have:

Proposition 3. *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* (33) *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* $1 - \sum_{j=1}^{\infty}\beta_j L^j = 1/(1-\theta_1 L-\theta_2 L^2-.....-\theta_q L^q)$.

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

7.1 *The (regular) autocorrelation function*

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

$$\rho(m) = \frac{cov(Y_t, Y_{t-m})}{var(Y_t)}, \; m = 0,1,2,...... \tag{38}$$

It is trivial that:

Proposition 4. *For an MA*(q) process, $\rho(m) = 0$ *for* $m > q$, *and* $\rho(q) \neq 0$.

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

$$\hat{\rho}(m) = \frac{(1/(n-m))\sum_{t=m+1}^{n}(Y_t-\bar{Y})(Y_{t-m}-\bar{Y})}{\sqrt{(1/(n-m))\sum_{t=1}^{n-m}(Y_t-\bar{Y})^2}\sqrt{(1/(n-m))\sum_{t=m+1}^{n}(Y_{t-m}-\bar{Y})^2}}, \; m = 0,1,2,...... \tag{39}$$

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 (8) satisfying condition (9). Then it follows from (18) 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 $m > 1$, as is demonstrated in Figure 2.. The same applies to more general AR processes.

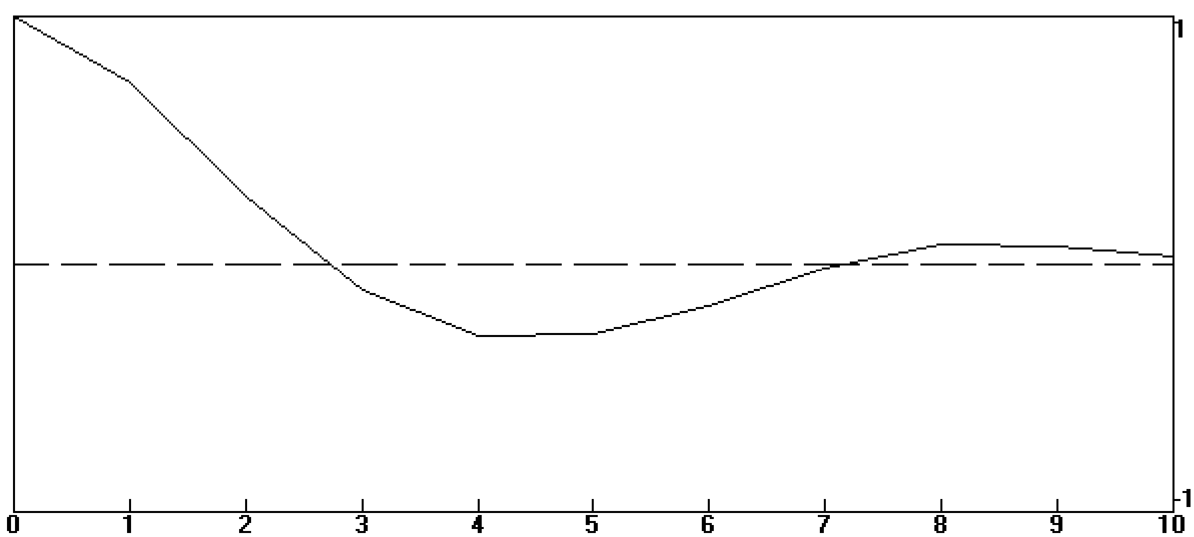

Figure 2: Estimated autocorrelation function $\hat{\rho}(m)$ of the AR(2) process (31).

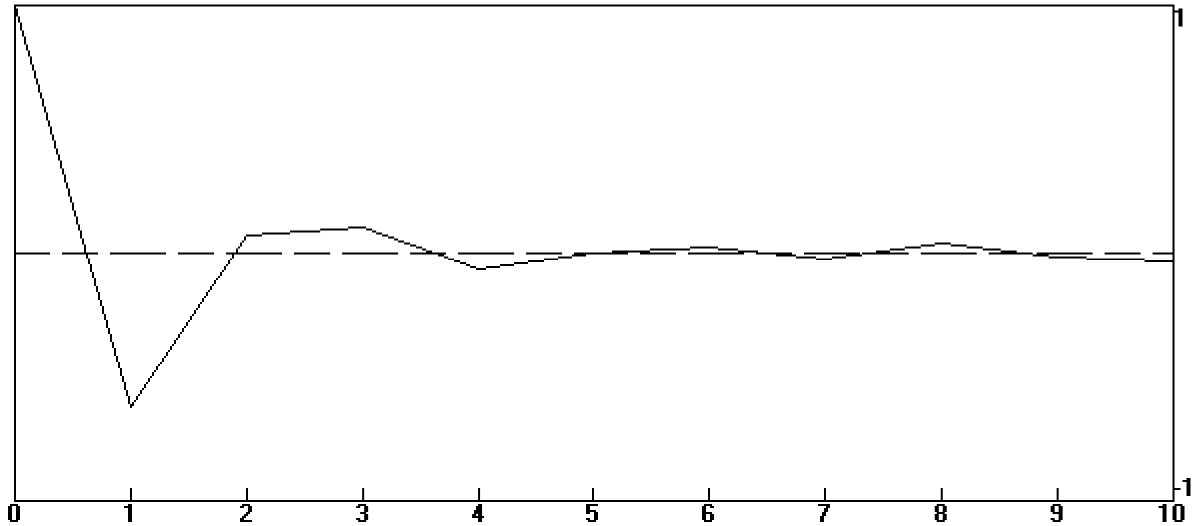

Figure 3: Estimated autocorrelation function $\hat{\rho}(m)$ of the MA(2) process (40)

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

$$Y_t = U_t - 1.4U_{t-1} + 0.5U_{t-2},\ U_t \sim i.i.d.\ N(0,1),\ t = 1,2,....,500 \quad (40)$$

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

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 $\bar{q} = 4$.

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. For example, the PAC(m) of the MA(2) process (40) does not drop sharply to zero for $m > 2$, as is demonstrated in Figure 4.

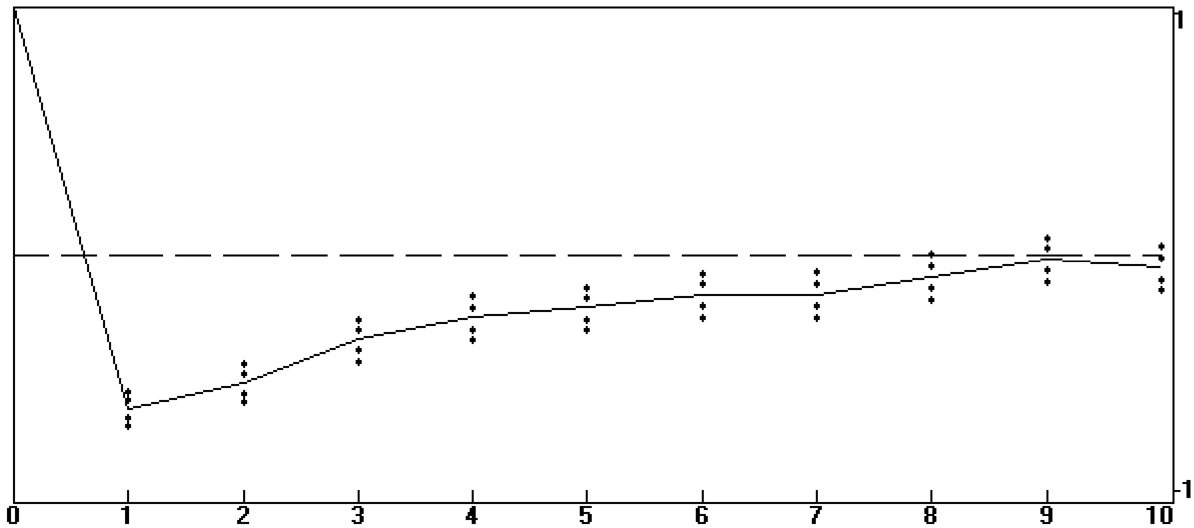

Figure 4: Partial autocorrelation function, PAC(*m*), of the MA(2) process (40)

7.2 *Information criteria*

The three information criteria, Akaike, Hannan-Quinn and Schwarz also apply to MA processes. Therefore, estimate the MA(q) model (33) 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.

7.3 *Wald test*

As a double check, estimate the MA model (33) 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 (33) 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: $\mu = 0$, $\theta_1 = 1.4$, $\theta_2 = -0.5$.

8. *ARMA models*

8.1 *Invertibility conditions*

Denote

$$\varphi_p(L) = 1 - \beta_1 L - \beta_2 L^2 - ..... - \beta_p L^p, \\ \psi_q(L) = 1 - \theta_1 L - \theta_2 L^2 - .... - \theta_q L^q, \qquad (41)$$

Then the ARMA(*p*,*q*) model (5) can be written more compactly as

$$\varphi_p(L)Y_t = \beta_0 + \psi_q(L)U_t. \tag{42}$$

Proposition 5. *A necessary condition for the stationarity of ARMA*(p,q) *process* (42) *is that* $\varphi_p(z_1) = 0$ *implies* $|z_1| > 1$. *If so, it has the MA*(∞) *representation* (*or equivalently, the Wold decomposition*) $Y_t = \beta_0/\varphi_p(1) + \varphi_p(L)^{-1}\psi_q(L)U_t$. *If in addition* $\psi_q(z_2) = 0$ *implies* $|z_2| > 1$ *then the ARMA*(p,q) *process* (42) *has an AR*(∞) *representation*: $\psi_q(L)^{-1}\varphi_p(L)Y_t = \beta_0/\psi_q(1) + U_t$.

I will demonstrate Proposition 5 for the case $p = q = 1$:

$$(1-\beta_1 L)Y_t = \beta_0 + (1-\theta_1 L)U_t. \tag{43}$$

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 1 that $\psi_1(L)^{-1} = (1-\theta_1 L)^{-1} = \sum_{j=0}^{\infty}\theta_1^j L^j$, hence

$$\begin{aligned}\sum_{j=0}^{\infty}\theta_1^j L^j(1-\beta_1 L)Y_t &= \psi_1(L)^{-1}(1-\beta_1 L)Y_t = \sum_{j=0}^{\infty}\theta_1^j L^j\beta_0 + \sum_{j=0}^{\infty}\theta_1^j L^j(1-\theta_1 L)U_t \\ &= \sum_{j=0}^{\infty}\theta_1^j\beta_0 + U_t = \beta_0/(1-\theta_1) + U_t\end{aligned} \tag{44}$$

and

$$\begin{aligned}\sum_{j=0}^{\infty}\theta_1^j L^j(1-\beta_1 L)Y_t &= \sum_{j=0}^{\infty}\theta_1^j L^j Y_t - \beta_1\sum_{j=0}^{\infty}\theta_1^j L^{j+1}Y_t \\ = Y_t + \sum_{j=1}^{\infty}\theta_1^j Y_{t-j} - \beta_1\sum_{j=0}^{\infty}\theta_1^j Y_{t-j-1} &= Y_t + \theta_1\sum_{j=0}^{\infty}\theta_1^j Y_{t-1-j} - \beta_1\sum_{j=0}^{\infty}\theta_1^j Y_{t-1-j} \\ &= Y_t - (\beta_1-\theta_1)\sum_{j=0}^{\infty}\theta_1^j Y_{t-1-j}.\end{aligned} \tag{45}$$

Combining these results yields

$$Y_t = \beta_0/(1-\theta_1) + (\beta_1-\theta_1)\sum_{j=0}^{\infty}\theta_1^j Y_{t-1-j} + U_t, \tag{46}$$

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

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

$$\begin{aligned} Y_t &= (1-\beta_1 L)^{-1}(1-\beta_1 L)Y_t = (1-\beta_1 L)^{-1}\beta_0 + (1-\beta_1 L)^{-1}(1-\theta_1 L)U_t \\ &= \beta_0/(1-\beta_1) + \textstyle\sum_{j=0}^{\infty}\beta_1^j L^j(1-\theta_1 L)U_t \\ &= \beta_0/(1-\beta_1) + \textstyle\sum_{j=0}^{\infty}\beta_1^j L^j U_t - \theta_1\sum_{j=0}^{\infty}\beta_1^j L^{j+1}U_t \\ &= \beta_0/(1-\beta_1) + U_t - (\theta_1-\beta_1)\textstyle\sum_{j=0}^{\infty}\beta_1^j U_{t-1-j}, \end{aligned} \tag{47}$$

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

8.2 *Common roots*

Observe from (46) and (47) 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 6. *Let the conditions in Proposition 5 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* (42) *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 5. The ARMA*(p,q) *process* (42) *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 (42) because $\beta_0 = (1-\delta)\beta^*_0$ for arbitrary δ. For example, let for $p = q = 2$,

$$\begin{aligned} \varphi_2(L) &= (1-\delta L)(1-\beta L) = 1-(\delta+\beta)L + \delta.\beta L^2 = 1-\beta_1 L-\beta_2 L^2 \\ \psi_2(L) &= (1-\delta L)(1-\theta L) = 1-(\delta+\theta)L + \delta.\theta L^2 = 1-\theta_1 L-\theta_2 L^2 \end{aligned} \tag{48}$$

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 , $\beta_1 = \delta+\beta$, $\beta_2 = -\delta.\beta$, $\theta_1 = \delta+\theta$, $\theta_2 = -\delta.\theta$ 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 6 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:

$$\text{plim}_{n\to\infty}\hat{\sigma}^2_{p,q} = \text{plim}_{n\to\infty}\hat{\sigma}^2_{p-1,q-1} = \sigma^2. \tag{49}$$

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 (49).

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

$$Y_t = 0.3 + 0.7Y_{t-1} + U_t + 0.5U_{t-1}, \quad U_t \sim i.i.d\ N(0,1), \tag{50}$$

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

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + U_t - \theta_1 U_{t-1} - \theta_2 U_{t-2}. \tag{51}$$

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:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + U_t - \theta_1 U_{t-1}. \tag{52}$$

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.

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

The AR(∞) representation (46) of the ARMA(1,1) process (50) is

$$\begin{aligned} Y_t &= \beta_0/(1-\theta_1) + (\beta_1-\theta_1)\textstyle\sum_{j=0}^{\infty}\theta_1^j Y_{t-1-j} + U_t \\ &= 0.3/(1+0.5) + 1.2\textstyle\sum_{j=0}^{\infty}(-0.5)^j Y_{t-1-j} + U_t \\ &= 0.2 + 1.2\textstyle\sum_{j=0}^{\infty}(-0.5)^j Y_{t-1-j} + U_t \\ &= 0.2 + 1.2Y_{t-1} - 0.6Y_{t-2} + 0.3Y_{t-3} - 0.15Y_{t-4} + 0.075Y_{t-5} + \ldots\ldots + U_t \end{aligned} \tag{53}$$

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:

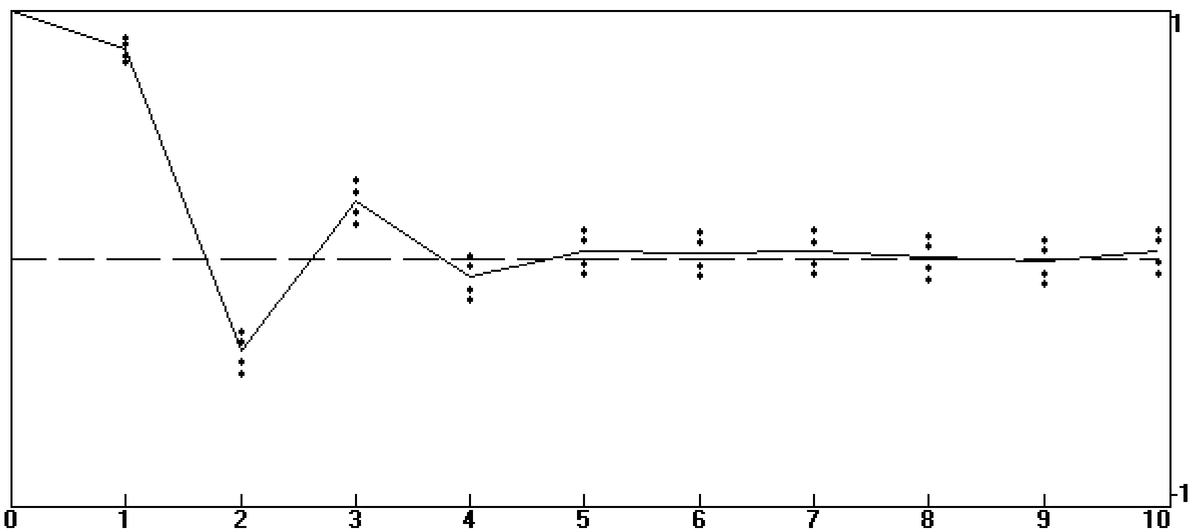

Figure 5: Partial autocorrelation function, PAC(*m*), of the ARMA(1,1) process (50)

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 corroborate this:

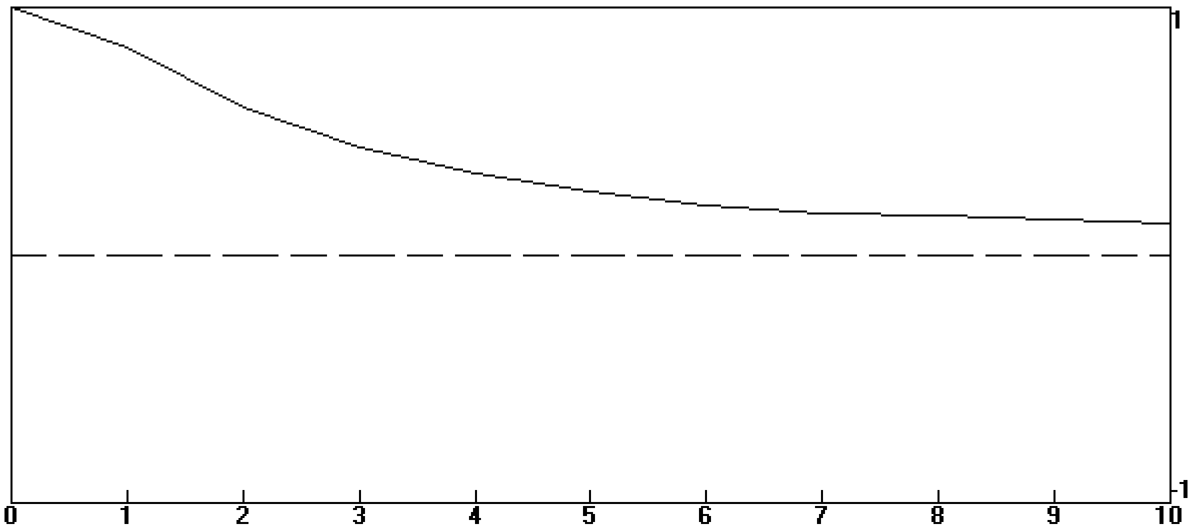

Figure 6: Estimated autocorrelation function $\hat{\rho}(m)$ of the ARMA(1,1) process (50)

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

So how to proceed? My recommendation is the following: Select an upper bound $\bar{p}$ of *p*, for example on the basis of the PAC plot, and select an upper bound $\bar{q}$ of *q*. The latter is a matter of guess work; there is no rule of thumb for this. Then try all ARMA(*p*,*q*) models with $p \le \bar{p}$, $q \le \bar{q}$, and pick the model with the lowest value of one of the information criteria. This can be done automatically in EasyReg, via Menu > Single equation models > ARIMA model selection via information criteria. For example in the case of (50), with $\bar{p} = \bar{q} = 4$, the Hannan-Quinn information criterion takes its smallest value for $p = q = 1$, which are the true values.

8.4 *Forecasting with an ARMA model*

In EasyReg the AR(∞) representation of an ARMA model is used as forecasting scheme, because for covariance stationary Gaussian processes 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

$$\hat{Y}_{n+1} = \beta_0/(1-\theta_1) + (\beta_1-\theta_1)\Sigma_{j=0}^{\infty}\theta_1^j Y_{n-j}, \tag{54}$$

where *n* is the last observed time period. Compare (46). 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 (54) by its sample mean $\bar{Y} = (1/n)\Sigma_{t=1}^{n} Y_t$. Thus, the actual forecast of Y_{n+1} is:

$$\begin{aligned}\tilde{Y}_{n+1|n} &= \hat{\beta}_0/(1-\hat{\theta}_1) + (\hat{\beta}_1-\hat{\theta}_1)\Sigma_{j=0}^{n-1}\hat{\theta}_1^j Y_{n-j} + (\hat{\beta}_1-\hat{\theta}_1)\Sigma_{j=n}^{\infty}\hat{\theta}_1^j \bar{Y} \\ &= \frac{\hat{\beta}_0}{1-\hat{\theta}_1} + (\hat{\beta}_1-\hat{\theta}_1)\Sigma_{j=0}^{n-1}\hat{\theta}_1^j Y_{n-j} + \frac{(\hat{\beta}_1-\hat{\theta}_1)\hat{\theta}_1^n \bar{Y}}{1-\hat{\theta}_1},\end{aligned} \tag{55}$$

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 (55) by n+1, and the unobserved Y_{n+1} by its forecast:

$$\tilde{Y}_{n+2|n} = \hat{\beta}_0/(1-\hat{\theta}_1) + (\hat{\beta}_1-\hat{\theta}_1)\tilde{Y}_{n+1|n} + (\hat{\beta}_1-\hat{\theta}_1)\Sigma_{j=1}^{n}\hat{\theta}_1^j Y_{n+1-j} + \frac{(\hat{\beta}_1-\hat{\theta}_1)\hat{\theta}_1^{n+1}\bar{Y}}{1-\hat{\theta}_1}. \tag{56}$$

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

$$\tilde{Y}_{n+h|n} = \frac{\hat{\beta}_0}{1-\hat{\theta}_1} + (\hat{\beta}_1-\hat{\theta}_1)\Sigma_{j=0}^{h-2}\hat{\theta}_1^j \tilde{Y}_{n+h-1-j|n} + (\hat{\beta}_1-\hat{\theta}_1)\Sigma_{j=h+1}^{n+h-2}\hat{\theta}_1^j Y_{n+h-1-j} + \frac{(\hat{\beta}_1-\hat{\theta}_1)\hat{\theta}_1^{n+h-1}\bar{Y}}{1-\hat{\theta}_1}. \tag{57}$$

Note however, that in this case

$$\begin{aligned}&\lim_{h\to\infty}\tilde{Y}_{n+h|n} = \hat{\beta}_0/(1-\hat{\theta}_1) + (\hat{\beta}_1-\hat{\theta}_1)\lim_{h\to\infty}\Sigma_{j=0}^{h-2}\hat{\theta}_1^j \tilde{Y}_{n+h-1-j|n} \\ &= \hat{\beta}_0/(1-\hat{\theta}_1) + (\hat{\beta}_1-\hat{\theta}_1)\Sigma_{j=0}^{\infty}\hat{\theta}_1^j \lim_{h\to\infty}\tilde{Y}_{n+h|n} = \hat{\beta}_0/(1-\hat{\theta}_1) + \left((\hat{\beta}_1-\hat{\theta}_1)/(1-\hat{\theta}_1)\right)\lim_{h\to\infty}\tilde{Y}_{n+h|n} \\ &= \hat{\beta}_0/(1-\hat{\beta}_1)\end{aligned} \tag{58}$$

which is just the estimate of $\mu = E[Y_t]$. Compare (47). Thus, if we choose the forecast horizon h too large, the recursive forecast $\tilde{Y}_{n+h|n}$ will be close to the expectation $\mu = E[Y_t]$.

The forecast issue will be treated in more detail separately. See Bierens (2008b).

9. *ARMA models for seasonal time series*

9.1 *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

$$E[Y_t] = \mu_0 + \mu_1 Q_{1,t} + \mu_2 Q_{2,t} + \mu_3 Q_{3,t} \tag{59}$$

or

$$E[Y_t] = \mu_1^* Q_{1,t} + \mu_2^* Q_{2,t} + \mu_3^* Q_{3,t} + \mu_4^* Q_{4,t}, \tag{60}$$

where the $Q_{s,t}$ ‘s are seasonal dummy variables:

$$Q_{s,t} = 1 \textit{ if the quarter of } t \textit{ is } s,\ Q_{s,t} = 0 \textit{ if not.} \tag{61}$$

The equivalence of (59) and (60) follows from the fact that $\sum_{s=1}^{4} Q_{s,t} = 1$, so that

$$\begin{aligned} E[Y_t] &= \mu_1^* Q_{1,t} + \mu_2^* Q_{2,t} + \mu_3^* Q_{3,t} + \mu_4^*\left(1 - Q_{1,t} - Q_{2,t} - Q_{3,t}\right) \\ &= \mu_4^* + (\mu_1^* - \mu_4^*)Q_{1,t} + (\mu_2^* - \mu_4^*)Q_{2,t} + (\mu_3^* - \mu_4^*)Q_{3,t}, \end{aligned} \tag{62}$$

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 (60) 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 (62) becomes

$$E[Y_t] = (\mu_0^* + \mu_4^*) + (\mu_1^* - \mu_4^*)Q_{1,t} + (\mu_2^* - \mu_4^*)Q_{2,t} + (\mu_3^* - \mu_4^*)Q_{3,t}, \tag{63}$$

which is also equivalent to (59). Hence

$$\mu_0 = \mu_0^* + \mu_4^*,\ \mu_1 = \mu_1^* - \mu_4^*,\ \mu_2 = \mu_2^* - \mu_4^*,\ \mu_3 = \mu_3^* - \mu_4^*, \tag{64}$$

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

$$Y_t = \mu_0 + \mu_1 Q_{1,t} + \mu_2 Q_{2,t} + \mu_3 Q_{3,t} + X_t \tag{65}$$

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

$$\begin{aligned} \overline{Y}_n &= \mu_0 + \mu_1(1/n)\textstyle\sum_{t=1}^{n} Q_{1,t} + \mu_2(1/n)\sum_{t=1}^{n} Q_{2,t} + \mu_3(1/n)\sum_{t=1}^{n} Q_{3,t} + (1/n)\sum_{t=1}^{n} X_t \\ &\approx \mu_0 + 0.25\mu_1 + 0.25\mu_2 + 0.25\mu_3 \end{aligned} \tag{66}$$

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 $\text{plim}_{n\to\infty}(1/n)\sum_{t=1}^{n}X_t = E[X_t] = 0$ by the law of large numbers. Then it it is not hard to show that there exists constants c_s, $s = 1,2,3,4,$ such that for $n \to \infty$,

$$\frac{1}{n-m}\sum_{t=m+1}^{n}(Y_t - \bar{Y}_n)(Y_{t-m} - \bar{Y}_n) \to \begin{cases} \gamma_x(m) + c_1 \ for\ m = 0,4,8,12,..... \\ \gamma_x(m) + c_2 \ for\ m = 1,5,9,13,..... \\ \gamma_x(m) + c_3 \ for\ m = 2,6,10,14,..... \\ \gamma_x(m) + c_4 \ for\ m = 3,7,11,15,..... \end{cases} \qquad (67)$$

in probability. It follows now from (39) and (67) that the estimated autocorrelation function $\hat{\rho}(m)$ will have spikes at distances of four lags, and will not die out to zero. For example consider the quarterly process

$$Y_t = 1 + 2Q_{1,t} - Q_{2,t} - 2Q_{3,t} + X_t,\ where\ X_t \sim i.i.d.\ N(0,1), \qquad (68)$$

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

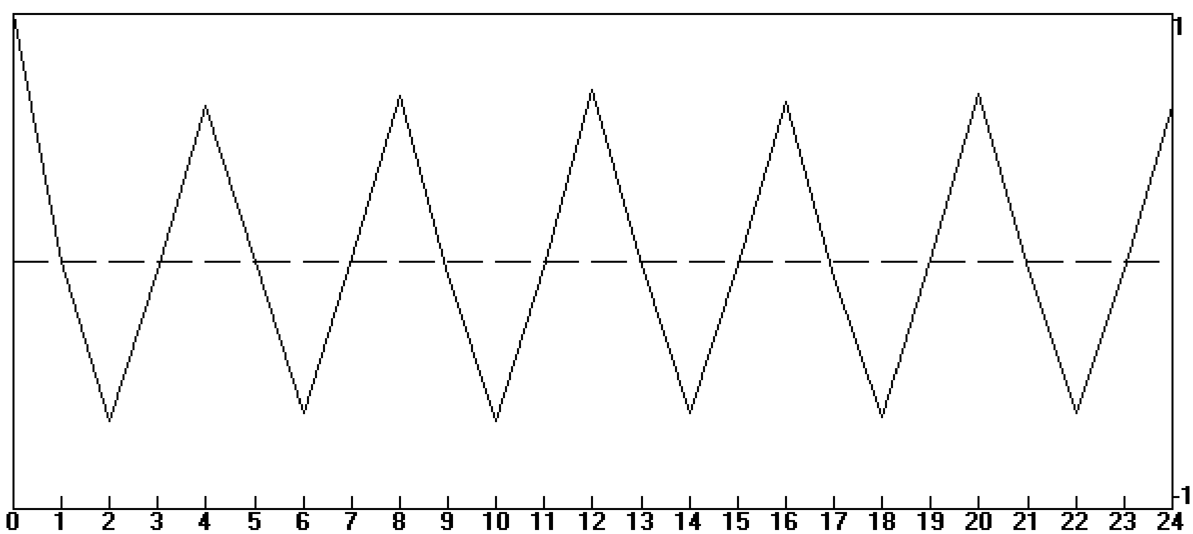

Figure 7: Estimated autocorrelation function $\hat{\rho}(m)$ of quarterly process (68)

9.2 *Seasonal lag polynomials*

Seasonality may also occur in the process X_t in (65) itself. For example, let X_t be a quarterly ARMA process

$$\varphi_p(L)\lambda_r(L^4)X_t = \psi_q(L)\eta_s(L^4)U_t, \qquad (69)$$

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.

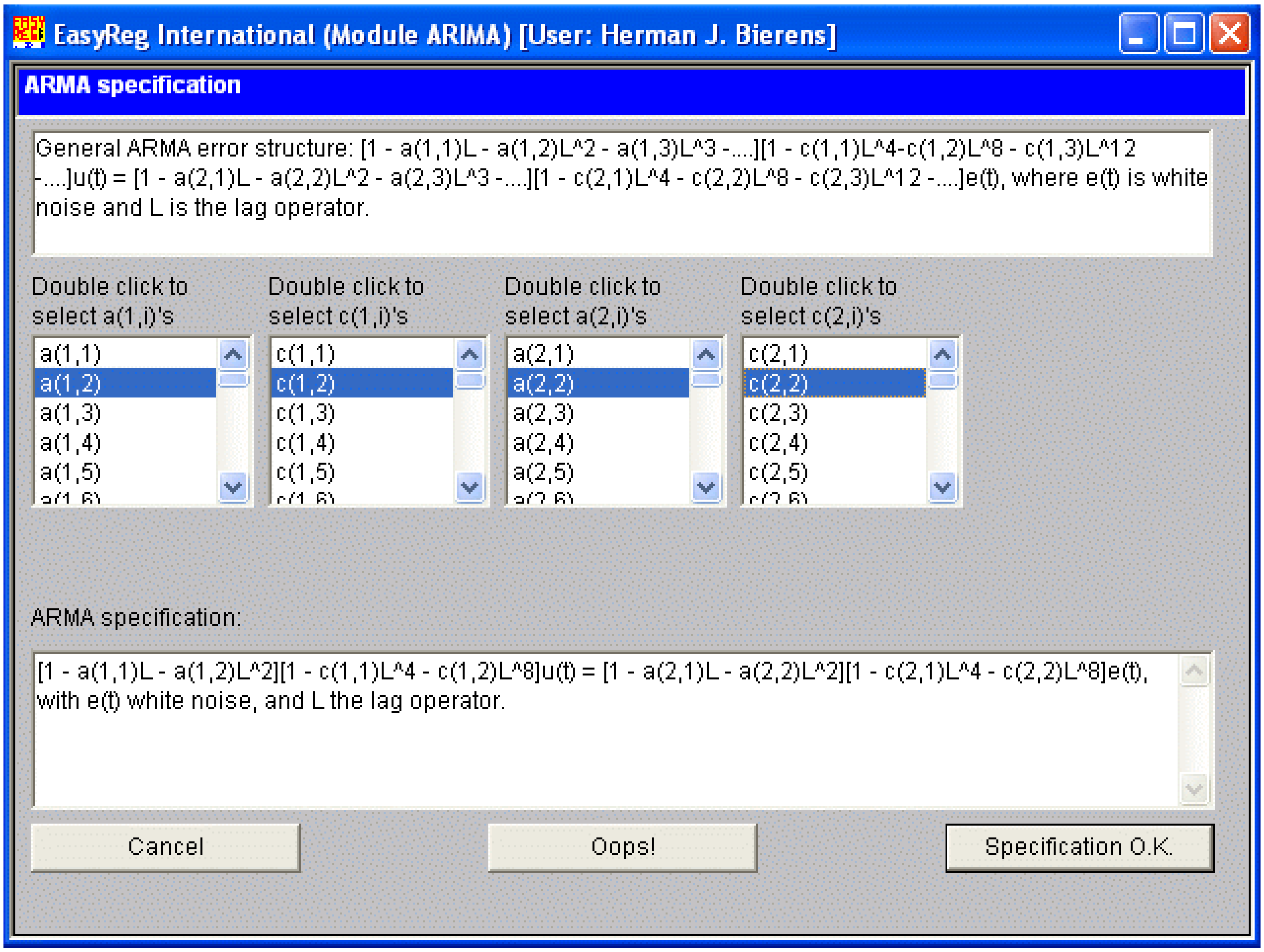

Figure 8: Specification of a seasonal ARMA model in EasyReg.

In EasyReg these polynomials are specified via the window displayed in Figure 8. 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.

References

Akaike, H. (1974): "A New Look at the Statistical Model Identification," *I.E.E.E. Transactions on Automatic Control, AC* 19, 716-723.

Akaike, H. (1976): "Canonical Correlation Analysis of Time Series and the Use of an Information Criterion," in R. K. Mehra and D. G. Lainotis (eds.), *System Identification*: *Advances and Case Studies*, Academic Press, New York, 52-107.

Bierens, H. J. (2004): *Introduction to the Mathematical and Statistical Foundations of Econometrics*, Cambridge University Press, Cambridge, UK.

Bierens, H. J. (2008a): *EasyReg International*, free econometrics software package, downloadable from http://econ.la.psu.edu/~hbierens/EASYREG.HTM

Bierens, H. J. (2008b): *Forecasting*, Lecture note, downloadable from http://econ.la.psu.edu/~hbierens/FORECAST.PDF

Hannan, E. J., and B. G. Quinn (1979): "The Determination of the Order of an Autoregression," *Journal of the Royal Statistical Society, B*, 41, 190-195.

Schwarz, G. (1978): "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461-464.