Economic Forecasting
Lecture 4: The ARMA Model

Richard G. Pierse

1 Introduction

The *ARMA* model was introduced by Box and Jenkins (1976) as a general way of explaining a variable in terms of its own past. The Box and Jenkins methodology has proved popular as a method of producing short-term forecasts.

2 Autoregressive processes

2.1 The first order autoregressive process

Let y_t be defined by the process

$$y_t = \phi y_{t-1} + \varepsilon_t \tag{1}$$

where ε_t is an independently identically distributed random variable of innovations with

$$\mathrm{E}(\varepsilon_t) = 0, \quad \mathrm{var}(\varepsilon_t) = \sigma^2, \quad \mathrm{cov}(\varepsilon_t \varepsilon_{t-s}) = 0, \ s \neq 0.$$

For the process to be covariance stationary we require that the parameter ϕ satisfies the restriction that

$$|\phi| < 1.$$

This process is known as a first order autoregressive or *AR(1)* process. By repeated substitution we can write y_t in terms of the innovations ε_t

$$y_t = \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + \cdots$$

The variable y_t has the properties that

$$\mathrm{E}(y_t) = \mathrm{E}(\varepsilon_t) + \phi\,\mathrm{E}(\varepsilon_{t-1}) + \phi^2\,\mathrm{E}(\varepsilon_{t-2}) + \phi^3\,\mathrm{E}(\varepsilon_{t-3}) + \cdots = 0$$

$$\begin{aligned} Var(y_t) &= \mathrm{E}(\varepsilon_t^2) + \phi^2\,\mathrm{E}(\varepsilon_{t-1}^2) + \phi^4\,\mathrm{E}(\varepsilon_{t-2}^2) + \phi^6\,\mathrm{E}(\varepsilon_{t-3}^2) + \cdots \\ &= (1 + \phi^2 + \phi^4 + \phi^6 + \cdots)\sigma^2 \\ &= \sigma^2/(1-\phi^2) \end{aligned}$$

and

$$\begin{aligned} Cov(y_t y_{t-s}) &= \phi^s Var(y_{t-s}) \\ &= \phi^s \sigma^2/(1-\phi^2). \end{aligned}$$

It can be seen that y_t has a constant mean of zero, a constant variance and autocovariances that depend only on the distance s between observations. This verifies that y_t is *covariance stationary.*

Recall from last week that the autocorrelation function (ACF) of any stationary process y_t is defined by

$$\rho_s = \frac{Cov(y_t, y_{t-s})}{Var(y_t)} \tag{2}$$

so, for the *AR(1)* process

$$\rho_s = \phi^s$$

and the ACF coefficients die away gradually (in absolute value) since $|\phi^s| < |\phi^{s-1}|$.

The partial autocorrelation function ($PACF$) of y_t, $p(s)$, measures the autocorrelation between y_t and y_{t-s} taking into account the effect of all lags in between. The $PACF$ is measured by the coefficient p_s in the regression equation

$$y_t = p_1 y_{t-1} + p_2 y_{t-2} + p_3 y_{t-3} + \cdots + p_s y_{t-s} + \varepsilon_t.$$

In the *AR(1)* model, clearly $p_1 = \phi$ and $p_s = 0$, $s > 1$. It can be seen that the $PACF$ coefficients cut off abruptly after $s = 1$.

2.2 Higher order processes

We can define the *AR(p)* process

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \tag{3}$$

where as before

$$\mathrm{E}(\varepsilon_t) = 0,\ \mathrm{E}(\varepsilon_t^2) = \sigma^2,\ \mathrm{E}(\varepsilon_t \varepsilon_s) = 0,\ s \neq t.$$

Introducing the *lag operator*, L, defined by

$$L^k x_t = x_{t-k} \quad , \quad L^0 x_t = 1$$

we can rewrite the model as

$$y_t - \phi_1 L y_t - \phi_2 L^2 y_t - \cdots - \phi_p L^p y_t = \varepsilon_t$$

or

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) y_t = \phi(L) y_t = \varepsilon_t$$

where $\phi(L)$ is a polynomial function in the lag operator. This polynomial can be factorised as the product of its roots

$$\phi(L) = \prod_{j=1}^{p} (1 - \alpha_j L) = (1 - \alpha_1 L)(1 - \alpha_2 L) \cdots (1 - \alpha_p L) \tag{4}$$

where the roots are

$$\frac{1}{\alpha_1}, \frac{1}{\alpha_2}, \cdots, \frac{1}{\alpha_p}.$$

These roots are either *real* numbers or *complex* numbers of the form $a+bi$ where a and b are real numbers and where i is the imaginary number defined by $i = \sqrt{-1}$. Where roots are complex, they must appear in *complex conjugate pairs* of the form

$$\alpha_j = a_j + b_j i \qquad \text{and} \qquad \alpha_{j+1} = a_j - b_j i$$

so that the sum of the pair of roots is

$$\alpha_j + \alpha_{j+1} = (a_j + b_j i) + (a_j - b_j i) = 2a_j$$

and the product of the two roots is

$$\alpha_j \alpha_{j+1} = (a_j + b_j i)(a_j - b_j i) = a_j^2 - b_j^2 i^2 = a_j^2 + b_j^2$$

are both *real* numbers. A real root corresponds to a damped exponential whereas a pair of complex conjugate roots corresponds to a damped cosine wave (a cycle).

For the *AR(p)* process to be stationary we require that

$$\|\alpha_j\| < 1, \quad j = 1, \cdots, p$$

where $\|\alpha_j\|$ is the *norm* of α_j defined for the *real* case as $\|\alpha_j\| = |\alpha_j|$ and for the *complex* case as

$$\|\alpha_j\| = \sqrt{(a_j + b_j i)(a_j - b_j i)} = \sqrt{a_j^2 + b_j^2}.$$

The condition for stationarity is sometimes stated as the condition that all the roots of the *AR(p)* process lie *outside* the unit circle since

$$\left\| \frac{1}{\alpha_j} \right\| > 1, \forall j.$$

.

The *AR(p)* model can be written in terms of the innovations ε_t as

$$y_t = \phi(L)^{-1}\varepsilon_t$$

where

$$\phi(L)^{-1} = \prod_{j=1}^{p}(1-\alpha_j L)^{-1}$$

is a polynomial in the lag operator. This polynomial will only exist when the *AR(p)* process satisfies the conditions for stationarity. In general it will be of infinite order so that y_t depends on the whole past history of ε_t.

The *AR(p)* process has the properties that

$$\mathrm{E}(y_t) = \mathrm{E}(\phi(L)^{-1}\varepsilon_t) = 0,$$

$$\begin{aligned} Var(y_t) &\equiv \gamma_0 = E[y_t(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t)] \\ &= \phi_1\gamma_1 + \phi_2\gamma_2 + \cdots + \phi_p\gamma_p + \sigma^2 \end{aligned}$$

$$\begin{aligned} Cov(y_t y_{t-s}) &\equiv \gamma_s = E[y_{t-s}(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t)] \\ &= \phi_1\gamma_{s-1} + \phi_2\gamma_{s-2} + \cdots + \phi_p\gamma_{s-p}. \end{aligned}$$

The coefficients of the *ACF*, ρ_s the autocovariances, will die out gradually as s increases but will never disappear completely. The *PACF* partial autocovariances p_s are non-zero up to p_p but then cut off completely.

3 Moving Average processes

Let y_t be defined by the process

$$y_t = \varepsilon_t + \theta\varepsilon_{t-1} \tag{5}$$

where ε_t is an independently identically distributed random variable with

$$\mathrm{E}(\varepsilon_t) = 0, \quad \mathrm{var}(\varepsilon_t) = \sigma^2, \quad \mathrm{cov}(\varepsilon_t\varepsilon_{t-s}) = 0, \ s \neq 0.$$

This process is known as a first order moving average or *MA(1)* process.

The variable y_t has the properties that

$$\mathrm{E}(y_t) = \mathrm{E}(\varepsilon_t) + \theta\,\mathrm{E}(\varepsilon_{t-1}) = 0,$$

$$\begin{aligned} Var(y_t) &= \mathrm{E}(\varepsilon_t + \theta\varepsilon_{t-1})^2 \\ &= \mathrm{E}(\varepsilon_t^2) + 2\theta\,\mathrm{E}(\varepsilon_t\varepsilon_{t-1}) + \theta^2\,\mathrm{E}(\varepsilon_{t-1}^2) \\ &= \sigma^2(1+\theta^2), \end{aligned}$$

and

$$\begin{aligned} Cov(y_t y_{t-s}) &= \mathrm{E}(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-s} + \theta\varepsilon_{t-s-1}) \\ &= \mathrm{E}(\varepsilon_t\varepsilon_{t-s}) + \theta\,\mathrm{E}(\varepsilon_t\varepsilon_{t-s-1}) + \theta\,\mathrm{E}(\varepsilon_{t-1}\varepsilon_{t-s}) + \theta^2\,\mathrm{E}(\varepsilon_{t-1}\varepsilon_{t-s-1}) \\ &= \sigma^2\theta \quad , \quad s=1 \qquad \text{and} \qquad = 0 \quad , \quad s>1. \end{aligned}$$

It can be seen that y_t has a constant mean of zero, a constant variance and autocovariances that depend only on the distance s between observations. Thus y_t is a covariance stationary variable and this is true for all values of θ, including the unit moving average root cases $\theta = \pm 1$.

The autocovarances ρ_s cut off after the first order whereas it can be shown that the partial autocovariances p_s will gradually damp down and die away.

3.1 Identification

For the autocorrelation function of the *MA(1)* process we have $\rho_1 = \theta/(1+\theta^2)$ and $\rho_s = 0$ for all $s > 1$. Consider an alternative *MA(1)* process with *MA* coefficient $\theta^* = 1/\theta$. The autocorrelation function for this process has

$$\rho_1^* = \frac{\theta^*}{(1+\theta^{*2})} = \frac{1}{\theta}\frac{1}{(1+\frac{1}{\theta^2})} = \frac{1}{\theta}\frac{\theta^2}{(1+\theta^2)} = \frac{\theta}{(1+\theta^2)} = \rho_1$$

and $\rho_s^* = 0$ for all $s > 1$. Thus an *MA(1)* process with coefficient $1/\theta$ has exactly the same autocorrelation function as a process with coefficient θ and it is impossible to distinguish between the two processes from their autocorrelations. This is called an *identification* problem and to resolve this we impose the identification restriction on the *MA(1)* model that

$$|\theta| \leq 1.$$

As before, note that the identification restriction does not exclude the unit root cases $\theta = \pm 1$.

3.2 Higher order processes

We can define the *MA(q)* process

$$y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}$$

where as before

$$\mathrm{E}(\varepsilon_t) = 0, \quad \mathrm{var}(\varepsilon_t) = \sigma^2, \quad \mathrm{cov}(\varepsilon_t \varepsilon_{t-s}) = 0, \ s \neq 0.$$

This process has the properties

$$\mathrm{E}(y_t) = \mathrm{E}(\varepsilon_t) + \theta_1 \mathrm{E}(\varepsilon_{t-1}) + \theta_2 \mathrm{E}(\varepsilon_{t-2}) + \cdots + \theta_q \mathrm{E}(\varepsilon_{t-q}) = 0,$$

$$\begin{aligned} Var(y_t) &= \mathrm{E}(\varepsilon_t^2) + \theta_1^2 \mathrm{E}(\varepsilon_{t-1}^2) + \theta_2^2 \mathrm{E}(\varepsilon_{t-2}^2) + \cdots + \theta_q^2 \mathrm{E}(\varepsilon_{t-q}^2) \\ &= (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)\sigma^2 \end{aligned}$$

and

$$\begin{aligned} Cov(y_t y_{t-s}) &= (\theta_s + \theta_1\theta_{s+1} + \cdots + \theta_{q-s}\theta_q)\sigma^2 \quad , \quad s \leq q \\ &= 0 \quad , \quad s > q. \end{aligned}$$

The autocovariances ρ_s are non-zero up to order q but cut off after this point. The partial autocovariances p_s will die away gradually.

4 The *ARMA* model

4.1 The Wold Representation Theorem

Wold (1954) proved the important result that any *covariance stationary* stochastic process y_t with mean μ and variance σ^2 can be written in the form

$$y_t - \mu = \psi_0\varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \cdots = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \tag{6}$$

where ε_t is a sequence of uncorrelated random variables with mean 0 and constant variance σ^2. This equation is called the *infinite moving average* representation of y_t, or the *Wold* representation. The moving average coefficients are subject to the condition that they are *absolutely summable*

$$\sum_{j=0}^{\infty} |\psi_j| < \infty \, .$$

Using the *lag operator* L, equation (6) can be rewritten as

$$y_t - \mu = (1 + \psi_1 L + \psi_2 L^2 + \cdots)\varepsilon_t = \psi(L)\varepsilon_t$$

where $\psi(L)$ is a polynomial function in the lag operator. Without loss of generality, we have imposed the normalisation restriction that $\psi_0 = 1$.

The polynomial $\psi(L)$ can be factorised as the product of its roots

$$\psi(L) = \prod_{j=1}^{\infty}(1 + \beta_j L) = (1 + \beta_1 L)(1 + \beta_2 L) \cdots \tag{7}$$

with roots given by

$$-\frac{1}{\beta_1}, -\frac{1}{\beta_2}, \quad \text{etc.}$$

These moving average roots must satisfy the condition of *identifiability* that

$$\|\beta_j\| \leq 1 \quad , \quad \forall j \, .$$

Note that this condition of identifiability does not rule out the possibility of unit moving average roots where $\|\beta_j\| = 1$.

In practice we can approximate the infinite order *MA* process by a *finite order MA process*. As we shall see, under certain conditions an autoregressive representation and a mixed *ARMA* representation will also exist.

4.2 Invertibility and the autoregressive representation

When there are no unit roots in (7) so that *all* the roots satisfy the stronger condition that $\|\beta_j\| < 1$, then the process is said to be *invertible* and y_t can be written in the *autoregressive* representation

$$\psi(L)^{-1} y_t = \varepsilon_t \, .$$

More generally, if *some* of the roots satisfy $\|\beta_i\| < 1$, then $\psi(L)$ can be factorised into two polynomials

$$y_t = \psi(L)\varepsilon_t = \phi(L)^{-1}\theta(L)\varepsilon_t$$

where the first polynomial $\phi(L)^{-1}$ has no unit roots and so is invertible and the second $\theta(L)$ may contain some unit roots. Inverting the first polynomial leads to the model

$$\phi(L) y_t = \theta(L)\varepsilon_t \, . \tag{8}$$

(8) is a *mixed ARMA(p,q)* model where

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$$

and

$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q \, .$$

4.3 Integrated Processes: the *ARIMA* model

Suppose that y_t is not stationary but instead is *integrated* of order d, ($y_t \sim I(d)$). Then, by definition, the *dth* order difference of y_t,

$$\Delta^d y_t = (1 - L)^d y_t$$

is stationary, and can be expressed as an *ARMA(p,q)* process. Therefore, it follows that we can write

$$\phi(L)\Delta^d y_t = \theta(L)\varepsilon_t. \tag{9}$$

Such a process is said to be an integrated *ARMA* process or an *ARIMA(p,d,q)* process where d denotes the order of differencing.

4.4 Mixed Processes

4.4.1 Advantages of mixed processes

Why is it necessary to consider mixed processes? Box and Jenkins (1976) stress *parsimony*. They argue that an *ARMA(p,q)* model with small values of p and q will do as well at explaining a process y_t as a high order pure $AR(p^*)$ or $MA(q^*)$ process. Allowing an MA component may be useful since it can give evidence of *over-differencing*. Suppose that the 'true' model is $y_t = \varepsilon_t$, but the forecaster mistakenly differences the process and estimates an *ARMA* model for Δy_t. From the true model, $\Delta y_t = \Delta\varepsilon_t = \varepsilon_t + \theta\varepsilon_{t-1}$, where $\theta = -1$. This is an *MA* process with a (negative) unit root which has been induced by over-differencing. The forecaster who estimates an *ARMA* model for Δy_t including an *MA(1)* component will find an estimated parameter $\widehat{\theta}$ close to -1. This should alert the forecaster to probable over-differencing and this could not be picked up in a *pure AR* model.

4.4.2 Problems with mixed processes

One problem with estimating mixed processes is that of *common factors*. Suppose the 'true' model is *ARMA(p,q)* but the investigator mistakenly estimates the model *ARMA(p+1,q+1)*. If the true model is given by $\phi(L)y_t = \theta(L)\varepsilon_t$, then the estimated model can be written

$$(1 - \alpha_{p+1}L)\phi(L)y_t = (1 + \beta_{q+1}L)\theta(L)\varepsilon_t$$

where α_{p+1} and β_{q+1} are extra (superfluous) roots. For *any* values of α_{p+1} and β_{q+1} satisfying $\alpha_{p+1} = -\beta_{q+1} = \gamma$, the extra roots cancel out so that the model reduces to *ARMA(p,q)*. This model is *not identified* since the parameters α_{p+1} and β_{q+1} can not be estimated.

This result implies that starting out from an over-parameterised $ARMA$ model and testing down to find a parsimonious representation, the so called *general-to-simple modelling strategy*, will not work with a mixed $ARMA$ model. It will still work with pure AR or pure MA models, however.

5 Choosing the order of the *ARIMA* model

5.1 Identifying the order of differencing *d*

The first step in choosing an appropriate $ARIMA$ model is to identify the correct order of differencing. This is a question of testing for unit roots and a natural test to use is the (augmented) Dickey-Fuller (ADF) test. The appropriate procedure would be to test down from an initial order of integration d^* that is at least as large as the (unknown) true value d. Then a sequence of Dickey-Fuller tests are computed testing the null hypothesis $\Delta^{d*} y_t \sim I(1)$ against the alternative hypothesis that $\Delta^{d*} y_t \sim I(0)$, reducing d^* each time, until the null hypothesis fails to be rejected. The final d^* then determines d.

5.2 Identifying the orders of *p* and *q*

Examining the correlogram and partial correlogram may help distinguish between pure AR and pure MA processes. In pure MA processes, we know that the simple autocorrelations should cut off after a certain point whereas in pure AR processes, the simple autocorrelations will damp down gradually but never disappear completely. Conversely, in pure AR models, the partial autocorrelations should cut off after a certain point whereas in pure MA models they will only damp down gradually. Examining the correlogram and partial correlogram for evidence of these features should allow us to distinguish a pure AR from a pure MA process. In practice however, noise may blur these distinctions and make it difficult to decide on the correct process from the correlogram and partial correlogram.

Information criteria such as the AIC and SIC can be used to select the most parsimonious model that fits the data. Sometimes, more than one model may fit a series equally well and different researchers can often disagree about the best $ARMA$ model for a particular series. For example, Box and Jenkins themselves identify two different processes for some of their test series.

6 An Example: Canadian Employment

As an example, let us look at building an $ARIMA$ model for the seasonally adjusted quarterly Canadian Employment index considered by Diebold (2004) and

illustrated in Figure 1. Although this series is not obviously trended, we should still first test for a unit root to determine whether the order of differencing d should be 0 or 1.

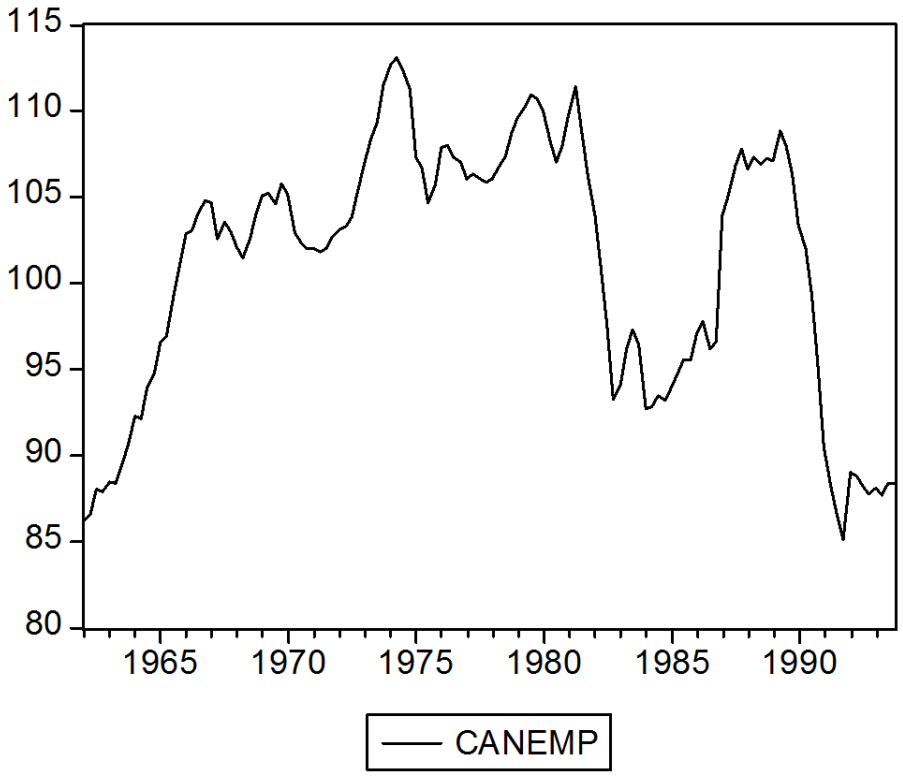

Figure 1: Canadian Employment Data: 1961-1994

The *Augmented Dickey-Fuller* test reported by *EViews* gives a test statistic of -2.2056 which is less than the 5% critical value of -2.8841 so the test fails to reject the null of a unit root. This suggests that the series is *integrated of order 1*, *I(1)*, and needs to be differenced before fitting an *ARMA* model. Despite this, Diebold chooses to fit an *ARMA* model to the level of the series.

Diebold reports the correlogram and partial correlograms illustrated in Figure 2 . The solid lines represent two-standard error bands. The graphs show that the simple autocorrelations fall off gradually whereas the partial autocorrelations cut off sharply. We have seen that this pattern is characteristic of an *autoregressive process*. Also, the fact that the autocorrelations die out is consistent with the series being covariance stationary and contradicts the result of the previous *ADF*

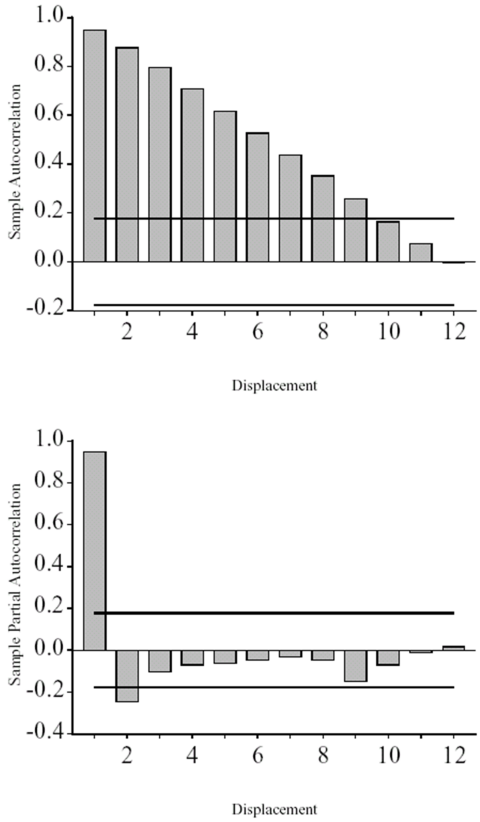

Figure 2: Sample *ACF* and *PACF* for Canadian Employment Index

test for a unit root. (This may be because of the low power of the unit root test to reject when the alternative has roots close to but less than one).

	AR(p) models		*MA(q) models*		*ARMA(p,q) models*		
Order	*AIC*	*SIC*	*AIC*	*SIC*	*Order*	*AIC*	*SIC*
1	3.85	3.89	5.70	5.74	1, 1	3.67	3.73
2	3.60	3.67	4.94	5.01	2, 1	3.61	3.70
3	3.61	3.70	4.46	4.54	1, 2	3.63	3.72
4	3.63	3.74	4.15	4.26	3, 1	3.60	3.71
8	3.70	3.91	3.78	3.98	1, 3	3.64	3.75

The table presents *AIC* and *SIC* information criteria for various different *ARMA(p,q)* models fitted to the Canadian Employment data. Note that the low order *AR* models perform better than the low order *MA* models with the *AR(2)* model doing best by both criteria. This is consistent with the patterns observed in the correlogram and partial correlogram which suggest that the form

of the autocorrelation is autoregressive and can be captured by a low order AR model. To approximate this by an MA process requires a much higher order process, for example it requires *MA(8)* to get close to the fit of an *AR(1)*. That the *AR(2)* seems to do best is consistent with the cut-off point in the partial correlogram which shows that partial autocorrelations beyond the second order are not significant. The estimated *AR(2)* model is

$$\widehat{y}_t = 101.2 + 1.44y_{t-1} - 0.48y_{t-2}$$

or

$$(1 - 0.92L)(1 - 0.52L)\widehat{y}_t = 101.2. \tag{10}$$

This process has two real roots, the first of which is quite close to unity which helps to explain why the hypothesis of a unit root was not rejected by the Dickey-Fuller test.

Can a mixed $ARMA$ model do better than the *AR(2)* model? The simplest $ARMA$ model, the *ARMA(1,1)* has the same number of parameters as the *AR(2)* but doesn't fit as well as can be seen from comparing the AIC and SIC criteria. Increasing the order of the MA part by considering *ARMA(1,2)* and *ARMA(1,3)* doesn't help. Increasing the order of the AR component to consider *ARMA(2,1)* and *ARMA(3,1)* does better but of course these models nest the *AR(2)* model. For the *ARMA(2,1)* the results are

$$\widehat{y}_t = 101.2 + 1.57y_{t-1} - 0.61y_{t-2} + \varepsilon_t - 0.18\varepsilon_{t-1}$$

or

$$(1 - 0.90L)(1 - 0.68L)\widehat{y}_t = 101.2 + (1 - 0.18L)\varepsilon_t$$

but the MA parameter estimate is insignificantly different from zero with p-value *0.33*. For the *ARMA(3,1)* the results are

$$\widehat{y}_t = 101.1 + 0.50y_{t-1} + 0.87y_{t-2} - 0.44y_{t-3} + \varepsilon_t + 0.97\varepsilon_{t-1}$$

or

$$(1 - 0.93L)(1 - 0.51L)(1 + 0.94L)\widehat{y}_t = 101.1 + (1 + 0.97L)\varepsilon_t.$$

Note that the third AR root almost cancels out with the MA root, suggesting that this is an example of an over-parameterised, unidentified $ARMA$ process with a common factor. Cancelling the redundant root reduces the model to the *AR(2)* model (10), which is thus the preferred model for this time series.

To verify that the *AR(2)* model is indeed adequate, we can use the *Box-Ljung Q-statistic* to test that the *residuals* from equation (10) are white noise with no evidence of serial correlation. Choosing a lag length of 12, the test statistic is *5.44* which has a p-value of 0.86, so supporting the null hypothesis of no residual serial correlation.

References

[1] Box, G. E. P, and G. M. Jenkins, (1976), *Time Series Analysis: Forecasting and Control*, (2nd ed.), Holden-Day, San Fransisco.

[2] Diebold, F. X. (2004), *Elements of Forecasting*, (3rd ed.), Thomson South-Western, Louiseville.

[3] Wold, H. O. (1954), *A Study in the Analysis of Stationary Time Series*, (2nd ed.), Almquist and Wicksell, Uppsala.