

Time Series Analysis

Forecasting with ARIMA models

Andrés M. Alonso Carolina García-Martos

Universidad Carlos III de Madrid

Universidad Politécnica de Madrid

June – July, 2012

7. Forecasting with ARIMA models

Outline:

- Introduction
- The prediction equation of an ARIMA model
- Interpreting the predictions
- Variance of the predictions
- Forecast updating
- Measuring predictability

Recommended readings:

- ▶ Chapters 5 and 6 of Brockwell and Davis (1996).
- ▶ Chapter 4 of Hamilton (1994).
- ▶ Chapter 5 of Peña, Tiao and Tsay (2001).

Introduction

- ▶ We will look at forecasting using a known ARIMA model. We will define the **optimal predictors** as those that minimize the **mean square prediction errors**.
- ▶ We will see that the prediction function of an ARIMA model has a simple structure:
 - The non-stationary operators, that is, if the differences and the constant exist they determine the long-term prediction.
 - The stationary operators, AR and MA, determine short-term predictions.
- ▶ Predictions are of little use without a measurement of their accuracy, and we will see how to obtain the distribution of the prediction errors and how to calculate prediction confidence intervals.
- ▶ Finally we will study how to revise the predictions when we receive new information.

The prediction equation of an ARIMA model

Conditional expectation as an optimal predictor

▷ Suppose we have observed a realization of length T in a time series $\mathbf{z}_T = (z_1, \dots, z_T)$, and we wish to forecast a future value $k > 0$ periods ahead, z_{T+k} .

▷ We let $\hat{z}_T(k)$ be a **predictor** of z_{T+k} obtained as a function of the T values observed, that is, with the forecast *origin* in T , and the forecast *horizon* at k .

▷ Specifically, we will study **linear predictors**, which are those constructed as a linear combination of the available data:

$$\hat{z}_T(k) = \alpha_1 z_T + \alpha_2 z_{T-1} + \dots + \alpha_T z_1.$$

▷ The predictor is well defined when the constants $\alpha_1, \dots, \alpha_T$ used to construct it are known.

Conditional expectation as an optimal predictor

- ▶ We denote by $e_T(k)$ the **prediction error** of this predictor, given by:

$$e_T(k) = z_{T+k} - \hat{z}_T(k)$$

and we want this error to be as small as possible.

- ▶ Since the variable z_{T+k} is unknown, it is impossible to know a priori the error that we will have when we predict it.
- ▶ Nevertheless, if we know its probability distribution, that is, its possible values and probabilities, with a defined predictor we can calculate the probability of its error being among the given values.
- ▶ To compare predictors we specify the desired objective by means of a criterion and select the best predictor according to this criterion.

Conditional expectation as an optimal predictor

- If the objective is to obtain small prediction errors and it does not matter if they are positive or negative we can eliminate the error sign using the criterion of minimizing the expected value of a **symmetric loss function** of the error, such as:

$$l(e_{T+k}) = ce_T^2(k) \quad \text{or} \quad l(e_T(k)) = c |e_T(k)|,$$

where c is a constant.

- Nevertheless, if we prefer the errors in one direction,, we minimize the average value of an **asymmetric loss function** of the error which takes into account our objectives. For example,

$$l(e_T(k)) = \begin{cases} c_1 |e_T(k)|, & \text{if } e_T(k) \geq 0 \\ c_2 |e_T(k)|, & \text{if } e_T(k) \leq 0 \end{cases} \quad (130)$$

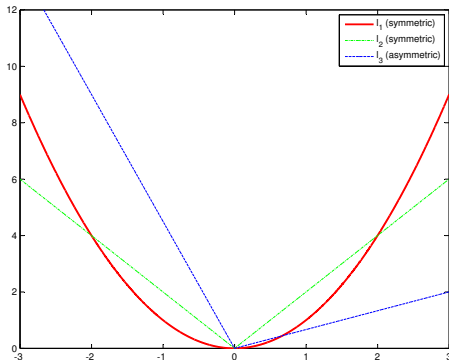
for certain positive c_1 and c_2 . In this case, the higher the c_2/c_1 ratio the more we penalize low errors.

Example 64

The figure compares the three above mentioned loss functions. The two symmetric functions

$$l_1(e_T(k)) = e_T^2(k), \quad y \quad l_2(e_T(k)) = 2|e_T(k)|,$$

and the asymmetric (130) have been represented using $c_1 = 2/3$ and $c_2 = 9/2$.



- ▶ Comparing the two symmetric functions, we see that the quadratic one places less importance on small values than the absolute value function, whereas it penalizes large errors more.
- ▶ The asymmetric function gives little weight to positive errors, but penalizes negative errors more.

Conditional expectation as an optimal predictor

- ▶ The most frequently utilized loss function is the **quadratic**, which leads to the criterion of minimizing the **mean square prediction error** (MSPE) of z_{T+k} given the information \mathbf{z}_T . We have to minimize:

$$MSPE(z_{T+k}|\mathbf{z}_T) = E [(z_{T+k} - \hat{z}_T(k))^2|\mathbf{z}_T] = E [e_T^2(k)|\mathbf{z}_T]$$

where the expectation is taken with respect to the distribution of the variable z_{T+k} conditional on the observed values \mathbf{z}_T .

- ▶ We are going to show that the predictor that minimizes this mean square error is the **expectation** of the variable z_{T+k} **conditional** on the available information.
- ▶ Therefore, if we can calculate this expectation we obtain the optimal predictor without needing to know the complete conditional distribution.

Conditional expectation as an optimal predictor

► To show this, we let

$$\mu_{T+k|T} = E[z_{T+k}|\mathbf{z}_T]$$

be the mean of this conditional distribution. Subtracting and adding $\mu_{T+k|T}$ in the expression of $MSPE(z_{T+k})$ expanding the square we have:

$$MSPE(z_{T+k}|\mathbf{z}_T) = E[(z_{T+k} - \mu_{T+k|T})^2|\mathbf{z}_T] + E[(\mu_{T+k|T} - \hat{z}_T(k))^2|\mathbf{z}_T] \quad (131)$$

since the double product is cancelled out.

► Indeed, the term $\mu_{T+k|T} - \hat{z}_T(k)$ is a constant, since $\hat{z}_T(k)$ is a function of the past values and we are conditioning on them and:

$$\begin{aligned} E[(\mu_{T+k|T} - \hat{z}_T(k))(z_{T+k} - \mu_{T+k|T})|\mathbf{z}_T] &= \\ &= (\mu_{T+k|T} - \hat{z}_T(k))E[(z_{T+k} - \mu_{T+k|T})|\mathbf{z}_T] = 0 \end{aligned}$$

thus we obtain (131).

Conditional expectation as an optimal predictor

- ▷ This expression can be written:

$$MSPE(z_{T+k}|\mathbf{z}_T) = \text{var}(z_{T+k}|\mathbf{z}_T) + E[(\mu_{T+k|T} - \hat{z}_T(k))^2|\mathbf{z}_T].$$

- ▷ Since the first term of this expression does not depend on the predictor, we minimize the $MSPE$ of the predictor setting the second to zero. This will occur if we take:

$$\hat{z}_T(k) = \mu_{T+k|T} = E[z_{T+k}|\mathbf{z}_T].$$

- ▷ We have shown that the predictor that minimizes the mean square prediction error of a future value is obtained by taking its conditional expectation on the observed data.

Example 65

Let us assume that we have 50 data points generated by an AR(1) process: $z_t = 10 + 0.5z_{t-1} + a_t$. The last two values observed correspond to times $t = 50$, and $t = 49$, and are $z_{50} = 18$, and $z_{49} = 15$. We want to calculate predictions for the next two periods, $t = 51$ and $t = 52$.

The first observation that we want to predict is z_{51} . Its expression is:

$$z_{51} = 10 + 0.5z_{50} + a_{51}$$

and its expectation, conditional on the observed values, is:

$$\hat{z}_{50}(1) = 10 + 0.5(18) = 19.$$

For $t = 52$, the expression of the variable is:

$$z_{52} = 10 + 0.5z_{51} + a_{52}$$

and taking expectations conditional on the available data:

$$\hat{z}_{50}(2) = 10 + 0.5\hat{z}_{50}(1) = 10 + 0.5(19) = 19.5.$$

The prediction equation of an ARIMA model

Prediction estimations

► Let us assume that we have a realization of size T , $\mathbf{z}_T = (z_1, \dots, z_T)$, of an ARIMA (p, d, q) process, where $\phi(B) \nabla^d z_t = c + \theta(B)a_t$, with known parameters. We are going to see how to apply the above result to compute the predictions.

► Knowing the parameters we can obtain all the innovations a_t fixing some initial values. For example, if the process is ARMA(1,1) the innovations, a_2, \dots, a_T , are computed recursively using the equation:

$$a_t = z_t - c - \phi z_{t-1} + \theta a_{t-1}, \quad t = 2, \dots, T$$

The innovation for $t = 1$ is given by:

$$a_1 = z_1 - c - \phi z_0 + \theta a_0$$

and neither z_0 nor a_0 are known, so we cannot obtain a_1 . We can replace it with its expectation $E(a_1) = 0$, and calculate the remaining a_t using this initial condition.

Prediction estimations

- ▶ As a result, we assume from here on that both the observations as well as the innovations are known up to time T .
- ▶ The prediction that minimizes the mean square error of z_{T+k} , which for simplicity from here on we will call the optimal prediction of z_{T+k} , is the expectation of the variable conditional on the observed values.
- ▶ We define

$$\begin{aligned}\hat{z}_T(j) &= E[z_{T+j} | \mathbf{z}_T] & j = 1, 2, \dots \\ \hat{a}_T(j) &= E[a_{t+j} | \mathbf{z}_T] & j = 1, 2, \dots\end{aligned}$$

where the subindex T represents the *forecast origin*, which we assume is fixed, and j represents the *forecast horizon*, which will change to generate predictions of different future variables from origin T .

Prediction estimations

▶ Letting $\varphi_h(B) = \phi_p(B) \nabla^d$ be the operator of order $h = p + d$ which is obtained multiplying the stationary AR(p) operator and differences, the expression of the variable z_{T+k} is:

$$z_{T+k} = c + \varphi_1 z_{T+k-1} + \dots + \varphi_h z_{T+k-h} + a_{T+k} - \theta_1 a_{T+k-1} - \dots - \theta_q a_{T+k-q}. \quad (132)$$

▶ Taking expectations conditional on \mathbf{z}_T in all the terms of the expression, we obtain:

$$\begin{aligned} \hat{z}_T(k) = c + \varphi_1 \hat{z}_T(k-1) + \dots + \varphi_h \hat{z}_T(k-h) \\ - \theta_1 \hat{a}_T(k-1) - \dots - \hat{a}_T(k-q) \end{aligned} \quad (133)$$

▶ In this equation some expectations are applied to observed variables and others to unobserved.

- When $i > 0$ the expression $\hat{z}_T(i)$ is the conditional expectation of the variable z_{T+i} that has not yet been observed.
- When $i \leq 0$, $\hat{z}_T(i)$ is the conditional expectation of the variable $z_{T-|i|}$, which has already been observed and is known, so this expectation will coincide with the observation and $\hat{z}_T(-|i|) = z_{T-|i|}$.

Prediction estimations

▷ Regarding innovations, the expectations of future innovations conditioned on the history of the series is equal to its absolute expectation, zero, since the variable a_{T+i} is independent of \mathbf{z}_T , and we conclude that for $i > 0$, $\hat{a}_T(i)$ are zero.

- When $i \leq 0$, the innovations $a_{T-|i|}$ are known, thus their expectations coincide with the observed values and $\hat{a}_T(-|i|) = a_{T-|i|}$.

▷ This way, we can calculate the predictions recursively, starting with $j = 1$ and continuing with $j = 2, 3, \dots$, until the desired horizon has been reached.

▷ We observe that for $k = 1$, subtracting (132) from (133) results in:

$$a_{T+1} = z_{T+1} - \hat{z}_T(1),$$

and the innovations can be interpreted as one step ahead prediction errors when the parameters of the model are known.

Prediction estimations

- ▶ Equation (133) indicates that after q initial values the moving average terms disappear, and the prediction will be determined exclusively by the autoregressive part of the model.
- ▶ Indeed, for $k > q$ all the innovations that appear in the prediction equation are unobserved, their expectancies are zero and they will disappear from the prediction equation.
- ▶ Thus the predictions of (133) for $k > q$ satisfy the equation:

$$\hat{z}_T(k) = c + \varphi_1 \hat{z}_T(k-1) + \dots + \varphi_h \hat{z}_T(k-h). \quad (134)$$

- ▶ We are going to rewrite this expression using the lag operator in order to simplify its use. In this equation the forecast origin is always the same but the horizon changes. Thus, we now say that the operator B acts on the forecast horizon such that:

$$B\hat{z}_T(k) = \hat{z}_T(k-1)$$

whereas B does not affect the forecast origin, T , which is fixed.

▷ With this notation, equation (134) is written:

$$(1 - \varphi_1 B - \dots - \varphi_h B^h) \hat{z}_T(k) = \phi(B) \nabla^d \hat{z}_T(k) = c, \quad k > q. \quad (135)$$

▷ This equation is called the **final prediction equation** because it establishes how predictions for high horizons are obtained when the moving average part disappears.

▷ We observe that the relationship between the predictions is similar to that which exists between autocorrelations of an ARMA stationary process, although in the predictions in addition to the stationary operator $\phi(B)$ a non-stationary operator ∇^d also appears, and the equation is not equal to zero, but rather to the constant, c .

Example 66

Let us assume an $AR(1)$ process $z_t = c + \phi_1 z_{t-1} + a_t$. The prediction equation for one step ahead is:

$$\hat{z}_T(1) = c + \phi_1 z_T,$$

and for two steps:

$$\hat{z}_T(2) = c + \phi_1 \hat{z}_T(1) = c(1 + \phi_1) + \phi_1^2 z_T.$$

▷ Generalizing, for any period $k > 0$ we have:

$$\hat{z}_T(k) = c + \phi_1 \hat{z}_T(k-1) = c(1 + \phi_1 + \dots + \phi_1^k) + \phi_1^k z_T.$$

▷ For large k since $|\phi_1| < 1$ the term $\phi_1^k z_T$ is close to zero and the prediction is $c(1 + \phi_1 + \dots + \phi_1^k) = c/(1 - \phi_1)$, which is the mean of the process.

▷ We will see that for any stationary $ARMA(p, q)$ process, the forecast for a large horizon k is the mean of the process, $\mu = c/(1 - \phi_1 - \dots - \phi_p)$.

▷ As a particular case, if $c = 0$ the long-term prediction will be zero.

Example 67

Let us assume a random walk: $\nabla z_t = c + a_t$. The one step ahead prediction with this model is obtained using:

$$z_t = z_{t-1} + c + a_t.$$

▷ Taking expectations conditional on the data up to T , we have, for $T + 1$:

$$\hat{z}_T(1) = c + z_T,$$

and for $T + 2$:

$$\hat{z}_T(2) = c + \hat{z}_T(1) = 2c + z_T,$$

and for any horizon k :

$$\hat{z}_T(k) = c + \hat{z}_T(k-1) = kc + z_T.$$

▷ Since the prediction for the following period is obtained by always adding the same constant, c , we conclude that the predictions will continue in a straight line with slope c . If $c = 0$ the predictions for all the periods are equal to the last observed value, and will follow a horizontal line.

▷ We see that the constant determines the slope of the prediction equation.

Example 68

Let us assume the model $\nabla z_t = (1 - \theta B)a_t$.

Given the observations up to time T the next observation is generated by

$$z_{T+1} = z_T + a_{T+1} - \theta a_T$$

and taking conditional expectations, its prediction is:

$$\hat{z}_T(1) = z_T - \theta a_T.$$

For two periods ahead, since $z_{T+2} = z_{T+1} + a_{T+2} - \theta a_{T+1}$, we have

$$\hat{z}_T(2) = \hat{z}_T(1)$$

and, in general

$$\hat{z}_T(k) = \hat{z}_T(k-1), \quad k \geq 2.$$

- ▷ To see the relationship between these predictions and those generated by simple exponential smoothing, we can invert operator $(1 - \theta B)$ and write:

$$\nabla(1 - \theta B)^{-1} = 1 - (1 - \theta)B - (1 - \theta)\theta B^2 - (1 - \theta)\theta^2 B^3 - \dots$$

so that the process is:

$$z_{T+1} = (1 - \theta) \sum_{j=0}^{\infty} \theta^j z_{T-j} + a_{T+1}.$$

- ▷ This equation indicates that the future value, z_{T+1} , is the sum of the innovation and an average of the previous values with weights that decrease geometrically.

- ▷ The prediction is:

$$\hat{z}_T(1) = (1 - \theta) \sum_{j=0}^{\infty} \theta^j z_{T-j}$$

and it is easy to prove that $\hat{z}_T(2) = \hat{z}_T(1)$.

Example 69

We are going to look at the seasonal model for monthly data

$$\nabla_{12} z_t = (1 - 0.8B^{12})a_t.$$

The forecast for any month in the following year, j , where $j = 1, 2, \dots, 12$, is obtained by taking conditional expectations of the data in:

$$z_{t+j} = z_{t+j-12} + a_{t+j} - 0.8a_{t+j-12}.$$

For $j \leq 12$, the result is:

$$\hat{z}_T(j) = z_{T+j-12} - 0.8a_{T+j-12-1}, \quad j = 1, \dots, 12$$

and for the second year, since the disturbances a_t are no longer observed:

$$\hat{z}_T(12 + j) = \hat{z}_T(j).$$

► This equation indicates that the forecasts for all the months of the second year will be identical to those of the first. The same occurs with later years.

► Therefore, the prediction equation contains 12 coefficients that repeat year after year.

▷ To interpret them, let

$$\bar{z}_T(12) = \frac{1}{12} \sum_{j=1}^{12} \hat{z}_T(j) = \frac{1}{12} \sum_{j=1}^{12} z_{T+j-12} - 0.8 \frac{1}{12} \sum_{j=1}^{12} a_{T+j-12-1}$$

denote the mean of the forecasts for the twelve months of the first year. We define:

$$S_T(j) = \hat{z}_T(j) - \bar{z}_T(12)$$

as seasonal coefficients. By their construction they add up to zero, and the predictions can be written:

$$\hat{z}_T(12k + j) = \bar{z}_T(12) + S_T(j), \quad j = 1, \dots, 12; k = 0, 1, \dots$$

▷ The prediction with this model is the sum of a constant level, estimated by $\bar{z}_T(12)$, plus the seasonal coefficient of the month.

▷ We observe that the level is approximately estimated using the average level of the last twelve months, with a correction factor that will be small, since $\sum_{j=1}^{12} a_{T+j-12-1}$ will be near zero since the errors have zero mean.

Non-seasonal processes

- ▶ We have seen that for non-seasonal ARIMA models the predictions carried out from origin T with horizon k satisfy, for $k > q$, the final prediction equation (135).
- ▶ Let us assume initially the simplest case of $c = 0$. Then, the equation that predictions must satisfy is:

$$\phi(B) \nabla^d \hat{z}_T(k) = 0, \quad k > q \quad (136)$$

- ▶ The solution to a difference equation that can be expressed as a product of prime operators is the sum of the solutions corresponding to each of the operators.
- ▶ Since the polynomials $\phi(B)$ and ∇^d from equation (136) are prime (they have no common root), the solution to this equation is the sum of the solutions corresponding to the stationary operator, $\phi(B)$, and the non-stationary, ∇^d .

Interpreting the predictions - Non-seasonal processes

▷ We can write:

$$\hat{z}_T(k) = P_T(k) + t_T(k) \quad (137)$$

where these components must satisfy:

$$\nabla^d P_T(k) = 0, \quad (138)$$

$$\phi(B) t_T(k) = 0. \quad (139)$$

▷ It is straightforward to prove that (137) with conditions (138) and (139) is the solution to (136). We only need to plug the solution into the equation and prove that it verifies it.

▷ The solution generated by the non-stationary operator, $P_T(k)$, represents the **permanent component** of the prediction and the solution to the stationary operator, $t_T(k)$, is the **transitory component**.

Interpreting the predictions - Non-seasonal processes

► It can be proved that the solution to (138), is of form:

$$P_T(k) = \beta_0^{(T)} + \beta_1^{(T)}k + \dots + \beta_{d-1}^{(T)}k^{(d-1)}, \quad (140)$$

where the parameters $\beta_j^{(T)}$ are constants to determine, which depend on the origin T , and are calculated using the last available observations.

► The **transitory component**, solution to (139), provides the **short-term predictions** and is given by:

$$t_T(k) = \sum_{i=1}^P A_i G_i^k,$$

with G_i^{-1} being the roots of the AR stationary operator. This component tends to zero with k , since all the roots are $|G_i| \leq 1$, thus justifying the name transitory component.

Interpreting the predictions - Non-seasonal processes

▷ As a result, the **final prediction equation** is valid for $k > \max(0, q - (p + d))$ and is the sum of two components:

- ① A permanent component, or predicted trend, which is a polynomial of order $d - 1$ with coefficients that depend on the origin of the prediction;
- ② A transitory component, or short-term prediction, which is a mixture of cyclical and exponential terms (according to the roots of $\phi(B) = 0$) and that tend to zero with k .

▷ To determine the coefficients $\beta_j^{(k)}$ of the predicted trend equation, the simplest method is to generate predictions using (133) for a large horizon k until:

- (a) if $d = 1$, they are constant. Then, this constant is $\beta_0^{(T)}$,
- (b) if $d = 2$, it is verified that the difference between predictions, $\hat{z}_T(k) - \hat{z}_T(k - 1)$ is constant. Then the predictions are in a straight line, and their slope is $\beta_1^{(T)} = \hat{z}_T(k) - \hat{z}_T(k - 1)$, which adapts itself according to the available observations.

Interpreting the predictions - Non-seasonal processes

- ▷ We have assumed that the stationary series has zero mean. If this is not the case, the final prediction equation is:

$$\phi(B) \nabla^d \hat{z}_T(k) = c. \quad (141)$$

- ▷ This equation is not equal to zero and we have seen that the solution to a difference equation that is not homogeneous is the solution to the homogeneous equation (the equation that equals zero), plus a particular solution.
- ▷ A particular solution to (141) must have the property whereby differentiating d times and applying term $\phi(B)$ we obtain the constant c .
- ▷ For this to occur the solution must be of type $\beta_d k^d$, where the lag operator is applied to k and β_d is a constant obtained by imposing the condition that this expression verifies the equation, that is:

$$\phi(B) \nabla^d (\beta_d k^d) = c.$$

Interpreting the predictions - Non-seasonal processes

▶ To obtain the constant β_d we are first going to prove that applying the operator ∇^d to $\beta_d k^d$ yields $d!\beta_d$, where $d! = (d-1) * (d-2) * \dots * 2 * 1$.

▶ We will prove it for the usual values of d . If $d = 1$:

$$(1 - B)\beta_d k = \beta_d k - \beta_d(k-1) = \beta_d$$

and for $d = 2$:

$$(1 - B)^2 \beta_d k^2 = (1 - 2B + B^2) \beta_d k^2 = \beta_d k^2 - 2\beta_d(k-1)^2 + \beta_d(k-2)^2 = 2\beta_d$$

▶ Since when we apply $\phi(B)$ to a constant we obtain $\phi(1)$ multiplying that constant, we conclude that:

$$\phi(B) \nabla^d (\beta_d k^d) = \phi(B) d! \beta_d = \phi(1) d! \beta_d = c$$

which yields:

$$\beta_d = \frac{c}{\phi(1) d!} = \frac{\mu}{d!}, \quad (142)$$

since $c = \phi(1) \mu$, where μ is the mean of the stationary series.

Interpreting the predictions - Non-seasonal processes

- ▷ Since a particular solution is permanent, and does not disappear with the lag, we will add it to the permanent component (140), thus we can now write:

$$P_T(k) = \beta_0^{(T)} + \beta_1^{(T)}k + \dots + \beta_d k^d.$$

- ▷ This equation is valid in all cases since if $c = 0$, $\beta_d = 0$, and we return to expression (140).
- ▷ There is a basic difference between the coefficients of the predicted trend up to order $d - 1$, which depend on the origin of the predictions and change over time, and the coefficient β_d , given by (142), which is constant for any forecast origin because it depends only on the mean of the stationary series.
- ▷ Since long-term behavior will depend on the term of the highest order, long-term forecasting of a model with constant are determined by that constant.

Example 70

We are going to generate predictions for the population series of people over 16 in Spain. Estimating this model, with the method that we will described in next session, we obtain: $\nabla^2 z_t = (1 - 0.65B)a_t$.

The predictions generated by this model for the twelve four-month terms are given the figure. The predictions follow a line with a slope of 32130 additional people each four-month period.

Dependent Variable: D(POPULATIONOVER16,2)

Method: Least Squares

Date: 02/07/08 Time: 10:59

Sample (adjusted): 1977Q3 2000Q4

Included observations: 94 after adjustments

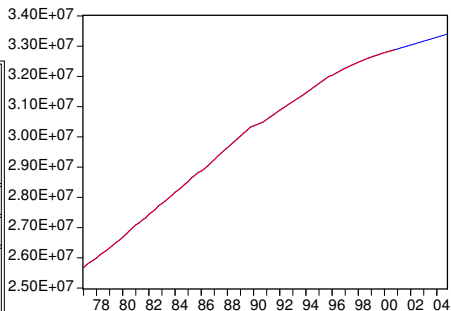
Convergence achieved after 10 iterations

Backcast: 1977Q2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	-0.652955	0.078199	-8.349947	0.0000

R-squared	0.271014	Mean dependent var	-698.8085
Adjusted R-squared	0.271014	S.D. dependent var	17165.34
S.E. of regression	14655.88	Akaike info criterion	22.03365
Sum squared resid	2.00E+10	Schwarz criterion	22.06071
Log likelihood	-1034.582	Durbin-Watson stat	1.901285

Inverted MA Roots .65



— POPULATIONF
— Spanish population over 16 years of age

Example 71

Compare the long-term prediction of a random walk with drift, $\nabla z_t = c + a_t$, with that of model $\nabla^2 z_t = (1 - \theta B) a_t$.

▷ The random walk forecast is, according to the above, a line with slope c which is the mean of the stationary series $w_t = \nabla z_t$ and is estimated using

$$\hat{c} = \frac{1}{T-1} \sum_{t=2}^{t=T} \nabla z_t = \frac{z_T - z_1}{T-1}.$$

▷ Since

$$\hat{z}_T(k) = \hat{c} + \hat{z}_T(k-1) = k\hat{c} + z_T,$$

the future growth prediction in a period, $\hat{z}_T(k) - \hat{z}_T(k-1)$, will be equal to the average growth observed in the whole series \hat{c} and we observe that this slope is constant.

▷ The model with two differences if $\theta \rightarrow 1$ will be very similar to $\nabla z_t = c + a_t$, since the solution to

$$\nabla^2 z_t = \nabla a_t$$

is

$$\nabla z_t = c + a_t.$$

▷ Nevertheless, when θ is not one, although it is close to one, the prediction of both models can be very different in the long-term. In the model with two differences the final prediction equation is also a straight line, but with a slope that changes with the lag, whereas in the model with one difference and a constant the slope is always constant.

▷ The one step ahead forecast from the model with two differences is obtained from

$$z_t = 2z_{t-1} - z_{t-2} + a_t - \theta a_{t-1}$$

and taking expectations:

$$\hat{z}_T(1) = 2z_T - z_{T-1} - \theta a_T = z_T + \hat{\beta}_T$$

where we have let

$$\hat{\beta}_T = \nabla z_T - \theta a_T$$

be the quantity that we add to the last observation, which we can consider as the slope appearing at time T .

- ▷ For two periods

$$\hat{z}_T(2) = 2\hat{z}_T(1) - z_T = z_T + 2\hat{\beta}_T$$

and we see that the prediction is obtained by adding that slope twice, thus z_T , $\hat{z}_T(1)$ and $\hat{z}_T(2)$ are in a straight line.

- ▷ In the same way, it is easy to prove that:

$$\hat{z}_T(k) = z_T + k\hat{\beta}_T$$

which shows that all the predictions follow a straight line with slope $\hat{\beta}_T$.

- ▷ We observe that the slope changes over time, since it depends on the last observed growth, ∇z_T , and on the last forecasting error committed a_T .

▷ Replacing $a_T = (1 - \theta B)^{-1} \nabla^2 z_T$ in the definition of $\hat{\beta}_T$, we can write

$$\hat{\beta}_T = \nabla z_T - \theta(1 - \theta B)^{-1} \nabla^2 z_T = (1 - \theta(1 - \theta B)^{-1} \nabla) \nabla z_T,$$

and the resulting operator on ∇z_T is

$$1 - \theta(1 - \theta B)^{-1} \nabla = 1 - \theta(1 - (1 - \theta)B - \theta(1 - \theta)B^2 - \dots)$$

so that $\hat{\beta}_T$ can be written

$$\hat{\beta}_T = (1 - \theta) \sum_{i=0}^{T-1} \theta^i \nabla z_{T-i}.$$

▷ This expression shows that the slope of the final prediction equation is a weighted mean of all observed growth, ∇z_{T-i} , with weights that decrease geometrically.

▷ We observe that if θ is close to one this mean will be close to the arithmetic mean, whereas if θ is close to zero the slope is estimated only with the last observed values.

► This example shows the greater **flexibility** of the model with two differences compared to the model with one difference and a constant.

Example 72

Compare these models with a deterministic one that also generates predictions following a straight line. Let us look at the deterministic linear model:

$$\hat{z}_T(k) = a + b_R(T + k)$$

▷ We assume to simplify the analysis that $T = 5$, and we write $t = (-2, -1, 0, 1, 2)$ such that $\bar{t} = 0$. The five observations are $(z_{-2}, z_{-1}, z_0, z_1, z_2)$. The slope of the line is estimated by least squares with

$$b_R = \frac{-2z_{-2} - z_{-1} + z_1 + 2z_2}{10}$$

and this expression can be written as:

$$b_R = .2(z_{-1} - z_{-2}) + .3(z_0 - z_{-1}) + .3(z_1 - z_0) + .2(z_2 - z_1)$$

which is a weighted mean of the observed growth but with minimum weight given to the last one.

▷ It can be proved that in the general case the slope is written

$$b_R = \sum_{t=2}^T w_t \nabla z_t$$

where $\sum w_t = 1$ and the weights are symmetric and their minimum value corresponds to the growth at the beginning and end of the sample.

▷ Remember that:

- The random walk forecast is a line with slope c which is the mean of the stationary series $w_t = \nabla z_t$ and is estimated using

$$\hat{c} = \frac{1}{T-1} \sum_{t=2}^{t=T} \nabla z_t.$$

- The two differences model forecast is $\hat{z}_T(k) = z_T + k\hat{\beta}_T$ where the slope $\hat{\beta}_T$ that changes over time.

► This example shows the **limitations** of the deterministic models for forecasting and the greater **flexibility** that can be obtained by taking differences.

Interpreting the predictions - Seasonal processes

- ▶ If the process is seasonal the above decomposition is still valid: the **final prediction equation** will have a **permanent component** that depends on the non-stationary operators and the mean μ of the stationary process, and a **transitory component** that encompasses the effect of the stationary AR operators.
- ▶ In order to separate the trend from seasonality in the permanent component, the operator associated with these components cannot have any roots in common. If the series has a seasonal difference such as:

$$\nabla_s = (1 - B^s) = (1 + B + B^2 + \dots + B^{s-1})(1 - B)$$

the seasonal operator $(1 - B^s)$ incorporates two operators:

- The **difference operator**, $1 - B$, and the **pure seasonal operator**, given by:

$$S_s(B) = 1 + B + \dots + B^{s-1} \quad (143)$$

which produces the sum of s consecutive terms.

Interpreting the predictions - Seasonal processes

- ▷ Separating the root $B = 1$ from the operator ∇_s , the seasonal model, assuming a constant different from zero, can be written:

$$\Phi(B^s) \phi(B) S_s(B) \nabla^{d+1} z_t = c + \theta(B) \Theta(B^s) a_t$$

where now the four operators $\Phi(B^s)$, $\phi(B)$, $S_s(B)$ and ∇^{d+1} have no roots in common.

- ▷ The **final prediction equation** then is:

$$\Phi(B^s) \phi(B) S_s(B) \nabla^{d+1} \hat{z}_t(k) = c \quad (144)$$

an equation that is valid for $k > q + sQ$ and since it requires $d + s + p + sP$ initial values (maximum order of B in the operator on the right), it can be used to calculate predictions for $k > q + sQ - d - s - p - sP$.

Interpreting the predictions - Seasonal processes

▷ Since the stationary operators, $S_s(B)$ and ∇^{d+1} , now have no roots in common, the permanent component can be written as the sum of the solutions of each operator.

▷ The solution to the homogeneous equation is:

$$\hat{z}_t(k) = T_T(k) + S_T(k) + t_T(k)$$

where $T_T(k)$ is the **trend component** that is obtained from the equation:

$$\nabla^{d+1} T_T(k) = 0$$

and is a polynomial of degree d with coefficients that adapt over time.

▷ The **seasonal component**, $S_T(k)$, is the solution to:

$$S_s(B) S_T(k) = 0 \tag{145}$$

whose solution is any function of period s with coefficients that add up to zero.

Interpreting the predictions - Seasonal processes

▷ In fact, an $S_T(k)$ sequence is a solution to (145) if it verifies:

$$\sum_{j=1}^s S_T(j) = \sum_{j=s+1}^{2s} S_T(j) = 0$$

and the coefficients $S_T(1), \dots, S_T(s)$ obtained by solving this equation are called seasonal coefficients.

▷ Finally, the **transitory component** will include the roots of polynomials $\Phi(B^s)$ and $\phi(B)$ and its expression is

$$t_T(k) = \sum_{i=1}^{p+P} A_i G_i^k$$

where the G_i^{-1} are the solutions of the equations $\Phi(B^s) = 0$ and $\phi(B) = 0$.

Interpreting the predictions - Seasonal processes

- ▶ A particular solution to equation (144) is $\beta_{d+1}k^{d+1}$, where the constant β_{d+1} is determined by the condition

$$\Phi(1)\phi(1)s(d+1)!\beta_{d+1} = c$$

since the result of applying the operator ∇^{d+1} to $\beta_{d+1}k^{d+1}$ is $(d+1)!\beta_{d+1}$ and if we apply the operator $S_s(B)$ to a constant we obtain it s times (or: s times this constant).

- ▶ Since the mean of the stationary series is $\mu = c/\Phi(1)\phi(1)$ the constant β_{d+1} is

$$\beta_{d+1} = \frac{c}{\Phi(1)\phi(1)s(d+1)!} = \frac{\mu}{s(d+1)!}.$$

which generalizes the results of the model with constant but without seasonality. This additional component is added to the trend term.

Interpreting the predictions - Seasonal processes

► To summarize, the general solution to the **final prediction equation** of a **seasonal process** has three components:

- 1 The **forecasted trend**, which is a polynomial of degree d that changes over time if there is no constant in the model and a polynomial of degree $d + 1$ with the coefficient of the highest order β_{d+1} deterministic and given by $\mu/s(d + 1)!$, with μ being the mean of the stationary series.
- 2 The **forecasted seasonality**, which will change with the initial conditions.
- 3 A short-term **transitory term**, which will be determined by the roots of the regular and seasonal AR operators.

► The general solution above can be utilized to obtain predictions for horizons $k > q + sQ - d - s - p - sP$.

Interpreting the predictions - Airline passenger model

- ▶ The most often used ARIMA seasonal model is called the **airline passenger model**:

$$\nabla \nabla_s z_t = (1 - \theta B)(1 - \Theta B^{12})a_t$$

whose equation for calculating predictions, for $k > 0$, is

$$\begin{aligned}\hat{z}_t(k) = & \hat{z}_t(k-1) + \hat{z}_t(k-12) - \hat{z}_t(k-13) - \\ & - \theta \hat{a}_t(k-1) - \Theta \hat{a}_t(k-12) + \theta \Theta \hat{a}_t(k-13).\end{aligned}$$

- ▶ Moreover, according to the above results we know that the prediction of this model can be written as:

$$\hat{z}_t(k) = \beta_0^{(t)} + \beta_1^{(t)}k + S_t^{(k)}. \quad (146)$$

- ▶ This prediction equation has 13 parameters. This prediction is the sum of a linear trend - which changes with the forecast origin - and eleven seasonal coefficients, which also change with the forecast origin, and add up to zero.

Interpreting the predictions - Airline passenger model

► Equation (146) is valid for $k > q + Qs = 13$, which is the future moment when the moving average terms disappear, but as we need $d + s = 13$ initial values to determine it, the equation is valid for $k > q + Qs - d - s = 0$, that is, for the whole horizon.

► To determine the coefficients $\beta_0^{(t)}$ and $\beta_1^{(t)}$ corresponding to a given origin and the seasonal coefficients, we can solve the system of equations obtained by setting the predictions for thirteen periods equal to structure (146), resulting in:

$$\begin{aligned}\hat{z}_t(1) &= \hat{\beta}_0^{(t)} + \hat{\beta}_1^{(t)} + S_1^{(t)} \\ \vdots &\quad \quad \quad \vdots \\ \hat{z}_t(12) &= \hat{\beta}_0^{(t)} + 12\hat{\beta}_1^{(t)} + S_{12}^{(t)} \\ \hat{z}_t(13) &= \hat{\beta}_0^{(t)} + 13\hat{\beta}_1^{(t)} + S_1^{(t)}\end{aligned}$$

and with these equations we can obtain $\beta_0^{(t)}$ and $\beta_1^{(t)}$ with the restriction $\sum S_j^{(t)} = 0$.

Interpreting the predictions - Airline passenger model

- ▶ Subtracting the first equation from the last and dividing by 12:

$$\hat{\beta}_1^{(t)} = \frac{\hat{z}_t(13) - \hat{z}_t(1)}{12} \quad (147)$$

and the monthly slope is obtained by dividing expected yearly growth, computed by the difference between the predictions of any month in two consecutive years.

- ▶ Adding the first 12 equations the seasonal coefficients are cancelled out, giving us:

$$\bar{z}_t = \frac{1}{12} \sum_{j=1}^{12} \hat{z}_t(j) = \hat{\beta}_0^{(t)} + \hat{\beta}_1^{(t)} \left(\frac{1 + \dots + 12}{12} \right)$$

which yields:

$$\hat{\beta}_0^{(t)} = \bar{z}_t - \frac{13}{2} \hat{\beta}_1^{(t)}.$$

- ▶ Finally, the seasonal coefficients are obtained by difference

$$S_j^{(t)} = \hat{z}_t(j) - \hat{\beta}_0^{(t)} - \hat{\beta}_1^{(t)}j$$

and it is straightforward to prove that they add up to zero within the year.

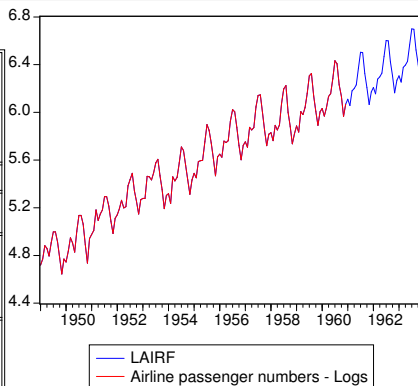
Example 73

We are going to calculate predictions using the airline passenger data from the book of Box and Jenkins (1976). The model estimated for these data is:

$$\nabla \nabla_{12} \ln z_t = (1 - .40B)(1 - .63B^{12})a_t$$

and we are going to generate predictions for three years assuming that this is the real model for the series.

Dependent Variable: D(LAIR,1,12) Method: Least Squares Date: 02/07/08 Time: 15:34 Sample (adjusted): 1950M02 1960M12 Included observations: 131 after adjustments Convergence achieved after 10 iterations Backcast: 1949M01 1950M01				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	-0.404855	0.080238	-5.045651	0.0000
SMA(12)	-0.631572	0.069841	-9.042955	0.0000
R-squared	0.371098	Mean dependent var	0.000291	
Adjusted R-squared	0.366223	S.D. dependent var	0.045848	
S.E. of regression	0.036500	Akaike info criterion	-3.767866	
Sum squared resid	0.171859	Schwarz criterion	-3.723970	
Log likelihood	248.7952	Durbin-Watson stat	1.933726	
Inverted MA Roots	.96 .48-.83i -.48+.83i -.96	.83-.48i .40 -.48-.83i -.83+.48i	.83+.48i .00+.96i -.83-.48i -.83+.48i	.48+.83i -.00-.96i -.83+.48i -.83+.48i



- ▷ The prediction function is a straight line plus seasonal coefficients. To calculate these parameters the table gives the predictions for the first 13 data points.

YEAR	J	F	M	A	M	J
61	6.11	6.05	6.18	6.19	6.23	6.36
YEAR	J	A	S	O	N	D
61	6.50	6.50	6.32	6.20	6.06	6.17
YEAR	J					
62	6.20					

- ▷ To calculate the yearly slope of the predictions we take the prediction for January, 1961, $\hat{z}_{144}(1) = 6.11$, and that of January, 1962, $\hat{z}_{144}(13) = 6.20$. Their difference is 0.09, which corresponds to an annual growth rate of 9.00%. The slope of the straight line is the monthly growth, which is $0.09/12 = 0.0075$.
- ▷ The seasonal factors are obtained by subtracting the trend from each prediction.

▷ The intercept is:

$$\hat{\beta}_0 = (6.11 + 6.05 + \dots + 6.17)/12 - \frac{13}{2}(0.0075) = 6.1904.$$

▷ The series trend follows the line: $P_{144}(k) = 6.1904 + 0.0075k$. Subtracting the value of the trend from each prediction we obtain the seasonal factors, which are shown in the table:

	<i>J</i>	<i>F</i>	<i>M</i>	<i>A</i>	<i>M</i>	<i>J</i>
$P_{144}(k)$	6.20	6.21	6.21	6.22	6.23	6.24
$\hat{z}_{144}(k)$	6.11	6.05	6.18	6.19	6.23	6.36
Seasonal coef.	-0.09	-0.16	-0.03	-0.03	0.00	0.12
	<i>J</i>	<i>A</i>	<i>S</i>	<i>O</i>	<i>N</i>	<i>D</i>
$P_{144}(k)$	6.24	6.25	6.26	6.27	6.27	6.28
$\hat{z}_{144}(k)$	6.50	6.50	6.32	6.20	6.06	6.17
Seasonal coef.	0.26	0.25	0.06	-0.07	-0.21	-0.11

▷ Notice that the lowest month is November, with a drop of 21% with respect to the yearly mean, and the two highest months are July and August, 26% and 25% higher than the yearly mean.

Variance of the predictions

- The **variance** of the predictions is easily calculated for MA(q) processes. Indeed, if $z_t = \theta_q(B) a_t$, we have:

$$z_{T+k} = a_{T+k} - \theta_1 a_{T+k-1} - \dots - \theta_q a_{T+k-q}$$

and taking conditional expectations on the observations up to time T and assuming that $q > k$, the unobserved innovations are cancelled due to having zero expectation, giving us

$$\hat{z}_T(k) = -\theta_k a_T - \theta_{k+1} a_{T-1} - \dots - \theta_q a_{T+k-q}.$$

- Subtracting these last two equations gives us the forecasting error:

$$e_T(k) = z_{T+k} - \hat{z}_T(k) = a_{T+k} - \theta_1 a_{T+k-1} - \dots - \theta_{k-1} a_{T+1} \quad (148)$$

Variance of the predictions

► Squaring the expression (148) and taking expectations we obtain the expected value of the square prediction error of z_{T+k} , which is equal to the variance of its distribution conditional on the observations until time T :

$$MSPE(e_T(k)) = E[(z_{T+k} - z_T(k))^2 | \mathbf{z}_T] = \sigma^2 (1 + \theta_1^2 + \dots + \theta_{k-1}^2) \quad (149)$$

► This idea can be extended to any ARIMA process as follows: Let $z_t = \psi(B)a_t$ be the $MA(\infty)$ representation of the process. Then:

$$z_{T+k} = \sum_0^\infty \psi_i a_{T+k-i} \quad (\psi_0 = 1), \quad (150)$$

► The optimal prediction is, taking conditional expectations on the first T observations,

$$\hat{z}_T(k) = \sum_0^\infty \psi_{k+j} a_{T-j} \quad (151)$$

Variance of the predictions

- ▷ Subtracting (151) from (150) the prediction error is obtained.

$$e_T(k) = z_{T+k} - \hat{z}_T(k) = a_{T+k} + \psi_1 a_{T+k-1} + \dots + \psi_{k-1} a_{T+1}$$

whose variance is:

$$\text{Var}(e_T(k)) = \sigma^2 (1 + \psi_1^2 + \dots + \psi_{k-1}^2). \quad (152)$$

- ▷ This equation shows that the uncertainty of a prediction is quite different for stationary and non-stationary models:

- In a stationary model $\psi_k \rightarrow 0$ if $k \rightarrow \infty$, and the long-term variance of the prediction converges at a constant value, the marginal variance of the process. This is a consequence of the long-term prediction being the mean of the process.
- For non-stationary models the series $\sum \psi_i^2$ is non-convergent and the long-term uncertainty of the prediction increases without limit. In the long-term we cannot predict the behavior of a non-stationary process.

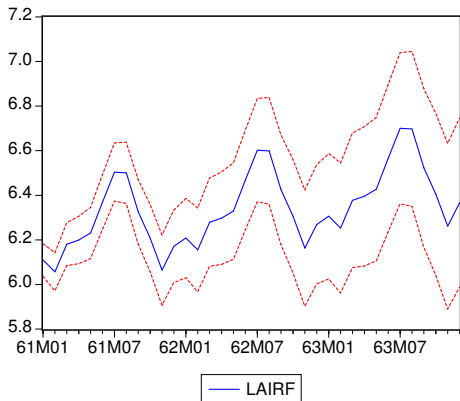
Distribution of the predictions

► If we assume that the distribution of the innovations is normal the above results allow us to calculate **confidence intervals** for the prediction.

► Thus z_{T+k} is a normal variable of mean $\hat{z}_T(k)$ and variance given by (152), and we can obtain the interval

$$\left(\hat{z}_T(k) \pm \lambda_{1-\alpha} \sqrt{\text{Var}(e_T(k))} \right)$$

where $\lambda_{1-\alpha}$ are the percentiles of the standard normal distribution.



Datafile airline.xls

Example 74

Given model $\nabla z_t = (1 - \theta B) a_t$, calculate the variance of the predictions to different periods.

▷ To obtain the coefficients of the $MA(\infty)$ representation, writing $z_t = \psi(B) a_t$, as in this model $a_t = (1 - \theta B)^{-1} \nabla z_t$, we have

$$z_t = \psi(B)(1 - \theta B)^{-1} \nabla z_t$$

that is,

$$(1 - \theta B) = \psi(B) \nabla = 1 + (\psi_1 - 1)B + (\psi_2 - \psi_1)B^2 + \dots$$

from which we obtain: $\psi_1 = 1 - \theta$, $\psi_2 = \psi_1$, ..., $\psi_k = \psi_{k-1}$.

▷ Therefore:

$$\text{Var}(e_T(k)) = \sigma^2(1 + (k-1)(1 - \theta^2))$$

and the variance of the prediction increases linearly with the horizon.

Example 75

Given the ARIMA model:

$$(1 - 0.2B) \nabla z_t = (1 - 0.8B) a_t$$

using $\sigma^2 = 4$, and assuming that $z_{49} = 30$, $z_{48} = 25$, $a_{49} = z_{49} - \hat{z}_{49/48} = -2$, obtain predictions starting from the last observed value for $t = 49$ and construct confidence intervals assuming normality.

▷ The expansion of the AR part is:

$$(1 - 0.2B)(1 - B) = 1 - 1.2B + 0.2B^2$$

and the model can be written:

$$z_t = 1.2z_{t-1} - 0.2z_{t-2} + a_t - 0.8a_{t-1}.$$

▷ Thus:

$$\hat{z}_{49}(1) = 1.2(30) - 0.2(25) - 0.8(-2) = 32.6,$$

$$\hat{z}_{49}(2) = 1.2(32.6) - 0.2(30) = 33.12,$$

$$\hat{z}_{49}(3) = 1.2(33.12) - 0.2(32.6) = 33.25,$$

$$\hat{z}_{49}(4) = 1.2(33.25) - 0.2(33.12) = 33.28.$$

▷ The confidence intervals of these forecasts require calculation of the coefficients $\psi(B)$. Equating the operators used in the $MA(\infty)$ we have $(1 - 1.2B + 0.2B^2)^{-1}(1 - 0.8B) = \psi(B)$, which implies:

$$(1 - 1.2B + 0.2B^2)(1 + \psi_1 B + \psi_2 B^2 + \dots) = (1 - 0.8B)$$

operating in the first member:

$$1 - B(1.2 - \psi_1) - B^2(1.2\psi_1 - 0.2 - \psi_2) - \\ - B^3(1.2\psi_2 - 0.2\psi_1 - \psi_3) - \dots = (1 - 0.8B).$$

and equating the powers of B , we obtain $\psi_1 = 0.4$, $\psi_2 = 0.28$, $\psi_3 = 0.33$.

▷ The variances of the prediction errors are:

$$\text{Var}(e_{49}(1)) = \sigma^2 = 4,$$

$$\text{Var}(e_{49}(2)) = \sigma^2 (1 + \psi_1^2) = 4 \times 1.16 = 4.64,$$

$$\text{Var}(e_{49}(3)) = \sigma^2 (1 + \psi_1^2 + \psi_2^2) = 4.95,$$

$$\text{Var}(e_{49}(4)) = \sigma^2 (1 + \psi_1^2 + \psi_2^2 + \psi_3^2) = 5.38,$$

▷ Assuming normality, the approximate intervals of the 95% for the four predictions are

$$\begin{array}{ll} (32.6 \pm 1.96 \times 2) & (33.12 \pm 1.96 \times \sqrt{4.64}) \\ (33.25 \pm 1.96 \times \sqrt{4.95}) & (33.28 \pm 1.96 \times \sqrt{5.38}). \end{array}$$

Forecast updating

► Let us assume that predictions are generated from time T for future periods $T + 1, \dots, T + j$. When value z_{T+1} is observed, we want to update our forecasts $\hat{z}_{T+2}, \dots, \hat{z}_{T+j}$ in light of this new information.

► According to (151) the prediction of z_{T+k} using information until T is:

$$\hat{z}_T(k) = \psi_k a_T + \psi_{k+1} a_{T-1} + \dots$$

whereas on observing value z_{T+1} and obtaining the prediction error, $a_{T+1} = z_{T+1} - \hat{z}_T(1)$, the **new prediction** for z_{T+k} , now from time $T + 1$, is:

$$\hat{z}_{T+1}(k-1) = \psi_{k-1} a_{T+1} + \psi_k a_T + \dots$$

textcolorbggreen► Subtracting these two expressions, we have:

$$\hat{z}_{T+1}(k-1) - \hat{z}_T(k) = \psi_{k-1} a_{T+1}.$$

Forecast updating

- Therefore, when we observe z_{T+1} and calculate a_{T+1} we can fit all the predictions by means of:

$$\hat{z}_{T+1}(k-1) = \hat{z}_T(k) + \psi_{k-1}a_{T+1}, \quad (153)$$

which indicates that the predictions are fitted by adding a fraction of the last prediction error obtained to the previous predictions.

- If $a_{T+1} = 0$ the predictions do not change.
- Equation (153) has an interesting interpretation:
- The two variables z_{T+1} and z_{T+k} have, given the information up to T , a joint normal distribution with expectations, $\hat{z}_T(1)$ and $\hat{z}_T(k)$, variances σ^2 and $\sigma^2(1 + \psi_1^2 + \dots + \psi_{k-1}^2)$ and covariance:

$$\begin{aligned} \text{cov}(z_{T+1}, z_{T+k}) &= E[(z_{T+1} - \hat{z}_T(1))(z_{T+k} - \hat{z}_T(k))] = \\ &= E[a_{T+1}(a_{T+k} + \psi_1 a_{T+k-1} + \dots)] = \sigma^2 \psi_{k-1}. \end{aligned}$$

- The **best prediction** of z_{T+k} given z_{T+1} and the information up to T can be calculated by regression, according to the expression:

$$E(z_{T+k}|z_{T+1}) = E(z_{T+k}) + \text{cov}(z_{T+1}, z_{T+k})\text{var}^{-1}(z_{T+1})(z_{T+1} - E(z_{T+1})) \quad (154)$$

where, to simplify the notation, we have not included the conditioning factor in all the terms for the information up to T , since it appears in all of them.

- Substituting, we obtain:

$$\hat{z}_{T+1}(k-1) = \hat{z}_T(k) + (\sigma^2\psi_{k-1})\sigma^{-2}a_{T+1}$$

which is equation (153).

Example 76

Adjust the predictions from Example 75 assuming that we observe the value z_{50} equal to 34. Thus:

$$a_{50} = z_{50} - \hat{z}_{49}(1) = 34 - 32.6 = 1.4,$$

and the new predictions are,

$$\hat{z}_{50}(1) = \hat{z}_{49}(2) + \psi_1 a_{50} = 33.12 + 0.4 \times 1.4 = 33.68,$$

$$\hat{z}_{50}(2) = \hat{z}_{49}(3) + \psi_2 a_{50} = 33.25 + 0.28 \times 1.4 = 33.64,$$

$$\hat{z}_{50}(3) = \hat{z}_{49}(4) + \psi_3 a_{50} = 33.28 + 0.33 \times 1.4 = 33.74,$$

with new confidence intervals $(33.68 \pm 1.96 \times 2)$, $(33.64 \pm 1.96 \times \sqrt{4.64})$ and $(33.74 \pm 1.96 \times \sqrt{4.95})$.

► *We observe that by committing an error of underpredicting in the prediction of z_{50} , the following predictions are revised upwardly.*

Measuring predictability

▷ Any stationary ARMA processes, z_t , can be decomposed in this way:

$$z_t = \hat{z}_{t-1}(1) + a_t,$$

which expresses the value of the series at each moment as a **sum** of the two **independent components**:

- the one step ahead prediction, knowing the past values and the parameters of the process, and
- the innovations, which are independent of the past of the series.

▷ As a result, we can write:

$$\sigma_z^2 = \sigma_{\hat{z}}^2 + \sigma^2$$

which decomposes the variance of the series, σ_z^2 , into two **independent sources of variability**:

- that of the predictable part, $\sigma_{\hat{z}}^2$, and
- that of the unpredictable part, σ^2 .

Measuring predictability

- ▶ Box and Tiao (1977) proposed measuring the **predictability** of a stationary series using the quotient between the variance of the predictable part and the total variance:

$$P = \frac{\sigma_{\hat{z}}^2}{\sigma_z^2} = 1 - \frac{\sigma^2}{\sigma_z^2}. \quad (155)$$

- ▶ This coefficient indicates the proportion of variability of the series that can be predicted from its history.

- ▶ As in an ARMA process $\sigma_z^2 = \sigma^2 \sum \psi_i^2$, coefficient P can be written as:

$$P = 1 - \left(\sum \psi_i^2 \right)^{-1}.$$

- ▶ For example, for an AR(1) process, since $\sigma_z^2 = \sigma^2 / (1 - \phi^2)$, we have:

$$P = 1 - (1 - \phi^2) = \phi^2,$$

and if ϕ is near zero the process will tend to white noise and the predictability will be close to zero, and if ϕ is near one, the process will tend to a random walk, and P will be close to one.

Measuring predictability

- ▶ The measure P is not helpful for integrated, non-stationary ARIMA processes because then the marginal variance tends to the infinite and the value of P is always one.
- ▶ A more general definition of predictability by relaxing the forecast horizon in the numerator and denominator:
 - instead of assuming a forecast horizon of one we can assume a general horizon k ;
 - instead of putting the prediction error with infinite horizon in the denominator we can plug in the prediction error with horizon $k + h$, for certain values of $h \geq 1$.
- ▶ Then, we define the **predictability** of a time series with horizon k obtained through h additional observations using:

$$P(k, h) = 1 - \frac{\text{var}(e_t(k))}{\text{var}(e_t(k + h))}.$$

Measuring predictability

▷ For example, let us assume an ARIMA process and for convenience sake we will write it as $z_t = \psi(B)a_t$, although in general the series $\sum \psi_i^2$ will not be convergent.

▷ Nevertheless, using (151) we can write:

$$P(k, h) = 1 - \frac{\sum_{i=0}^k \psi_i^2}{\sum_{i=0}^{k+h} \psi_i^2} = \frac{\sum_{i=k+1}^{k+h} \psi_i^2}{\sum_{i=0}^{k+h} \psi_i^2}.$$

▷ This statistic $P(k, h)$ measures the advantages of having h additional observations for the prediction with horizon k . Specifically, for $k = 1$ and $h = \infty$ the statistic P defined in (155) is obtained.

Example 77

Calculate the predictability of the process $\nabla z_t = (1 - .8B)a_t$ for one and two steps ahead as a function of h .

▷ The one step ahead prediction is:

$$P(1, h) = \frac{\sum_{i=2}^{h+1} \psi_i^2}{\sum_{i=0}^{h+1} \psi_i^2} = \frac{.04h}{1 + .04(h+1)}.$$

▷ This function indicates that if $h = 1$, $P(1, 1) = .04/1.08 = .037$, and having an additional observation, or going from two steps to one in the prediction, reduces the prediction error 3.7%.

▷ If we have 10 more observations, the error reduction for $h = 10$ is of $P(1, 10) = .4/(1.44) = .2778$, 27.78%. If $h = 30$, then $P(1, 30) = 1.2/(2.24) = .4375$, and when $h \rightarrow \infty$, then $P(1, h) \rightarrow 1$.