

DYNAMIC MODELS FOR STATIONARY TIME SERIES

ECONOMETRICS C ♦ LECTURE NOTE 4

HEINO BOHN NIELSEN

FEBRUARY 6, 2012

Lecture note 2 considered the statistical analysis of regression models for time series data, and gave sufficient conditions for the ordinary least squares (OLS) principle to provide consistent, unbiased, and asymptotically normal estimators. The main assumptions imposed on the data were stationarity and weak dependence, and the main assumption on the model was some degree of exogeneity of the regressors. This lecture note introduces some popular classes of dynamic time series models and goes into details with the mathematical structure and the economic interpretation of the models. To introduce the ideas we first consider the moving average (MA) model, which is probably the simplest dynamic model. Next, we consider the popular autoregressive (AR) model and present conditions for this model to generate stationary time series. We also discuss the relationship between AR and MA models and introduce mixed ARMA models allowing for both AR and MA terms. Finally we look at the class of autoregressive models with additional explanatory variables, the so-called autoregressive distributed lag (ADL) models, and derive the dynamic multipliers and the steady-state solution.

OUTLINE

§1	Estimating Dynamic Effects	2
§2	Univariate Moving Average Models	3
§3	Univariate Autoregressive Models	6
§4	ARMA Estimation and Model Selection	13
§5	Univariate Forecasting	18
§6	The Autoregressive Distributed Lag Model	20
§7	Concluding Remarks	26

1 ESTIMATING DYNAMIC EFFECTS

A central feature of time series data is a pronounced time dependence, and it follows that shocks may have dynamic effects. Below we present a number of popular time series models for the estimation of dynamic responses after an impulse to the model. As a starting point it is useful to distinguish univariate from multivariate models.

1.1 UNIVARIATE MODELS

Univariate models consider a single time series, y_1, y_2, \dots, y_T , and model the systematic variation in y_t as a function of its own past, i.e. $E[y_t \mid y_{t-1}, y_{t-2}, \dots]$. Although the univariate approach is limited it nevertheless serves two important purposes.

The first purpose is to serve as a *descriptive tool* to characterize the dynamic properties of a time series. We may for example be interested in the strength of the time dependence, or *persistence*, of the time series, and we often want to assess whether the main assumption of stationarity is likely to be fulfilled. In empirical applications the univariate description often precedes a more elaborate multivariate analysis of the data.

The second purpose is *forecasting*: To forecast y_{T+1} at time T based on a multivariate model for y_t conditional on x_t it is obviously necessary to know x_{T+1} , which is generally not in the information set at time T . Univariate models offer the possibility of forecasts of y_t based solely on its own past. These forecasts are simple extrapolations based on the systematic variation in the past.

Below we look at two classes of univariate models: We first introduce the *moving average* model in §2. This is the simplest class of dynamic models and the conditions for stationarity is straightforwardly obtained. We next present the *autoregressive* model in §3 and derive the stationarity condition by referring to MA case. We emphasize the relationship between the two models and present the ARMA class of mixed models allowing for both autoregressive and moving-average terms. We continue in §4 and §5 to discuss estimation and forecasting.

1.2 MULTIVARIATE MODELS

An alternative to univariate models is to consider a model for y_t given an information set including other explanatory variables, the vector x_t say. These models are obviously more interesting from an economic point of view, and they allow the derivation of dynamic multipliers,

$$\frac{\partial y_t}{\partial x_t}, \frac{\partial y_{t+1}}{\partial x_t}, \frac{\partial y_{t+2}}{\partial x_t}, \dots$$

An important example could be the analysis of monetary policy in which case x_t could be the policy interest rate and y_t the variable of interest e.g. unemployment or inflation. In §6 we look at a model for y_t conditional on x_t and the past, i.e. $E[y_t \mid y_{t-1}, y_{t-2}, \dots, x_t, x_{t-1}, x_{t-2}, \dots]$. This is known as the *autoregressive distributed lag* (ADL) model and is the workhorse in single-equation dynamic modelling.

2 UNIVARIATE MOVING AVERAGE MODELS

Possibly the simplest class of univariate time series models is the *moving average* (MA) model. To discuss the MA model we first define an IID error process ϵ_t with mean zero and constant variance σ^2 . An error term with these properties, which we will write as $\epsilon_t \sim \text{IID}(0, \sigma^2)$, is often referred to as a *white noise* process and in the time series literature ϵ_t is sometimes called an *innovation* or a *shock* to the process. Recall that a (weakly) stationary stochastic process is characterized by constant mean, variance, and autocovariances (unconditionally), and the white noise process, ϵ_t , is obviously stationary.

2.1 FINITE MA-MODELS

Next we define the moving average model of order q , MA(q), by the equation

$$y_t = \mu + \epsilon_t + \alpha_1 \epsilon_{t-1} + \alpha_2 \epsilon_{t-2} + \dots + \alpha_q \epsilon_{t-q}, \quad t = 1, 2, \dots, T, \quad (1)$$

which states that y_t is a moving average of q past shocks. The equation in (1) defines y_t for $t = 1, 2, \dots, T$, which means that we need q initial values for the unobserved error process, and it is customary to assume that $\epsilon_{-(q-1)} = \epsilon_{-(q-2)} = \dots = \epsilon_{-1} = \epsilon_0 = 0$. The equation in (1) includes a constant term, but the deterministic specification could be made more general and the equation could include e.g. a linear trend or seasonal dummies. We note that α_i is the response on y_t of an impulse to ϵ_{t-i} , and the sequence of α 's are often referred to as the *impulse response function*.

The stochastic process y_t can be characterized directly from (1). The expectation is given by

$$E[y_t] = E[\mu + \epsilon_t + \alpha_1 \epsilon_{t-1} + \alpha_2 \epsilon_{t-2} + \dots + \alpha_q \epsilon_{t-q}] = \mu,$$

which is just the constant term in (1). The variance can be derived by inserting (1) in the definition,

$$\begin{aligned} \gamma_0 := V(y_t) &= E[(y_t - \mu)^2] \\ &= E[(\epsilon_t + \alpha_1 \epsilon_{t-1} + \alpha_2 \epsilon_{t-2} + \dots + \alpha_q \epsilon_{t-q})^2] \\ &= E[\epsilon_t^2] + E[\alpha_1^2 \epsilon_{t-1}^2] + E[\alpha_2^2 \epsilon_{t-2}^2] + \dots + E[\alpha_q^2 \epsilon_{t-q}^2] \\ &= (1 + \alpha_1^2 + \alpha_2^2 + \dots + \alpha_q^2) \sigma^2. \end{aligned} \quad (2)$$

The third line in the derivation follows from the independent distribution of ϵ_t so that all covariances are zero, $E[\epsilon_t \epsilon_{t-h}] = 0$ for $h \neq 0$. This implies that the variance of the sum is the sum of the variances. The autocovariances of y_t can be found in a similar way, noting

again that all cross terms are zero in expectation, i.e.

$$\begin{aligned}
\gamma_1 := \text{Cov}(y_t, y_{t-1}) &= E[(y_t - \mu)(y_{t-1} - \mu)] \\
&= E[(\epsilon_t + \alpha_1\epsilon_{t-1} + \dots + \alpha_q\epsilon_{t-q})(\epsilon_{t-1} + \alpha_1\epsilon_{t-2} + \dots + \alpha_q\epsilon_{t-q-1})] \\
&= (\alpha_1 + \alpha_2\alpha_1 + \alpha_3\alpha_2 + \dots + \alpha_q\alpha_{q-1})\sigma^2, \\
\gamma_2 := \text{Cov}(y_t, y_{t-2}) &= E[(y_t - \mu)(y_{t-2} - \mu)] \\
&= E[(\epsilon_t + \alpha_1\epsilon_{t-1} + \dots + \alpha_q\epsilon_{t-q})(\epsilon_{t-2} + \alpha_1\epsilon_{t-3} + \dots + \alpha_q\epsilon_{t-q-2})] \\
&= (\alpha_2 + \alpha_3\alpha_1 + \alpha_4\alpha_2 + \dots + \alpha_q\alpha_{q-2})\sigma^2, \\
&\vdots \\
\gamma_q := \text{Cov}(y_t, y_{t-q}) &= \alpha_q\sigma^2,
\end{aligned}$$

while the autocovariances are zero for larger lag lengths, $\gamma_k := \text{Cov}(y_t, y_{t-k}) = 0$ for $k > q$.

We note that the mean and variance are constant, and the autocovariance γ_k depends on k but not on t . By this we conclude that the MA(q) process is stationary by construction. The intuition is that y_t is a linear combination of stationary terms, and with constant weights the properties of y_t are independent of t .

A common way to characterize the properties of the time series is to use the autocorrelation function, ACF. It follows from the covariances that the ACF is given by the sequence

$$\rho_k := \frac{\gamma_k}{\gamma_0} = \frac{\alpha_k + \alpha_{k+1}\alpha_1 + \alpha_{k+2}\alpha_2 + \dots + \alpha_q\alpha_{q-k}}{1 + \alpha_1^2 + \alpha_2^2 + \dots + \alpha_q^2}, \quad \text{for } k \leq q,$$

while $\rho_k = 0$ for $k > q$. We note that the MA(q) process has a memory of q periods.

To illustrate the appearance of MA processes, Figure 1 reports some simulated series and their theoretical and estimated autocorrelation functions. We note that the appearance of the MA(q) process depends on the order of the process, q , and on the parameters, $\alpha_1, \dots, \alpha_q$, but the shocks to the white noise process are in all cases recognizable, and the properties do not change fundamentally. Also note that the MA(q) process has a memory of q periods, in the sense that it takes q time periods before the effect of a shock ϵ_t has disappeared.

A memory of exactly q periods may be difficult to rationalize in many economic settings, and the MA model is not used very often in econometric applications.

2.2 INFINITE MA-MODELS

The arguments for the finite MA-process can be extended to the *infinite moving average process*, MA(∞). In this case, however, we have to ensure that the variance of y_t is bounded, $V(y_t) = (1 + \alpha_1^2 + \alpha_2^2 + \dots)\sigma^2 < \infty$, which requires that $\sum_{j=0}^{\infty} \alpha_j^2 < \infty$ so that the infinite sum converges.

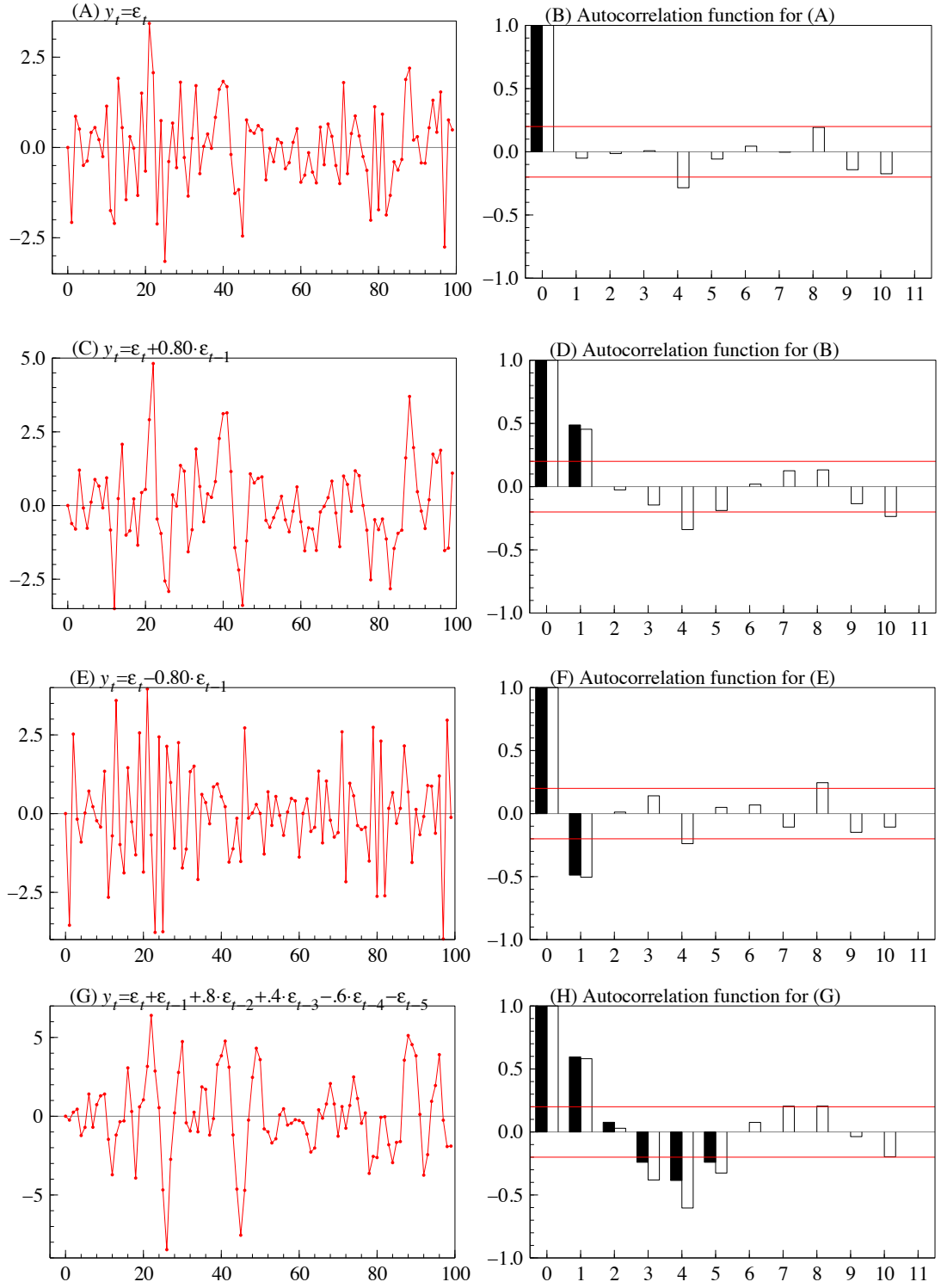


Figure 1: Examples of simulated $MA(q)$ processes. ■ indicates the theoretical ACF while □ indicates the estimated ACF. Horizontal lines are the 95% confidence bounds for zero autocorrelations derived for an IID process.

3 UNIVARIATE AUTOREGRESSIVE MODELS

The autoregressive (AR) model with p lags is defined by the equation

$$y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} + \epsilon_t, \quad t = 1, 2, \dots, T, \quad (3)$$

where $\epsilon_t \sim \text{iID}(0, \sigma^2)$ is a white noise error process. We use again the convention that the equation in (3) holds for observations y_1, y_2, \dots, y_T , which means that we have observed also the p previous values, $y_{-(p-1)}, y_{-(p-2)}, \dots, y_{-1}, y_0$; they are referred to as *initial values* for the equation.

The assumptions imply that we can think of the systematic part as the best linear prediction of y_t given the past,

$$E[y_t \mid y_{t-1}, y_{t-2}, \dots] = E[y_t \mid y_{t-1}, y_{t-2}, \dots, y_{t-p}] = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p}.$$

We note that the first p lags capture all the information in the past, and that the conditional expectation is a linear function of the information set.

3.1 THE AR(1) MODEL

The analysis of the autoregressive model is more complicated than the analysis of moving average models and to simplify the analysis we focus on the autoregressive model for $p = 1$ lag, the so-called first order autoregressive, AR(1), model:

$$y_t = \delta + \theta y_{t-1} + \epsilon_t. \quad (4)$$

In this case the first lag captures all the information in the past.

The only exogenous (or forcing) variable in (4) is the error term and the development of the time series y_t is determined solely by the sequence of innovations $\epsilon_1, \dots, \epsilon_T$. To make this point explicit, we can find the solution for y_t in terms of the innovations and the initial value. To do this we recursively substitute the expressions for y_{t-k} , $k = 1, 2, \dots$, to obtain the solution

$$\begin{aligned} y_t &= \delta + \theta y_{t-1} + \epsilon_t \\ &= \delta + \theta(\delta + \theta y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &= (1 + \theta) \delta + \epsilon_t + \theta \epsilon_{t-1} + \theta^2 y_{t-2} \\ &= (1 + \theta) \delta + \epsilon_t + \theta \epsilon_{t-1} + \theta^2 (\delta + \theta y_{t-3} + \epsilon_{t-2}) \\ &= (1 + \theta + \theta^2) \delta + \epsilon_t + \theta \epsilon_{t-1} + \theta^2 \epsilon_{t-2} + \theta^3 y_{t-3} \\ &\vdots \\ &= (1 + \theta + \theta^2 + \theta^3 + \dots + \theta^{t-1}) \delta + \epsilon_t + \theta \epsilon_{t-1} + \theta^2 \epsilon_{t-2} + \dots + \theta^{t-1} \epsilon_1 + \theta^t y_0. \end{aligned} \quad (5)$$

We see that y_t is given by a deterministic term, a moving average of past innovations, and a term involving the initial value. Due to the moving average structure the solution is often referred to as the *moving average representation* of y_t .

If we for a moment make the abstract assumption that the process y_t started in the remote infinite past, then we may state the solution as an infinite sum,

$$y_t = (1 + \theta + \theta^2 + \theta^3 + \dots) \delta + \epsilon_t + \theta \epsilon_{t-1} + \theta^2 \epsilon_{t-2} + \dots, \quad (6)$$

which we recognize as an infinite moving average process. It follows from the result for infinite MA processes that the AR(1) process is stationary if the MA terms converge to zero so that the infinite sum converges. That is the case if $|\theta| < 1$ which is known as the *stationarity condition* for an AR(1) model. In the analysis of the AR(1) model below we assume stationarity and impose this condition. We emphasize that while a finite MA process is always stationary, stationarity of the AR process requires conditions on the parameters.

The properties of the time series y_t can again be found directly from the moving average representation. The expectation of y_t given the initial value is

$$E[y_t | y_0] = (1 + \theta + \theta^2 + \theta^3 + \dots + \theta^{t-1}) \delta + \theta^t y_0,$$

where the last term involving the initial value converges to zero for t increasing, $\theta^t y_0 \rightarrow 0$. The unconditional mean is the expectation of the convergent geometric series in (6), i.e.

$$E[y_t] = \frac{\delta}{1 - \theta} =: \mu.$$

This is *not* the constant term of the model, δ , it also depends on the autoregressive parameter, θ . We hasten to note that the constant unconditional mean is not defined if $\theta = 1$, but that case is ruled out by the stationarity condition.

The unconditional variance can also be found from the solution in (6). Using the definition we obtain

$$\begin{aligned} \gamma_0 := V(y_t) &= E[(y_t - \mu)^2] \\ &= E[(\epsilon_t + \theta \epsilon_{t-1} + \theta^2 \epsilon_{t-2} + \theta^3 \epsilon_{t-3} + \dots)^2] \\ &= \sigma^2 + \theta^2 \sigma^2 + \theta^4 \sigma^2 + \theta^6 \sigma^2 + \dots \\ &= (1 + \theta^2 + \theta^4 + \theta^6 + \dots) \sigma^2, \end{aligned}$$

which is again a convergent geometric series with the limit

$$\gamma_0 = \frac{\sigma^2}{1 - \theta^2}.$$

The autocovariances can also be found from (6):

$$\begin{aligned} \gamma_1 := Cov(y_t, y_{t-1}) &= E[(y_t - \mu)(y_{t-1} - \mu)] \\ &= E[(\epsilon_t + \theta \epsilon_{t-1} + \theta^2 \epsilon_{t-2} + \dots)(\epsilon_{t-1} + \theta \epsilon_{t-2} + \theta^2 \epsilon_{t-3} + \dots)] \\ &= \theta \sigma^2 + \theta^3 \sigma^2 + \theta^5 \sigma^2 + \theta^7 \sigma^2 + \dots \\ &= \theta \gamma_0, \end{aligned}$$

where we again use that the covariances are zero: $E[\epsilon_t \epsilon_{t-h}] = 0$ for $h \neq 0$. Likewise it follows that $\gamma_k := \text{Cov}(y_t, y_{t-k}) = \theta^k \gamma_0$. The autocorrelation function, ACF, of a stationary AR(1) is given by

$$\rho_k := \frac{\gamma_k}{\gamma_0} = \frac{\theta^k \gamma_0}{\gamma_0} = \theta^k,$$

which is an exponentially decreasing function if $|\theta| < 1$. It is a general result that the autocorrelation function goes exponentially to zero for a stationary autoregressive time series.

Graphically the results imply that a stationary time series will fluctuate around a constant mean with a constant variance. Non-zero autocorrelations imply that consecutive observations are correlated and the fluctuations may be systematic, but over time the process will not deviate too much from the unconditional mean. This is often phrased as the process being *mean reverting* and we also say that the process has an *attractor*, defined as a steady state level to which it will eventually return: In this case the unconditional mean, μ .

Figure 2 shows examples of first order autoregressive processes. Note that the appearance depends more fundamentally on the autoregressive parameter. For the non-stationary case, $\theta = 1$, the process wanders arbitrarily up and down with no attractor; this is known as a random walk. If $|\theta| > 1$ the process for y_t is explosive. We note that there may be marked differences between the true and estimated ACF in small samples.

3.2 THE GENERAL CASE

The results for a general AR(p) model are most easily presented in terms of the lag-polynomial introduced in Box 1. We rewrite the model in (3) as

$$\begin{aligned} y_t &= \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} + \epsilon_t \\ y_t - \theta_1 y_{t-1} - \theta_2 y_{t-2} - \dots - \theta_p y_{t-p} &= \delta + \epsilon_t \\ \theta(L)y_t &= \delta + \epsilon_t, \end{aligned} \tag{7}$$

where $\theta(z) := 1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_p z^p$ is the *autoregressive polynomial*.

The AR(p) model can be written as an infinite moving average model, MA(∞), if the autoregressive polynomial can be inverted. As presented in Box 1 we can write the inverse polynomial as

$$\theta^{-1}(z) = 1 + c_1 z + c_2 z^2 + c_3 z^3 + c_4 z^4 + \dots,$$

where the coefficients converge to zero. It follows that we can write the AR(p) model as

$$\begin{aligned} y_t &= \theta^{-1}(L) (\delta + \epsilon_t) \\ &= (1 + c_1 L + c_2 L^2 + c_3 L^3 + c_4 L^4 + \dots) (\delta + \epsilon_t) \\ &= (1 + c_1 + c_2 + c_3 + c_4 + \dots) \delta + \epsilon_t + c_1 \epsilon_{t-1} + c_2 \epsilon_{t-2} + c_3 \epsilon_{t-3} + \dots, \end{aligned} \tag{8}$$

BOX 1: LAG POLYNOMIALS AND CHARACTERISTIC ROOTS

A useful tool in time series analysis is the *lag-operator*, L , that has the property that it lags a variable one period, i.e. $Ly_t = y_{t-1}$ and $L^2 y_t = y_{t-2}$. The lag operator is related to the well-known first difference operator, $\Delta = 1 - L$, and $\Delta y_t = (1 - L)y_t = y_t - y_{t-1}$.

Using the lag-operator we can write the AR(1) model as

$$\begin{aligned} y_t - \theta y_{t-1} &= \delta + \epsilon_t \\ (1 - \theta L)y_t &= \delta + \epsilon_t, \end{aligned}$$

where $\theta(z) := 1 - \theta z$ is a (first order) polynomial. For higher order autoregressive processes the polynomial will have the form

$$\theta(z) := 1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_p z^p.$$

Note that the polynomial evaluated in $z = 1$ is $\theta(1) = 1 - \theta_1 - \theta_2 - \dots - \theta_p$, which is one minus the sum of the coefficients.

The *characteristic equation* is defined as the polynomial equation, $\theta(z) = 0$, and the p solutions, z_1, \dots, z_p , are denoted the *characteristic roots*. For the AR(1) model the characteristic root is given by

$$\theta(z) = 1 - \theta z = 0 \quad \Rightarrow \quad z = \theta^{-1},$$

which is just the inverse coefficient. The polynomial of an AR(p) model has p roots, and for $p \geq 2$ they may be complex numbers of the form $z_j = a_j \pm b_j \sqrt{-1}$. Recall, that the roots can be used to factorize the polynomial, and the characteristic polynomial can be rewritten as

$$\theta(z) = 1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_p z^p = (1 - \phi_1 z)(1 - \phi_2 z) \cdots (1 - \phi_p z), \quad (\text{B1-1})$$

where $\phi_j = z_j^{-1}$ denotes an *inverse root*.

The usual results for convergent geometric series also hold for expressions involving the lag-operator, and if $|\theta| < 1$ it holds that

$$1 + \theta L + \theta^2 L^2 + \theta^3 L^3 + \dots = \frac{1}{1 - \theta L} =: \theta^{-1}(L), \quad (\text{B1-2})$$

where the right hand side is called the *inverse polynomial*. The inverse polynomial is infinite and it exists if the terms on the left hand side converges to zero, i.e. if $|\theta| < 1$. Higher order polynomials may also be inverted if the infinite sum converges. Based on the factorization in (B1-1) we note that the general polynomial $\theta(z)$ is invertible if each of the factors are invertible, i.e. of all the inverse roots are smaller than one in absolute value, $|\phi_j| < 1$, $j = 1, 2, \dots, p$. The general inverse polynomial has infinitely many terms,

$$\theta^{-1}(z) = 1 + c_1 z + c_2 z^2 + c_3 z^3 + c_4 z^4 + \dots,$$

where the coefficients can be found as complicated functions of the autoregressive parameters:

$$c_1 = \theta_1, \quad c_2 = c_1 \theta_1 + \theta_2, \quad c_3 = c_2 \theta_1 + c_1 \theta_2 + \theta_3, \quad c_4 = c_3 \theta_1 + c_2 \theta_2 + c_1 \theta_3 + \theta_4,$$

etc., where we insert $\theta_k = 0$ for $k > p$. Technically, the MA-coefficients can be derived from the multiplication of the inverse factors of the form (B1-2). These calculations are straightforward but tedious, and will not be presented here.

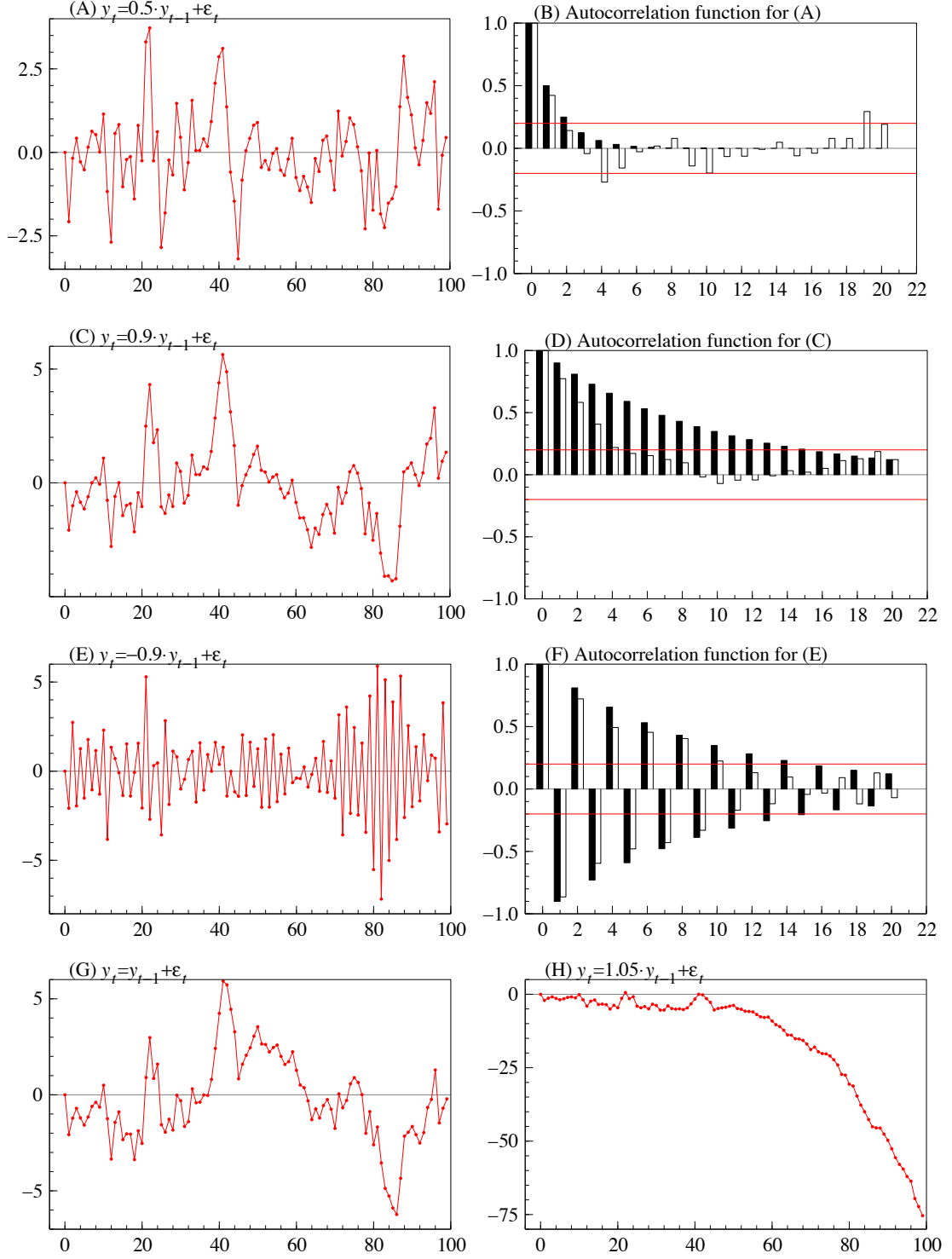


Figure 2: Examples of simulated $AR(p)$ processes. ■ indicates the theoretical ACF while □ indicates the estimated ACF. Horizontal lines are the 95% confidence bounds for zero autocorrelations derived for an IID process.

where we have used that δ is constant and $L\delta = \delta$. We recognize (8) as a moving average representation. We note that the AR(p) model is stationary if the moving average coefficients converge to zero; and that is automatically the case for the inverse polynomial. We can therefore state the *stationarity condition* for an AR(p) model as the requirement that the autoregressive polynomial is invertible, i.e. that the p roots of the autoregressive polynomial are larger than one in absolute value, $|z_j| > 1$, or that the inverse roots are smaller than one, $|\phi_j| < 1$.

Note that the moving average coefficients measure the dynamic impact of a shock to the process,

$$\frac{\partial y_t}{\partial \epsilon_t} = 1, \quad \frac{\partial y_t}{\partial \epsilon_{t-1}} = c_1, \quad \frac{\partial y_t}{\partial \epsilon_{t-2}} = c_2, \dots,$$

and the sequence of MA-coefficients, c_1, c_2, c_3, \dots , is also known as the *impulse-responses* of the process. Note that the stationarity condition implies that the impulse response function dies out eventually, and the process is *weakly dependent*.

Again we can find that properties of the process from the MA-representation. As an example, the constant mean is given by

$$\mu := E[y_t] = (1 + c_1 + c_2 + c_3 + c_4 + \dots) \delta \rightarrow \frac{\delta}{\theta(1)} = \frac{\delta}{1 - \theta_1 - \theta_2 - \dots - \theta_p}.$$

For this to be defined we require that $z = 1$ is not a root of the autoregressive polynomial, but that is ensured by the stationarity condition. The variance is given by $\gamma_0 := V(y_t) = (1 + c_1^2 + c_2^2 + c_3^2 + c_4^2 + \dots) \sigma^2$.

3.3 ARMA AND ARIMA MODELS

At this point it is worth emphasizing the duality between AR and MA models. We have seen that a stationary AR(p) model can always be represented by a MA(∞) model because the autoregressive polynomial $\theta(z)$ can be inverted. We may also write the MA model using a lag polynomial,

$$y_t = \mu + \alpha(L)\epsilon_t,$$

where $\alpha(z) := 1 + \alpha_1 z + \dots + \alpha_q z^q$ is a polynomial. If $\alpha(z)$ is invertible (i.e. if all the inverse roots of $\alpha(z) = 0$ are smaller than one), then the MA(q) model can also be represented by an AR(∞) model. As a consequence we may approximate a MA model with an autoregression with many lags; or we may alternatively represent a long autoregression with a shorter MA model.

The AR(p) and MA(q) models can also be combined into a so-called autoregressive moving average, ARMA(p,q), model,

$$y_t = \delta + \theta_1 y_{t-1} + \dots + \theta_p y_{t-p} + \epsilon_t + \alpha_1 \epsilon_{t-1} + \dots + \alpha_q \epsilon_{t-q}.$$

This is a very flexible class of models that is capable of representing many different patterns of autocovariances. Again we can write the model in terms of lag-polynomials as

$$\theta(L)(y_t - \mu) = \alpha(L)\epsilon_t,$$

BOX 2: AUTOCORRELATIONS FROM YULE-WALKER EQUATIONS

The presented approach for calculation of autocovariances and autocorrelations is totally general, but it is sometimes difficult to apply by hand because it requires that the MA-representation is derived. An alternative way to calculate autocorrelations is based on the so-called *Yule-Walker equations*, which are obtained directly from the model. To illustrate, this box derives the autocorrelations for an AR(2) model.

Consider the AR(2) model given by

$$y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \epsilon_t. \quad (\text{B2-1})$$

First we find the mean by taking expectations, $E[y_t] = \delta + \theta_1 E[y_{t-1}] + \theta_2 E[y_{t-2}] + E[\epsilon_t]$. Assuming stationarity, $E[y_t] = E[y_{t-1}]$, it follows that

$$\mu := E[y_t] = \frac{\delta}{1 - \theta_1 - \theta_2}.$$

Next we define a new process as the deviation from the mean, $\tilde{y}_t := y_t - \mu$, so that

$$\tilde{y}_t = \theta_1 \tilde{y}_{t-1} + \theta_2 \tilde{y}_{t-2} + \epsilon_t. \quad (\text{B2-2})$$

Now remember that $V(y_t) = E[(y_t - \mu)^2] = E[\tilde{y}_t^2]$, and we do all calculations on (B2-2) rather than (B2-1). If we multiply both sides of (B2-2) with \tilde{y}_t and take expectations, we find

$$\begin{aligned} E[\tilde{y}_t^2] &= \theta_1 E[\tilde{y}_{t-1} \tilde{y}_t] + \theta_2 E[\tilde{y}_{t-2} \tilde{y}_t] + E[\epsilon_t \tilde{y}_t] \\ \gamma_0 &= \theta_1 \gamma_1 + \theta_2 \gamma_2 + \sigma^2, \end{aligned} \quad (\text{B2-3})$$

where we have used the definitions and that $E[\epsilon_t \tilde{y}_t] = E[\epsilon_t (\theta_1 \tilde{y}_{t-1} + \theta_2 \tilde{y}_{t-2} + \epsilon_t)] = \sigma^2$. If we multiply instead with \tilde{y}_{t-1} , \tilde{y}_{t-2} , and \tilde{y}_{t-3} , we obtain

$$E[\tilde{y}_t \tilde{y}_{t-1}] = \theta_1 E[\tilde{y}_{t-1} \tilde{y}_{t-1}] + \theta_2 E[\tilde{y}_{t-2} \tilde{y}_{t-1}] + E[\epsilon_t \tilde{y}_{t-1}] \Leftrightarrow \gamma_1 = \theta_1 \gamma_0 + \theta_2 \gamma_1 \quad (\text{B2-4})$$

$$E[\tilde{y}_t \tilde{y}_{t-2}] = \theta_1 E[\tilde{y}_{t-1} \tilde{y}_{t-2}] + \theta_2 E[\tilde{y}_{t-2} \tilde{y}_{t-2}] + E[\epsilon_t \tilde{y}_{t-2}] \Leftrightarrow \gamma_2 = \theta_1 \gamma_1 + \theta_2 \gamma_0 \quad (\text{B2-5})$$

$$E[\tilde{y}_t \tilde{y}_{t-3}] = \theta_1 E[\tilde{y}_{t-1} \tilde{y}_{t-3}] + \theta_2 E[\tilde{y}_{t-2} \tilde{y}_{t-3}] + E[\epsilon_t \tilde{y}_{t-3}] \Leftrightarrow \gamma_3 = \theta_1 \gamma_2 + \theta_2 \gamma_1. \quad (\text{B2-6})$$

The set of equations (B2-3)-(B2-6) is known as the Yule-Walker equations. To find the variance we can substitute γ_1 and γ_2 into (B2-3) and solve. This is a bit tedious, however, and will not be done here. To find the autocorrelations, $\rho_k := \gamma_k / \gamma_0$, just divide the Yule-Walker equations with γ_0 to obtain

$$\rho_1 = \theta_1 + \theta_2 \rho_1, \quad \rho_2 = \theta_1 \rho_1 + \theta_2, \quad \text{and} \quad \rho_k = \theta_1 \rho_{k-1} + \theta_2 \rho_{k-2}, \quad k \geq 3,$$

or by collecting terms, that

$$\rho_1 = \frac{\theta_1}{1 - \theta_2}, \quad \rho_2 = \frac{\theta_1^2}{1 - \theta_2} + \theta_2, \quad \text{and} \quad \rho_k = \theta_1 \rho_{k-1} + \theta_2 \rho_{k-2}, \quad k \geq 3.$$

and there may exist both an AR-presentation, $\alpha^{-1}(L)\theta(L)(y_t - \mu) = \epsilon_t$, and a MA-representation, $(y_t - \mu) = \theta^{-1}(L)\alpha(L)\epsilon_t$, both with infinitely many terms. In this way we may think of the ARMA model as a *parsimonious* representation of a given autocovariance structure, i.e. the representation that uses as few parameters as possible.

For y_t to be stationary it is required that the inverse roots of the characteristic equation are all smaller than one. If there is a root at unity, a so-called *unit root* $\phi_1 = 1$, then the factorized polynomial can be written as

$$\theta(z) = (1 - z)(1 - \phi_2 z) \cdots (1 - \phi_p z). \quad (9)$$

Noting that $\Delta = 1 - L$ is the first difference operator, the model can be written in terms of the first differenced variable

$$\theta^*(L)\Delta y_t = \delta + \alpha(L)\epsilon_t, \quad (10)$$

where the polynomial $\theta^*(z)$ is defined as the last $p - 1$ terms in (9). The general model in (10) is referred to as an integrated ARMA model or an ARIMA(p,d,q), with p stationary autoregressive roots, d first differences, and q MA terms.

We return to the properties of unit root processes and the testing for unit roots later in the course.

EXAMPLE 1 (DANISH HOUSE PRICES): To illustrate the construction of ARIMA models we consider the real Danish house prices, 1972:1-2004:2, defined as the log of the house price index divided with the consumer price index. Estimating a second order autoregressive model yields

$$p_t = 0.0034 + 1.545 \cdot p_{t-1} - 0.565 \cdot p_{t-2},$$

with the autoregressive polynomial given by $\theta(z) = 1 - 1.545 \cdot z + 0.565 \cdot z^2$. The $p = 2$ inverse roots of the polynomial are given by $\phi_1 = 0.953$ and $\phi_2 = 0.592$ and we can factorize the polynomial as

$$\theta(z) = 1 - 1.545 \cdot z + 0.565 \cdot z^2 = (1 - 0.953 \cdot z)(1 - 0.592 \cdot z).$$

We do not want to test for unit roots at this point and we assume without testing that the first root is unity and use $(1 - 0.953 \cdot L) \approx (1 - L) = \Delta$. Estimating a AR(1) model for Δp_t , which is the same as an ARIMA(1,1,0) model for p_t , we get the following results

$$\Delta p_t = 0.0008369 + 0.544 \cdot \Delta p_{t-1},$$

where the second root is basically unchanged. ◆

4 ARMA ESTIMATION AND MODEL SELECTION

If we are willing to assume a specific distributional form for the error process, ϵ_t , then it is natural to estimate the parameters using *maximum likelihood*. The most popular

assumption in macro-econometrics is the assumption of normal errors. In this case the log-likelihood function has the well-known form

$$\log L(\theta, \sigma^2) = -\frac{T}{2} \log(2\pi\sigma^2) - \sum_{t=1}^T \frac{\epsilon_t^2}{2\sigma^2}, \quad (11)$$

where θ contains the parameters to be estimated in the conditional mean and σ^2 is the error variance. Other distributional assumptions can also be used, and in financial econometrics it is often preferred to rely on error distributions with more probability mass in the tails of the distribution, i.e. with a higher probability of extreme observations. A popular choice is the student $t(v)$ -distribution, where v is the number of degrees of freedom. Low degrees of freedom give a heavy tail distribution and for $v \rightarrow \infty$, the $t(v)$ -distribution approaches the normal distribution. In a likelihood analysis we may even treat v as a parameter and estimate the degrees of freedom in the t -distribution.

The autoregressive model is basically a linear regression and estimation is particularly simple. We can write the error in terms of the observed variables

$$\epsilon_t = y_t - \delta - \theta_1 y_{t-1} - \theta_2 y_{t-2} - \dots - \theta_p y_{t-p}, \quad t = 1, 2, \dots, T,$$

insert in (11) and maximize the likelihood function over $(\delta, \theta_1, \theta_2, \dots, \theta_p, \sigma^2)'$ given the observed data. We note that the analysis is conditional on the initial values and that the ML estimator coincides with the OLS estimator for this case.

For the moving average model the idea is the same, but it is more complicated to express the likelihood function in terms of observed data. Consider for illustration the MA(1) model given by

$$y_t = \mu + \epsilon_t + \alpha \epsilon_{t-1}.$$

To express the sequence of error terms, $\epsilon_1, \dots, \epsilon_T$, as a function of the observed data, y_1, \dots, y_T , we solve recursively for the error terms

$$\begin{aligned} \epsilon_1 &= y_1 - \mu \\ \epsilon_2 &= y_2 - \mu - \alpha \epsilon_1 = y_2 - \alpha y_1 - \mu + \alpha \mu \\ \epsilon_3 &= y_3 - \mu - \alpha \epsilon_2 = y_3 - \alpha y_2 + \alpha^2 y_1 - \mu + \alpha \mu - \alpha^2 \mu \\ \epsilon_4 &= y_4 - \mu - \alpha \epsilon_3 = y_4 - \alpha y_3 + \alpha^2 y_2 - \alpha^3 y_1 - \mu + \alpha \mu - \alpha^2 \mu + \alpha^3 \mu \\ &\vdots \end{aligned}$$

where we have assumed a zero initial value, $\epsilon_0 = 0$. The resulting log-likelihood is a complicated non-linear function of the parameters $(\mu, \alpha, \sigma^2)'$ but it can be maximized using numerical algorithms to produce the ML estimates.

4.1 MODEL SELECTION

In empirical applications it is necessary to choose the lag orders, p and q , for the ARMA model. If we have a list of potential models, e.g.

$$\begin{aligned} \text{ARMA}(1,1) &: y_t - \theta y_{t-1} = \delta + \epsilon_t + \alpha \epsilon_{t-1} \\ \text{AR}(1) &: y_t - \theta y_{t-1} = \delta + \epsilon_t \\ \text{MA}(1) &: y_t = \delta + \epsilon_t + \alpha \epsilon_{t-1} \end{aligned}$$

then we should find a way to choose the most relevant one for a given data set. This is known as model selection in the literature. There are two different approaches: *general-to-specific* (GETS) testing and *model selection* based on information criteria.

If the models are nested, i.e. if one model is a special case of a larger model, then we may use standard likelihood ratio (LR) testing to evaluate the reduction from the large to the small model. In the above example, it holds that the AR(1) model is nested within the ARMA(1,1), written as $\text{AR}(1) \subset \text{ARMA}(1,1)$, where the reduction imposes the restriction $\alpha = 0$. This restriction can be tested by a LR test, and we may reject or accept the reduction to an AR(1) model. Likewise it holds that $\text{MA}(1) \subset \text{ARMA}(1,1)$ and we may test the restriction $\theta = 0$ using a LR test to judge the reduction from the ARMA(1,1) to the MA(1).

If both reductions are accepted, however, it is not clear how to choose between the AR(1) model and the MA(1) model. The models are not nested (you cannot impose a restriction on one model to get the other) and standard test theory will not apply. In practice we can estimate both models and compare the fit of the models and the outcome of misspecification testing, but there is no valid formal test between the models.

An alternative approach is called model selection and it is valid also for non-nested models with the same regressand. We know that the more parameters we allow in a model the smaller is the residual variance. To obtain a parsimonious model we therefore want to balance the model fit against the complexity of the model. This balance can be measured by a so-called *information criteria* that takes the log-likelihood and subtract a penalty for the number of parameters, i.e.

$$IC = \log \hat{\sigma}^2 + \text{penalty}(T, \#parameters).$$

A small value indicates a more favorable trade-off, and model selection could be based on *minimizing* the information criteria. Different criteria have been proposed based on different penalty functions. Three important examples are the Akaike, the Hannan-Quinn, and Schwarz' Bayesian criteria, defined as, respectively,

$$\begin{aligned} AIC &= \log \hat{\sigma}^2 + \frac{2 \cdot k}{T} \\ HQ &= \log \hat{\sigma}^2 + \frac{2 \cdot k \cdot \log(\log(T))}{T} \\ BIC &= \log \hat{\sigma}^2 + \frac{k \cdot \log(T)}{T}, \end{aligned}$$

where k is the number of estimated parameters, e.g. $k = p + q + 1$. The idea of the model selection is to choose the model with the smallest information criteria, i.e. the best combination of fit and parsimony.

It is worth emphasizing that different information criteria will not necessarily give the same preferred model, and the model selection may not agree with the GETS testing. In practice it is therefore often difficult to make firm choices, and with several candidate models it is a sound principle to ensure that the conclusions you draw from an analysis are robust to the chosen model.

A final and less formal approach to identification of p and q in the ARMA(p, q) model is based directly on the shape of the autocorrelation function. Recall from Lecture Note 1 that the partial autocorrelation function, PACF, is defined as the autocorrelation conditional on intermediate lags, $Corr(y_t, y_{t-k} \mid y_{t-1}, y_{t-2}, \dots, y_{t-k+1})$. It follows directly from the definition that an AR(p) model will have p significant partial autocorrelations and PACF is zero for lags $k > p$; at the same time it holds that the ACF exhibits an exponential decay. For the MA(q) model we observe the reverse pattern: We know that the first q entries in the ACF are non-zero while ACF is zero for $k > q$; and if we write the MA model as an AR(∞) we expect the PACF to decay exponentially. The AR and MA model are therefore mirror images and by looking at the ACF and PACF we could get an idea on the appropriate values for p and q . This methodology is known as the *Box-Jenkins identification* procedure and it was very popular in times when ARMA models were hard to estimate. With today's computers it is probably easier to test formally on the parameters then informally on their implications in terms of autocorrelation patterns.

EXAMPLE 2 (DANISH CONSUMPTION-INCOME RATIO): To illustrate model selection we consider the log of the Danish quarterly consumption-to-income ratio, 1971:1-2003:2. This is also the inverse savings rate. The time series is illustrated in Figure 3 (A), and the autocorrelation functions are given in (B). The PACF suggests that the first autoregressive coefficient is strongly significant while the second is more borderline. Due to the exponential decay of the ACF, implied by the autoregressive coefficient, it is hard so assess the presence of MA terms; this is often the case with the Box-Jenkins identification. We therefore estimate an ARMA(2,2)

$$y_t - \theta_1 y_{t-1} - \theta_2 y_{t-2} = \delta + \epsilon_t + \alpha_1 \epsilon_{t-1} + \alpha_2 \epsilon_{t-2},$$

and all sub-models obtained by imposing restrictions on the parameters $(\theta_1, \theta_2, \alpha_1, \alpha_2)'$. Estimation results and summary statistics for the models are given in Table 1. All three information criteria are minimized for the purely autoregressive AR(2) model, but the value for the mixed ARMA(1,1) is by and large identical. The reduction from the ARMA(2,2) to these two models are easily accepted by LR tests. Both models seem to give a good description of the covariance structure of the data and based on the output from Table 1 it is hard to make a firm choice. ♦

ARMA(p,q)	(2,2)	(2,1)	(2,0)	(1,2)	(1,1)	(1,0)	(0,2)	(0,1)	(0,0)
θ_1	1.418 (4.28)	0.573 (1.70)	0.536 (6.36)	0.833 (12.2)	0.857 (15.2)	0.715 (11.7)
θ_2	-0.516 (-1.82)	0.224 (0.89)	0.251 (2.95)
α_1	-0.899 (-2.80)	-0.040 (-0.11)	...	-0.304 (-2.78)	-0.301 (-3.06)	...	0.577 (6.84)
α_2	0.308 (2.13)	0.085 (0.95)	0.397 (5.75)	0.487 (8.36)	...
μ	-0.094 (-11.1)	-0.094 (-9.77)	-0.094 (-9.87)	-0.094 (-9.87)	-0.094 (-9.44)	-0.094 (-12.6)	-0.094 (-52.3)	-0.094 (-24.9)	-0.095 (-30.4)
log-likelihood	300.822	300.395	300.389	300.428	299.993	296.174	287.961	274.720	249.826
BIC	-4.441	-4.472	-4.509	-4.472	-4.503	-4.482	-4.318	-4.152	-3.806
HQ	-4.506	-4.524	-4.548	-4.525	-4.542	-4.508	-4.357	-4.178	-3.819
AIC	-4.551	-4.560	-4.575	-4.560	-4.569	-4.526	-4.384	-4.196	-3.828
Normality	[0.70]	[0.83]	[0.84]	[0.81]	[0.84]	[0.80]	[0.44]	[0.12]	[0.16]
No-autocor.	[0.45]	[0.64]	[0.72]	[0.63]	[0.66]	[0.20]	[0.00]	[0.00]	[0.00]

Table 1: Estimation results for ARMA(2,2) and sub-models. Figures in parentheses are t -ratios. Figures in square brackets are p -values for misspecification tests. Estimation is done using the ARFIMA package in PcGive, which uses a slightly more complicated treatment of initial values than that presented in the present text.

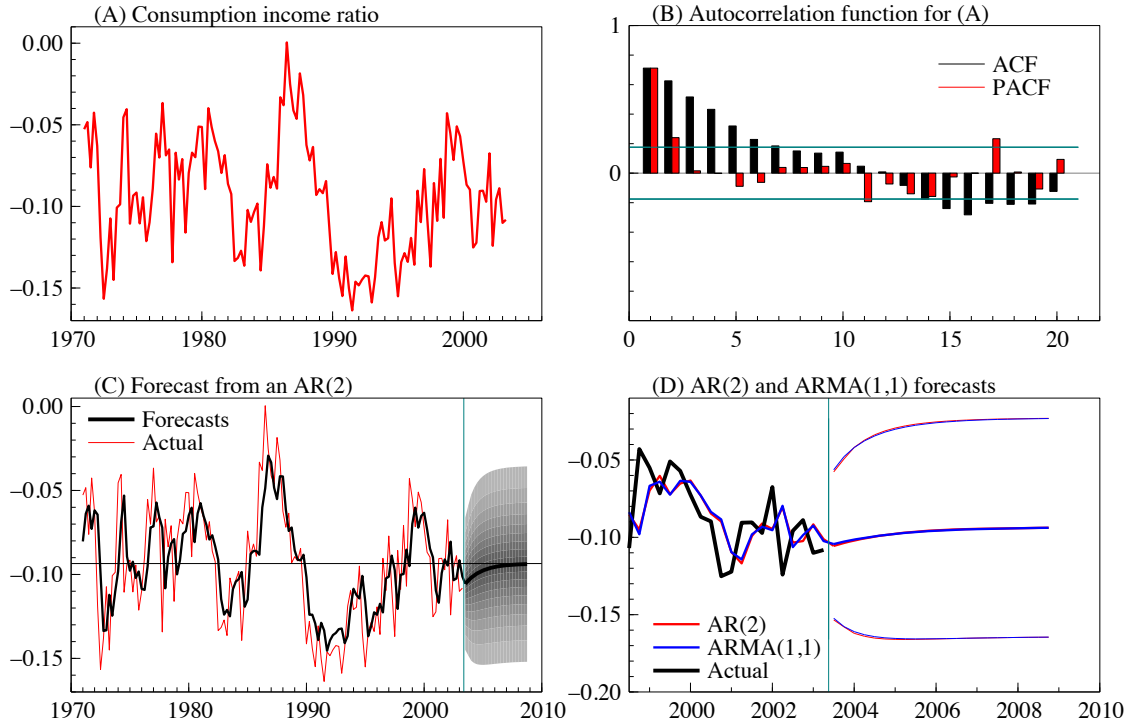


Figure 3: Example based on Danish consumption-income ratio.

5 UNIVARIATE FORECASTING

It is straightforward to forecast with univariate ARMA models, and the simple structure allows you to produce forecasts based solely on the past of the process. The obtained forecasts are plain extrapolations from the systematic part of the process and they will contain no economic insight. In particular it is very rarely possible to predict turning points, i.e. business cycle changes, based on a single time series. The forecast may nevertheless be helpful in analyzing the direction of future movement in a time series, all other things equal.

The object of interest is a prediction of y_{T+k} given the information up to time T . Formally we define the *information set* available at time T as $\mathcal{I}_T = \{y_{-\infty}, \dots, y_{T-1}, y_T\}$, and we define the optimal predictor as the conditional expectation

$$y_{T+k|T} := E[y_{T+k} \mid \mathcal{I}_T].$$

To illustrate the idea we consider the case of an ARMA(1,1) model,

$$y_t = \delta + \theta y_{t-1} + \epsilon_t + \alpha \epsilon_{t-1},$$

for $t = 1, 2, \dots, T$. To forecast the next observation, y_{T+1} , we write the equation

$$y_{T+1} = \delta + \theta y_T + \epsilon_{T+1} + \alpha \epsilon_T,$$

and the best prediction is the conditional expectation of the right-hand-side. We note that y_T and $\epsilon_T = y_T - \delta - \theta y_{T-1} - \alpha \epsilon_{T-1}$ are in the information set at time T , while the best prediction of future shocks are zero, $E[\epsilon_{T+k} \mid \mathcal{I}_T] = 0$ for $k > 0$. We find the predictions

$$\begin{aligned} y_{T+1|T} &= E[\delta + \theta y_T + \epsilon_{T+1} + \alpha \epsilon_T \mid \mathcal{I}_T] &= & \delta + \theta y_T + \alpha \epsilon_T \\ y_{T+2|T} &= E[\delta + \theta y_{T+1} + \epsilon_{T+2} + \alpha \epsilon_{T+1} \mid \mathcal{I}_T] &= & \delta + \theta E[y_{T+1} \mid \mathcal{I}_T] = \delta + \theta y_{T+1|T} \\ y_{T+3|T} &= E[\delta + \theta y_{T+2} + \epsilon_{T+3} + \alpha \epsilon_{T+2} \mid \mathcal{I}_T] &= & \delta + \theta E[y_{T+2} \mid \mathcal{I}_T] = \delta + \theta y_{T+2|T} \\ &\vdots & & \end{aligned}$$

We note that the error term, ϵ_T , affects the first period forecast due to the MA(1) structure, and after that the first order autoregressive process takes over and the forecasts will converge exponentially towards the unconditional expectation, μ .

In practice we replace the true parameters with the estimators, $\hat{\theta}$, $\hat{\alpha}$, and $\hat{\delta}$, and the true errors with the estimated residuals, $\hat{\epsilon}_1, \dots, \hat{\epsilon}_T$, which produces the feasible forecasts, $\hat{y}_{T+k|T}$.

5.1 FORECAST ACCURACY

The forecasts above are point forecasts, i.e. the best point predictions given by the model. Often it is of interest to assess the variances of these forecasts, and produce confidence bounds or distributions of the forecasts.

To analyze forecast errors, consider first a MA(q) model

$$y_t = \mu + \epsilon_t + \alpha_1 \epsilon_{t-1} + \alpha_2 \epsilon_{t-2} + \dots + \alpha_q \epsilon_{t-q}.$$

The sequence of forecasts are given by

$$\begin{aligned} y_{T+1|T} &= \mu + \alpha_1 \epsilon_T + \alpha_2 \epsilon_{T-1} + \dots + \alpha_q \epsilon_{T-q+1} \\ y_{T+2|T} &= \mu + \alpha_2 \epsilon_T + \dots + \alpha_q \epsilon_{T-q+2} \\ y_{T+3|T} &= \mu + \alpha_3 \epsilon_T + \dots + \alpha_q \epsilon_{T-q+3} \\ &\vdots \\ y_{T+q|T} &= \mu + \alpha_q \epsilon_T \\ y_{T+q+1|T} &= \mu, \end{aligned}$$

where the information set improves the predictions for q period. The corresponding forecast errors are given by the error terms not in the information set

$$\begin{aligned} y_{T+1} - y_{T+1|T} &= \epsilon_{T+1} \\ y_{T+2} - y_{T+2|T} &= \epsilon_{T+2} + \alpha_1 \epsilon_{T+1} \\ y_{T+3} - y_{T+3|T} &= \epsilon_{T+3} + \alpha_1 \epsilon_{T+2} + \alpha_2 \epsilon_{T+1} \\ &\vdots \\ y_{T+q} - y_{T+q|T} &= \epsilon_{T+q} + \alpha_1 \epsilon_{T+q-1} + \dots + \alpha_{q-1} \epsilon_{T+1} \\ y_{T+q+1} - y_{T+q+1|T} &= \epsilon_{T+q+1} + \alpha_1 \epsilon_{T+q} + \dots + \alpha_q \epsilon_{T+1}. \end{aligned}$$

We can find the variance of the forecast as the squared forecast errors, i.e.

$$\begin{aligned} C_1 &= E[\epsilon_{T+1}^2 | \mathcal{I}_T] &= \sigma^2 \\ C_2 &= E[(\epsilon_{T+2} + \alpha_1 \epsilon_{T+1})^2 | \mathcal{I}_T] &= (1 + \alpha_1^2) \sigma^2 \\ C_3 &= E[(\epsilon_{T+3} + \alpha_1 \epsilon_{T+2} + \alpha_2 \epsilon_{T+1})^2 | \mathcal{I}_T] &= (1 + \alpha_1^2 + \alpha_2^2) \sigma^2 \\ &\vdots \\ C_q &= E[(\epsilon_{T+q} + \alpha_1 \epsilon_{T+q-1} + \dots + \alpha_{q-1} \epsilon_{T+1})^2 | \mathcal{I}_T] &= (1 + \alpha_1^2 + \alpha_2^2 + \dots + \alpha_{q-1}^2) \sigma^2 \\ C_{q+1} &= E[(\epsilon_{T+q+1} + \alpha_1 \epsilon_{T+q} + \dots + \alpha_q \epsilon_{T+1})^2 | \mathcal{I}_T] &= (1 + \alpha_1^2 + \alpha_2^2 + \dots + \alpha_q^2) \sigma^2. \end{aligned}$$

Note that the forecast error variance increases with the forecast horizon, and that the forecast variances converge to the unconditional variance of y_t , see equation (2). This result just reflects that the information set, \mathcal{I}_T , is useless for predictions in the remote future: The best prediction will be the unconditional mean, μ , and the uncertainty is the unconditional variance, $C_\infty = \gamma_0$.

Assuming that the error term is normally distributed, we may produce 95% confidence bounds for the forecasts as $y_{T+h|T} \pm 1.96 \cdot \sqrt{C_h}$, where 1.96 is the quantile of the normal distribution. Alternatively we may give a full distribution of the forecasts as $N(y_{T+h|T}, C_h)$.

To derive the forecast error variance for AR and ARMA models we just write the models in their MA(∞) form and use the derivations above. For the case of a simple AR(1) model the infinite MA representation is given in (6), and the forecast error variances are found to be

$$C_1 = \sigma^2, \quad C_2 = (1 + \theta^2) \sigma^2, \quad C_3 = (1 + \theta^2 + \theta^4) \sigma^2, \quad \dots$$

where we again note that the forecast error variances converge to the unconditional variance, γ_0 .

EXAMPLE 3 (DANISH CONSUMPTION-INCOME RATIO): For the Danish consumption-income ratio, we saw that the AR(2) model and the ARMA(1,1) gave by and large identical in-sample results. Figure 3 (C) shows the out-of-sample forecast for the AR(2) model. We note that the forecasts are very smooth, just describing an exponential convergence back to the unconditional mean—the attractor in the stationary model. The shaded area is the distribution of the forecast, and the widest band corresponds to 95% confidence. Graph (D) compares the forecasts from the AR(2) and the ARMA(1,1) models as well as their 95% confidence bands. The two sets of forecasts are very similar and for practical use it does not matter which one we choose; this is reassuring as the choice between the models was very difficult. \blacklozenge

6 THE AUTOREGRESSIVE DISTRIBUTED LAG MODEL

The models considered so far were all univariate in the sense that they focus on a single variable, y_t . In most situations, however, we are primarily interested in the interrelationships between variables. As an example we could be interested in the dynamic responses to an intervention in a given variables, i.e. in the dynamic multipliers

$$\frac{\partial y_t}{\partial x_t}, \quad \frac{\partial y_{t+1}}{\partial x_t}, \quad \frac{\partial y_{t+2}}{\partial x_t}, \dots$$

A univariate model will not suffice for this purpose, and the object of interest is the conditional mean of y_t given $y_{t-1}, y_{t-2}, \dots, x_t, x_{t-1}, x_{t-2}, \dots$. Assuming linearity of the conditional mean, and assuming that only one lag of the information set is necessary, we have the so-called *autoregressive distributed lag* (ADL) model:

$$y_t = \delta + \theta_1 y_{t-1} + \phi_0 x_t + \phi_1 x_{t-1} + \epsilon_t, \tag{12}$$

for $t = 1, 2, \dots, T$, with $E[\epsilon_t \mid y_{t-1}, y_{t-2}, \dots, x_t, x_{t-1}, x_{t-2}, \dots] = 0$. Often we make the stronger assumption that $\epsilon \sim \text{iID}(0, \sigma^2)$, ruling out also heteroskedasticity. To simplify the notation we consider the model with one explanatory variable and one lag in y_t and x_t , but the model can easily be extended to more general cases.

By referring to the results for an AR(1) process, we immediately see that the process y_t is stationary if $|\theta_1| < 1$ and x_t is a stationary process. The first condition excludes

unit roots in the equation (12), while the second condition states that the forcing variable, x_t , is also stationary. The results from Lecture Note 2 holds for estimation and inference in the model; here we are interested in the dynamic properties of the model under the stationarity condition.

Before we go into details with the dynamic properties of the ADL model, we want to emphasize that the model is quite general and contains other relevant models as special cases. The AR(1) model analyzed above prevails if $\phi_0 = \phi_1 = 0$. A static regression with IID errors is obtained if $\theta_1 = \phi_1 = 0$. A static regression with AR(1) autocorrelated errors is obtained if $\phi_1 = -\theta_1\phi_0$; this is the common factor restriction discussed in Lecture Note 2. Finally, a model in first differences is obtained if $\theta_1 = 1$ and $\phi_0 = -\phi_1$. Whether these special cases are relevant can be analyzed with standard Wald or LR tests on the coefficients.

6.1 DYNAMIC- AND LONG-RUN MULTIPLIERS

To derive the dynamic multiplier, we write the equations for different observations,

$$\begin{aligned} y_t &= \delta + \theta_1 y_{t-1} + \phi_0 x_t + \phi_1 x_{t-1} + \epsilon_t \\ y_{t+1} &= \delta + \theta_1 y_t + \phi_0 x_{t+1} + \phi_1 x_t + \epsilon_{t+1} \\ y_{t+2} &= \delta + \theta_1 y_{t+1} + \phi_0 x_{t+2} + \phi_1 x_{t+1} + \epsilon_{t+2}, \end{aligned}$$

and state the *dynamic multipliers* as the derivatives:

$$\begin{aligned} \frac{\partial y_t}{\partial x_t} &= \phi_0 \\ \frac{\partial y_{t+1}}{\partial x_t} &= \theta_1 \frac{\partial y_t}{\partial x_t} + \phi_1 = \theta_1 \phi_0 + \phi_1 \\ \frac{\partial y_{t+2}}{\partial x_t} &= \theta_1 \frac{\partial y_{t+1}}{\partial x_t} = \theta_1 (\theta_1 \phi_0 + \phi_1) \\ \frac{\partial y_{t+3}}{\partial x_t} &= \theta_1 \frac{\partial y_{t+2}}{\partial x_t} = \theta_1^2 (\theta_1 \phi_0 + \phi_1) \\ &\vdots \\ \frac{\partial y_{t+k}}{\partial x_t} &= \theta_1^{k-1} (\theta_1 \phi_0 + \phi_1). \end{aligned}$$

Under the stationarity condition, $|\theta_1| < 1$, shocks have only transitory effects $\frac{\partial y_{t+k}}{\partial x_t} \rightarrow 0$ as $k \rightarrow \infty$. We think of the sequence of multipliers as the impulse-reponses to a temporary change in x_t and for stationary variables it is natural that the long-run effect is zero. To illustrate the flexibility of the ADL model with one lag, Figure 4 (A) reports example of dynamic multipliers. In all cases the contemporaneous impact is $\frac{\partial y_t}{\partial x_t} = 0.8$, but the dynamic profile can be fundamentally different depending on the parameters.

Now consider a permanent shift in x_t , so that $E[x_t]$ changes. To find the *long-run*

multiplier we take expectations to obtain

$$\begin{aligned}
y_t &= \delta + \theta_1 y_{t-1} + \phi_0 x_t + \phi_1 x_{t-1} + \epsilon_t \\
E[y_t] &= \delta + \theta_1 E[y_{t-1}] + \phi_0 E[x_t] + \phi_1 E[x_{t-1}] \\
E[y_t](1 - \theta_1) &= \delta + (\phi_0 + \phi_1) E[x_t] \\
E[y_t] &= \frac{\delta}{1 - \theta_1} + \frac{\phi_0 + \phi_1}{1 - \theta_1} E[x_t].
\end{aligned} \tag{13}$$

In the derivation we have used the stationarity of y_t and x_t , i.e. that $E[y_t] = E[y_{t-1}]$ and $E[x_t] = E[x_{t-1}]$. Based on (13) we find the long-run multiplier as the the derivative

$$\frac{\partial E[y_t]}{\partial E[x_t]} = \frac{\phi_0 + \phi_1}{1 - \theta_1}. \tag{14}$$

It also holds that the long-run multiplier is the sum of the short-run effects, i.e.

$$\begin{aligned}
\frac{\partial y_t}{\partial x_t} + \frac{\partial y_{t+1}}{\partial x_t} + \frac{\partial y_{t+2}}{\partial x_t} + \dots &= \phi_0 + (\theta_1 \phi_0 + \phi_1) + \theta_1 (\theta_1 \phi_0 + \phi_1) + \theta_1^2 (\theta_1 \phi_0 + \phi_1) + \dots \\
&= \phi_0 (1 + \theta_1 + \theta_1^2 + \dots) + \phi_1 (1 + \theta_1 + \theta_1^2 + \dots) \\
&= \frac{\phi_0 + \phi_1}{1 - \theta_1}.
\end{aligned}$$

For stationary variables we may define a steady state as $y_t = y_{t-1}$ and $x_t = x_{t-1}$. Inserting that in the ADL model with get the so-called *long-run solution*:

$$y_t = \frac{\delta}{1 - \theta_1} + \frac{\phi_0 + \phi_1}{1 - \theta_1} x_t = \mu + \beta x_t, \tag{15}$$

which is the steady state version of the equation in (14), and we see that the steady state derivative is exactly the long-run multiplier.

Figure 4 (B) reports examples of cumulated dynamic multipliers, which picture of the convergence towards the long-run solutions. The first period impact is in all cases $\frac{\partial y_t}{\partial x_t} = 0.8$, while the long-run impact is β which depends on the parameters of the model.

EXAMPLE 4 (DANISH INCOME AND CONSUMPTION): To illustrate the impact of income changes on changes in consumption, we consider time series for the growth from quarter to quarter in Danish private aggregate consumption, $c_t = \Delta \log(\text{CONS}_t)$ say, and private disposable income, $y_t = \Delta \log(\text{INC}_t)$, for 1971 : 1 – 2004 : 2, see Figure 4 (C). To analyze the interdependencies between the variables we estimate and ADL model with one lag and obtain the equation

$$\hat{c}_t = \underset{(2.09)}{0.003} - \underset{(-3.88)}{0.328} \cdot c_{t-1} + \underset{(3.80)}{0.229} \cdot y_t + \underset{(1.48)}{0.092} \cdot y_{t-1}, \tag{16}$$

where the numbers in parentheses are t -ratios. Apart from a number of outliers, the model appears to be well specified and the misspecification tests for no-autocorrelation cannot be rejected. The residuals are reported in graph (D). The long-run solution is given by (15) with coefficients given by

$$\mu = \frac{\delta}{1 - \theta_1} = \frac{0.003}{1 + 0.328} = 0.0024 \quad \text{and} \quad \beta = \frac{\phi_0 + \phi_1}{1 - \theta_1} = \frac{0.229 + 0.092}{1 + 0.328} = 0.242.$$

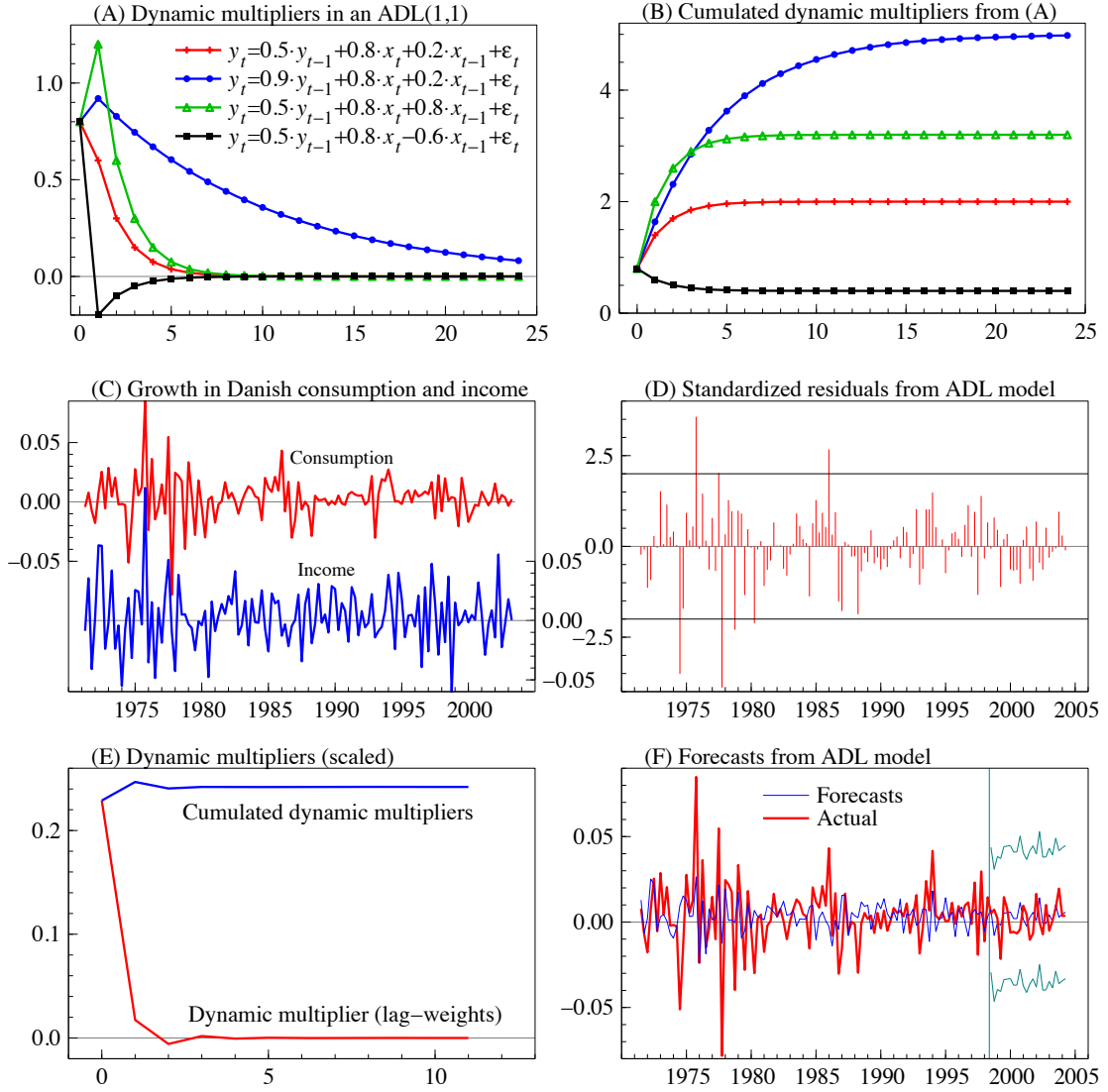


Figure 4: (A)-(B): Examples of dynamic multipliers for an ADL model with one lag. (C)-(D): Empirical example based on Danish income and consumption

The standard errors of μ and β are complicated functions of the original variances and covariances, but PcGive supply them automatically and we find that $t_{\mu=0} = 2.10$ and $t_{\beta=0} = 3.31$ so both long-run coefficients are significantly different from zero at a 5% level. The impulse responses, $\partial c_t / \partial y_t, \partial c_{t+1} / \partial y_t, \dots$, are reported in Figure 4 (E). We note that the contemporaneous impact is 0.229 from (16). The cumulated response is also presented in graph (E); it converges to the long-run multiplier, $\beta = 0.242$, which in this case is very close to the first period impact. \blacklozenge

6.2 GENERAL CASE

To state the results for the general case of more lags, it is again convenient to use lag-polynomials. We write the model with p and q lags as

$$\theta(L)y_t = \delta + \phi(L)x_t + \epsilon_t,$$

where the polynomials are defined as

$$\theta(L) := 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_p L^p \quad \text{and} \quad \phi(L) := \phi_0 + \phi_1 L + \phi_2 L^2 + \dots + \phi_q L^q.$$

Under stationarity we can write the model as

$$y_t = \theta^{-1}(L)\delta + \theta^{-1}(L)\phi(L)x_t + \theta^{-1}(L)\epsilon_t, \quad (17)$$

in which there are infinitely many terms in both x_t and ϵ_t . Writing the combined polynomial as

$$c(L) = \theta^{-1}(L)\phi(L) = c_0 + c_1 L + c_2 L^2 + \dots,$$

the dynamic multipliers are given by the sequence c_0, c_1, c_2, \dots . This is parallel to the result for the AR(1) model stated above.

If we take expectations in (17) we get the equivalent to (13),

$$E[y_t] = \theta^{-1}(L)\delta + \theta^{-1}(L)\phi(L)E[x_t].$$

Now recall that $E[y_t] = E[y_{t-1}] = LE[y_t]$ and $E[x_t] = E[x_{t-1}] = LE[x_t]$ are constant due to stationarity, so that $\theta^{-1}(L)\phi(L)E[x_t] = \theta^{-1}(1)\phi(1)E[x_t]$ and the long-run solution can be found as the derivative

$$\frac{\partial E[y_t]}{\partial E[x_t]} = \frac{\phi(1)}{\theta(1)} = \frac{\phi_0 + \phi_1 + \phi_2 + \dots + \phi_q}{1 - \theta_1 - \theta_2 - \dots - \theta_p}.$$

This is also the sum of the dynamic multipliers, $c_0 + c_1 + c_2 + \dots$

6.3 ERROR-CORRECTION MODEL

There exists a convenient formulation of the model that incorporates the long-run solution directly. In particular, we can rewrite the model as

$$\begin{aligned} y_t &= \delta + \theta_1 y_{t-1} + \phi_0 x_t + \phi_1 x_{t-1} + \epsilon_t \\ y_t - y_{t-1} &= \delta + (\theta_1 - 1)y_{t-1} + \phi_0 x_t + \phi_1 x_{t-1} + \epsilon_t \\ y_t - y_{t-1} &= \delta + (\theta_1 - 1)y_{t-1} + \phi_0(x_t - x_{t-1}) + (\phi_0 + \phi_1)x_{t-1} + \epsilon_t \\ \Delta y_t &= \delta + (\theta_1 - 1)y_{t-1} + \phi_0 \Delta x_t + (\phi_0 + \phi_1)x_{t-1} + \epsilon_t. \end{aligned} \quad (18)$$

The idea is that the levels appear only once, and the model can be written in the form

$$\Delta y_t = \phi_0 \Delta x_t - (1 - \theta_1)(y_{t-1} - \mu - \beta x_{t-1}) + \epsilon_t, \quad (19)$$

where μ and β refer to (15). This model in (19) is known as the *error-correction model* (ECM) and it has a very natural interpretation in terms of the long-run steady state solution and the dynamic adjustment towards equilibrium.

First note that

$$y_{t-1} - y_{t-1}^* = y_{t-1} - \mu - \beta x_{t-1}$$

is the deviation from steady state the previous period. For the steady state to be sustained the variables have to eliminate the deviations and move towards the long-run solution. This tendency is captured by the coefficient $-(1 - \theta_1) < 0$, and if y_{t-1} is above the steady state value then Δy_t will be affected negatively, and there is a tendency for y_t to move back towards the steady state. We say that y_t *error-corrects* or *equilibrium-corrects* and $y_t^* = \mu + \beta x_t$ is the *attractor* for y_t .

We can estimate the ECM model and obtain the alternative representation directly. We note that the models (12), (18), and (19) are equivalent, and we can always calculate from one representation to the other. It is important to note, however, that the version in (18) is a linear regression model that can be estimated using OLS while the solved ECM in (19) is non-linear and requires a different estimation procedure (e.g. maximum likelihood). The choice of which model to estimate depends on the purpose of the analysis and on the used software. The analysis in PcGive focus on the standard ADL model and the long-run solution and the dynamic multipliers are automatically supplied. In other software packages it is sometimes more convenient to estimate the ECM form directly.

EXAMPLE 5 (DANISH INCOME AND CONSUMPTION): If we estimate the linear error-correction form for the Danish income and consumption data series we obtain the equation,

$$\Delta \hat{c}_t = \underset{(3.80)}{0.229} \cdot \Delta y_t - \underset{(-15.7)}{1.328} \cdot c_{t-1} + \underset{(3.19)}{0.321} \cdot y_{t-1} + \underset{(2.09)}{0.0032}.$$

Note that the results are equivalent to the results from the ADL model in (16). We can find the long-run multiplier as $\beta = \frac{0.321}{1.328} = 0.242$. Using maximum likelihood (with a normal error distribution) we can also estimate the solved ECM form to obtain

$$\Delta \hat{c}_t = \underset{(3.80)}{0.229} \cdot \Delta y_t - \underset{(-15.7)}{1.328} \cdot \left(c_{t-1} - \underset{(2.10)}{0.0024} - \underset{(3.31)}{0.242} \cdot y_{t-1} \right),$$

where we again recognize the long-run solution. ◆

6.4 CONDITIONAL FORECASTS

From the conditional ADL model we can also produce forecasts of y_{T+k} . This is parallel to the univariate forecasts presented above and we will not go into details here. Note, however, that to forecast y_{T+k} we need to know x_{T+k} . Since this is rarely in the information set at time T , a feasible forecast could forecast x_{T+k} first, e.g. using a univariate time series model and then forecast y_{T+k} conditional on $\hat{x}_{T+k|T}$.

EXAMPLE 6 (DANISH INCOME AND CONSUMPTION): To illustrate dynamic forecasting with the ADL model we reestimate the ADL model for the Danish consumption conditional income. We now estimate for the sample 1971 : 2 – 1998 : 2 and retain the most recent 24 observations (6 years of quarterly data) for post sample analysis. We forecast the observations (conditional on the observed observations for y_t) and compare the forecasts with the actual observations for c_t . For the reduced sample we obtain the results

$$\hat{c}_t = \underset{(1.95)}{0.004} - \underset{(-3.65)}{0.342} \cdot c_{t-1} + \underset{(3.66)}{0.250} \cdot y_t + \underset{(1.31)}{0.094} \cdot y_{t-1},$$

which are very similar to the full sample results. Figure 4 (F) reports the forecasts and the actual observations. The forecast do not seem to be very informative in this case, which just reflect that the noise in the time series is very large compared to the systematic variation. That makes forecasting very difficult. For more persistent time series it is often easier to predict the direction of future movements. \blacklozenge

7 CONCLUDING REMARKS

In this note we have discussed single-equation tools for stationary time series. The single equation approach maintains the important assumption of no contemporaneous feedback from x_t to y_t , i.e. that the causality runs from x_t to y_t within one time period. Statistically this was formulated as $E[\epsilon_t x_t] = 0$, where ϵ_t is the error term in the equation for y_t . In many cases that assumption may be reasonable, but in other cases the single equation model is overly restrictive.

If there are causal effects in both direction, the analysis of y_t given x_t is not valid and the natural extension is to consider a joint model for of y_t and x_t given their joint past. Assuming again a linear structure we can write the model using the *two* equations

$$\begin{aligned} y_t &= \delta_1 + \pi_{11}y_{t-1} + \pi_{12}x_{t-1} + \epsilon_{1t} \\ x_t &= \delta_2 + \pi_{21}y_{t-1} + \pi_{22}x_{t-1} + \epsilon_{2t}. \end{aligned}$$

To simplify notation we have assumed only one lag. Using vectors and matrices we can stack the equations to obtain

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} + \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}.$$

Note that the variables are treated on equal footing and both appear on the left hand side. Defining the vector $z_t = (y_t, x_t)'$ for the variables we can write the model as

$$z_t = \delta + \pi z_{t-1} + \epsilon_t. \tag{20}$$

This is a straightforward generalization of the AR(1) model, and (20) is known as a *vector autoregressive* (VAR) model. The simultaneous effect, i.e. the relationship between y_t and

x_t within a time period, is captured by the covariance of the error terms,

$$V(\epsilon_t) = E[\epsilon_t \epsilon_t'] = E \left[\begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} (\epsilon_{1t}, \epsilon_{2t}) \right] = \begin{pmatrix} E[\epsilon_{1t} \epsilon_{1t}] & E[\epsilon_{1t} \epsilon_{2t}] \\ E[\epsilon_{2t} \epsilon_{1t}] & E[\epsilon_{2t} \epsilon_{2t}] \end{pmatrix} =: \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_{22}^2 \end{pmatrix},$$

where σ_{12} is the covariance between shocks to y_t and x_t .

If we assume a (multivariate) distribution for ϵ_t , we can estimate the parameters of the VAR model using maximum likelihood. With normal errors it can be shown that the ML estimators of the unrestricted VAR coincide with estimators obtained by applying single-equation OLS to the equations one-by-one.

EXAMPLE 7 (DANISH INCOME AND CONSUMPTION): For the Danish income and consumption we estimate a VAR model with one lag and obtain the estimated model

$$\begin{pmatrix} c_t \\ y_t \end{pmatrix} = \begin{pmatrix} 0.004 \\ (2.65) \\ 0.004 \\ (1.98) \end{pmatrix} + \begin{pmatrix} -0.279 & 0.0085 \\ (-3.17) & (0.138) \\ 0.214 & -0.367 \\ (1.75) & (-4.30) \end{pmatrix} \begin{pmatrix} c_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix},$$

where the numbers in parentheses are t -ratios. The correlation between the estimated residuals is given by $\text{corr}(\epsilon_{1t}, \epsilon_{2t}) = 0.319$, indicating a strong correlation between changes in consumption and changes in income. In the conditional analysis we interpret this correlation as an effect from y_t to c_t . ♦

7.1 FURTHER READINGS

The analysis of univariate time series is treated in most textbooks. Enders (2004) gives an excellent treatment of the univariate models based on the theory for difference equations. He also extends the analysis to VAR models and non-stationary time series. A Danish introduction to the Box-Jenkins methodology is given in Milhøj (1994). A very detailed analysis of the ADL is model is given in Hendry (1995), who also goes through many special cases of the ADL model. There is a detailed coverage of VAR models for both stationary and non-stationary variables in Lütkepohl and Krätzig (2004) and Lütkepohl (2005) but the technical level is higher than this note.

REFERENCES

- ENDERS, W. (2004): *Applied Econometric Time Series*. John Wiley & Sons, 2nd edn.
- HENDRY, D. F. (1995): *Dynamic Econometrics*. Oxford University Press, Oxford.
- LÜTKEPOHL, H. (2005): *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin.

- LÜTKEPOHL, H., AND M. KRÄTZIG (2004): *Applied time Series Econometrics*. Cambridge University Press, Cambridge.
- MILHØJ, A. (1994): *Tidsrækkeøkonometri for Økonomer*. Akademisk Forlag, Copenhagen, 2nd edn.
- VERBEEK, M. (2008): *A Guide to Modern Econometrics*. John Wiley & Sons, 3rd edn.