EC821: Time Series Econometrics, Spring 2003

Notes Section 3

ARMA models

We have spoken of a univariate white noise process, ϵ_t, a zero–mean covariance stationary process with no serial correlation:

$$E\left[\epsilon_t\right] = 0, E\left[\epsilon_t^2\right] = 0 < \sigma^2 < \infty,\ E\left[\epsilon_t \epsilon_{t-j}\right] = 0\ \Upsilon\ j \neq 0.$$

A very important class of covariance–stationary processes, called **linear processes**, can be created by taking a **moving average** of a white noise process. Let us consider that the white noise process is defined for all integers, $\{t = 0, \pm 1, \pm 2, ...\}$ so that we may assume that the process started in the distant past, and its mean and autocovariance function have stabilized to time–invariant constants.

The simplest linear process that exhibits serial correlation is a finite–order moving average process. $\{y_t\}$ is said to be a **qth-order moving average** process, $MA(q)$, if it can be written as a weighted average of the current and most recent q values of a white noise process:

$$y_t = \mu + \theta_0 \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q}, \theta_0 = 1. \qquad (1)$$

A moving average process is covariance stationary with mean μ. It is easy to show that the jth-order autocovariance is $\gamma_j = \sigma^2 \sum_{k=0}^{q-j} \theta_{j+k}\theta_k$ for $j = 0, 1, ...q$, and zero for $j > q$, with σ^2 denoting the variance of the ϵ white noise process. Thus, if $q = 1$, the zero-order autocovariance is $(1 + \theta_1)\sigma^2$, the first-order autocovariance is $\theta_1 \sigma^2$, and the second-order

(and all higher-order) autocovariances are zero. The entire autocovariance function is described by $(q+1)$ parameters: $\{\theta_1, ...\theta_q, \sigma^2\}$. The autocorrelation function depends only upon the q parameters in the θ sequence.

We may also consider an $MA(\infty)$ process, in which the effects of past shocks do not abruptly dissipate after q periods. Thus, we might consider that the finite set of terms in (1) could be replaced by an infinite set of terms $\sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$. For this to make sense, it must be the case that this sum of an infinite set of random variables is well defined—that is, that the partial sum $\sum_{j=0}^{n} \psi_j \epsilon_{t-j}$ converges to a finite random variable as $n \to \infty$. A condition under which this convergence will occur is that of absolute summability: $\sum_{j=0}^{n} |\psi_j| < \infty$. Intuitively, this requires that the effects of the past shocks represented by ψ_j eventually die away. If the sequence ψ_j is absolutely summable, then an infinite–order MA process ($MA(\infty)$) for y_t converges in mean square, and the process is covariance stationary, with mean μ and autocovariances $\gamma_j = \sigma^2 \sum_{k=0}^{\infty} \psi_{j+k}\psi_k$. The autocovariances will themselves be absolutely summable. If the ϵ process is iid, then the y process is strictly stationary and ergodic.

We may now define a **linear filter**. Let $\{x_t\}$ be a covariance stationary process and $\{h_j\}$ be a sequence of real numbers that is absolutely summable. Then the infinite sum $y_t = \sum_{j=0}^{\infty} h_j x_{t-j}$ converges in mean square, such that the y process is covariance stationary. If the autocovariances of the x process are absolutely summable, then so are the autocovariances of the y process.

The operation of taking a weighted average of (possibly

infinitely many) successive values of the x process is called **filtering**, and we can write it using lag operator notation, with a filter being represented by a lag polynomial $\alpha(L) = \alpha_0 + \alpha_1 L + \alpha_2 L^2 ...$ The object $y_t = \alpha(L)x_t$ is a well-defined random variable forming a covariance stationary process if the sequence $\{\alpha_j\}$ is absolutely summable and if the input process x_t is covariance stationary.

Properties of linear filters

For a given sequence of real numbers $\{\alpha_0, \alpha_1, ...\}$ define a **filter** as the lag polynomial $\alpha(L) = \alpha_0 + \alpha_1 L + \alpha_2 L^2 + ...$, which may be applied to an input process $\{x_t\}$ to yield $\alpha(L)x_t = \sum_{j=0}^{\infty} \alpha_j x_{t-j}$. The filter could be finite, such that $\alpha_j = 0$ for $j > p$, which then defines a $p^{th}-$ order lag polynomial. When applied to an input process, this finite filter creates a weighted average of the current and p most recent values of the process. If the input process is covariance stationary and the sequence $\{\alpha_j\}$ is absolutely summable, the filtered series will be a covariance stationary process.

Let $\{\alpha_j\}$ and $\{\beta_j\}$ be two arbitrary sequences of real numbers, and define the **convolution** of these sequences $\{\delta_j\}$ as

$$\begin{aligned} \alpha_0\beta_0 &= \delta_0, \quad (2) \\ \alpha_0\beta_1 + \alpha_1\beta_0 &= \delta_1, \\ \alpha_0\beta_2 + \alpha_1\beta_1 + \alpha_2\beta_0 &= \delta_2, ... \\ \alpha_0\beta_j + \alpha_1\beta_{j-1} + \alpha_2\beta_{j-2} + ... + \alpha_{j-1}\beta_1 + \alpha_j\beta_0 &= \delta_j \end{aligned}$$

Product: the convolution may be viewed as the product of two

filters:

$$\delta(L) = \alpha(L)\beta(L)$$

which may be computed with the same techniques as the product of two ordinary polynomials, e.g.

$$(1 + \alpha_1 L)(1 + \beta_1 L) = 1 + (\alpha_1 + \beta_1) L + \alpha_1 \beta_1 L^2.$$

If each of these sequences are absolutely summable, and the input series is covariance stationary, then $\alpha(L)\beta(L)x_t = \delta(L)x_t$ is a well–defined random variable, also covariance stationary, and the sequence $\{\delta_j\}$ will also be absolutely summable.

Inverse: if $\delta(L) = 1$, so that $\alpha(L)\beta(L) = 1$, we may say that $\beta(L)$ is the inverse of $\alpha(L)$, denoted as $\alpha(L)^{-1}$. As long as $\alpha_0 \neq 0$, the inverse of $\alpha(L)$ may be defined for any arbitrary sequence $\{\alpha_j\}$ by successively solving the equations (2). For instance, $(1 - L)^{-1} = 1 + L + L^2 + L^3 + ...$ which may not be absolutely summable.

To work with ARMA processes, we must calculate the inverse of a finite–order ($p^{th}-$ degree) lag polynomial $\Theta(L)$:

$$\Theta(L) = 1 - \Theta_1 L - \Theta_2 L^2 - ... - \Theta_p L^p$$

Since $\Theta_0 \neq 0$, the inverse polynomial $\Psi(L) = \Theta(L)^{-1}$ may be defined. We may determine the coefficients of the inverse polynomial using the convolution formula in (2); for $j > p$, the coefficients of $\Psi(L)$ may be solved recursively, knowing the first p coefficients. The coefficients in the sequence $\{\Psi_j\}$ will eventually decline if the stability condition on the $\Theta(L)$ polynomial—relating to its eigenvalues—holds, and those coefficients will then be absolutely summable. As an example,

consider the lag polynomial $1 - \phi L$. The root of the associated polynomial $1 - \phi z = 0$ is ϕ^{-1}, and the stability condition requires that $|\phi^{-1}| > 1$ or, alternatively, that $|\phi| < 1$. The inverse of the filter is, given stability, $1 + \phi L + \phi^2 L^2 + \phi^3 L^3 + ...$, an infinite sequence, but one which is bounded and finite.

The $AR(1)$ process

A first-order autoregressive process ($AR(1)$) satisfies the stochastic difference equation

$$\begin{aligned} y_t &= c + \phi y_{t-1} + \epsilon_t \\ (1 - \phi L) y_t &= c + \epsilon_t \end{aligned}$$

where the ϵ process is white noise. For $\phi \neq 1$, may rewrite this equation as

$$\begin{aligned} (1 - \phi L)(y_t - \mu) &= \epsilon_t, \\ \mu &= c/(1 - \phi) \end{aligned}$$

If y_t is covariance stationary, μ is the mean of the process, and the moving average is formed from the successive values of y_t. We seek a covariance–stationary solution to this stochastic difference equation. If $|\phi| < 1$, we may use the inverse given above, so that we may define

$$\begin{aligned} (1 - \phi L)^{-1}(1 - \phi L)(y_t - \mu) &= (1 - \phi L)^{-1}\epsilon_t \\ (y_t - \mu) &= (1 - \phi L)^{-1}\epsilon_t \\ y_t &= \mu + \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}. \end{aligned}$$

In this case, y_t has the moving average representation shown here:an infinite–order moving average process, or $MA(\infty)$. The

stability condition that allows us to form the inverse is known as the stationarity condition for the autoregressive lag polynomial, which states that shocks to the system will be damped.

What about the case where that condition is violated, and $|\phi| > 1$? Then there is a "forward solution" where the unique covariance–stationary process y_t is an infinite-order moving average of *future* values of ϵ :

$$y_t = \mu - \sum_{j=0}^{\infty} \phi^{-j} \epsilon_{t+j}$$

and this infinite sum is well defined, since the sequence $\{\phi^{-j}\}$ is absolutely summable if $|\phi| > 1$.

Last, let us consider the case where there is no covariance–stationary solution, either in terms of past values of ϵ nor future values of ϵ : the case of a **unit root**, where $\phi = 1$. In this case, the $j-$ period change in y becomes

$$y_t - y_{t-j} = c \cdot j + \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} + ... + \epsilon_{t-j+1},$$

which we call a **random walk with drift**.

The $AR(p)$ process

All that we have derived here for the $AR(1)$ process may equally well be expressed in terms of the $p^{th}-$ order autoregressive process, which satisfies the $p^{th}-$ order stochastic difference equation $\Phi(L)y_t = c + \epsilon_t$ with $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p$. If the process is CS, then y_t has a mean of $\mu = c\Phi(1)^{-1}$, and an $MA(\infty)$ representation $y_t = \mu + \Psi(L)\epsilon_t$, with $\Psi(L) = \psi_0 + \psi_1 L + \psi_2 L^2 ...$, where $\Psi(L) = \Phi(L)^{-1}$.

The $ARMA(p, q)$ process

We may combine finite–order $AR(p)$ and $MA(q)$ processes to yield an $ARMA(p,q)$ process:

$$\begin{aligned} y_t &= c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \theta_0 \epsilon_t + \\ &\quad \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + ... + \theta_q \epsilon_{t-q}, \end{aligned}$$

or

$$\Phi(L) y_t = c + \Theta(L) \epsilon_t.$$

If $\Phi(1) \neq 1$, we may set $\mu = c\Phi(1)^{-1}$, and write the model in the deviation–from–mean form

$$\Phi(L) \left(y_t - \mu \right) = \Theta(L) \epsilon_t$$

which is still a $p^{th}-$ order stochastic difference equation, but with a serially correlated forcing process $\Theta(L)\epsilon_t$ rather than the white noise process ϵ_t. If the equation satisfies the stationarity conditions, then its unique covariance–stationary solution has the $MA(\infty)$ representation

$$\begin{aligned} y_t &= \mu + \Psi(L) \epsilon_t. \\ \Psi(L) &= \Phi(L)^{-1} \Theta(L). \end{aligned}$$

The mean of the $ARMA(p,q)$ process is again $c\Phi(1)^{-1}$.

If there are common roots in the two polynomials, then they could be factorized to represent the $ARMA(p,q)$ process with a lower order process $ARMA(p-1, q-1)$. We generally assume that there are no common roots in the AR and MA polynomials. If the $\Theta(L)$ polynomial satisfies the stability condition, then the $ARMA$ process is said to be **invertible**, and that stability condition is often termed the invertibility condition.

If an $ARMA$ process is invertible, then we may also express it in an infinite–order AR representation

$$\Theta(L)^{-1}\Phi(L)y_t = c\Theta(1)^{-1} + \epsilon_t.$$

Note that a process can be invertible yet not be stationary, and vice versa; both of its lag polynomials have to satisfy their respective stability condition if the process is to be termed **stationary and invertible**. If the process is both stationary and invertible, then the finite–order $ARMA(p,q)$ process also possesses both an $AR(\infty)$ and a $MA(\infty)$ representation.

The autocovariance–generating function

The entire set of autocovariances of a covariance stationary process y_t may be summarized by the **autocovariance generating function**:

$$\begin{aligned} g_y(z) &= \sum_{j=-\infty}^{\infty} \gamma_j z^j \\ &= \sum_{j=1}^{\infty} \gamma_j \left(z^j + z^{-j}\right) \end{aligned}$$

with z a complex scalar. Note that for $|z| = 1$, the unit circle, this implies that

$$\begin{aligned} g_y(1) &= \sum_{j=-\infty}^{\infty} \gamma_j \\ &= \gamma_0 + 2\sum_{j=1}^{\infty} \gamma_j \end{aligned}$$

which will represent a convergent sum given that the autocovari-

ances are absolutely summable. If we transform this function g_y by dividing by 2π and setting z, a complex number on the unit circle, to $e^{-i\omega} = \cos(\omega) - i\sin(\omega)$, we have defined the **spectral density function** or **power spectrum** of y_t :

$$\begin{aligned} s_y(\omega) &= (2\pi)^{-1} g_y\left(e^{-i\omega}\right) \\ &= (2\pi)^{-1} g_y\left(\cos(\omega) - i\sin(\omega)\right) \end{aligned}$$

where ω represents the **frequency** (the inverse of the period) of the cyclical component in y. The time series may be represented in the frequency domain, as the sum of an infinite number of sines and cosines, or in the time domain, as a process with an infinite number of autocovariances. The two representations are equivalent, and we can represent y in the frequency domain as the **Fourier transform** of the timeseries data. We will discuss the frequency domain representation at greater length when we discuss fractionally–integrated processes. (See Hamilton, Chapter 6).

Extension to vector processes

A **vector white noise process** ε_t is a jointly covariance–stationary process satisfying

$$\begin{aligned} E(\varepsilon_t) &= 0, \\ E(\varepsilon_t \varepsilon_t') &= \Omega, p.d., \\ E\left(\varepsilon_t \varepsilon_{t-j}'\right) &= 0, \ for\ j \neq 0. \end{aligned}$$

Since Ω need not be diagonal, there may be contemporaneous covariance among the elements of ε_t. Perfect correlation is ruled out, since Ω must be positive definite. A **vector** $MA(\infty)$ **process**

may be expressed as:

$$y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}.$$
$$\psi_0 = I.$$

This implies that if Γ_j is the $j^{th}-$ order autocovariance matrix $E\left[(y_t - \mu)(y_{t-j} - \mu)'\right]$, the autocovariance function may be written as

$$\Gamma_j = \sum_{k=0}^{\infty} \psi_{j+k} \Omega \psi_k'$$

Multivariate filters may also be written as $H(L) = H_0 + H_1 L + H_2 L^2 + ...$, with vector $y_t = H(L)x_t$. Products of filters may now be computed with linear algebra: for instance, the product of two filters $A(L)$ and $B(L)$ can be written in terms of sums and products of their coefficient matrices.

We may also consider the stability conditions in terms of lag matrix polynomials. For instance, if we have

$$\Phi(L) = I_r - \Phi_1 L - \Phi_2 L^2 - ... - \Phi_p L^p$$

where Φ_j are $r \cdot r$ matrices, we may write the stability condition for this polynomial as

$$|I - \Phi_1 z - ... - \Phi_p z^p| = 0$$

all of the roots of which must lie outside the unit circle. For instance, if we consider a two–variable form of a first–order lag polynomial $\Phi(L) = I_r - \Phi_1 L$ with coefficient matrix $\begin{matrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{matrix}$, the determinantal equation becomes $1 - (\phi_{11} + \phi_{22}) z + (\phi_{11}\phi_{22} - \phi_{12}\phi_{21}) z^2 = 0$.

The $p^{th}-$ order **vector autoregressive process**, $VAR(p)$, will be the solution to the vector stochastic difference equation

$$\begin{aligned} \Phi(L)\left(y_t-\mu\right) &= \varepsilon_t, \\ \mu &= \Phi(1)^{-1}c. \end{aligned}$$

The $VAR(p)$ is itself a special case of a multivariate $ARMA$ model, the $VARMA(p,q)$, where each of the variables in the system are considered to follow an $ARMA$ process of order (p,q).

The autocovariance generating function for a vector covariance–stationary process y_t may be written as

$$\begin{aligned} G_Y(z) &= \sum_{j=-\infty}^{\infty}\Gamma_j z^j \\ &= \Gamma_0+\sum_{j=1}^{\infty}\left(\Gamma_j z_j+\Gamma_j' z^{-j}\right) \end{aligned}$$

which leads to the definition of the spectrum of the vector process as

$$s_Y(\omega)=(2\pi)^{-1}G_y\left(e^{-i\omega}\right).$$

Special cases of this for vector processes are:

$$\begin{aligned} VWN &: G_Y(z)=\Omega \\ VMA(\infty) &: G_Y(z)=\Psi(z)\,\Omega\Psi\left(z^{-1}\right)' \\ VAR(p) &: G_Y(z)=\left[\Phi(z)^{-1}\right]\Omega\left[\Phi(z^{-1})^{-1}\right]' \end{aligned}$$

which thus allows us to define the autocovariances of these processes at all lags.

Estimating autoregressions, VARs and $ARMA$ processes

Pure AR models may be consistently estimated with OLS, with the assumption that the error process is iid. How should the order of an AR process be chosen? A sensible criterion is the general–to–specific sequential t rule, which starts by estimating a model with $p_{\max}$ lags, where that parameter is selected to a.s. overfit the process. If the high–order lag term is significant at some prespecified level, then $p_{\max}$ should be increased. Then, the autoregression is reestimated, dropping the high–order lag, until the high–order lag is significant. Since the $t-$ test is consistent, this rule will never underfit the model; it is biased toward overfitting the model, and thus may be criticized on its lack of parsimony. It should be noted that all of the models described above should be fit over a common sample, to prevent the inclusion of additional data points from modifying the judgement.

A second approach is to apply the Akaike information criterion (AIC) or the Schwartz information criterion (SIC), also known as the Bayesian information criterion (BIC). In general terms, these criteria seek the model which minimizes

$$\log\left(\frac{SSR_j}{n}\right) + \frac{(j+1)\,C(n)}{n}$$

where n is the sample size, SSR_j is the sum of squared residuals (the least squares criterion) for the $j^{th}-$ order autoregression, and $C(n)$ is set to 2 for the AIC and $\log(n)$ for the SIC (BIC). Either of the information criteria strike a balance between a better fit (in the first term) and the parsimony of the model (in the

second term). Ng and Perron suggest, as above, the use of a fixed sample period to compare models with these criteria. Just as with the general–to–specific rule, the AIC has a positive probability of overfitting the model. In contrast, the BIC is a consistent estimator of j.

VARs may equally easily be estimated by single–equation OLS. Since the right–hand side of each equation of the VAR contains the same set of regressors, there is no gain in applying a system estimator, conditional on iid errors. Likewise, there is a multivariate generalization of the AIC and BIC that may be applied to search for the appropriate lag length of the VAR (presuming that the same number of lags are to be used in each equation).

The estimation of $ARMA(p,q)$ models is more challenging, in that the model may be written as

$$\begin{aligned} u_t &= \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q}, \\ z_t &= (1, y_{t-1}, y_{t-2}, ..., y_{t-p})', \\ \delta &= \left(c, \phi_1, \phi_2, ..., \phi_p\right)'. \\ y_t &= z_t' \delta + u_t \end{aligned}$$

which poses several problems. The errors are serially correlated (since they are $MA(q)$), and since the lagged dependent variables included in z_t are correlated, by construction, with lags of ϵ included in the error term, the regressors z_t are not orthogonal to the error term u_t. The second problem could be solved by using suitably lagged values of the dependent variable (at lags higher than q) as instruments. But the efficient estimation

of $ARMA(p,q)$ processes requires a different modelling strategy, relying on the maximum likelihood principle to estimate "Box-Jenkins" models.

For instance, consider the $MA(1)$ process. If we condition on initial values for the ϵ 's, this becomes straightforward. The model:

$$\begin{aligned} Y_t &= \mu + \epsilon_t + \theta\epsilon_{t-1}, \\ \epsilon_t &\tilde{} N(0, \sigma^2). \end{aligned}$$

Let $\Theta = (\mu, \theta, \sigma^2)'$ denote the population parameters of interest. If ϵ_{t-1} was known with certainty, $Y_t|\ \epsilon_{t-1} \tilde{} N(\mu + \theta\epsilon_{t-1}, \sigma^2)$ and we could write

$$f(Y_t|\epsilon_{t-1}; \Theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(y_t - \mu - \theta\epsilon_{t-1})^2}{2\sigma^2}\right]$$

If we knew for certain that $\epsilon_0 = 0$, we would know the value of ϵ_1 with certainty as well, and could in turn calculate the value of ϵ_2 conditional on ϵ_1, and so on. The entire sequence of ϵ values can then be recursively calculated from

$$\epsilon_t = y_t - \mu - \theta\epsilon_{t-1}$$

given knowledge of μ and θ. Thus the conditional loglikelihood may be written as

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

for a particular value of Θ. Numerical optimization may then be used to vary elements of Θ and search for a maximum of

the loglikelihood function. No analytical solution exists—even for the $MA(1)$ —but the solution technique should be reliable. This approach conditions the estimates on the specification that $\epsilon_0 = 0$; other approaches, which may be found in more sophisticated software, may use other techniques for generating the initial conditions for optimization. If $|\theta|$ is substantially less than unity, the effect of imposing $\epsilon_0 = 0$ will quickly die out, and the conditional likelihood will be a good approximation to the unconditional likelihood for a reasonably large sample size. In contrast, if $|\theta| > 1$, the consequences of imposing $\epsilon_0 = 0$ accumulate over time. If numerical optimization yields a value in this region, the results must be discarded. An appropriate starting value for numerical optimization in that case would be $\hat{\theta}^{-1}$.

In particular, a $MA(q)$ model will require the specification of q starting values for pre–sample elements of the ϵ vector. Just as above, we may generate conditional likelihood estimates on the assumption that $\epsilon_0 = \epsilon_{-1} = ... = \epsilon_{-q+1} = 0$. The conditional loglikelihood is only useful if all values of z for which

$$1 + \theta_1 z + \theta_2 z^2 + ... + \theta_q z^q = 0$$

lie outside the unit circle.

Numerical optimization to estimate the parameters of a general $ARMA(p, q)$ process require two sets of assumed starting values: the q presample values of the ϵ sequence, and p presample values of the y process. Of course, one may have the prior values of y (that is, you may start the estimation at least p periods after the start of the data). Box and Jenkins, in their classic text, recommended setting the presample ϵ 's to zero,

but setting the presample y 's to their actual values. The same caution applies with respect to stability of the resulting estimates: if the MA polynomial does not satisfy the stability condition (i.e. invertibility), the estimates should not be trusted.