Handout 6

February 24, 2010

1 Some Basic Time-Series Concepts

You can find many good introductions to time series. See, for example, chapters 1 to 3 and 5 of Hamilton's "Time Series Analysis."

1.1 Recursive Substitution

Consider the process

$$y_t = \rho y_{t-1} + \epsilon_t.$$

The equation above summarizes the evolution of y_t through time using a first-order difference equation. The order is given by the fact that only the first lag of y appears in the equation. Notice also that the equation is linear. For most of the course we'll be concerned with linear processes, only touching on non-linear stochastic processes towards the end of the course. Notice that the relationship in equation (1.1) implies

$$y_1 = \rho y_0 + \epsilon_1$$
$$y_2 = \rho y_1 + \epsilon_2$$
$$y_3 = \rho y_2 + \epsilon_3$$

$$\vdots$$

$$y_T \; = \; \rho y_{T-1} + \epsilon_T$$

Knowing the starting value of y, y_0, we could back out the path ϵ through time. Notice also that, by repeated substitution, we can express y_t as:

$$y_t = \rho^t y_0 + \rho^{t-1}\epsilon_1 + \rho^{t-2}\epsilon_2 + \ldots + \rho\epsilon_{t-1} + \epsilon_t.$$

From the equation above, we can also notice that if ϵ is uncorrelated through time, the partial effect of a change in ϵ_1 on y_t is given by:

$$\frac{\partial y_t}{\partial \epsilon_1} = \rho^{t-1},$$

which also has the interpretation as the dynamic multiplier on y_t of the innovation ϵ_1. Notice that tracing the dynamic multipliers for an innovation through time corresponds to computing an impulse response to the innovation. Thus think of an impulse response function as a collection of dynamic multipliers:

$$IRF(\epsilon_1) = \left\{ \frac{\partial y_1}{\partial \epsilon_1}, \frac{\partial y_2}{\partial \epsilon_1}, \frac{\partial y_2}{\partial \epsilon_1}, \ldots \right\}$$

Notice that if $|\rho| < 1$ then $\lim_{t \to \infty} \frac{\partial y_t}{\partial \epsilon_1} = 0$. In other words, the effect of any one innovation is temporary if ρ lies within the unit circle. By contrast, if ρ lies on or outside the unit circle, an innovation has non-zero effects that extend into the infinite future.

1.2 Higher order difference equations

The method of repeated substitutions extends readily to higher-order processes. Consider, for example

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \rho_3 y_{t-3} + \epsilon_t.$$

In this case, rewrite the equation in companion form:

$$\xi_t = \rho\xi_{t-1} + \begin{pmatrix} \epsilon_t \\ 0 \\ 0 \end{pmatrix}, \tag{1}$$

where

$$\xi_t = \begin{pmatrix} y_t \\ y_{t-1} \\ y_{t-2} \end{pmatrix} \quad \text{and} \quad \rho = \begin{pmatrix} \rho_1 & \rho_2 & \rho_3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \tag{2}$$

Again, knowing the starting values for y, say y_{-1} and y_0, allows us to back out a series for ϵ from date 1 till the end of the observed sample. Notice that by repeated substitution,

$$\xi_t = \rho^t\xi_0 + \rho^{t-1}\begin{pmatrix} \epsilon_1 \\ 0 \\ 0 \end{pmatrix} + \rho^{t-2}\begin{pmatrix} \epsilon_2 \\ 0 \\ 0 \end{pmatrix} + ... + \rho\begin{pmatrix} \epsilon_{t-1} \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ 0 \\ 0 \end{pmatrix}. \tag{3}$$

From the equation above, one can see that

$$\frac{\partial y_t}{\partial \epsilon_1} = \rho_{1,1}^{t-1},$$

where $\rho_{1,1}^{t-1}$ denotes the $1,1$ element of the matrix ρ taken to the $t-1$ power. In this case, investigating whether or not $\lim_{t\to\infty} \frac{\partial y_t}{\partial \epsilon_1}$ is finite, will depend on the eigenvalues of the matrix ρ. Let's consider this question a little more thoroughly.

Remember that if a square matrix ρ of dimension n has n distinct eigenvalues, the Jordan canonical decomposition of ρ takes the form

$$\rho = VDV^{-1}, \tag{4}$$

where D is a matrix with the eigenvalues of ρ along its diagonal and zeros elsewhere. V is a matrix whose columns are the eigenvectors of ρ. From equation (4), one can see that:

$$\rho^{t-1} = VD^{t-1}V^{-1}. \tag{5}$$

But since D is diagonal, the $t-1$ power of D obtains by simply raising to the power of $t-1$ each of its diagonal element individually. Thus, one can see that if all the eigenvalues of ρ are within the unit circle, $\lim_{t\to\infty} \frac{\partial y_t}{\partial \epsilon_1} = 0$.

For an extension of the argument above to the case in which ρ does not have n distinct eigenvalues see Chapter 1 of Hamilton's "Time Series Analysis."

2   Unconditional Mean and Variance of AR Processes

Extending the argument used in the solution of AR processes by repeated substitution, one can rewrite an AR process as a function of innovations for the infinite past, call this the MA(∞) representation. We can use this representation to find the mean and autocovariances as long as the roots of the process lie within the unit circle. To fix ideas we shall work with an AR(1) process, but the arguments presented readily extend to higher-order processes.

Consider again

$$y_t = \rho y_{t-1} + \epsilon_t$$

where ϵ_t is process independently and identically distributed through time with mean zero and variance σ^2. By repeated substitution:

$$y_t = \epsilon_t + \rho \epsilon_{t-1} + \rho^2 \epsilon_{t-2} + ...$$

Taking expectations:

$$E[y_t] = E[\epsilon_t + \rho \epsilon_{t-1} + \rho^2 \epsilon_{t-2} + ...].$$

But given $|\rho| < 1$ the expression on the RHS is absolutely summable and distributing the expectation operator to each term

$$E[y_t] = E[\epsilon_t] + \rho E[\epsilon_{t-1}] + \rho^2 E[\epsilon_{t-2}] + ... = 0.$$

One can also use the MA(∞) representation to facilitate computing the unconditional variance of the process. Thus,

$$VAR(y_t) = VAR(\epsilon_t + \rho \epsilon_{t-1} + \rho^2 \epsilon_{t-2} + ...)$$

From the fact that the innovations are independently distributed through time

$$VAR(y_t) = VAR(\epsilon_t) + VAR(\rho\epsilon_{t-1}) + VAR(\rho^2\epsilon_{t-2}) + ...$$

With a little manipulation, as long as $|\rho| < 1$,

$$VAR(y_t) = \sigma^2 + \rho^2\sigma^2 + \rho^4\sigma^2 + ... = \frac{\sigma^2}{1 - \rho^2}$$

2.1 Autocovariances, autocorrelations, and the correlogram

Autocovariances are covariances of a process with its own lag. Trivially, the autocovariance at lag zero is the variance. The autocovariance for lag 1 is $COV(y_t, y_{t-1})$, and so on. The MA(∞) representation of an AR process can also come in handy in computing the autocovariances. Again, taking as an example and AR(1) process:

$$y_t \quad = \quad \epsilon_t + \rho\epsilon_{t-1} + \rho^2\epsilon_{t-2} + \rho^3\epsilon_{t-3}... \tag{6}$$

$$y_{t-1} \quad = \quad \quad \epsilon_{t-1} + \rho\ \epsilon_{t-2} + \rho^2\epsilon_{t-3}... \tag{7}$$

Then autocovariance at lag 1 can be found as

$$COV(y_t, y_{t-1}) = E[\epsilon_t + \rho\epsilon_{t-1} + \rho^2\epsilon_{t-2} + ...] \cdot [\epsilon_{t-1} + \rho^2\epsilon_{t-2} + ...]$$

Collecting terms, as long as $|\rho| < 1$,

$$COV(y_t, y_{t-1}) = \rho\sigma^2 + \rho^3\sigma^2 + \rho^5\sigma^2 + ... = \rho\frac{\sigma^2}{1 - \rho^2}.$$

By extension,

$$COV(y_t, y_{t-j}) = \rho^j\frac{\sigma^2}{1 - \rho^2}.$$

Normalizing the covariance by the variance produces the autocorrelation. Thus, for an AR(1) process the autocorrelation at lag j is given by

$$\frac{COV(y_t, y_{t-j})}{VAR(y_t)} = \frac{\rho^j\frac{\sigma^2}{1-\rho^2}}{\frac{\sigma^2}{1-\rho^2}} = \rho^j.$$

A plot of the autocorrelation function with the value of the autocorrelation along the ordinates and the lag on the abscissae produces the correlogram, as demonstrated by the Matlab program

plot_ar1.m in the zipped file corresponding to this handout. The pattern of exponential decay of the correlogram is a feature that extends to autoregressive processes of higher orders and cannot be used to distinguish visually among autocorrelation processes of different order.

2.2 Some definitions

- **Weak Stationarity**. If the mean and the autocovariances at all lags do not depend on the date t, then a time series process is said to be (weakly) stationary.

- **Strong stationarity**. If the joint distribution of $X(t_1), ..., X(t_k)$ is the same as the joint distribution of $X(t_{1+\tau}), ..., X(t_{k+\tau})$ for all k and τ, then the time series process governing X is said to be strictly stationary.

- **Gaussianity**. If the joint distribution $X(t_1), ..., X(t_k)$ is jointly normal for any k, then the process is said to be Gaussian. It follows that a covariance-stationary Gaussian process is also strictly stationary.

- **White Noise**. Consider the sequence $\{\epsilon_t\}_{t=-\infty}^{\infty}$. If all elements have mean 0 and variance σ and if $E(\epsilon_t, \epsilon_{t+\tau}) = 0$ for all τ except $\tau = 0$, then the process governing the ϵ terms is said to be white noise.

3 Estimating autoregressive processes

A typical assumption of the multiple regression model is that the regressors do not covary with the error term of the regression at any lag of time. That assumption is clearly violated in the case of an autoregressive process. As a consequence, the OLS estimator is not the best linear unbiased estimator (it will be biased, in fact) and there is no particular reason for sticking with it. What other methods are available then for the estimation of an autoregressive process? Let's consider two alternatives: conditional and unconditional maximum likelihood estimation.

3.1 Maximum likelihood estimation

To fix ideas we shall consider how to set up the likelihood of an AR(1) process, but the procedure easily extends to higher-order processes. Consider, again, the process

$$y_t = \rho y_{t-1} + \epsilon_t.,$$

where ϵ_t is NID with variance σ^2. Suppose that there are only two observations, then their joint density function can be written as

$$P(y_2, y_1) = P(y_2|y_1)P(y_1) \tag{8}$$

Similarly for 3 observations

$$P(y_3, y_2, y_1) = P(y_3|y_2, y_1)P(y_2, y_1) \tag{9}$$

Substituting (8) into (10)

$$P(y_3, y_2, y_1) = P(y_3|y_2, y_1)P(y_2|y_1)P(y_1) \tag{10}$$

Extending this reasoning to T observations, you can see that the likelihood can be written as:

$$L(Y_T, \rho, \sigma^2) = P(y_1) \prod_{t=2}^{T} P(y_t|Y_{t-1}) \tag{11}$$

where Y_t represents the vector of observations from 1 through y_{t-1}. Remember that if ϵ is distributed as Normal with mean μ and variance σ^2, then the probability density function for epsilon is given by:

$$PDF(\epsilon) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(\epsilon - \mu)^2}{2\sigma^2}\right)$$

In turn, for the AR(1) process in equation (3.1), the log-likelihood function takes the form:

$$\log L(Y_T, \rho, \sigma^2) = -\frac{T-1}{2}\log(2\pi) - \frac{T-1}{2}\log\sigma^2 - \sum_{t-2}^{T}\frac{(y_t - \rho y_{t-1})^2}{2\sigma^2} + \log P(y_1) \tag{12}$$

While the evaluation of (??) seems straight-forward, the term $\log P(y_1)$ deserves some further discussion. If this term can be thought of as being fixed in repeated draws of the observed sample, then it simply drops out of the likelihood:

$$\log L(Y_T, \rho, \sigma^2) = -\frac{T-1}{2}\log(2\pi) - \frac{T-1}{2}\log\sigma^2 - \sum_{t-2}^{T}\frac{(y_t - \rho y_{t-1})^2}{2\sigma^2} \tag{13}$$

Maximizing the remaining terms yields the **conditional** maximum likelihood estimates for the AR(1) process. Conveniently, there is no need to numerically optimize the function above, as the maximum likelihood estimate can also be thought of as the estimate produced by minimizing the residual sum of square, as performed by the OLS estimator.

An alternative way of proceeding for stationary series (in this case, when $|\rho| < 1$) is to think that y_1 was drawn for the unconditional distribution for y. As seen above, y_1 then has mean 0 and variance $\frac{\sigma^2}{1-\rho^2}$. Thus, equation (**??**) becomes

$$\log L(Y_T, \rho, \sigma^2) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\sigma^2 + \frac{1}{2}\log(1-\rho^2) - \sum_{t-2}^{T} \frac{(y_t - \rho y_{t-1})^2}{2\sigma^2} - \frac{y_1^2}{\frac{\sigma^2}{1-\rho^2}} \qquad (14)$$

which is referred to as the **unconditional** likelihood. Unfortunately, in this case, numerical optimization of the likelihood cannot be avoided.

4   Bootstrap Standard Errors and Confidence Intervals

So far, we have not considered the question of how to retrieve the standard errors for the parameter estimates of the AR model. The inversion of the information matrix is still a valid way of retrieving the variance covariance matrix for the estimates and has asymptotic justification. An alternative is the bootstrap method, that in practice seems to have better properties in small samples, even though its justification is asymptotic in nature.

Here is a cookbook recipe for the application of the bootstrap method to an AR(1) process:

$y_t = \rho y_{t-1} + \epsilon_t$

1. Estimate ρ by $\hat{\rho}$

2. Form the residuals $\hat{\epsilon}_t = y_t - \hat{\rho} y_{t-1}$

3. Sample from the residuals with replacement, and form new data $y_{i,t}$

4. With the new data, form the estimate $\hat{\rho}_i$

5. Repeat the replication and re-estimation of ρ an appropriately large number of times

6. The variance of $\hat{\rho}_i$ can be taken to be the estimate of the variance of $\hat{\rho}$.

The procedure above also yields a way to compute confidence intervals for the parameter estimates. Inverting the normal distribution, a 90% confidence interval can be constructed as $\hat{\rho} - 1.6 * \sqrt{VAR(\hat{\rho}_i)} < \rho < \hat{\rho} + 1.6 * \sqrt{VAR(\hat{\rho}_i)}$.

5 Moving Average Processes

Consider the process

$$y_t = \mu + \epsilon_t + \rho\epsilon_{t-1}, \tag{15}$$

where μ and ρ could be any constants and ϵ is governed by a white noise process. The process for y is called a moving average process of first-order, or MA(1).

5.1 Mean, Variance, and Covariances

The expectation of y is given by:

$$E(y_t) = \mu + E(\epsilon_t) + \rho E(\epsilon_{t-1}).$$

Which leads to:

$$E(y_t) = \mu.$$

The unconditional variance of y is given by:

$$VAR(y_t) = E(y_t - \mu)^2 = E(\epsilon_t + \rho\epsilon_{t-1})^2.$$

But is ϵ is governed by a white noise process:

$$VAR(y_t) = 1 + \rho^2\sigma^2.$$

The covariance at lag one is given by:

$$E[(y_t - \mu)(y_{t-1} - \mu)] = E[(\epsilon_t + \rho\epsilon_{t-1})(\epsilon_{t-1} + \rho\epsilon_{t-2})] = \rho\sigma^2$$

For the MA(1) process, the covariances at all lags higher than 1 are zero.

5.2 The nth order moving average process

Consider the process

$$y_t = \mu + \epsilon_t + \sum_{i=1}^{n} \rho_i \epsilon_{t-i}, \tag{16}$$

where ϵ is governed by a white noise process. The process for y is MA(n). The variance of the process is given by:

$$VAR(y_t) = (1 + \sum_{i=1}^{n} \rho_i^2)\sigma^2.$$

One can easily see that the covariance at lag $j > 0$, but $j <= n$, takes the form:

$$COV(y_t, y_{t-j}) = (\rho_j + \rho_{j+1}\rho_1 + \rho_{j+2}\rho_2 + ... + \rho_n\rho_{n-1})\sigma^2.$$

For all lags greater than n, the covariance is zero. To give a concrete example, for an MA(2) process, the formulae above imply:

$$VAR(y_t) = (1 + \rho_1^2 + \rho_2^2)\sigma^2$$

$$COV(y_t, y_{t-1}) = (\rho_1 + \rho_2\rho_1)\sigma^2$$

$$COV(y_t, y_{t-2}) = \rho_2\sigma^2$$

$$COV(y_t, y_{t-3}) = 0.$$

Notice that from the formulas above, we can also infer that there are no restrictions on the parameters governing an MA process to ensure that the process be covariance stationary.