EC821: Time Series Econometrics, Spring 2003

Notes Section 1

Data types

In the narrowest sense, this course is concerned with time series data—those in which the individual observations are indexed by some notation of calendar time. What is a time series? In a very simple sense, merely a set of data indexed by some regular increment of time, which may or not be regular in the sense of the calendar. For instance, an interest rate series containing observations from the last day of each month will be unequally spaced in calendar time—and is likely that the "last day" will be other than that in the presence of weekends and holidays. Also, we often work with time series of "business daily" data, generated (at most) five days per week. We assume that the T observations available are a finite segment of a doubly infinite sequence, which goes back into the infinite past, and forward to infinity. We might consider several specific time series: for instance, a time trend is merely $y_t = t$; a constant series is $y_t = c$, and a Gaussian white noise process is $y_t = \epsilon_t$.

We speak of time series data to contrast with the other major form of data organization: the cross section, in which each observation is indexed by an identifier such as a person's Social Security number, a company's CUSIP, a country's ISO abbreviation, or a survey respondent's questionnaire ID. But there are two other forms of data organization which we often encounter, and which are very relevant for much of the research carried out with time series data. These are pooled cross sections, and longitudinal or panel data. Pooled cross sections, even if the cross sections are of the same size (e.g., 400 respondents to a weekly poll of presidential popularity) do not give rise directly to time series, since the individual observations are not linked across time. However, summary statistics from each of these cross sections may be computed, and assembled into time series to illustrate temporal changes in the sample's measures of central tendency and dispersion.

One of the most rapidly expanding areas in applied econometrics is the use of longitudinal, or panel, data—data indexed by both i and t subscripts, essentially repeated measurements on the same individuals over time. Panel data in common use in economics and finance have tended to appear in two common forms: the "small T, large N" panel, exemplified by Compustat firm–level data, or the Panel Study of Income Dynamics household data. These datasets have relatively few time–series observations but thousands of individuals. Many of the econometric techniques developed in this area make use of the "small T, large N" setting: for instance, Arellano and Bond's dynamic panel data GMM estimation technique (cf. Stata command `xtabond`). The econometric theory underlying these estimators is based on N$\rightarrow \infty$ for T fixed. In this context, there are few time–series aspects of the data that may be modelled, since the number of time series observations is quite limited. The other common form is the "small N, large T" panel, exempified by daily data from the financial markets (G–7 exchange rates or inflation rates, stock price series for a limited number of firms, etc.)

In this context, we generally have hundreds or thousands of time series observations on each unit, but a relatively small number of units. This allows many more of the time series properties of the data to be exploited in the estimation. Similar to the first form, the underlying asymptotic theory for estimators applied to these data is based on the assumption that $T \rightarrow \infty$ for N fixed. Datasets of these form are often used to perform panel unit root tests, or analyses of cointegration.

Dimensionality

When we utilise time series data, it is important to determine the dimensionality of a variable: e.g., whether it refers to a level, stock, flow, or rate of change. An interest rate, or a consumer price index, or an index of industrial production at a point in time is a level. The capital stock employed by the firms in an industry at a point in time is just that—a stock, in units of number of machines, or their value in real terms. That stock is accumulated or decumulated over time via capital investment, which is a flow—and must refer to the unit of time over which the flow is measured. If a flow is converted into percentage terms, it becomes a dimensionless rate of change, such as the rate of change of the consumer price index, which we term inflation. It is customary to measure rates of change at an annual rate, so that inflation might be described as reaching three percent last month: for most economies, this does not reflect the monthly change in the price level, but rather the monthly change compounded to an annual rate.

Time series operators

We make extensive use in time series models of various time series operators, as shorthand to specify how a variable may be transformed. For instance, the multiplication operator might be used: $y_t = \beta x_t$. Unlike ordinary algebraic manipulation, this represents applying the operator to each element of the time series (for all defined t), analogous to filling a formula through all the rows of a spreadsheet, or using Stata's `generate` command to define a mappng between x and y. Since the multiplication or addition operators refer to element–wise operations, they follow all the rules of algebra: e.g. $z_t = \beta\,(y_t + x_t)$ can be implemented either by adding each period's values and multiplying the result by β, or by multiplying through the parens and then summing. Multiplication and addition are commutative.

The commonly employed first difference operator, Δ, transforms a series into its increments: for instance, ΔK_t, the first difference of the capital stock, will by convention generate the change between the previous period's value, K_{t-1}, and the current value, K_t. This involves computing the lagged value of K_t, which may itself be expressed using the lag operator as LK_t : that is, the lag operator applied to a set of observations "backshifts" them by one time unit. For this reason, many textbooks and articles express L as the B, or backshift operator. Note that the lag operator may be expressed as L^j; if $j = 1$, we consider the previous period's lag, but may express any lag. For quarterly data, $L^4 y_t$ would express the value of y four quarters ago, while the lead of y, $L^{-4} y_t$, would refer to the value of y four periods later. A negative power applied to the lag generates the lead, and just as any algebraic quantity raised to the zero power is itself, $L^0 y_t = y_t$. The lag operator also satisfies the law of exponents, so that $L^i L^j y_t = L^{i+j} y_t = y_{t-(i+j)}$; it is commutative with the multiplication operator, and distributive over the addition operator, so that $L(y_t + x_t) = Ly_t + Lx_t$.

One must be cautious in using the difference operator to note that Δ^j refers to the j^{th}difference: that is, the difference operator applied repeatedly. Thus $\Delta y_t = y_t - y_{t-1,}$ and $\Delta^2 y_t = \Delta\Delta y_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}$, and so on. One may define Δ^j for any positive j, but this should not be mistaken for, e.g., $y_t - L^4 y_t$, which is not the “fourth difference” of y_t. It should be remembered that the difference operator is the discrete equivalent of $\frac{dy}{dt}$, so that the second difference is the analog to $\frac{d^2y}{dt^2}$: the acceleration of y.

In Stata, the time series operators $D.$, $L.$ and $F.$ when prefixing a variable are the first difference, lag, and lead (forward) of the variable, respectively. The number of times that the operator is to be applied may also be specified in this syntax, so that one may say $L4.y$ or $D2.price$ to refer to the fourth lag of y, or the second difference of the price level, respectively.

Lag polynomials

In working with these time series operators, we often employ lag polynomials; e.g. the difference operator Δ may be defined as $(1 - L)$, a first–order polynomial in the lag operator, while $\Delta^2 = (1 - L)^2$. This allows easier access to the definition of the higher–order difference; since the expression $(1 - L)^2$ may be expanded as $(1 - 2L + L^2)$, it may immediately be seen that $\Delta^2 y_t = y_t - 2y_{t-1} + y_{t-2}$, where the coefficients of the expanded lag polynomial appear in each term. As shorthand, we often will write $A(L) = (1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3)$ as a finite–order lag polynomial The shorthand $A(1)$ refers to the steady state representation, in which all lags are set to zero: $A(1) = (1 - \sum \phi_i)$. We might also write a factored polynomial, such as $(1 - \lambda_1 L)(1 - \lambda_2 L) x_t = x_t - (\lambda_1 + \lambda_2) x_{t-1} + \lambda_1 \lambda_2 x_{t-2}$.

The univariate dynamic model

The simplest model of a stochastic process y_t could be written as a linear univariate dynamic model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \epsilon_t$$

which is a p^{th} order stochastic difference equation with constant coefficients. The process is said to be autoregressive, since the value of y depends on its own past values, as well as the innovation process ϵ. This may be written in lag polynomial notation as $\phi(L) y_t = \epsilon_t$ where the polynomial will be $\left(1 - \phi_1 L + \phi_2 L^2 + ... + \phi_p L^p\right)$. We will consider this model (or its more general counterpart which incorporates a constant term) as an $AR(p)$ model of y_t.

A dynamic bivariate model

Let us consider a dynamic model that contains two variables: an implementation of the partial adjustment model (PAM) that might be applied to the stock of consumer durables, or a firm's capital stock. The PAM contains a target value for y_t, denoted y_t^*, which in equilibrium will be realised. Due to adjustment costs, economic agents do not reach equilibrium immediately following a revision of the target value. Thus the law of motion for y_t may be expressed as

$$\Delta y_t = y_t - y_{t-1} = \delta \left(y_t^* - y_{t-1} \right) + \epsilon_t, |\delta| < 1$$

where the parameter δ indicates the speed of adjustment; the closer is δ to unity, the more rapidly is equilibrium restored. The model may be rewritten as

$$y_t = \delta y_t^* + (1 - \delta) y_{t-1} + \epsilon_t \tag{1}$$

so that the current value of y_t is a convex combination of the desired stock and last period's value. If we now define the desired stock as depending upon an exogenous factor x_t, $y_t^* = \lambda_0 x_t$, that may be substituted into (1) to yield, with some redefinition of parameters,

$$(1 - \gamma_1 L) y_t = \beta_0 x_t + \epsilon_t$$

The resulting dynamic model has one lag on the dependent variable and no lags on the explanatory variable. The model may be generalised, of course, to include richer dynamics in the determination of the desired stock.

The autoregressive distributed lag (ADL) model[1]

The model as given above is a special case of the autoregressive distributed lag (ADL) model. We can express the general ADL model as $ADL(p, q)$:

$$\begin{aligned} \left(1 - \sum_{i=1}^{p} \gamma_i L^i\right) y_t &= \alpha_0 + \left(1 - \sum_{j=0}^{q} \beta_j L^j\right) x_t + \epsilon_t \\ \gamma(L) y_t &= \alpha_0 + \beta(L) x_t + \epsilon_t \end{aligned}$$

[1] This section is taken from Davidson and MacKinnon, Estimation and Inference in Econometrics, section 19.4.

where ϵ_t is assumed to be IID($0,\sigma^2$). If we consider the $ADL(1,1)$ special case,

$$y_t = \alpha_0 + \gamma_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \epsilon_t \tag{2}$$

we may note that many models are nested within it. For instance, if $\gamma_1 = \beta_1 = 0$, we have a static regression. A univariate $AR(1)$ model for y sets $\beta_0 = \beta_1 = 0$. A model in first differences is a special case with $\gamma_1 = 1$ and $\beta_1 = -\beta_0$; α_0 in that context implies the absence of a time trend. Since all of these models are nested within the $ADL(1,1)$ model, they may readily be tested as restrictions upon that more general form.

We may also consider the steady state of this model, and the long run effect of x upon y. By removing all time subscripts and solving, we may derive that

$$y = \frac{\alpha_0}{1-\gamma_1} + \lambda x, \ \lambda = \frac{\beta_0 + \beta_1}{1-\gamma_1} \tag{3}$$

where the long run multiplier contains the sum of the lag coefficients, amplified by the term $(1-\gamma_1)^{-1}$. The stability condition requires that $|\gamma_1| < 1$. If y and x were measured in logarithms (as often is the case in macroeconomic models) then λ will be a long–run elasticity.

An interesting feature of the ADL model is that it may be rewritten in many different forms without affecting the model's ability to explain the data, or changing the least squares estimates of the coefficients of interest. For instance, we may rewrite (2) as:

$$\Delta y_t \ = \ \alpha_0 + (\gamma_1 - 1)\, y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \epsilon_t$$

$$
\begin{aligned}
\Delta y_t &= \alpha_0 + (\gamma_1 - 1)\, y_{t-1} + \beta_0 \Delta x_t + (\beta_0 + \beta_1)\, x_{t-1} + \epsilon_t \\
\Delta y_t &= \alpha_0 + (\gamma_1 - 1)\, y_{t-1} - \beta_1 \Delta x_t + (\beta_0 + \beta_1)\, x_t + \epsilon_t \\
\Delta y_t &= \alpha_0 + (\gamma_1 - 1)\, (y_{t-1} - \lambda x_{t-1}) + \beta_0 \Delta x_t + \epsilon_t
\end{aligned}
$$

These reformulations are useful, in that they make it possible to receive direct point and interval estimates of, say, the sum of the coefficients on x (as in the second and third forms). The most celebrated form is the last, which is known as the error correction form or error correction model (ECM), as defined by Hendry and Anderson (1977) and Davidson et al. (DHSY, 1978). The ECM, as we shall later discuss, expresses the revision of y in terms of the most recently observed disequilibrium in the system—since the parenthesized expression is the degree to which the long run equilibrium of the system was perturbed in the previous time period. Unlike the other forms of the model, the ECM form introduces a nonlinearity into the model, but it merely represents a linear model reparameterized in a nonlinear fashion. The error correction term is implicitly present in any of the other specifications, since its coefficient may be recovered from all of them. If the restriction $\lambda = 1$ is imposed (which may be sensible if y and x are similar in magnitude), the ECM form becomes linear in the parameters. This restriction, given the long run multiplier, is equivalent to the restriction that $\gamma_1 + \beta_0 + \beta_1 = 1$, which may readily be tested from any of the other forms of the model.

An excellent reference for these topics is Hendry, Pagan, Sargan, *Dynamic specification*, chapter 18 in the *Handbook of Econometrics*, volume II (1984).

Vector time series models

These ADL models may be straightforwardly extended to the multivariate context by considering y_t as a vector of several variables, which then are determined by their own lags as well as potentially influenced by the current and lagged values of additional variables. A pure autoregressive form of this structure is the VAR, or vector autoregression, which we will discuss at length later in the course. The VAR model may be augmented with additional exogenous variables, in which case the counterpart to the steady state of the univariate ADL model (3) may be defined in terms of a matrix lag polynomial. In the simplest case—that of a two–variable "pure VAR"—we have $y_t = (y_{1t}, y_{2t})'$. If we consider a fourth–order VAR, we may write

$$\left(I - \Pi_1 L - \Pi_2 L^2 - \Pi_3 L^3 - \Pi_4 L^4\right) y_t = \mu + \epsilon_t$$

where μ is a 2x1 vector of constant terms, and ϵ is a 2x1 vector of error terms. This model can be rewritten as

$$A(L)y_t = \mu + \epsilon_t$$

Just as in the univariate case, we may evaluate the polynomial at $A(1)$ in order to compute the sum of the lag weights.

We may also consider a VAR with additional exogenous variables—the multivariate analogue to the ADL model—in which an additional polynomial $B(L)x_t$ appears, so that now

$$A(L)y_t = \mu + B(L)x_t + \epsilon_t$$

The long-run value of the system may be computed, assuming that the $A(L)$ polynomial is invertible, as

$$y_t = A(L)^{-1}\mu + A(L)^{-1}B(L)x_t + A(L)^{-1}\epsilon_t$$

which is sometimes termed the final form of the system. This in turn may be expressed as

$$y_t = \mu^* + C(L)x_t + v_t$$

where the polynomial $C(L)$ is a "rational lag"—a ratio of two finite-order polynomials, which is an expression of infinite order, $C(L) = \sum_{i=0}^{\infty} C_i L^i$. The long-run relationship is given by $C(1) = A(1)^{-1}B(1)$, which will only be defined when the matrix $A(1)$ is non-singular. This long–run equilibrium may be achieved iff the model is dynamically stable; for a vector process, the requirement is that the eigenvalues of the matrix polynomial $A(L)$ have modulus less than one in absolute value. Since $A(L)$ is not in general a symmetric matrix, its eigenvalues will be complex, so that the moduli of these complex quantities must be evaluated (see Stata routine `geneigen`).

We will next take up the definitions of stationary and nonstationary random variables.