EC327: Financial Econometrics, Spring 2010

Wooldridge, Introductory Econometrics (4th ed, 2009)

Chapter 11: OLS with time series data

Stationary and weakly dependent time series

The notion of a *stationary process* is an important one when we consider econometric analysis of time series data. A stationary process is one whose probability distribution is stable over time, in the sense that any set of values (or *ensemble*) will have the same joint distribution as any other set of values measured at a different point in time. The stationary process is *identically distributed*, in the sense that its mean and variance will be the same whenever we measure it. It surely need not be *independently* distributed, and in fact most time series processes are far from independent. But (strict) stationarity requires that any correlation between elements of the process must not be time-varying.

A stochastic process which fails to satisfy these requirements is said to be a *nonstationary process*. This covers a very wide range of observed processes because the requirements of stationarity literally require that all moments of the process are constant over time. For instance, any process exhibiting a trend will fail to meet that requirement in that its mean is changing over time, even if its variance and covariances are not.

We often refer to a weaker form of stationarity known as *covariance stationarity* (CS). A process with a finite second moment, or variance, is CS if $E(x_i)$ is constant, $Var(x_i)$ is constant, and for any $t, h \geq 1$, $Cov(x_t, x_{t+h})$ depends only on h and not on t. CS focuses on only the first two moments of the stochastic process, and is in that sense a less restrictive condition than strict stationarity. Nevertheless, a trending time series will not be CS.

If we want to analyze the relationship between two or more time series variables, we must assume some sort of stability over time. Their joint distribution may include trending behavior, but if a model of that behavior is to be sensible there must be some stable relationship between them.

We also must be familiar with the concept of *weak dependence*, which places restrictions on the relationship between x_t and x_{t+h} as h gets large. A stationary time series is said to be weakly dependent (WD) if x_t and x_{t+h} are "almost independent" as h increases without bound. For a CS time series, this corresponds to the correlation between x_t and x_{t+h} going to zero "sufficiently quickly" as $h \to \infty$. Such a series is said to be *asymptotically uncorrelated*.

The simplest example of a WD series (beyond the trivial case of an $i.i.d.$ series) is a *moving average process* (MA). The simplest form, MA(1), is

$$x_t = e_t + \alpha_1 e_{t-1},\ t = 1, 2, \ldots \tag{1}$$

where $e_t \sim (0, \sigma_e^2)$. In this process, x_t is a weighted average of e_t and e_{t-1}. Adjacent terms in the x process are correlated because $Cov(x_t, x_{t-1}) = \alpha_1 Var(e)$, and $Corr(x_t, x_{t-1}) = \alpha_1/(1+\alpha_1^2)$. But if we consider elements of the sequence two (or more) periods apart, they are uncorrelated. The MA(1) process is said to have a *finite memory* of one period, in that the prior period's value matters, but observations prior to that have no effect on the current value of the process. Given that e_t is $i.i.d.$, x_t

is a stationary process, and exhibits weak dependence.

The other main building block of time series processes is the *autoregressive process* (AR). The simplest such process, AR(1), is

$$y_t = \rho_1 y_{t-1} + e_t,\ t = 1, 2, \ldots \qquad (2)$$

With the initial condition $y_0 = 0$ and $e_t \sim (0, \sigma_e^2)$, y_t will be a stable stochastic process if we satisfy the condition $|\rho_1| < 1$. Unlike the MA(1) process, the AR(1) process has *infinite memory* because every past value of the process affects the current value. However, as the effects of past values are weighted by powers of the fraction ρ_1, those effects damp to zero, and the process can be shown to be asymptotically uncorrelated. First let us show that it is CS. Given the assumption that e_t is distributed independently of y_0, we can see that $Var(y_t) = \rho_1^2 Var(y_{t-1}) + Var(e_t)$, or $\sigma_y^2 =$

$\rho_1^2\sigma_y^2 + \sigma_e^2$. Given the stability condition, this implies that

$$\sigma_y^2 = \frac{\sigma_e^2}{(1-\rho_1^2)} \tag{3}$$

where it may be seen that if $\rho_1 = 1$ this variance goes to infinity.

By substitution, we may calculate the covariance between y_t and $y_{t+h},\ h \geq 1$ as $\rho_1^h E(y_t^2) = \rho_1^h \sigma_y^2$. For a CS series, the standard deviation of y_t is a constant, so that $Corr(x_t, x_{t+h}) = \rho_1^h$, implying that successive terms in the sequence are correlated by ρ_1. Even if ρ_1 is large, e.g., 0.9, powers of that fraction approach zero rapidly because they follow a geometric progression.

A trending series, although not CS, can be weakly dependent, and a series that is stationary around a deterministic time trend is said to be *trend stationary*. We can use such series in regression models as long as we take proper account of the presence of a trend: for instance, it would be a mistake to "explain" a series that has an obvious trend by regressing it on a series devoid of trend.

Asymptotic properties of OLS

To work with OLS regression in a time series context, we add several assumptions to those made in the context of a cross section model. First, we assume that the set of variables $[\mathbf{x}, y]$ is stationary and weakly dependent. Stationary is not critical in this context, but weak dependence is very important.

The standard zero conditional mean assumption is modified for time series data to state that the explanatory variables $[\mathbf{x}]$ are contemporaneously exogenous of the error process:

$E(u_t|\mathbf{x}) = 0$. This assumption does not restrict the relationship between the regressors and the error process at other points in time; it merely states that their current values are independent at each point in time. For instance, we may have a model such as

$$y_t = \beta_0 + \beta_1 z_{t1} + \beta_2 z_{t2} + u_t \qquad (4)$$

The contemporaneous exogeneity condition will hold if the conditional mean $E(u_t|z_{t1}, z_{t2}) = 0$. This condition does not rule out, for instance, correlation between u_{t-1} and z_{1t}, as might arise if z_1 is a policy variable that reacts to past shocks.

Under these assumptions—plus the assumption that there is no perfect collinearity in the regressor matrix—the OLS estimators are *consistent*, but not necessarily unbiased.

A *finite distributed lag* model, such as

$$y_t = \alpha_0 + \beta_0 z_t + \beta_1 z_{t-1} + \beta_2 z_{t-2} + u_t \qquad (5)$$

is associated with the zero conditional mean assumption that the error term is independent of current and all lagged values of z.

In these two contexts, the regressors *could* be strictly exogenous (for instance, in the absence of feedback that would arise in a policy context). But in the AR(1) model, strict exogeneity is not possible:

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t \tag{6}$$

where we assume that the error has a zero expected value given the entire history of y:

$$E(u_t | y_{t-1}, y_{t-2}, \ldots) = 0 \tag{7}$$

and these equations can be combined to yield

$$E(y_t | y_{t-1}, y_{t-2}, \ldots) = E(y_t | y_{t-1}) = \beta_0 + \beta_1 y_{t-1} \tag{8}$$

This last expression summarizes the "first order" nature of the AR(1) model: once the first lag of y is included, no additional lags of y

affect the expected value of y in this linear relationship. But in this model, the strict exogeneity assumption cannot hold, since that would require that u_t is independent of all values of the regressor, past, present and future. This condition obviously does not hold because u_t is a component of the following period's y_{t-1}. In fact, $Cov(y_t, u_t) = Var(u_t) > 0$. For weak dependence to hold, the stability condition $|\beta_1| < 0$ must be satisfied.

To use standard inference procedures, we must impose assumptions of homoskedasticity and no serial correlation on the error process. For homoskedasticity, we assume that this holds in a contemporaneous fashion: $Var(u_t|\mathbf{x}_t) = \sigma^2$. For no serial correlation, we assume that for all $t \neq s$, $E(u_t u_s|\mathbf{x}_t \mathbf{x}_s) = 0$.

Just as in the cross section case, under the assumptions we have made (including contemporaneous homoskedasticity and serial independence), the OLS estimators are asymptotically normally distributed. This implies that the usual OLS standard errors, t statistics and F statistics are asymptotically valid.

We can apply these techniques to test a version of the *efficient markets hypothesis*: in the strict form, if markets are efficient, past information on stock returns should not be informative about current returns: the market is not forecastable. In the terminology we have used here,

$$E(y_t|y_{t-1}, y_{t-2}, \ldots) = E(y_t) \tag{9}$$

Under the EMH, the best predictor of market returns (at, say, weekly frequency) will be the mean of market returns over the sample. Under the alternative hypothesis that the market is forecastable (i.e. that technical analysis has some value), an AR model should have some explanatory power. In the simplest case, we fit an AR(1) model to weekly NYSE returns data, and find that the prior week's return is not a significant predictor of current returns. Even if we consider an expanded model: the AR(2),

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + u_t \qquad (10)$$

the coefficients on the two lagged terms are individually and jointly insignificant. Thus, for the NYSE as a whole, it would appear that there is no clear evidence against the EMH. We might find different results if we focused on a specific company's stock, but for the entire market efficiency appears to reign.

As another example, Wooldridge demonstrates how we may estimate an *expectations-augmented Phillips curve* (EAPC):

$$\dot{p}_t - \dot{p}_t^e = \beta_1 (U_t - \mu_0) + e_t \qquad (11)$$

where $\dot{p}$ represents inflation, $\dot{p}^e$ refers to expected inflation, U is the unemployment rate and parameter μ_o is a constant representing the natural rate of unemployment.

We must specify a model of expectations formation. An *adaptive expectations* model specifies that expectations today depend on recently observed values of the variable. Under the naïve model that $\dot{p}_t^e = \dot{p}_{t-1}$, we can write

$$\Delta \dot{p}_t = \beta_0 + \beta_1 U_t + e_t \tag{12}$$

where $\beta_0 = -\beta_1 \mu_0$. The EAPC relates the change in inflation to the level of unemployment and a supply shock, e_t. If supply shocks are contemporaneously uncorrelated with the unemployment rate, we can consistently estimate this equation with OLS. The coefficient on U from annual data is -0.543 with a standard error of 0.230, suggesting that there is a statistically significant relationship between unemployment and changes in inflation. We may use `nlcom` to estimate the natural rate of unemployment:

```
. nlcom _b[_cons]/-_b[unem]
```

which leads to a point estimate of 5.585 with a standard error of 0.657, leading to a 95% confidence interval of (4.262, 6.908).

Using highly persistent time series

We have discussed how OLS estimation techniques may be employed when we have time series exhibiting weak dependence. Unfortunately, many economic and financial time series do not possess that property, but are rather characterized as *highly persistent* or *strongly dependent*. We consider now how those series may be analyzed in econometric models.

In the simple AR(1) model

$$y_t = \rho_1 y_{t-1} + e_t \tag{13}$$

the assumption that $|\rho_1| < 1$ is crucial in determining weak dependence. Many time series are better characterized by an AR(1) model with $\rho_1 = 1$: the so-called *random walk* model

$$\begin{aligned} y_t &= y_{t-1} + e_t \\ \Delta y_t &= e_t \end{aligned} \tag{14}$$

We assume that e_t is $i.i.d., \sim (0, \sigma_e^2)$, and independently distributed of the initial value y_0. In this model, the *differences* of y_t will be independent of each other (and thus weakly dependent as well), but the *levels* of y_t are not. Since by back-substitution we may write

$$y_t = \sum_{j=1}^{t} e_j + y_0 \tag{15}$$

the current level of y_t may be written as its initial condition plus the unweighted sum of all errors in the interim: so that $Ey_t = Ey_0$. We often assume that $y_0 = 0$, so that the expectation of y_t is also zero. Thus, the mean of y_t is not time-varying if y follows a random walk.

The variance does change over time, though:

$$Var(y_t) = \sum_{j=0}^{t} Var(e_{t-j} = \sigma_e^2 t \tag{16}$$

showing that the variance of y increases linearly with time. If we assume that the process began long ago, the variance tends toward infinity. The process cannot be stationary or covariance stationary as a consequence.

Furthermore, the random walk process displays *highly persistent* behavior in the sense that the *memory* of the process is not only infinite—as it is for any stable AR process—but that values of the process long ago have as much importance for today's value as more recent values. In a stable AR process, the effect of past values is damped so that an innovation long ago no longer has a meaningful effect on today's value of the process. In the random walk process, that is not so: today's value fully incorporates every past innovation to the process.

Assuming $y_0 = 0$, this implies that

$$E(y_{t+h}|y_t) = y_t, \ \forall h \geq 1 \tag{17}$$

so that our best forecast of all future values of the process is today's realized value. In the stable AR case, this optimal forecast becomes

$$E(y_{t+h}|y_t) = \rho_1^h y_t, \ \forall h \geq 1 \tag{18}$$

which approaches the unconditional expected value, $Ey_t = y_0 = 0$ as h increases.

The random walk model also implies that the correlations of the process are very large when y_t follows a random walk:

$$Corr(y_t, y_{t+h}) = \sqrt{t/(t+h)} \tag{19}$$

This covariance changes over time with t, so that the process cannot be covariance stationary. For fixed t, the correlation tends to zero as $h \to \infty$, but it does so very slowly.

Despite the peculiar properties of the random walk process, it is not so evident that an observed process actually is a random walk. An AR(1) process with a very large ρ_1 (e.g. 0.99) will mimic a random walk process quite closely, even though it is a stationary process with finite variance.

We often speak of random walk processes as possessing a *unit root*, since using the lag operator L we may write equation (17) as

$$(1 - L)y_t = e_t \tag{20}$$

or in more general terms as

$$(1 - \lambda L)y_t = e_t \tag{21}$$

where we are searching for a root of the polynomial

$$(1 - \lambda x) = 0, \tag{22}$$

implying a root of $1/\lambda$. For stability of the AR(1) model, $1/\lambda$ must lie outside the unit circle, which requires that λ itself lie inside the unit circle. If λ lies on the unit circle, we have a *unit root* in the equation.

In economic terms, it is very important to know whether a particular time series is highly persistent (unit root process) or not. For instance, one implication of the random walk process is that the history of the process has no relevance for its future course. If stock returns follow a random walk, for example, then technical analysis—examining patterns in recent movements of the stock price—is absolutely worthless. At the macro level, if GDP is weakly dependent, then shocks will have a time-limited effect on its future values. If GDP follows a random walk, on the other hand, the effects of a particular shock will be long-lasting.

Part of the difficulty in distinguishing between random walk processes and weakly dependent processes is that the former process may be the *random walk with drift*:

$$
\begin{aligned}
y_t &= a_0 + y_{t-1} + e_t \\
\Delta y_t &= a_0 + e_t
\end{aligned} \tag{23}
$$

so that the changes in y_t will be on average a_0 units. This is known as a *stochastic trend* process, in that the level of the process will be increasing by a_0 units on average, but varying stochastically around that increment. In contrast, the *deterministic trend* process

$$
y_t = \beta_0 + \beta_1 t + \rho_1 y_{t-1} + e_t \tag{24}
$$

will change by a fixed β_1 units per period.

As we will discuss, if we misclassify an observed process as a deterministic trend when it is really a stochastic trend (random walk) process, we run the risk of running a *spurious regression*. If we are fairly certain that a process is a random walk process (often termed an

$I(1)$ process, for "integrated of order one") we may transform it to a weakly dependent process by taking its first differences (as above). The first difference of such an $I(1)$ process will be stationary, or $I(0)$: integrated of order zero. The differencing transformation will also remove any linear trend; if applied to Equation (24), for instance, we will have the equation

$$\Delta y_t = \beta_1 + \rho_1 \Delta y_{t-1} + \Delta e_t \qquad (25)$$

Notice, however, that this transformation induces a correlation between regressor and error by including Δy_{t-1}: a term that should not be present if the y_t process was a random walk.

The problem of classification of an observed series as a deterministic trend versus a stochastic trend is complicated by the difficulty of testing that $\rho_1 = 1$: a topic which we will take up in the next section of the course. For now, we simply note that there may be considerable risk in working with models expressed in levels (or log-levels) of economic and financial series that may indeed be highly persistent (or unit root) processes.

Dynamically complete models

If we have a dynamic model with a certain number of lags on both the response variable and the regressor:

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 y_{t-1} + \beta_3 z_{t-1} + e_t \quad (26)$$

as the data generating process, then if we assume that

$$E(u_t | z_t, y_{t-1}, z_{t-1}, y_{t-2}, z_{t-2}, \ldots) = 0 \quad (27)$$

this implies that

$$E(y_t | z_t, y_{t-1}, z_{t-1}, y_{t-2}, z_{t-2}, \ldots) = E(y_t | z_t, y_{t-1}, z_{t-1}) \quad (28)$$

or that once we have controlled for current and past z and the first lag of y, no further lags of either z or y affect current y. Once we have included sufficient lags of the response variable and regressors so that further lags do not matter, we say that the model is *dynamically complete*. Since the assumption of Equation (27) implies that

$$E(u_t|z_t, u_{t-1}, z_{t-1}, u_{t-2}, z_{t-2}, \ldots) = 0 \qquad (29)$$

it follows that the error process will be serially uncorrelated. On the other hand, if a model is not dynamically complete, then there are omitted dynamics which will cause bias and inconsistency in the coefficients of the included regressors as well as serial correlation in the errors, which will include the net effect of the omitted variables.

Chapter 12.6: ARCH models

Heteroskedasticity can occur in time series models, just as it may in a cross-sectional context. It has the same consequences: the OLS point estimates are unbiased and consistent, but their standard errors will be inconsistent, as will hypothesis test statistics and confidence intervals. We may prevent that loss of consistency by using heteroskedasticity-robust standard errors. The "Newey–West" or HAC standard errors available from `newey` in the OLS context or `ivreg2` in the instrumental variables context will be robust to arbitrary heteroskedasticity in the error process as well as serial correlation.

The most common model of heteroskedasticity employed in the time series context is that of *autoregressive conditional heteroskedasticity*, or ARCH. As proposed by Nobel laureate Robert Engle in 1982, an ARCH model starts from the premise that we have a static regression model

$$y_t = \beta_0 + \beta_1 z_t + u_t \tag{30}$$

and all of the Gauss–Markov assumptions hold, so that the OLS estimators are BLUE. This implies that $Var(u_t|Z)$ is constant. But even when this unconditional variance of u_t is constant, we may have time variation in the *condiitional variance* of u_t:

$$E(u_t^2|u_{t-1}, u_{t-2}, \ldots) = E(u_t^2|u_{t-1}) = \alpha_0 + \alpha_1 u_{t-1}^2 \tag{31}$$

so that the conditional variance of u_t is a linear function of the squared value of its predecessor. If the original u_t process is serially uncorrelated, the variance conditioned on a single lag is identical to that conditioned on the entire history of the series. We can rewrite this as

$$h_t = \alpha_0 + \alpha_1 u_{t-1}^2 \tag{32}$$

where $u_t = \sqrt{h_t}\ v_t,\ v_t \sim (0,1)$. This formulation represents the ARCH(1) model, in which a single lagged u^2 enters the ARCH equation. A higher-order ARCH equation would include additional lags of u^2. To ensure a positive variance, $\alpha_0 > 0$ and $\alpha_1 > 0$. When $\alpha_1 > 0$, the squared errors are positively serially correlated even though the u_t themselves are not.

Since we could estimate Equation (30) and derive OLS b which are BLUE, why should we be concerned about ARCH? First, we could derive consistent estimates of b which are asymptotically more efficient than the OLS estimates, since the ARCH structure is no longer a linear model. Second, the dynamics of the conditional variance are important in many contexts:

particularly financial models, in which movements in volatility are themselves important. Many researchers have found "ARCH effects" in higher-frequency financial data, and to the extent to which they are present, we may want to take advantage of them. We may test for the existence of ARCH effects in the residuals of a time series regression by using the command `estat archlm`. The null hypothesis is that of no ARCH effects; a rejection of the null implies the existence of significant ARCH effects, or persistence in the squared errors.

The ARCH model is inherently nonlinear. If we assume that the u_t are distributed Normally, we may use a maximum likelihood procedure such as that implemented in Stata's `arch` command to jointly estimate Equations (30) and (33).

The ARCH model has been extended to a generalized form which has proven to be much more appropriate in many contexts. In the simplest example, we may write

$$h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \gamma_1 h_{t-1} \qquad (33)$$

which is known as the GARCH(1,1) model since it involves a single lag of both the ARCH term and the conditional variance term. We must impose the additional constraint that $\gamma_1 > 0$ to ensure a positive variance.

We may also have a so-called ARCH-in-mean model, in which the h_t term itself enters the regression equation. This sort of model would be relevant if we had a theory that suggests that the level of a variable might depend on its variance, which may be very plausible in financial markets contexts or in terms of, say, inflation, where we often presume that the level of inflation may be linked to inflation volatility. In such instances we may want to specify a ARCH- or GARCH-in-mean model and consider interactions of this sort in the conditional mean (level) equation.

Alternative GARCH specifications

A huge literature on alternative $GARCH$ specifications exists; many of these models are preprogrammed in Stata's `arch` command, and references for their analytical derivation are given in the Stata manual. One of particular interest is Nelson's (1991) exponential $GARCH$, or $EGARCH$. He proposed:

$$\log h_t = \eta + \sum_{j=1}^{\infty} \lambda_j \left(\left| \epsilon_{t-j} \right| - E \left| \epsilon_{t-j} \right| + \theta \epsilon_{t-j} \right)$$

which is then parameterized as a rational lag of two finite–order polynomials, just as in Bollerslev's $GARCH$. Advantages of the $EGARCH$ specification include the positive nature of h_t irregardless of the estimated parameters, and the asymmetric nature of the impact of innovations: with $\theta \neq 0$, a positive shock will have a different effect on volatility than will a negative shock, mirroring findings in equity market research about the impact of "bad news" and "good news" on market volatility. For instance, a simple $EGARCH(1,1)$ model will provide a variance equation such as

$$\log h_t = -\delta_0 + \delta_1 z_{t-1} + \delta_2 \left| z_{t-1} - \sqrt{2/\pi} \right| + \delta_3 \log h_{t-1}$$

where $z_t = \epsilon_t / \sigma_t$, which is distributed as $N(0,1)$.

Nelson's model is only one of several extensions of $GARCH$ that allow for asymmetry, or consider nonlinearities in the process generating the conditional variance: for instance, the threshold $ARCH$ model of Zakoian (1990) and the Glosten et al. model (1993).