More Time Series Analysis

Econometric Methods, ECON 370

We will now examine the circumstances which permit us to use OLS, that is under which situations would our Gauss-Markov Assumptions for our Time Series Data be reasonable.

1 Stationary & Weakly Dependent Time Series

A stationary process as we had noted prior is one where the probability distributions are stable over time, i.e. the joint distribution from which we draw a set of random variables in any set of time periods remains unchanged. Formally, a stochastic process $\{x_t : t = 1, 2, ...\}$ is stationary if for a set of time indices, $1 \leq t_1 < t_2 < ...$, the joint distribution of a draw $\{x_{t_1}, x_{t_2}, ...\}$ is the same as that $\{x_{t_1+h}, x_{t_2+h}, ...\}$ for $h \geq 1$, or in words, the sequence is **identically distributed**. A process that is not stationart is said to be a **Nonstationary Process**. In general it is difficult to tell whether a process is stationary given the definition above, but we do know that seasonal and trending data are not stationary.

Essentially, what we are trying to do with these assumptions is to justify the use of OLS. There is another problem as you would recall regarding the correlation of independent variables across time, a problem which does not occur in cross sectional analysis since it is often difficult to think of individuals being correlated with one another especially when they come from different families. We need assumptions then to allow us to use OLS, of course the violations of which means that we won't be able to use OLS. In time series analysis, the concept of weak dependence relates to how strongly related x_t, and x_{t+h} is related to each other. The assumption of weak dependence in a sense constrains the degree of dependence, and says that the correlation between the same independent variable Nonetheless, sometimes we can rely on a weaker form of stationarity call **Covariance Stationarity**. A stochastic process is covariance stationary if,

1. $E(x_t) = c$, where c is a constant.
2. $var(x_t) = k$, where k is a constant.
3. $cov(x_t, x_{t+h}) = p_h$, where t, $h \geq 1$ and p_h depends on h, and not t.

Covariance Stationarity focuses only on the first two moments of the stochastic process, the mean and variance. Note that point number 3 above implies that the correlation between x_t, and x_{t+h} is nonzero and depends on h. When a stationary process has a finite second moment, as in the above, it must be covariance stationary, however the converse is not true. We generally refer to this also as weak stationarity, and the original discussion of stationarity as strict stationarity. But this in and of itself is not sufficient for us to use OLS. We need it to be **Weakly Dependent** which occurs when the correlation between x_t, and x_{t+h} tends towards zero sufficiently quickly as $h \to \infty$. What we need formally is that $corr(x_t, x_{t+h}) \to 0$ as $h \to 0$, and we say that the stochastic process is asymptotically uncorrelated. The significance of this assumption is that it replaces the assumption of random sampling through the use of **The Law of Large Numbers** and **Central Limit Theorem**.

We will examine two simple commonly cited weakly dependent time series which had glossed over rather quickly earlier in our introduction to time series analysis.

1. **Moving Average Process of Order One,** $MA(1)$:

$$x_t = e_t + \alpha e_{t-1}$$

for $t = 1, 2, ...,$ and where e_t is a i.i.d. sequence with a zero mean and variance σ_e^2. Note that we can in general describe a moving average process for any variable, be it the dependent on the independent variable, or even the errors. We will use the above form for the moment, i.e. for the independent variable. This discussion is more general than the brief introduction prior.

When a random variable follows the above process, we describe it as "x_t follows a moving average process of order one". Basically, the process is a weighted average of e_t, and e_{t-1}. This is an example of a weakly dependent process because of the following reasons,

$$var(x_t) = cov(e_t + \alpha e_{t-1}, e_t + \alpha e_{t-1}) = (1 + \alpha^2)\sigma_e^2$$

and

$$cov(x_t, x_{t+1}) = cov(e_t + \alpha e_{t-1}, e_{t+1} + \alpha e_t) = \alpha var(e_t) = \alpha \sigma_e^2$$

therefore,

$$corr(x_t, x_{t+1}) = \frac{cov(x_t, x_{t+1})}{\sqrt{var(x_t)}\sqrt{var(x_{t+1})}} = \frac{\alpha}{1 + \alpha^2}$$

Further, since

$$cov(x_t, x_{t+2}) = cov(e_t + \alpha e_{t-1}, e_{t+2} + \alpha e_{t+1}) = 0 = cov(x_t, x_{t+h})$$

for $h \geq 2$,

$$\Rightarrow corr(x_t, x_{t+h}) = 0$$

Therefore since e_t is i.i.d, $\{x_t\}$ is a stationary, weakly dependent process, and the law of large numbers, and central limit theorem applies. **How about an $MA(2)$ process or higher, for a relationship for a independent variable such as the above?**

2. **Autoregressive Process of Order One,** $AR(1)$: As you should recall, since it was just examined,

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

where for $t = 1, 2, ...,$ and it is referred to as a **autoregressive process of order one, AR(1)**. We usually assume that e_t is i.i.d. as before, and in addition that it is independent of y_0 and y_t, and that $E(y_0) = 0$. In addition, you will recall is the assumption that $|\rho| < 1$. Only when the latter assumption is true, can we say that the process is stable.

We assume that the process is covariance stationary, which in turn implies that $E(y_t) = E(y_{t-1})$, but since $\rho \neq 1$, this can happen if and only if $E(y_t) = 0$. Since e_t and y_{t-1} are uncorrelated,

$$var(y_t) = \rho^2 var(y_{t-1}) + var(e_t)$$
$$\Rightarrow \sigma_y^2 = \rho^2 \sigma_y^2 + \sigma_e^2$$
$$\Rightarrow s_y^2 = \frac{\sigma_e^2}{1 - \rho^2}$$

Next, note that

$$y_{t+h} = \rho y_{t+h-1} + e_{t+h} = \rho^2 y_{t+h-2} + \rho e_{t+h-1} + e_{t+h}$$
$$\Rightarrow y_{t+h} = \rho^h y_t + \sum_{i=0}^{h} \rho^i e_{t+h-i}$$

Therefore,

$$cov(y_t, y_{t+h}) = cov(y_t, \rho^h y_t + \sum_{i=0}^{h} \rho^i e_{t+h-i}) = \rho^h \sigma_y^2$$

which means that,

$$corr(y_t, y_{t+h}) = \frac{\rho^h \sigma_y^2}{\sigma_y^2} = \rho^h$$

which in turn implies that,

$$\lim_{h \to \infty} \rho^h = 0$$

for $|\rho| < 1$, which thus implies that y_t is weakly dependent.

2 Asymptotic Properties of OLS

We can now justify OLS in more general terms.

1. **Linearity and Weak Dependence**: Just as the linearity assumption in parameters, the structure of the general model remains the same. However, we add the assumption that, $\{\mathbf{x}_t, y_t\}$ is stationary and weakly dependent so that the law of large numbers and central limit theorems can be applied. The significance of adding this additional assumption of weak dependence allows us then to use lags of both dependent and independent variables besides the contemporaneous ones.

2. **No Perfect Collinearity**

3. **Zero Conditional Mean**: Which implies that $\mathbf{x}_t$ is contemporaneously exogenous. $E(e_t|\mathbf{x}_t) = 0$. By stationarity, if contemporaneous exogeneity holds for one time period, it holds for all periods.

   Like before, the first three assumptions here yield consistent OLS estimators, that is $p\lim \widehat{\beta}_j = \beta_j$. Note that unlike in the original discussion before, the estimator is just consistent. That is it may be biased.

4. **Homoskedasticity**: Errors are contemporaneously homoskedastic, $var(e_t|\mathbf{x}_t) = \sigma^2$. Note the slight difference where previously, the condition was stronger since $var(e_t|\mathbf{X}) = \sigma^2$.

5. **No Serial Correlation**: As above this condition is weaker than before, $E(e_t e_s|\mathbf{x}_t, \mathbf{x}_s) = 0$

   Then as long as the above five assumptions hold, the OLS estimators are asymptotically normally distributed, and the usual OLS standard errors, t statistics, F statistics, and LM statistics are asymptotically valid.

We will now examine some examples to see how these assumptions work.

- **Statis Model**: Consider,

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + e_t$$

  Under weak dependence,

$$E(e_t|x_{t,1}, x_{t,2}) = 0$$

Recall and keep in mind that if the model is misspecified, or if independent variables contain measurement etc, the assumptions that justify the use of OLS fails. Note the generality of the assumptions actually allow feedback between y_{t-1}, and $x_{t,1}$ and $x_{t,1}$. Consider the idea that we are examining interest rates, $x_{t,1}$ effect on economic activity, say, a stock exchange index in any period. However, it is not incorrect to believe that the stock exchange index in the previous period, y_{t-1}, having an effect on the interest rate, which is dependent say on the central bank (think of the interest rate as some average in an economy), i.e.

$$x_{t,1} = \alpha_0 + \alpha_1 y_{t-1} + \nu_t$$

This implies then that,

$$cov(e_{t-1}, x_{t,1}) = cov(e_{t-1}, \alpha_0 + \alpha_1 y_{t-1} + \nu_t)$$

$$= cov(e_{t-1}, \alpha_0 + \alpha_1(\beta_0 + \beta_1 x_{t-1,1} + \beta_2 x_{t-1,2} + e_{t-1}) + \nu_t)$$

$$= \alpha_1 var(e_{t-1}) = \alpha_1 \sigma_e^2 \neq 0$$

In the original assumptions, this kind of feedback wouldn't have been permitted.

- **Finite Distributed Lag Model**: Consider the following model,

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + e_t$$

Here the natural assumption that would allow us to use OLS is,

$$E(e_t | x_t, x_{t-1}, x_{t-2}, x_{t-3}, ...) = 0$$

which means that once we have controlled for x_t, x_{t-1}, and x_{t-2}, no further lags of x will affect $E(y_t|x_t, , x_{t-1}, x_{t-2}, x_{t-3}, ...)$. If not, all we need to do is to include more lags of the dependent variable, noting that as we include more lags, we are reducing the number of observations. To relate to assumption 3, think of $\widetilde{x}_t = \{x_t, , x_{t-1}, x_{t-2}\}$. As before, this assumption does not rule out feedback as described before.

- $AR(1)$ **Model**: Consider,

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

and we need to assume,

$$E(e_t | y_{t-1}, y_{t-2}, ...) = 0$$

Combining them,

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

Note that this says that as long as you have lagged variables, the standard strict exogeneity assumption does not hold, and consequently we can't use OLS by those assumptions. To see, note again that strict exogeneity requires that all dependent variables must be uncorrelated with the error term. Yet,

$$cov(y_t, e_t) = cov(\beta_0 + \beta_1 y_{t-1} + e_t, e_t) = \sigma_e^2 > 0$$

Note further that we also require that $|\beta_1| < 1$ for us to use OLS. However, the estimator is biased, and this bias is large for small sample sizes, but should be a good estimator under moderate to large samples.

Note further that the errors are not serially correlated. To see,

$$E(e_t, e_s|y_{t-1}, y_{s-1}) = E(e_t, y_s - \beta_0 - \beta_1 y_{s-1}|y_{t-1}, y_{s-1})$$

But by weak exogeneity,

$$E(e_t|e_s, y_{t-1}, y_{s-1}) = 0$$

$$\Rightarrow E(e_t e_s|e_s, y_{t-1}, y_{s-1}) = e_s E(e_t|y_{t-1}, y_{s-1}) = 0$$

$$E(e_s E(e_t|y_{t-1}, y_{s-1})) = E(e_s e_t|y_{t-1}, y_{s-1}) = 0$$

3 Using Highly Persistent Time Series

Although the previous section says that under weak assumptions, OLS is still valid. However this needn't and typically isn't true in Time Series Data. It is very typical that time series are not weakly dependent, but exhibit strong dependence, or **High Persistence**. You will recall that we can transform these data still so that we may still use OLS.

3.1 Highly Persistent Time Series

From the $AR(1)$ model,

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

we have learnt that for weak dependence to hold, $|\beta_1| < 1$. However, it turns out that time series data is typically characterized by $\beta_1 = 1$, or

$$y_t = \beta_0 + y_{t-1} + e_t \tag{2}$$

where assume that the error term is i.i.d with mean 0, and variance σ_e^2, which is nothing but the random walk with drift model. Let's for now focus on a random walk model instead.

$$y_t = y_{t-1} + e_t \tag{3}$$

Let the initial value be y_0, then given the above we know it can also be written as,

$$y_t = e_t + e_{t-1} + e_{t-2} + ... + e_1 + y_0$$

And taking expectations, we get

$$E(y_t) = E(e_t) + E(e_{t-1}) + E(e_{t-2}) + ... + E(e_1) + E(y_0) = E(y_0)$$

which means that the mean is not dependent on time, i.e. is time invariant. But we also the variance is not time invariant.

$$var(y_t) = var(e_t) + var(e_{t-1}) + var(e_{t-2}) + ... + var(e_1) + var(y_0) = t\sigma_e^2$$

It is clear also that random walk exhibits persistent behavior. Consider y_{t+h}, it is easy to see that y_t would have an effect on its value, that is,

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

In other words, what we see today is the best predictor of what we see tomorrow. Another way to see this is,

$$corr(y_t, y_{t+h}) = \frac{cov(y_t, y_{t+h})}{\sqrt{t\sigma_e^2(t+h)\sigma_e^2}} = \sqrt{\frac{t}{t+h}} \tag{4}$$

That is the correlation depends on the starting point in time, t. Although it may be said that for a given t, the correlation falls as $h \to \infty$ the rate is slow, and the rate is slower the higher t is. Thus a random walk does not satisfy the requirement of an asymptotically uncorrelated sequence. See figure 11.2 for an example of a time series that is typically thought to follow a random walk, the three month T-bill rate.

As we had noted in the quick introduction, the random walk model is a special example of a **unit walk process** where the process is similarly expressed as

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

The key difference is that the sequence of error terms, $\{\epsilon_t\}$ is allowed to be weakly dependent, i.e. as was mentioned, $\{\epsilon_t\}$ can follow a $MA(1)$ or a stable $AR(1)$ process. (What do we mean by a stable $AR(1)$ process?) Note that once the error term is not i.i.d. the properties of a random walk noted bove does not hold. Can you see?

Suppose the error terms follow a $MA(1)$ process, say $\epsilon_t = \phi u_{t-1} + u_t$, where suppose $\{u_t\}$ is i.i.d with mean 0, and variance σ_u^2, then

$$
\begin{aligned}
y_t = y_{t-1} + \epsilon_t &= y_{t-1} + \phi u_{t-1} + u_t \\
&= \sum_{i=1}^{t} \epsilon_i \\
&= \sum_{i=1}^{t} (\phi u_{i-1} + u_i) \\
\Rightarrow var(y_t) &= \sum_{i=1}^{t} ((\phi^2 + 1)\sigma_u^2) = (\phi^2 + 1) t \sigma_u^2
\end{aligned}
$$

Because the values in each period has such persistence in a random walk model, if indeed it were true, it will have a substantial effect on policies. Imagine the following, if the Bank of Canada knew that their choice of interest rates had persistent effect on the economy beyond a decade, would you imagine that they would exercise caution. Further, you should not mistake trending with the idea of persistence in time series, since the former says that the time series is inevitably rising or falling with **time**, and not with previous values or realizations of the dependent variable. On the other hand, persistence says that the effect of one realization of the variable under examination has a long run effect on the same variable, i.e. into the future. And you should recall that to account for trend, we can always use the **random walk with drift model**. But the random walk with drift model merely adds a constant term, how does it acts as a trending term that we typically use. To see this

$$
\begin{aligned}
y_t = \alpha + y_{t-1} + \epsilon_t &= \sum_{i=1}^{t} (\alpha + \epsilon_i) + y_0 \\
&= t\alpha + \sum_{i=1}^{t} \epsilon_i + y_0 \\
\Rightarrow E(y_t) &= t\alpha
\end{aligned}
$$

The last equality says that the expected value y_t will increase (decrease) with time if α is monotonically increasing (decreasing).

3.2 Transformations on Highly Persistent Time Series

As we have noted, when the time series sequence is persistent, our assumptions for OLS are misleading. And you might recall that we can transform a unit root process by differencing so that we make the model weakly dependent, and thereby allowing us to use OLS.

Weakly Dependent processes are said to be **integrated of order zero**, or $I(0)$. Such series can be used without transformation since as we have found with the weaker OLS assumptions, we can use the standard OLS estimators. However, processes such as the random walk, or the random walk with drift needs to be first differenced before regression analysis can be performed. The two types of series are said to be **integrated of order one**, or $I(0)$. After first differencing, the series then becomes weakly dependent and often stationary as well. Assume that the error of the random walk with drift model is i.i.d. Then it is easy to see that

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

Typically, time series data which are strictly positive are such that $\log(y_t)$ is $I(1)$. In which case first differencing gives,

$$\Delta \log(y_t) = \log(y_t) - \log(y_{t-1}) \approx \frac{y_t - y_{t-1}}{y_{t-1}}$$

That as we can either use the differenced log dependent variable, or proportionate or percentage change of dependent variable as the dependent variable in the regression. Another useful outcome in differencing integrated time series is that upon differencing, we remove time trends as we had noted before. You can see this by taking expectations of

$$E(\Delta y_t) = \alpha$$

which is not dependent on time as required (assuming ϵ_t has a mean of zero and variance σ_ϵ^2. In other words, this also say that instead of including trend variables, we can always just first difference (or choose the differencing order depending on the type of trend modelled.

The next question you should have is how can we decide whether the time series you see is a $I(0)$ or a $I(1)$ process? The test used for examining unit root is call the **Dickey Fuller Test** which will not be covered in this course, which you can find out more about from chapter 18 of your test. Meanwhile, a possible to gauge is simply to perform a regression for a $AR(1)$. That is perform

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

Then we have a unit root, ρ should be very close to 1. However, we can use the law of large numbers only in the situation where $|\rho| < 1$, and even when the latter is true, you would recall that the estimator for ρ is only consistent, but is not unbiased. However, you can imagine that if ρ were really a unit root process, the sampling distribution would be quite different, rendering the estimates of ρ very imprecise. There are no hard and fast rules, but some economist would perform first differencing if $\widehat{\rho}$ is greater than 0.9, while others would do so when it is greater than 0.8.

Also, if you know the time series data has a trend, it is sensible to detrend the series before using the data, since trends would affect the bias of the estimate for ρ, or more accurately increases the likelihood of you finding a ρ near one, since trends creates a positive bias.