Introduction to Time Series Analysis

Econometric Methods, ECON 370

Our discussions thus far has been pertaining to the use of Cross Sectional Data, and consequently, it is typically termed as Cross Sectional Analysis. In such analysis, the observations occur at a specific time. What we can do with the data is essentially to examine if there are significant differences between the observations, and thereby arrive at a particular story we wish to tell. What element in time that we can glean is possible if and only if the observations has been alive for differing periods of time. For example, we could use individuals with different ages to say how time might have caused differential in the dependent variable we are examining. If our observations are countries, we could examine time from independence or some such delineation. Ultimately, your imagination and care will determine what can and cannot be done with the data.

We will now turn to a direct way of examining observations across time using time series data, where observations are arranged by the timing in which they occurred. We have at the onset of this course noted this key differential. However, there are important differences in the two data structures.

1 The Nature of Time Series Data

In cross sectional analysis, we have viewed each sample drawn from a population as a random sample since each draw from the population will yield a different set of dependent and independent variables. Consequently, the estimates generated will be different, and thus we consider OLS estimators as random variables.

In time series analysis, the outcome of any variable in the following period is not known in the current. This then also allows us to think of the variables as random variables. A sequence of random variables indexed by time is called a **stochastic process** or a **time series process**. We can see a particular realization at the time of the draw or observing the realization, but we cannot go back in time to restart the process to get another realization of the same process. Further, if the historical circumstance can be different, we are not able to examine how the outcome would have been, that is we cannot get a counterfactual realization. We think of the set of all possible realizations of a time series process as the population parallel in cross sectional analysis. The sample size is the number of time periods over which we observe a variable of interest. Each period could be seconds, minutes, hours, day, months, quarters, years, decades, ...

2 Time Series Regression Models: An Overview of What is Available

1. **Static Models**: When we believe the relationship between a dependent variable, y, and independent variables, x, occurs immediately in any period in time, i.e. such that there is no effect lags, and if we wish to examine the magnitude of the effect, we can structure the model as,

   $$y_t = \beta_0 + \beta_1 x_t + \epsilon_t$$

   where $t = 1, 2, ..., T$. We call such a model a Static Model. Note that

   $$\Delta y_t = \beta_1 \Delta x_t$$

   where $\Delta y_t = y_t - y_{t-1}$, and the same is true for x_t.It is also a useful model when we wish to examine the tradeoff between the two variables in question. A more concrete example would be the effect on homicide of an increase in police recruitment rate, and inward migration rate,

   $$homicide_t = \beta_0 + \beta_1 police_t + \beta_2 migration_t + \epsilon_t$$

   Note that such a model only examines the contemporaneous effect of the two independent variable. What would you do to confirm causality?

2. **Finite Distributed Lag Model**: This model is a modification of the Static Model, where we use one or more lagged variables (lagged in time) as independent variables. Using the homicide example,

   $$homicide_t = \beta_0 + \beta_1 police_t + \beta_2 police_{t-1} + \beta_3 migration_t + \beta_4 migration_{t-1} + \epsilon_t$$

   In this case, we can see how previous levels of the police force, and migration rates might affect current rate, thereby perhaps say which variable might be causing the change in the other. But does performing this single regression do the trick to solve problems of causality intuitively?

To see the *ceteris paribus* effect, i.e. the partial effect from such a regression, we will set the error term in each period to zero. Consider the general model above,

$$y_t = \beta + \alpha_0 x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \epsilon_t$$

If we substract the equation for the change in period t with that in period $t-1$,

$$\Delta y_t = \alpha_0 \Delta x_t + \alpha_1 \Delta x_{t-1} + \alpha_2 \Delta x_{t-2} + \Delta \epsilon_t$$

Suppose all changes in this economy or object of interest has short lived temporary effects. An example might be that y is an individuals consumption pattern, and x is the individuals income across time. We would tend to think of her consumption pattern as stable an constant as long as there are no changes in her income. If we perturb the model, and imagine this individual bought a winning lottery ticket in period t. Suppose the winning is small in relation to her **permanent income**, so that this winning is nothing but a transient increase in income. We can imagine that this increase would have only temporary effect. This increase in income in t is represented by Δx_t. It is then clear from the above that a_0 is the immediate change in y due to the unit change in x at time t, the lottary winning. We call α_0 the **Impact Propensity** or **Impact Multiplier**. We can similarly think of the lottery as being in period $t-1$, then the effect a_1 is the change in y two period after the initial change in period $t-2$, and a_2 we can interpret α_2 in a similar fashion.

Of course, there are life altering winnings, or imagine being promoted on account of superb work performance and receives a quadruple jump in rank. In such a scenario, there would be a increase in the individuals permanent income. Such that x is higher in each and every period there after. Then the total effect of a unit increase in the individuals income as the sum of all the partial effects of the unit increase in each period, i.e. a dollar in increase in permanent income in period $t-2$ and thereafter raises consumption in t by $\alpha_0 + \alpha_1 + \alpha_2$, which reflects the *long run* change in y from a permanent change in x, and we call this **Long Run Propensity** or **Long Run Multiplier**.

More generally, a finite distributed lag model of order q can be written as,

$$y_t = \beta + \alpha_0 x_t + \alpha_1 x_{t-1} + ... + \alpha_q x_{t-q} + \epsilon_t$$

And in this case, the long run propensity is.

$$\alpha_0 + \alpha_1 + ... + \alpha_q$$

You can imagine that all the a_j, $j \in \{0, 1, ...q\}$ would be highly correlated (think of the income example), and consequently we have learned from before that when the correlation between the independent variables are high, we obtain very imprecise estimates of the coefficient of interest, a_j. However, it turns out that we do get good estimates of the **Long Run Propensity**.

3. **Moving Average Models**: You would have heard of such models. What do they mean by moving average? Moving Average of what? In MA models we assume that there are random shocks (innovations) that affect a particular outcome. Think of a country's GDP, and the shocks as business cycle shocks. The simplest moving average model is $MA(1)$, or first order moving average model, where the random shocks has a maximum effect that last for one period. We write the model as,

$$y_t = \beta_0 + \beta_1 \epsilon_{t-1} + \epsilon_t$$

How would the effect of a single period shock look like? Suppose the shock occurs in period t, but at period prior and thereafter, there were no shocks, that is $\epsilon_{t-1} = \epsilon t + 1 = \epsilon_{t+2} = 0$, of course this equality could be stretched further into the future. This then gives us the following;

$$y_{t-1} = \beta_0$$

$$y_t = \beta_0 + \epsilon_t$$

$$y_{t+1} = \beta_0 + \beta_1 \epsilon_t$$

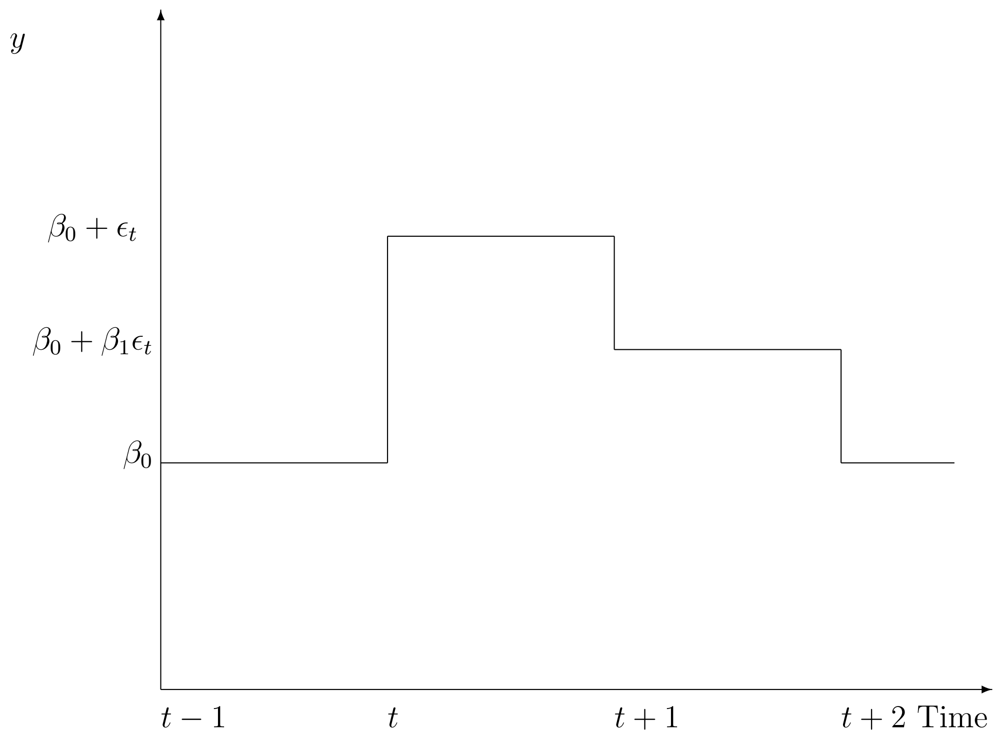
$$y_{t+2} = \beta_0$$

That is the effect of the shock despite occurring for only one period, has more than a one period effect.

More generally, a $MA(q)$ model is of the following form,

$$y_t = \beta_0 + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} + ... + \beta_q \epsilon_{t-q} + \epsilon_t$$

Figure 1: Effect of an MA(1)



4. **Autoregressive Models**: On the other hand, the Autoregressive Model says that current realization of the dependent variable today, is dependent on the previous periods' realization of the variable. The simplest case being the $AR(1)$ model,

$$y_t = \beta + \phi_1 y_{t-1} + \epsilon_t$$

Although the model seems to say that the realization has immediate effect, the model has a longer memory. To see this, suppose only $\epsilon_t \neq 0$ while all other random shock variables are zero. Further, for simplicity suppose without loss of generality that $\beta = 0$ and that $y^s = \phi_1 y_{t-1} = 0$, where y^s is a steady state equilibrium value of y. Then,

$$y_{t-1} = \phi_1 y_{t-2} + \epsilon_{t-1} = 0$$

$$y_t = \phi_1 y_{t-1} + \epsilon_t = \epsilon_t$$

$$y_{t+1} = \phi_1 y_t = \phi_1 \epsilon_t$$

$$y_{t+2} = \phi_1 y_{t+1} = \phi_1^2 \epsilon_t$$

and by induction,

$$y_{t+k} = \phi_1^k \epsilon_t$$

It should be obvious that as long as $\phi < 1$, the sequence of y will eventually converge back to the steady value. However, the rate is dependent on the value of ϕ_1, where the closer it is to 1, the slower is the rate of convergence.

5. **Autoregressive Moving Average (ARMA) Models**: The last two models are not mutually exclusive, since it is perfectly possible that the process under examination exhibits both. In the simplest case, we have a $ARMA(1,1)$, where we merge a $AR(1)$ with a $MA(1)$ process to get,

$$y_t = \alpha + \phi_1 y_{t-1} + \beta_1 \epsilon_{t-1} + \epsilon_t$$

and more generally, a $ARMA(q,p)$ can be written as,

$$y_t = \alpha + \phi_1 y_{t-1} + ... + \phi_q y_{t-q} + \beta_1 \epsilon_{t-1} + ... + \beta_p \epsilon_{t-p} + \epsilon_t$$

6. **Unit Roots & Random Walk**: In the previous discussion on $AR(1)$ model, we noted that as long as ϕ_1 is less than 1, the series would converge back to steady state. When $\phi_1 = 1$, we have the following,

$$y_t = \phi_1 y_{t-1} + \epsilon_t$$

and we call this model a **Random Walk Model**, or we say that the $AR(1)$ model has a **Unit Root**. This model exhibits the pattern that any shock to the series will have permanent effects. To see this, suppose as before that all period prior to t, there are no shocks, and that there are no shocks there after, and as before suppose that $y^s = 0$ is the steady state. Then the following is derived,

$$y_{t-1} = \phi_1 y_{t-2} + \epsilon_{t-1} = 0$$

$$y_t = \phi_1 y_{t-1} + \epsilon_t = \epsilon_t$$

$$y_{t+1} = \phi_1 y_t = \phi_1 \epsilon_t = \epsilon_t$$

$$y_{t+2} = \phi_1 y_{t+1} = \phi_1^2 \epsilon_t = \epsilon_t$$

and by induction,

$$y_{t+k} = \phi_1 y_{t+k-1} = \phi_1^k \epsilon_t = \epsilon_t$$

. This means then that once a shock occurs, the dependent variable gets pushed permanently away from its steady state, and consequently the importance of unit root to the idea of the evolution of the time series. Further, when ϕ_1 is greater than 1, we would have an exploding series.

The above example is of course very stylized, since it is perfectly possible that shocks can occur in each and every period. However, suppose the shocks have a mean of zero, then what we get is $E(y_t) = E(y_{t-1} + \epsilon_t) = y_{t-1}$, for all $t \in 1, 2, ..., T$. Which exhibits the idea of random walk since the movement from one period to the next is purely random, since the best predictor of where y would be in the next period, would be what had occurred in the current.

We could also include a trending element to allow for a constant growth pattern, such as the following,

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

which means that we wish to study the deviations from a general trend. This model says that ignoring the shocks, the value of today's measure is that of yesterday, plus a trend (which if positive is upward trending, while negative is downward trending). This model is known as a **Random Walk with Drift**. Note that as long as $\phi_1 = 1$ shocks will have permanent effects.

7. **Stationarity and the** $ARIMA$ **Models**: In cross sectional analysis, since observations are drawn at a particular time from a population we have noted it is natural to think of the sample as a random sample, and the variables as random. However, in time series, an important question is whether whether all the time series observations are drawn as random variables from the same distribution. If they are, we can then use OLS of before. Otherwise, our approach needs to be modified.

   When the observations are drawn from the same underlying distributions, we say that the time series process is **Stationary**. Technically, what we want is $f(X_t, X_{t+1}, ..., X_{t+k}) = f(X_{t+m}, X_{t+m+1}, ..., X_{t+m+k})$, for all m and k. This says that we want ideally to have the p.d.f. to be invariant to any displacement in time.

   A process is said to be *Weakly Stationary* if

   - $E(X_t) = \mu$, that is the mean is constant over time.
   - $var(X_t) = E(X_t - \mu)^2 = \sigma^2$, that is variance is constant over time.

- $cov(X_t, X_{t+k}) = E(X_t - \mu)(X_{t+k} - \mu) = \pi_k$, that is the covariance of X_t, and X_{t+k} is dependent on the distance between them in time, but not the particular period in time.

A stationary process will satisfy the weak stationarity conditions. A random walk process is not a stationary process.

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

$$\Rightarrow y_t = y_0 + \sum_{i=1}^{t} \epsilon_i$$

Let $y_0 = 0$ and $E(\epsilon_t) = \mu$, that is the shocks are stationary, then

$$E(y_t) = E(\sum_{i=1}^{t} \epsilon_i) = t\mu$$

Then only if $\mu = 0$ would the series, y_t seem to be stationary. However,

$$var(y_t) = \sum_{i=1}^{t} \epsilon_i = t\sigma^2$$

which means that the variance will never be constant. Consequently, the conditions for weak stationarity will never be satisfied. That is a random walk model with or without drift is never stationary. Note that we are not really done for since there are transformations that can alter the series into a stationary one. Consider a first difference, where we take the difference between the variables as the independent variable, $\Delta y_t = \epsilon_t$, since by assumption in the example, ϵ_t is stationary. We say that the variable, y_t is **difference stationary process**.

You can imagine that just as the random walk is not stationary, so to can an a general $ARMA$ model. However, it might be that after differencing, we can achieve stationarity. For the $ARMA$ model, when the process achieves stationarity after differencing d times, we say that the $ARMA$ model is an **Integrated Process of Order** d. Note that the stationarity of the series is dependent only on the autoregressive part of the model. We call such a model an $ARIMA(p, d, q)$.

This is a mere introduction. We will be more precise in the later sections abou the techniques required in Time Series Analysis.

3 OLS Assumptions for Time Series Analysis

In order to use the OLS estimators we need to modify some of the assumptions. The assumptions both new and old are restated below.

1. **Linearity in Parameter**: Let the stochastic process be $\mathbf{x}_t = \{x_{t,1}, x_{t,2}, ..., x_{t,k}, y_t\}$ where $t \in 1, 2, ..., T$. Then as usual, the linearity in parameters condition is just,

   $$y_t = \beta_0, \beta_1 x_{t,1} + ... + \beta_k x_{t,k} + \epsilon_t$$

   where ϵ_t is the sequence of error terms, and where T is the total number of observations.

2. **No Perfect Collinearity**: No independent variable is constant nor a perfect linear combination of the others. Recall that all this rules out only perfect correlation, but not imperfect correlation.

3. **Zero Conditional Mean**: $E(\epsilon_t|\mathbf{X}) = 0$ for $t \in \{1, 2, ..., T\}$. Note that this assumption works if we correctly specify the functional form of the regression equation. We can also write the condition as, $E(\epsilon_t|\mathbf{x}_t) = 0$. When the assumption holds, we say that each of the independent variables are **contemporaneously exogenous**, or that the independent variables are **strictly exogenous**. Note that the assumption says nothing about how the independent variables are correlated or how the errors are correlated across time. **It only says that the error terms are uncorrelated with all independent variables in all time periods**.

   This assumptions can fail for several reasons,

   - Omitted Variables.
   - Measurement Error.
   - The assumption fails when the error terms are correlated with the independent variables in other periods. When this is with lagged independent variables, the assumption of strict exogeneity is violated. In general we are not concerned with the correlation between the error term, and past realizations of the independent variable since the latter has been controlled in the model. But it is the feedback to the future that creates the problem. In the social sciences, many explanatory variables would likely violate the strict exogeneity assumption.

Under the first three assumptions, OLS estimators are unbiased conditional on X, and unconditionally as well. Note that the assumption that has been dropped is the random sampling assumption.

4. **Homoskedasticity**: $var(\epsilon_t|\mathbf{X} = var(\epsilon_t) = \sigma^2$, for $t \in \{1, 2, ..., T\}$.

5. **No Serial Correlation**: $corr(\epsilon_t, \epsilon_s|\mathbf{X}) = 0, \forall t \neq s$ or $corr(\epsilon_t, \epsilon_s) = 0$. When this assumption fails, we say that the errors suffer from **Serial Correlation** or **Autocorrelation**. It is very typical for time series analysis to suffer from serial correlation. Consider the stock prices in several periods. It is very likely that unobservables between periods relating to stock prices are driving the current period prices the same direction. A strong price for one period, is likely to keep the prices similarly high in the next period, quite apart from independent variables, perhaps due to investors' bullish sentiments which cannot be observed. Note that in cross sectional analysis, the problem with serial correlation does not show up since the sample is randomly drawn from the population.

The above five assumptions constitutes the Gauss-Markov assumptions for Time Series Analysis. The key outcomes from the assumptions are the following,

- Under the five assumptions, the variance of the parameter estimates are now,

$$var\left(\widehat{\beta}_j|\mathbf{X}\right) = \frac{\sigma^2}{TSS_j\left(1 - R_j^2\right)}$$

  for $j = 1, 2, ..., k$. Where all the variables are as in the cross sectional case.

- Under the five assumptions, the estimator $\sigma^2 = \frac{RSS}{df}$ is an unbiased estimator of σ^2, where $df = n - k - 1$.

- Under the five assumptions, the OLS estimators are B.L.U.E. conditional on $\mathbf{X}$

- **Normality**: If we assume in addition that the error terms, ϵ_t are independent of $\mathbf{X}$ and are independently and identically distributed as $N(0, \sigma^2)$, then all the assumptions in the aggregate implies that the OLS estimators are normally distributed, conditional on $\mathbf{X}$. Then under the null hypothesis, each t statistic has a t distribution, and each F statistic has a F distribution. This then allows us to perform our usual inferences regarding the estimates, and construct our confidence intervals.

4 Trends and Seasonality

Many economic variables would exhibit a growing or declining pattern over some periods just as civilizations rise and wane. Consequently, we must realize and account for the general trending pattern in order to make inferences regarding causality from the independent to the dependent variables. Why is that so? Often times two variables we wish to examine may have individual independent trends which may be positive or negatively trending for reasons other than each other. If we ignore the trend for each variable, we would risk attributing the seeming correlation to each other. This false attribution of a relationship is know as **Spurious Regression Problem**. Fortunately, this can be easily contained by accounting for trends.

4.1 Models including Trend Variable

The inclusion or the accounting of trends can be done in several ways,

1. **Linear Time Trend**: Here we account for a linear trend by creating a variable $t = \{1, 2, ..., T\}$, that is an integer that corresponds to each period, with the first being 1. We can then perform the following regression assuming there are no other factors that affect the dependent variable,

$$y_t = \alpha_0 + \alpha_1 t + \epsilon_t$$

   or more generally,

$$y_t = \alpha_0 + \alpha_1 x_{1t} + ... + \alpha_k x_{kt} + \phi t + e_t$$

2. **Exponential Trend**: It needn't always be true that trends are linear in nature, in fact most economic time series actually exhibit an exponential growth pattern. In which case, a more appropriate model is,

$$\log y_t = \alpha_0 + \alpha_1 t + \epsilon_t$$

3. **Quadratic Trend**: Another possible approximation is to use a quadratic time trend,

$$y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \epsilon_t$$

Although it seems possible to use other polynomials, we must be careful in being over zealous about controlling for trends (in fact the functional form we use should be informed by simply examining the time series plot of the variables) since the greater the order of the polynomial time trend, the more explanatory power the trend variables would draw away from the dependent variable we wish to examine. Essentially, you can track any time series by including larger and larger orders of the polynomial in trend! What we want to to capture the ″rough″pattern.

Including a time trend in a regression is equivalent to detrending (removal of the trending element in the original data set. It turn out that the estimators to the following regression equation,

$$y_t = \alpha_0 + \alpha_1 x_{1,t} + ... + \alpha_2 x_{k,t} + \phi t + \epsilon_t$$

which although can be directly obtained through OLS, can also be obtained in the following two step procedure (this procedure works even for the other often used method of modelling trends),

1. Regress each of the variables y_t, $x_{1,t}$,...,$x_{k,t}$ on a constant term (intercept) and the time trend variable t. Then save the residuals, and denote them as $\widehat{y}_t$, $\widehat{x}_{1,t}$,...,$\widehat{x}_{k,t}$. Then we can think of each of the residuals as detrended incarnation of the original variables.

2. Now perform the regression (OLS) without trend using the new variables.

$$\widehat{y}_t = \alpha_1 \widehat{x}_{1,t} + ... + \alpha_k \widehat{x}_{k,t} + e_t$$

   Note that we did not include the intercept. This is because we have modelled the intercept in the first stage.

Typically when performing time series analysis, the explanatory power of the time series regression model is a very good fit, particularly since most of the usual time series data are aggregated data, such as GDP of countries across time. This typically means that the goodness of fit measures are much higher. In fact this is especially true when the dependent variable itself follows a trend (either upward or downward). To understand this, recall,

$$R^2 = 1 - \frac{RSS}{TSS}$$

As long as we include the trend in the regresion, we are assured the RSS is unbiased. However, when the dependent variable is trending, TSS would be overestimated, i.e. is biased. Consequently, artificially inflating the goodness of fit.

Consider the general model with a linear trend (the method generalizes to other ways in which you may want to account for trend),

$$y_t = \alpha_0 + \alpha_1 x_{1,t} + ... + \alpha_k x_{k,t} + \phi t + \epsilon_t$$

To adjust for the problem, we have to first obtain the detrended dependent variable, $\widehat{y}_t$. Only then do we perform the regression,

$$\widehat{y}_t = \alpha_0 + \alpha_1 x_{1,t} + ... + \alpha_k x_{k,t} + \phi t + \epsilon_t$$

The consequent formula for the usual goodness of fit is,

$$R^2 = \left(1 - \frac{RSS}{\sum_{t=1}^{n} \widehat{y}_t} \right)$$

where RSS is the residual sum of squares in the full regression including the trending element(s), but TSS should be calculated as that from detrended dependent variable, $\widehat{y}_t$. In calculating $\overline{R}^2$, we just need to divide RSS by the degrees of freedom left after the original regression, which in the general case in $n - (k+1) - p$, where p is the number of trend parameters estimated, and we also have to divide TSS by $n - p$.

4.2 Seasonality

In time series, besides trending of the variables, we also have to be concerned with the changes with the season. For example, if we were to consider sales of winter coats by Canada Goose in North America, then we have to be cognizant of the fact that their sales would typically pick up during the winter months, quite apart from any dependent variables we might be more concerned with. This sought of behavior by variables is known as **seasonality** and occurs in all types of time series, including monthly, quarterly, weekly, daily and hourly. To see the reason for hourly variation, at which time of the day in the Toronto Stock Exchange would you think the greatest amount of trading would occur? For daily data, consider the local pub in this town, at which days of the week would you think the alcohol consumption be the highest? But this is not to say that all time series would exhibit seasonality effects. We have to be guided by our priors.

The adjustment for **seasonality** effects is so prevalent today that it is quite difficult to find data that are not seasonally adjusted. Nonetheless, should we encounter it, we can always include a set of **seasonal dummy or indicator variables**. Just as we did in examining detrended data, we could also think of the usual OLS regression we have done so far as running OLS on **deseasonalized** data. In that case, we can similarly calculate a more accurate goodness of fit measure using the deseasonalized dependent variable.