Introduction to Time series

1. What is the difference between randomness in cross-sectional/panel data and time series data? What does this mean for trying to recover causal effects?

Re: With cross-sectional data we can envision taking repeat samples from the population (e.g. take repeat samples of size 400 from the population of people in the US to calculate average income per capita). However, with time-series data it's harder to think about doing this (e.g., how would one draw different samples of the US GDP in each ear since 1950). The issue is that we only have one realized sequence of annual GDPs for the US. How do we think about randomness in this context? Because there is a natural time sequence, we call the sequence of GDPs a time series process (or stochastic process). Why? It's hard to think about the realized value in 2010 without know how GDPs evolved over the past decade.

Because in time series we often only have one observation per unit of time, there is difficulty in distinguishing the effects of time versus the effect of other factors changing over time. In particular, this would make recovering causal effects from a static [time series] model difficult.

2. If we are only following one unit over time (e.g. the US), can we run a regression with time series data? If so, is it OLS? What would that mean for our standard MLR1-6?

Re: Yes, we can run OLS regressions with time series data. The assumptions for unbiasedness of beta change; now we only require:

(i) Linearity in parameters
(ii) No perfect collinearity
(iii) Zero conditional mean assumption.

We exclude random sampling because we don't have a population to randomly sample. While there are fewer required assumptions, the zero conditional mean assumption is likely much more difficult to satisfy.

The assumptions for calculating unbiased standard errors for hypothesis testing also change. We require:

(i) Homoskedasticity
(ii) No serial correlation (see next question)

The assumptions for hypothesis testing remain basically the same; we still require that the errors are independent of X (the covariates) and independently and identically distributed Normal (0, sigma2). (See Ch. 10.3 of Wooldridge 4e).

3. What is serial correlation?

Re: Serial correlation is when the error from one period is correlated with the error from another period. One can test for this by looking at the residuals. There is serial correlation if there is some pattern to the residuals over time.

4. Can we use the same types of functional forms as we did with cross-sectional data? Which ones are most common?

Re: Of course. Time series model are pretty common in macroeconomics and in that setting it is common to take the natural logarithm of the dependent variable (e.g. ln(gdp)). If you regress a logged outcome variable on a linear time trend then the coefficient on *t* is an estimated growth rate.

5. What are some common types of time series models? (discuss 5 types)

Note: You could always add controls/other explanatory variables to these models.

Model with trends:

$$y_t = \beta_0 + \beta_1 t + u_t$$

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + u_t$$

Model with seasonality:

$$y_t = \beta_0 + \beta_1 I_t^{Spring} + \beta_1 I_t^{Summer} + \beta_1 I_t^{Fall} + u_t$$

Static model:

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

Finite distributed lag model:

$$y_t = \beta_0 + \beta_1 x_{t-1} + u_t$$

$$y_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-2} + u_t$$

Moving average models

Example using a four-year moving average of outcome and then a moving average of the variable *x*:

$$y_t = \beta_0 + \beta_1 \bar{y}_{t-1 \text{ to } t-4} + u_t$$

$$y_t = \beta_0 + \beta_1 \bar{x}_{t-1 \text{ to } t-4} + u_t$$

Autoregressive models

AR(1) or autoregressive model of order 1:

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

AR(2) or autoregressive model of order 2:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-1} + u_t$$

6. Can we combine time series methods to cross-sectional data?

Re: Yes. For example, what if I think there is cheating on the exam based on who you sit next too. Like time series data, there is a natural ordering to the data. This would introduce serial correlation which might be modeled with a lag-term (who sits to your right).

7. Can we combine time series methods to panel data?

Re: Yes. At a minimum, most panel data models are similar to static models. We could add a time trend, a lagged explanatory variable, a lagged dependent variable or a moving average to the model.

8. Is time series data a good tool for prediction/forecasting?

Re: Often times, yes! Think of an autoregressive model like the following:

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

In many settings the outcome of today is very predictive of the outcome tomorrow even if we don't know what causes the outcome (e.g. think of daily temperature).