Testing for Serial Correlation

Tyler Ransom

Univ of Oklahoma

Feb 26, 2019

Today's plan

1. Review reading topics
    1.1 Basics of Time Series Data
    1.2 Serial Correlation
    1.3 Testing for Serial Correlation
2. In-class activity: practice with time series data, testing for serial correlation

Basics of Time Series Data

Time Series Data

- Time series data are quite different from cross-sectional
- Think: group of spaghetti noodles vs. one strand of spaghetti
- Time series: temporally ordered (earliest to latest)
- Example: US Quarterly GDP from 1901:Q1 until 2018:Q2

Differences relative to cross-sectional data

- Now, we don't have a notion of random sampling
- Instead, need to worry about **serial correlation**:

  2018:Q1 GDP likely correlated with 2018:Q2 GDP
- Also need to worry about trends and seasonality (cover after midterm)
- For today, we focus on serial correlation

Example of a time series regression

- Suppose you're interested in how trade deficits affect the interest rate
- Can specify the following model:

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

  where now we index observations by t to emphasize the time dimension

- y_t: Federal funds rate in year t
- x_t: US trade deficit (or US unemp. rate, or US inflation rate, ...) in year t
- This is called a **static model**

More detailed example of a time series regression

- We can also estimate a more elaborate model:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \beta_3 x_{t-2} + u_t$$

- now: current interest rate depends on the past 3 years of trade deficit
- This is called a **distributed lag model**

Other time series models

- Nothing stops us from having more independent variables:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \beta_3 z_t + \beta_4 z_{t-1} + u_t$$

- where z is nominal GDP

- This model might perform better at **forecasting** future interest rate

Lagged dependent variables

- We can also include lags of the dependent variable:

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

- now: current interest rate only depends on the previous interest rate
- This is called an **autoregressive** or **AR(1)** model
- if we put $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$ on RHS, we have **AR(p)** model
- Also sometimes called a **lagged dependent variable model**

Most general time series model

- We can put everything together and have an even bigger model:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \beta_3 z_t + \beta_4 z_{t-1} + \beta_5 y_{t-1} + u_t$$

- now, current interest rate depends on:
    - current and previous trade deficit
    - current and previous nominal GDP
    - previous interest rate

Estimating the β's

- Here, estimating the β's is pretty much same as before
- Formula for $\hat{\beta}$ is the same
- Formula for $\hat{\sigma}_u^2$ is the same
- Formula for $se\left(\hat{\beta}\right)$ still same
- Gauss-Markov assumptions are slightly different (see next slide)

Gauss-Markov in time series models

- Slight tweaks to Gauss-Markov relative to cross-sectional data:
- Replace A(2) [random sampling] with *no serial correlation*:

$$E\left(u_t u_s \middle| \mathbf{X}\right) = 0 \text{ for all } t \neq s$$

- Homoskedasticity and $E\left(u \middle| \mathbf{x}\right)$ are contemporaneous:

$$\begin{aligned} Var\left(u_t \middle| \mathbf{X}\right) &= \sigma^2 \\ E\left(u_t \middle| \mathbf{X}\right) &= 0 \end{aligned}$$

Serial Correlation

Serial correlation

- when $Corr(u_t, u_s) \neq 0$ for some (t, s) pairs
- say that errors exhibit **serial correlation** or **autocorrelation**
- No need to worry about it in cross-sectional data
- Why? $Corr(u_i, u_j) = 0$ in a random sample

Three different kinds of correlation in TS models

1. x's correlated over time (e.g. GDP_t correlated with GDP_{t-1})
   - This is fine, so long as no perfect collinearity
2. x's correlated with u_t (this implies $E\left(u|\mathbf{x}\right) \neq 0$!)
   - We assume away this possibility
3. u's correlated over time
   - This is serial correlation

Consequences of serial correlation

- Serial correlation is a lot like heteroskedasticity
- $\hat{\beta}$ still unbiased under serial correlation
- But can't do hypothesis testing without correcting for it
- TL;DR use robust standard errors
- This time, robust to serial correlation *and* heteroskedasticity
- We'll cover this next time

Testing for Serial Correlation

Testing for serial correlation

- Idea: specify a simple alternative model where u's can be serially correlated

  then use the model to test H_0 : no serial correlation

- Simplest example:

$$u_t = \rho u_{t-1} + e_t$$

  where e_t is serially uncorrelated, zero mean, constant variance

- Then the null is

$$H_0 : \rho = 0$$

Using residuals

- Easiest way to test H_0 is to regress u_t on u_{t-1}
- Problem: we don't observe the u's
- Solution: use $\hat{u}$'s

1. Run your main regression and extract the residuals
2. Run a regression of $\hat{u}_t$ on its lag, $\hat{u}_{t-1}$
3. Look at the default t-stat or p-value

Higher-order serial correlation

- We might have that u_t is correlated with both u_{t-1} and u_{t-2}

- In this case, modify the steps:

1. Run your main regression and extract the residuals

2. Run a regression of $\hat{u}_t$ on its lags $\hat{u}_{t-1}$ and $\hat{u}_{t-2}$

3. Look at the default F-stat for joint significance of $\hat{u}_{t-1}$ and $\hat{u}_{t-2}$

- Can do this for even higher-order models: $u_{t-1}, u_{t-2}, \ldots, u_{t-p}$

Exogeneity of regressors

- The previous slides' steps only apply when x's are **strictly exogenous**
- Needs to be modified for models where u_{t-p} correlated with x's, including $AR(p)$ models (i.e. lagged dependent variable models)

1. Run your main regression and extract the residuals
2. Run a regression of $\hat{u}_t$ on its lag, $\hat{u}_{t-1}$ and all x's
3. Look at the default t-stat or p-value