EEP/IAS 118 - Introductory Applied Econometrics, Lecture 14

Gregory Lane

August 2017

Time Series: Intro

Time series data is just another data type: it follows a single unit over many periods of time. Examples:

- U.S. yearly unemployment and GDP per capita for the years 1980-2010
- Global average temperature

We might use this type of data if we're interested in how aggregate quantities relate to each other (like unemployment and GDP_{pc}).

Time Series: Differences

Time series data is markedly different from what we've done so far, and this has some econometric consequences.

Primary Differences:

1. Data has a temporal ordering: we know that the data from 1980 comes right before the data for 1981
2. The past can affect the future: what happened in 1980 could influence what we observe in 1981 (lose independence of observations)
3. A time series dataset is one possible outcome of the stochastic process, it is impossible to go back in time and start the process again
    - The single time series is analogous to the single sample we draw in a cross-section

Time Series Analysis: Static Model

The static model is simply relating y and x at the same time period:

$$y_t = \beta_0 + \beta_1 x_t + u_t \qquad t = 1, 2, \cdots, T$$

- Static model is proposed when a change in x at time t is believed to have an immediate effect on $y : \Delta y_t = \beta_1 \Delta x_t$
- An example of a static model is the static Phillips curve:

$$interest_t = \beta_0 + \beta_1 unemp_t + u_t$$

- Illustrates the contemporaneous tradeoff between inflation and unemployment

Time Series Analysis: Finite Distributed Lag (FDL) Models

Allow one or more variables to affect y with a lag.

- For example, for annual observations, consider the model:

  $$y_t = \alpha_0 + \delta_0 x_t + \delta_1 x_{t-1} + \delta_2 x_{t-2} + u_t$$

  which is an FDL of order 2.

- We have related y_t to the values of x in the current time period and the previous two time-periods
- Note how subscripts have become important

Time Series Analysis: Finite Distributed Lag (FDL) Models

How to interpret the coefficients in this model?

- Need to distinguish between *temporary* and *permanent* changes in x

Temporary Change in x

- Suppose at time t, x increases by one unit to and then reverts to its previous level at time $t+1$. That is, the increase in x is **temporary**
- Important thing to realize is that the effect of this temporary change in x will *change* depending on when we are examining y

Time Series Analysis: Finite Distributed Lag (FDL) Models

The effect of a **temporary** unit increase in x at time t. Assume x always has constant value c otherwise:

$$y_{t-1} = \alpha_0 + \underbrace{\delta_0 c}_{t-1} + \underbrace{\delta_1 c}_{t-2} + \underbrace{\delta_2 c}_{t-3} \tag{1}$$

$$y_t \quad = \alpha_0 + \underbrace{\delta_0(c+1)}_{t} + \underbrace{\delta_1 c}_{t-1} + \underbrace{\delta_2 c}_{t-2} \tag{2}$$

$$y_{t+1} = \alpha_0 + \delta_0 c + \delta_1(c+1) + \delta_2 c \tag{3}$$

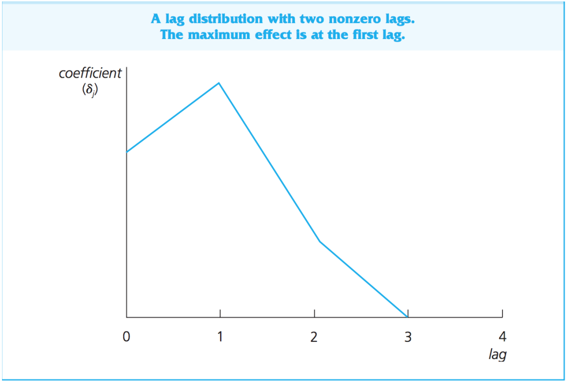
$$y_{t+2} = \alpha_0 + \delta_0 c + \delta_1 c + \delta_2(c+1) \tag{4}$$

$$\tag{5}$$

- δ_0 is the **immediate** change to the one unit increase in x
- δ_1 is the change **one period after** the temporary increase
- δ_2 is the change **two periods after** the temporary increase

Time Series Analysis: Finite Distributed Lag (FDL) Models

- At time $t+3$, y has reverted back to its initial level: $y_{t+3} = y_{t-1}$. This is because we have assumed that only two lags of x will appear in our model
- We can then plot the δ_j as a function of j. Example:



Time Series Analysis: Finite Distributed Lag (FDL) Models

Permanent Change in x

- Suppose that before time period t, x equals the constant c. At time t, suppose x increases permanently to $(c+1)$

$$y_{t-1} = \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c \tag{1}$$
$$y_t = \alpha_0 + \delta_0(c+1) + \delta_1 c + \delta_2 c \tag{2}$$
$$y_{t+1} = \alpha_0 + \delta_0(c+1) + \delta_1(c+1) + \delta_2 c \tag{3}$$
$$y_{t+2} = \alpha_0 + \delta_0(c+1) + \delta_1(c+1) + \delta_2(c+1) \tag{4}$$

- δ_0: contemporaneous effect
- $\delta_0 + \delta_1$: effect after one period
- $\delta_0 + \delta_1 + \delta_2$: effect after two periods

Time Series Analysis: Finite Distributed Lag (FDL) Models

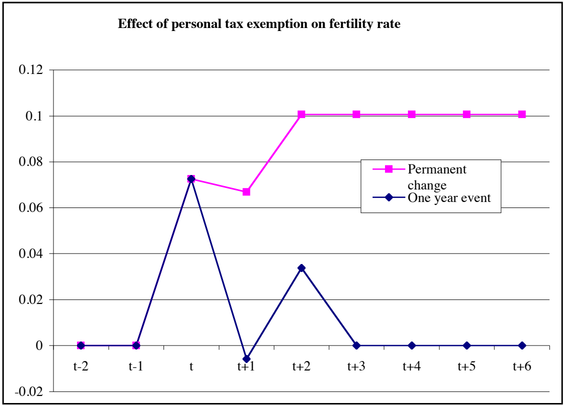
Example: Fertility rates on personal tax-exemption with two lags:

```
reg gfr pe pe1 pe2 ww2 pill

      Source |       SS       df       MS              Number of obs =      70
-------------+------------------------------           F(  5,    64) =   12.73
       Model |  12959.7886     5  2591.95772           Prob > F      =  0.0000
    Residual |  13032.6443    64  203.635067           R-squared     =  0.4986
-------------+------------------------------           Adj R-squared =  0.4594
       Total |  25992.4329    69  376.701926           Root MSE      =   14.27

------------------------------------------------------------------------------
         gfr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          pe |   .0726718   .1255331     0.58   0.565    -.1781094     .323453
         pe1 |  -.0057796   .1556629    -0.04   0.970     -.316752    .3051929
         pe2 |   .0338268   .1262574     0.27   0.790    -.2184013     .286055
         ww2 |   -22.1265   10.73197    -2.06   0.043    -43.56608   -.6869196
        pill |  -31.30499   3.981559    -7.86   0.000    -39.25907   -23.35091
       _cons |    95.8705   3.281957    29.21   0.000     89.31403     102.427
```

Time Series Analysis: Finite Distributed Lag (FDL) Models



Trends: Spurious Correlation

Some time series will contain a time trend: i.e. some variables will be naturally increasing or decreasing over time:

- Examples: Height of children, prices, GDP
- If two sequences are trending in the same or opposite directions, we might falsely conclude that one variable is causing a change in another
- Failing to account for the trend is a type of omitted variable bias
- The assumption for unbiasedness for time series is:

$$E(u_t|X) = 0, \quad t = 1, 2, \cdots, n$$

which says that the error term at time t (u_t) is uncorrelated with each explanatory variable in *every* time period.

Trends: Dealing with the problem

There are two possible solutions to deal with this potential source of endogeneity:

1. **Include a linear time trend:**

$$y_t = \alpha_0 + \delta_0 x_t + \cdots + \alpha_1 t + e_t$$

   α_1 measures the change in y_t from one period to the next due to the passage of time

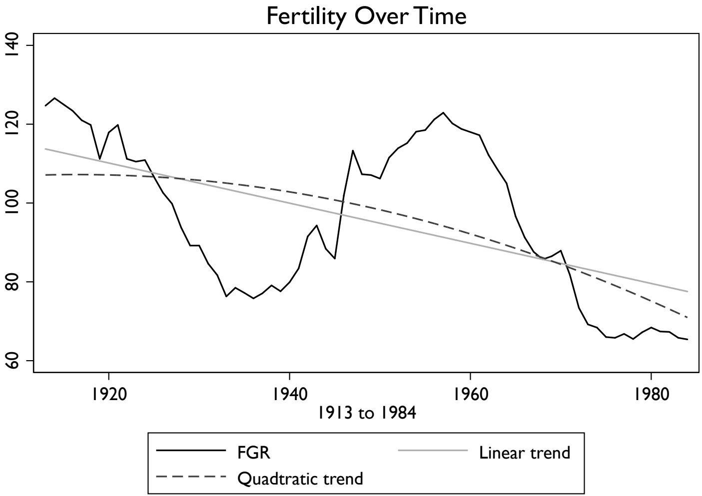
2. **Use first differences:**

$$\Delta y_t = \beta \Delta x_t + u_t$$

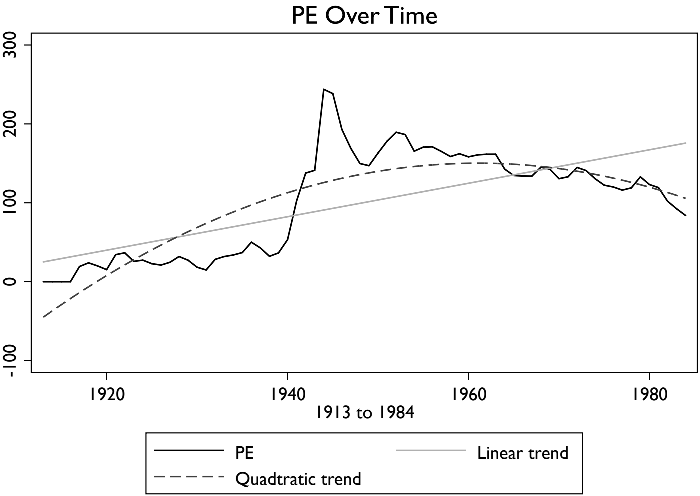
   We have re-written the model to measure *changes* from one time-period to the next Taking the first difference allows you to parse out a common time pattern

Trends: Example

Let's go back to the example of the relationship between fertility and the personal tax exemption. We observe the following trends in the data:



Trends: Example



Example

Let's say we run the following regression:

$$gfr_t = \beta_0 + \beta_1 pe_t + \beta_2 ww2_t + \beta_3 pill_t + u_t$$

What problems is this regression vulnerable to?

Example

Let's say we run the following regression:

$$gfr_t = \beta_0 + \beta_1 pe_t + \beta_2 ww2_t + \beta_3 pill_t + u_t$$

What problems is this regression vulnerable to?

- Spurious correlations because we don't control for time
- $\hat{\beta}_1$ may be biased if the effect of personal exemptions takes/lasts more than one period (i.e. the static model isn't appropriate)

Question: What would you expect the direction of bias to be on β_1?

Example: Naive Regression

```
. eststo: reg gfr pe ww2 pill
      Source |       SS       df       MS              Number of obs =      72
-------------+------------------------------           F(  3,    68) =   20.38
       Model |  13183.6215     3  4394.54049           Prob > F      =  0.0000
    Residual |  14664.2739    68  215.651087           R-squared     =  0.4734
-------------+------------------------------           Adj R-squared =  0.4502
       Total |  27847.8954    71  392.223879           Root MSE      =  14.685
------------------------------------------------------------------------------
         gfr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          pe |     .08254   .0296462     2.78   0.007     .0233819    .1416981
         ww2 |   -24.2384   7.458253    -3.25   0.002    -39.12111   -9.355684
        pill |  -31.59403   4.081068    -7.74   0.000    -39.73768   -23.45039
       _cons |   98.68176   3.208129    30.76   0.000     92.28003    105.0835
```

Example: Add Time Trend

```
. eststo: reg gfr pe ww2 pill t
      Source |       SS       df       MS              Number of obs =      72
-------------+------------------------------           F(  4,    67) =   32.84
       Model |  18441.2357     4  4610.30894           Prob > F      =  0.0000
    Residual |  9406.65967    67  140.397905           R-squared     =  0.6622
-------------+------------------------------           Adj R-squared =  0.6420
       Total |  27847.8954    71  392.223879           Root MSE      =  11.849
------------------------------------------------------------------------------
         gfr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          pe |   .2788778   .0400199     6.97   0.000     .1989978    .3587578
         ww2 |  -35.59228   6.297377    -5.65   0.000     -48.1619   -23.02266
        pill |   .9974479    6.26163     0.16   0.874    -11.50082    13.49571
           t |  -1.149872   .1879038    -6.12   0.000    -1.524929   -.7748145
       _cons |   111.7694   3.357765    33.29   0.000     105.0673    118.4716
```

Time Series: Summary

Summary of approach for basic time-series:

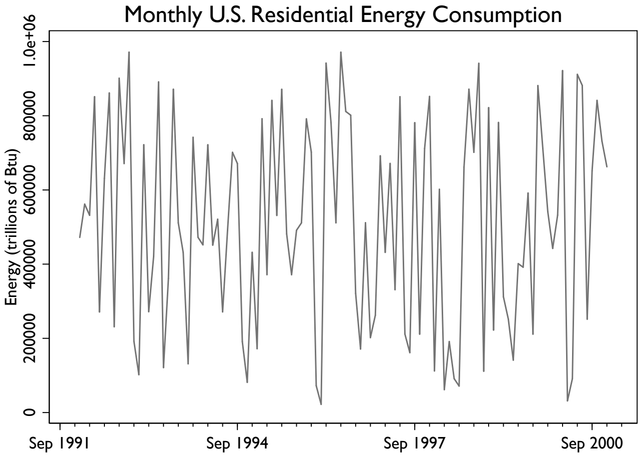
1. Do your variables have significant time trends?
   - Include a measure of time in my regression model or use first differences
2. Do you think variable of interest x_{kt} will only have an immediate effect on y_t?
   - Try a "static model", where x_{kt} doesn't affect y_t in other time periods
3. Is there a reason why my variable of interest x_{kt} might take a few time periods to affect y_t? Or do I think the effect might linger for more than one time period?
   - Include lagged values of x_{kt}, like $x_{k,t-1}$ and $x_{k,t-2}$, in my regression model

Seasonality

The same spurious correlation we observed in time series can also occur if our data has a strong seasonal pattern. The approaches to dealing with seasonality are very similar to how we dealt with linear time trends:

1. Control for seasonality by including the appropriate dummy variables (month, season, quarter) in the regression
2. Use seasonally adjusted data

Seasonality: Example



Seasonality: Process

Here is the basic process to seasonally adjust data:

1. Run the regression: $y_t = \beta_1 jan_t + \beta_2 feb_t + ... + \beta_1 2dec_t + u_t$
2. Compute: $\bar{\beta} = \frac{1}{N} \sum_{k=1}^{12} \hat{\beta}_k$
3. Create a new variable: $y_t^{sa} = y_t - \hat{\beta}_k + \bar{\beta}$ if y_t is from month k

This process is equivalent to including month fixed effects