Ch16: Time Series

- **HW8 due tonight**
- **Please download: 22-TheFed.xls**

24 Nov 2011
BUSI275
Dr. Sean Ho

Trinity Western University

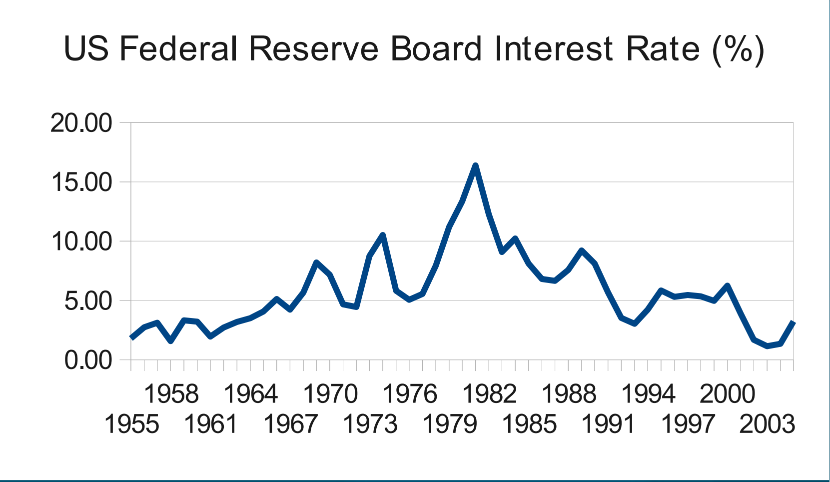
Outline for today

- Time series data:
  - Dependent observations
- Trend-based approach:
  - Trends, cycles, seasons
  - Additive vs. multiplicative model
- Autoregressive approach:
  - Autocorrelation
  - Correlogram
  - Finite differencing and the ARIMA model
- Combining trends with ARIMA

Time series data

- Time is one of the independent variables
  - Often only 1 DV and 1 IV (time)
  - But can also have other time-varying IVs
- Why not just use regression with time as the IV?
  - Assumptions of regression: in particular, observations need to be independent!
- Two (complementary) approaches:
  - Model the time-varying patterns and factor them out to leave residuals that are independent (uncorrelated)
  - Model the conditional dependence of the present value on past values

Patterns



- Patterns to look for:
  - Trend: linear growth/loss
    - Also non-linear trends: t^λ, $\ln(t)$, etc.
  - Cycle: multi-year repeating pattern
  - Season: pattern that repeats each year
    - e.g., if data is quarterly, use dummy vars for the seasons: $b_2S_2 + b_3S_3 + b_4S_4$
- Additive model:
  - $Y_t = (b_0 + b_1t)$ + (cyclical component) + (seasonal component) + (residual)
- Assumes residuals are independent, normally distributed, with constant variance

Additive vs. multiplicative

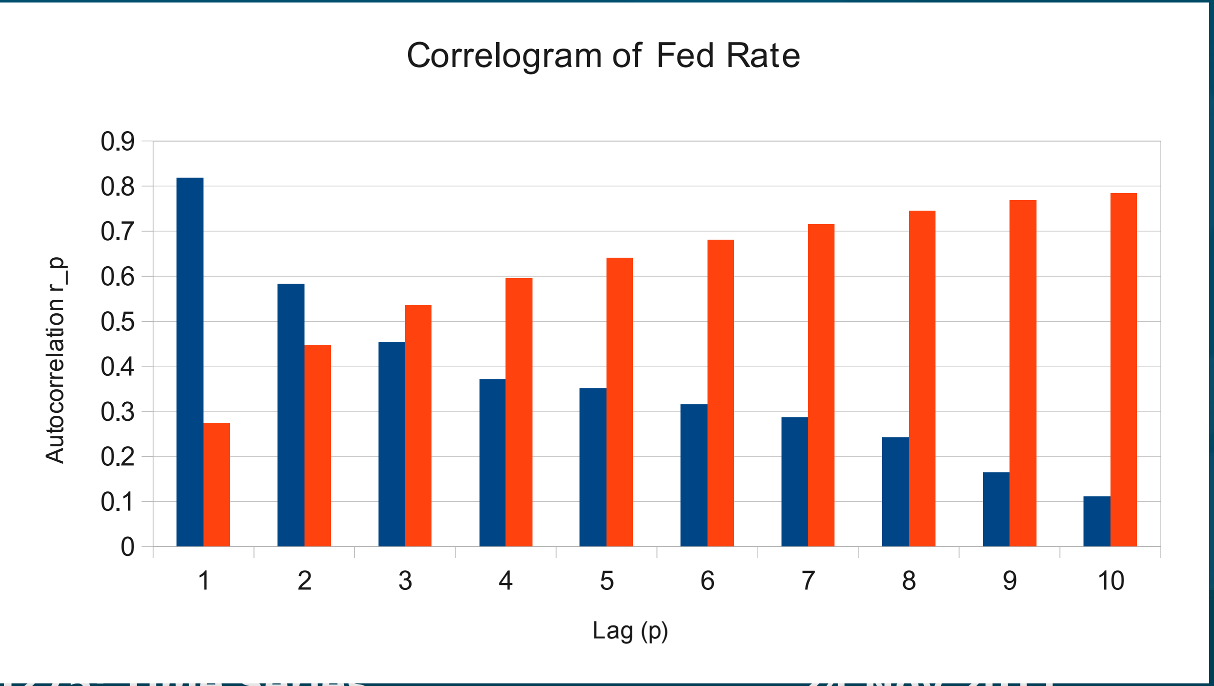
- Homoscedasticity of residuals is often an issue
- Plot resids vs. predicted value
  - Look for systematic variation in residual SD
  - "Spread vs. level" plot: √(std resids) vs. predicted value
- If you see a distinct "fan" shape,
  - i.e., the SD of the random variation grows with the level of the variable
- Then apply a log transform to the variable:
  - $\ln(Y_t)$ = (linear) + (cyclic) + (seasonal)
- This is equivalent to a multiplicative model:
  - Y_t = (linear) * (cyclic) * (seasonal)

Autocorrelation

- Another approach models the correlation of the current value against past values:
  - $P(Y_t \mid Y_{t-1})$
  - Or in general: $P(Y_t \mid \{Y_s : \text{all } s<t\})$
- The autocorrelation (ACF) r_p of a variable Y is the correlation of the variable against a time-shifted version of itself:
  - Let $\text{Covar}(x, y) = (1/n)\,\Sigma\,(x - \bar{x})(y - \bar{y})$
  - Then $r_p = \text{Cov}(Y_t, Y_{t-p}) / \text{Var}(Y_t)$
  - p is the lag (always positive)
- e.g., quarterly seasonal data may have large r_4

Correlogram

- The correlogram is a column chart illustrating the autocorrelation for various lags
- Statistical software will also show the critical value for each autocorrelation
    - Autocorrelations that are significant suggest an autoregressive model with lag p: AR(p)
- TheFed data: AR(2) model



Differencing

- Another tool to reduce dependencies of consecutive values is finite differencing:
  - Look at $Y_t - Y_{t-d}$, where d is the lag
  - Year-over-year change on annual data: d=1
  - Year-over-year change on quarterly: d=4
- A model that combines finite differencing (integration) with autoregression and moving averages is called an ARIMA(p,d,q) model:
  - p = lag for autocorrelation
  - d = lag for differencing
  - q = lag for moving average
    - Use partial correlogram (PACF)

Combining approaches

- The trend-based approach and the autoregressive approach can be combined:
- First fit broad trends/cycles/seasons
  - Resulting residuals (de-trended, de-seasonalized data) may still be auto-correlated
- Use correlograms to choose an ARIMA model for the residuals
- Goal is to get the residuals to be small, independent, normally distributed, and with constant variance

TODO

- HW8 (ch15,12): due tonight
- Projects:
  - Presentations next week!
  - Final paper due Wed 7Dec