Choosing a ‘best’ model (set of explanatory variables):

1. AIC & BIC model selection criteria – lowest is best (closest to ).
2. THINK! Think about the underlying DGP.
3. Predictive accuracy test – which model predicts best. Compute the RMSPE and compare. Should compare out-of-sample prediction, not in-sample. ‘Press statistic?’
4. t-statistics, i.e. p-values (magnitude of the coefficients).
5. Subset F-test (NOT the overall F-test).
6. Omitted variables bias vs. (inefficiency due to) inclusion of irrelevant variables

Time series

BEFORE we estimate any models, we transform ALL the variables into stationary form.

Why?

The spurious regression problem – nonstationary variables all appear related because the all vary with time.

**A variables is stationary if:**

1. **It has constant mean**
2. **It has constant variance**
3. [all the autocovariances (autocorrelations) are independent of time]

**How to decide if a variable is stationary?**

1. **Visual inspection**: plot the variable, its first difference, its second difference. –   
   Does the time plot vary within a constant band (confidence interval) around a constant mean. Does the variable zig-zag around zero (suggests overdifferencing).
2. **Standard deviations**: Compute the SD of the variable, its first difference, its second difference. – If SD drops substantially (more the 50%?!) when differencing, that suggests to difference.
3. **ACFs** (autocorrelation function): Does the ACF die out fairly quickly, if not, then not stationary.
4. **ADF test** (Augmented Dickey-Fuller test): testing if the variable is stationary or not. Known as a “unit root test”.
5. **Think about the DGP**. Often we can tell if a variable is stationary or not!
6. **“If in doubt, difference!”**

This is known as “determining the **order of integration** of a variable”.

If is I(0) (integrated of order 0), then is stationary (“difference zero times”).

If is I(1) (integrated of order 1), then the first difference, , is stationary (“difference once”).

If is I(2) (integrated of order 2), then the second difference, , is stationary (“difference twice”).

**Autocorrelation function**

Autocorrelation = serial correlation.

However, in econometrics when we say “serial correlation” we usually really mean serial correlation *in the error term* of a model.

Autocorrelation aka serial correlation = correlation in a variable over time, i.e.

We know what corr(x,y) .

is. An autocorrelation = corr( = 1st order autocorrelation.

corr( = 2nd order autocorrelation.

corr( = *k*th order autocorrelation.

ACF(1) = corr(.

ACF(2) = corr(

…

ACF(k) = corr(

**ADF test**

**1st round** is comparing I(0) with not I(0) (stationary or not).

If the test statistic is 1) negative, and 2) statistically sig., i.e. more negative than -2.6 (10% critical value), then the variable is stationary. If not sig., then not stationary.

The null hypothesis is that the variable is nonstationary!

**2nd round** is comparing I(1) with I(2) (I(1) or I(2)).

Same test, just looking at the difference of the variable.

WARNING! ADF Test has low power. Do not rely on the test.

Order of int first, qu. 2 at least 200 obs., an explosive variable

Linear trend model

Suppose is stationary!