**DO NOT EVER CORRECT FOR SERIAL CORRELATION** (G. Mizon J. of Ectrics, A simple message for autocorrelation correctors: Don't (1995)

To capture potential serial correlation/autocorrelation in the error, include lags, especially of the dependent variable, as explanatory variable.

**Cointegration**

Two (or more) variables are cointegrated IF

1. All the variables are integrated of the same order and nonstationary, i.e. all I(d), d> 0
2. There exist a linear combination of the variables that is stationary, i.e. I(0).

E.g., suppose . If

To determine whether two variables are cointegrated

Determine that all variables are integrated of same order and not I(0). To do this we use something like the ADF test – the Bayesian ADF is best!!

**DF test**: ,

So test vs.

Augmented DF test: just add lags of dependent variable on RHS to capture serial correlation in the model error (so errors are more likely to satisfy white noise assumption: )

**ADF test**: Add enough lags (choose ) to capture the dynamics of the process.

Problem: If is not stationary, then the OLS estimator of is nonstationary, hence the ADF test has different critical values than the usual t-test, and **has low power**.

[If we include a trend on the RHS of the ADF test equation (e.g., include ),

, then **we are assuming** that is nonstationary, and are testing for TS vs. DS! Don’t do this either!]

Same hypothesis, same test, just somewhat more powerful because we have made the error more likely to satisfy the assumptions.

**A Bayesian version of ADF test**

Same equation to test, and same hypotheses,

vs.

Just compute (from the posterior for prob(.

Small sample distribution (no asymptotic assumptions) and a one-tail test, so more powerful. ***Same distribution regardless of whether is nonstationary or not.***

Note that is always cointegrated with (and any of its lags) … Mary had a little lamb.. !

**Cointegration**

Let’s use two variables, as above, as an example,

Suppose we find .

**Engle-Granger methodology**

Estimate the linear regression,

Get the residuals,

and test whether the residuals are stationary or not,

if using the above test for stationarity, with as the variable (in place of ).

Problem: this is a two-step process, we have to use estimates of the cointegrating relationship to perform the test because we don’t know the “true” parameter values – so less efficient, less powerful test + all the drawbacks of the frequentist ADF test.

**Serial correlation vs. Stationarity testing**

IF is the model error from our model interest, then **the test for serial correlation** in the error is,

vs.

IF is just some variable (not a model error), say , then **the test for stationarity** of this variable is,

vs.

IF is the cointegrating relationship – our above - (or estimate thereof), then **the test for stationarity** in this error, i.e., the **test for cointegration** is,

vs.

**Bayesian one-step approach!**

We want to test if in the following,

1. (cointegrating relationship)
2. (test equation)

–we could do the ADF test equation:

(alternative test equation)

Rearrange (1),

Lag (3) once,

Substitute (3) and (4) into (2),

Rearranging,

Are all these parameters (in the constant) **identified**?

Looks like and are not identifiable (cannot **separately** estimate them).

For example, (assuming is stationary) if we specify a simple model as,

What wrong with this?

Infinite set of combinations of that give the same , so they are **not identifiable**.

However, we can write equation (5) as,

and, since we are not interested in separate estimates of and , we can estimate (6) with no problems using HMC. It is a nonlinear equation, but so what!?

Then our cointegration test is,

vs.

We compute the probability of each hypothesis from the posterior of .