**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 .

**Suppose you find that two variables ARE cointegrated.** So what? What difference does this make?

What do you do?

Cointegration is a **LONG RUN** relationship. The variables could be unrelated in the short run, but move together in the long run (they have a “common trend”).

Models with AR and MA components are all **SHORT RUN** models. The strength of the relationship between the variables decays over time and the process has a finite memory.

When we difference (or detrend) we are removing the long run relationship. Cointegration gives us a way to model the long run relationship explicitly. We do this by including an ERROR CORRECTION term, to give us an **error correction model**.

**Error correction models**

An error correction = dynamic regression model (so lags of all variables) + the error correction term.

The error correction = one lag (and only one lag) of the cointegrating relationship.

Suppose we have two variables that are both I(1), i.e. .

**A dynamic regression model** would be:

Assuming the variables are cointegrated, so , where

we include one lag of this in the dynamic regression to capture long run adjustment (“error correction”) between the two variables, so the **error correction model** is:

If the two variables are NOT cointegrated, then we estimate a dynamic regression model (i.e., just don’t include the error correction term).

To estimate an ECM we can either

1. Use a two-step process: estimate

And plug in one lag of the estimates of (include – this is the error correction term.

[drawback: ignoring the uncertainty about the parameters].

1. Substitute the cointegrating relationship into the ECM equation:

A one-step approach would be to estimate:

A nonlinear likelihood function. Can we easily identify and estimate each parameter?

Simulation experiment:

DGP: dx = randn(n); x = cumsum(dx)

y = 1.0 .+ 0.5.\*x .+ 0.6.\*randn(n)

Two-step model parameter estimates (other than cointegrating relationship, for which we just estimate the linear model)

Cointegrating relationship parameter estimates: 0.9897, 0.500

parameters mean std mcse ess\_bulk ess\_tail rhat

delta -0.2309 0.2924 0.0074 1566.1349 1444.9613 0.9999

alpha -5.7980 0.1874 0.0043 1944.5271 1609.5130 0.9995

sig 4.2550 0.1367 0.0032 1779.4414 1425.0606 1.0005

beta 0.1748 0.2273 0.0052 1898.4820 1723.0748 0.9997

phi 0.0646 0.2633 0.0072 1355.5236 1491.0494 0.9996

One-step estimation:

parameters mean std mcse ess\_bulk ess\_tail rhat

gamma 11.5024 4.8009 0.1750 683.1458 676.6405 1.0035

delta -0.0544 0.0237 0.0010 678.8116 649.0039 1.0036

alpha 0.9787 0.0720 0.0019 1402.6637 1390.3605 0.9996

sig 0.7979 0.0261 0.0006 1645.1040 990.8581 1.0029

beta -0.0544 0.0410 0.0011 1387.8337 1110.8888 1.0009

phi -0.0270 0.0418 0.0012 1236.0268 1188.8858 1.0044

The **Johansen VECM approach** to estimating and testing for cointegration:

Estimate a VAR (see below) to determine lag length.

Then estimate a VECM (vector error correction models) - so similar to above, only one equation for each variable)

Reduced rank likelihood estimation.

Vector Autoregressive Model (VAR)

Just a system of dynamic regression model equations, one equation for each variable.

IF we use exactly the same number of lags of all variables in all equations, then we can estimate as a standard linear regression (“OLS”).

VAR(p) (VAR of order p, i.e. p lags) for two variables:

IF you allow for different lag lengths for different and in different equations, THEN the error terms are correlated across equations (i.e. and are correlated), so we should estimate as a system allowing for nonzero covariance in the covariance matrix, so a seemingly unrelated regression (SUR) model), i.e., instead of assuming , which can be written as,

we allow a more general contemporaneous covariance structure:

A VECM = VAR + add the error correction to each equation:

VECM(p) (VECM of order p, i.e. p lags) for two variables:

**Drawbacks of the Engle-Granger approach:**

1. Ignoring uncertainty in the parameter estimates (two-steps to test, so greater chance of error).
2. If more than two variables, then potentially many cointegrating relationships to consider, so multiple testing!

E.g., Suppose we have 4 variables, all I(1). The following all are possible cointegrating relationships:

all cointegrated (and which one should you use as the dependent variable – doesn’t matter asymptotically, but can matter with a finite sample).

or

? etc.

Must then look at combinations of 3 variables:

Must then look at combinations of 2 variables:

…

**ARCH-GARCH and stochastic volatility models**

Modeling the error **variance** as correlated over time

AutoRegressive Conditional Heteroskedasticity = ARCH. GARCH = Generalized ARCH

The equation variance is a function of the past value of the residual (one lag), or the past value of the variance (one lag of the variance, so )

Some model )

Then model the heteroskedastic error variance as an AR process (with either lags of the variance, or lags of “best estimate” of the variance at each time point = squared residual), e.g.,

+ …

**CAUSALITY**

Who is Judea Pearl?

What is a do operator?

What is a DAG (directed acyclic graph)?

Does economic or finance data come from a DGP that is a DAG?

What, if anything, do mediators and moderators have to do with this?

Is statistics all correlation not causation? How do we get causation into the picture?

Deterministic or probabilistic?

All or nothing vs. magnitude of effect?

Should we do diff in diff? If not, what should we do?

PANEL MODELS