POL 606 Time Series Analysis
Weeks 7 & 8
Unit-Roots, Cointegration, Error Correction, Fractional Integration, Near Integration and Fractional Cointegration
March 10th and 12th 2008

Review – Weak Stationarity

A series is weakly stationary if the expected value of the mean and the variance is constant for all time periods.
And, the expected covariance between any two periods of t is constant.

Random Walk:

$$y_t = 1.0y_{t-1} + \varepsilon_t$$

y_t is the sum of all previous shocks.

This is the defining characteristic of a nonstationary series: it "never forgets."

Since the ϕ_1 coefficient equals 1, stochastic shocks are not discounted as the process evolves.

They accumulate over time at their full value.

By recursive substitution for y_{t-1} we could show that:

$$y_t = y_{t=0} + \sum \varepsilon_i$$

With $y_{t=0}$ telling us the initial value for *y*.

This process has a non-constant variance and is said to exhibit a *stochastic trend*.

Random walk with drift:

$$y_t = y_{t=0} + R * t + \sum \varepsilon_t$$

y_t is the sum of the initial value of *y* plus some value R that accumulates for each time point plus the sum of shocks.

The R is the "drift" component that is a deterministic trend.

This process also creates nonstationary variables.

Note that a random walk is not a deterministic trend, it is stochastic only in that it is the sum of stochastic shocks since the initial period.

We can also have a random walk with noise:

$$y_t = \mu_t + v_t \qquad v_t \sim N(0, \sigma^2)$$

where:

$$\mu_t = \mu_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma^2)$$

That is, μ_t is a random walk.

And local linear trend model:

$$y_t = \mu_t + v_t \qquad v_t \sim N(0, \sigma_1^2)$$

where $\mu_t = a_t + \mu_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma_2^2)$

$$a_t = a_{t-1} + \delta * t$$

These are all types of non-stationary variables.

If we suspect nonstationarity, the answer is to difference the variable prior to using it in regression analyses.

Spurious Regressions

Dealing with non-stationary data is a problem that was not properly dealt with until the 1970s.

The biggest consequences of not doing so are spurious regression results.

This is not a new discovery, but was noticed by Yule in 1926.

A spurious regression occurs when a common factor causes variables and we mistakenly believe that one variable is causing the other.

In time series, this common factor can often be either the unit-root or some other aspect of memory.

Assume 2 time series that are completely independent from each other.

If we generate both of them as random walks so that:

$$y_t = y_{t-1} + \varepsilon_t$$

and:

$$x_t = x_{t-1} + \varepsilon_t^*$$

And cov($\varepsilon_t , \varepsilon_t^*$)=0 for all values of t.

Monte Carlo studies have shown that a regression of :

$$y_t = a + bx_t + u_t$$

will produce a statistically significant estimate of *b* as often as 75% of the time in samples of 100-150 (Granger and Newbold 1974).

This will get even worse as the length of the two series increases, but occurrence of Type I errors is incredibly high with any value of *t*.

This is the result of non-stationarity and is the classic problem of spurious regression.

From Davidson and Mackinnon 1993, p. 672.

Table 19.1 Spurious Rejections and Sample Size

n	Random Walk	Lag Added	Drift	Trend
25	0.530	0.146	0.645	0.066
50	0.662	0.154	0.825	0.431
75	0.723	0.162	0.905	0.987
100	0.760	0.162	0.945	1.000
250	0.847	0.169	0.997	1.000
500	0.890	0.167	1.000	1.000
750	0.916	0.170	1.000	1.000
1000	0.928	0.169	1.000	1.000
2000	0.947	0.168	1.000	1.000

These are Monte Carlo results and show the proportion of times in 10,000 regressions where equation (0) gives a statistically significant coefficient when:

Column 1: both x_t and y_t are generated by independent random walks.

Column 2: both x_t and y_t are the same as column 1 but a lagged endogenous variable is included on the right-hand-side.

Column 3: Both x_t and y_t are generated by independent random walks with drift with the drift parameter being 1/5th of the size of the standard error.

Column 4: Both x_t and y_t are independent trend-stationary series with the trend coefficient being 1/25th the size of the standard error.

Column 3 and 4 are not surprising – since both are trending upward.

Rejections of the null become assured very quickly.

Column 1 and 2 – series are completely independent and neither has a trend.

Why do we get these false rejections so often?

To have a false rejection means that *t* is significant meaning that we reject the null hypothesis that $b=0$.

The false rejection is telling us that the null hypothesis, $b=0$, must be false.

Because *b* cannot be 0 in $y_t = a + bx_t + u_t$ or it would reduce to $y_t = a + u_t$.

That is, y_t is generated by only a constant plus a stationary error – which is certainly not true when y_t is a random walk.

Thus, when we test the null hypothesis, even against an alternative that is also false, we often reject it.

The error term cannot satisfy the Gauss-Markov assumptions if the null hypothesis is true.

If the null hypothesis is true, and we know it to be true, the error term is itself a random walk – or without the constant in the equation **it is y_t itself**.

There is very high auto-correlation in the errors.

So, the standard errors and the *t*-statistics are invalidated.

Davidson and Mackinon say that this result is not entirely satisfactory, but it's a good start. See Phillips (1986) for the classic discussion.

Traditionally, this is handled by differencing the series and then regressing the differenced versions of the variables:
$\Delta y_t = a + b\Delta x_t + u_t$

In Box-Jenkins we difference not only to avoid spurious regressions but also to identify the AR and MA processes.

Taking account of non-stationarity helps to remove these threats to inference **but** we are throwing away the chance to look at the long run relationships between the variables.

This is the problem known as: "The second face of spuriousness."

We'll see that if we can find that two non-stationary series are cointegrated we can look at the short- and long-run dynamics of non-stationary models.

Formal Tests for Unit Roots

Begin with the Dickey-Fuller tests.

$$y_t = by_{t-1} + \varepsilon_t \qquad (1)$$

if $b < 1$ this is a stationary AR process.

if $b = 1$ it is a random walk. Since b=1, that is the "unit" that is the "root" of the series.

We begin by adding a constant and subtracting y_{t-1} from both sides of (1) to get:

$$(1-L)Y_t = \beta_0 + \gamma Y_{t-1} + \varepsilon_t \qquad (2)$$

The Dickey-Fuller test has a null hypothesis of non-stationarity.

The basic test has the null hypothesis of a random walk.

If so, $\gamma = 0$. (γ is gamma)

Additional parameters for a constant and/or a deterministic trend may be added to (2) depending upon what assumptions are made about the data generating process.

Whatever the form of the test, rejection of the null implies stationarity.

A key point is that the *t* distributions for the Dickey-Fuller test is nonstandard and special critical values must be employed.

Like any other regression we can look at the residuals.

They should be white noise – problematic if the ε's have non-zero correlations.

If diagnostics suggest the presence of such correlation, lags of the dependent variable, which is differenced, are included in a respecified model and the model parameters are reestimated.

The resulting test is called an "augmented Dickey-Fuller" test.

$$\Delta Y_t = \beta_0 + \gamma Y_{t-1} + \beta_1 \Delta Y_{t-1} + \beta_2 \Delta Y_{t-2} ... + \varepsilon_t$$

We will run the Dickey-Fuller test at multiple lags and with and without a trend and a constant.

With a constant means we are hypothesizing a random walk with drift.

Without a constant → no drift.

With a constant in the model the critical value is:
-2.58 at the .10 level.
-2.89 at the .05 level
-3.51 at the .01 level.

These values are the same for the ADF test.

See Enders Page 183 for specific critical values depending on our type of Dickey-Fuller test.

Below (more negative) these critical values we would reject the null hypothesis.

Above we would conclude a unit-root.

There are many problems with the Dickey-Fuller test specifically and with tests of stationarity in general.

First, structural breaks in otherwise stationary processes may lead us to falsely conclude that a series is nonstationary.

For the possibility of a structural break use a "Perron Test."

Second, unit-root tests have low power.

It is difficult to distinguish a series where d=1 from a series that is "near integrated" or "fractionally integrated."

Near Integration refers to a series with a very strong autoregressive component. Such a series will have long memory that deteriorates at an exponential rate.

A fractionally integrated series becomes a possibility when we relax the assumption that d must be confined to integers. That is, d may be somewhere in between 0 and 1.

More about these possibilities soon.

Not only do unit-root tests have trouble distinguishing between series that are unit-roots in the face of these two near and fractional alternatives, they may not be very good at distinguishing between series that are d=0 or d=1. Diebold et al. (1991).

There are many unit-root tests and a big reason for that is that no test has shown itself to be good enough at testing the unit-root hypothesis (Stock 1994).

Third, in the case of Dickey-Fuller, its strange to have a null hypothesis of nonstationarity.

There may be implications to this – makes it easier to find unit-roots in the absence of hard proof against them.

And, just because a series isn't a unit root, it may not be strictly stationary either.

In RATS:

```
source(noecho) dfunit.src
@dfunit(ttest) varname
```

or can do:

```
@dfunit(ttest,trend) varname
```

or:

```
@dfunit(ttest,lags=12) varname 1980:01 1990:12
```

How do we know which types of DF tests to do?

Do them all, use the first lagged value that gives you white noise residuals.

There are other tests for stationarity all of which including the Dickey-Fuller test are available at the Estima website for downloading, along with many other source files.

For good, quick descriptions, see Box-Steffensmeier and Smith, 1996. "The Dynamics of Aggregate Partisanship."

Also, Lebo, Walker and Clarke 2000.

KPSS Test

Null hypothesis is that the series is a "strong mixing" process – or stationarity.

This is the best strength of the KPSS test.

A series is strong mixing if the rate at which dependence between past and future observations goes to zero as the distance between them grows is “fast” enough.

That is, the memory is mixed away quickly in a fast mixing process.

Stationary series, which decay at a geometric rate are strong mixing processes.

Fractionally integrated series which decay at a hyperbolic rate and unit-root processes which do not decay are not strong mixing.

KPSS decomposes the series, by assumption, into a deterministic trend, a random walk, and stationary errors.

A series can be tested for strong mixing properties through a score test:

$$\eta_\mu = T - 2\sum_{t=1}^{T} \frac{S^2}{S^2(l)} \quad \text{where } S_t = \sum_{i=1}^{t} \varepsilon_i \qquad (3)$$

T is the size of the sample, $S^2(l)$ is an estimate of the disturbance variance, and S_t is the partial sum of the residuals ε_i.

See: Kwiatkowski, Phillips, Schmidt and Shin (1992).

Two versions of this test:

Using a base equation consisting only of an intercept and the stationary errors, the first test, η_τ tests for the strong mixing that is characteristic of stationary series against the alternative of either fractionally integrated or unit-root processes.

Adding a linear trend to the intercept and the errors gives the second test, η_μ which tests the null same null and alternative hypotheses.

Use various “lag truncation parameters” depending on the length of series – usually 4 or 8.

Use η_τ unless you have reason to believe a trend exists.

Variance Ratio Test

Another test for unit root processes but has the advantage of a null hypothesis of a random walk with drift against an alternative hypothesis of pure fractional noise. See Diebold 1989.

Pure fractional noise is formally described as: $(1-L)^d x_t = \varepsilon_t$

Remembering that the random walk never forgets a random shock, this test relies on the assumption that the variance at time *k* should be *k* times the variance in period 1.

The test statistic is:

$$R(k) = \frac{k\hat{\sigma}_1^2}{\hat{\sigma}_k^2}, k = 2,3,...K \quad \text{where} \quad \hat{\sigma}_k^2 = \frac{1}{T-k+1}\sum_{t=k}^{T}(x_t - x_{t-k} - k\hat{\mu})^2 \qquad (4)$$

for $k = 1,2,...K$ and $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T}(x_t - x_{t-1})$.

k is the differencing interval.

Studies that employ multiple unit root tests often find contradictory results – that is, one test says unit root and another says stationary.

For example: Box-Steffensmeier and Smith 1996; Lebo, Walker and Clarke 2000.

This indicates at least two possibilities:

1st the tests have low power.

2nd the all or nothing distinction may not be sufficient.

Using multiple tests allows us to weigh the evidence – may be a lot of evidence one way or another or enough contradictory evidence to consider another alternative.

Cointegration and Error Correction

If we find that our series are nonstationary we difference them in order to overcome the threats to inference that plague nonstationary data.

One effect of this is that we lose the ability to study long term relationships among the variables of interest.

Following Engle and Granger (1987 *Econometrica*) political scientists began in the early 1990s to study long-run relationships using the concepts of cointegration and error correction.

Two non-stationary series are cointegrated when there exists a stationary linear combination of the variables.

That is, if X_t and Y_t are both I(1) but we can create a series that is I(0) by combining them somehow.

This new combination of variables will not require differencing.

In such a case we can look at both the short-term and long-term relationships between the variables using an error correction mechanism (ECM).

Cointegration among three of more nonstationary variables is defined the same way, although it is possible that the variables will form "multiple cointegrating variables."

Cointegration cannot be assumed – it is an empirical question.

Just having two series that are I(1) does not necessarily mean that there will be cointegration between them.

Cointegration indicates an especially close relationship – the linear combination of the variables is mean stationary. That is, if may move up or down but it always returns to some equilibrium level.

There are many ways to test for cointegration but lets looks at three, one proposed by Engle and Granger (1987), the second by Johansen (1991), and the third by Pesaran, Shin, and Smith (2001).

Engle and Granger Representation Theorem

This is called a 2-step procedure, though it has a few more than 2 steps.

First, use tests to establish that the series of interest are nonstationary.

That is, to be candidates for cointegration all the series must be I(1).

Then create a linear combination of the variables by regressing the level-form (undifferenced) Y_t on level-form X_t – this is called a cointegrating regression.

$$Y_t = a + bX_t \qquad (5)$$

From this we will get $\hat{a}, \hat{b}$ such that $y_t - \hat{a} - \hat{b}x_t = \hat{\varepsilon}_t$

The residuals, $\hat{\varepsilon}_t$, are by definition a linear combination of y_t and x_t.

Next, we want to test if these residuals are stationary.

We perform a Dickey-Fuller test (and other tests of stationarity) on the residuals $\hat{\varepsilon}_t$.

If the residuals are stationary we can conclude that y_t and x_t are cointegrated and we will model them in error correction form.

To do this, we use differenced y_t and x_t and the residuals from the cointegrating regression, lagged back one period: $y_{t-1} - \hat{a} - \hat{b}x_{t-1} = \hat{\varepsilon}_{t-1}$. ← This is the ECM.

$$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + \beta_2 (y_{t-1} - \hat{a} - \hat{b}x_{t-1}) + \varepsilon_t^* \qquad (6)$$

We could rewrite that as:

$$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + \beta_2 \hat{\varepsilon}_{co(t-1)} + \hat{\varepsilon}_t^* \quad \text{or as: } \Delta y_t = \beta_0 + \beta_1 \Delta x_t + \beta_2 ECM_{(t-1)} + \hat{\varepsilon}_t^* \qquad (7)$$

We are looking at the effect of the combination of y_t and x_t in the previous period on y_t in the present period.

$\widehat{\beta}_2$ should vary between 0 and -1.

If y_t and x_t do not cointegrate $\widehat{\beta}_2$ will not be significant.

As y_t and x_t move together through time, the difference between them in the previous period will erode as they come back to equilibrium.

For example, an estimate of $\widehat{\beta}_2$ = -0.6 indicates that if a shock at *t* moves the two variables apart, 60% of the distance between them will be gone by *t+1* and 84% of it at *t+2* and 96% at *t+3* and so on.

The absolute value of $\widehat{\beta}_2$ calibrates the speed with which shocks to the system are reequilibriated by the cointegrating relationship between y_t and x_t.

That is, the absolute value of the ECM value tells us what % of a shock will erode in every subsequent month.

The ECM can be called a “negative feedback term.”

If short-term shocks move one of two variables away from long term trend, the negative feedback term brings it back into line.

This is a restricted version of what we can do in Box Jenkins.

Says every shock dies down at the same rate rather than having several delta parameters in a Box Jenkins model.

An ECM coefficient close to -1 indicates a very close relationship.

Rearranging terms shows that models such as (6) above are variants of the more familiar autoregressive distributed lag form.

We can rewrite (6) as:

$$y_t = \beta_0 + (1-\beta_2)y_{t-1} + \beta_1 x_t + (\beta_2 \hat{b} - \beta_1)x_{t-1} + \hat{\varepsilon}_t^* \qquad (8)$$

This is a type of distributed lag model.

Important points for analyses and understanding:

1. If the two variables cointegrate they should have a high R^2 in the cointegrating regression and a very high *t* statistic – of course, because the two series are non-stationary the *t* statistic should be suspicious to us.

2. The sign on the cointegrating regressors should be intuitively plausible.

3. The residuals should be stationary.

When doing a cointegration test we need a new set of critical values, different from regular Dickey Fuller values.

"Mackinnon Values" about -3.44 at the .05 level.

4. ECM `resids{1}` should be negative and between 0 and -1 and statistically significant.

Error correction models are an important part of economics but haven't been used too extensively in political science.

ECMs are attractive because they address the threat of spurious regressions while enabling the analyst to study long-term relationships among a set of variables.

If there is more than one independent variable involved in the cointegrating relationship we may have multiple cointegrating vectors.

This two-step procedure can be followed by a "third step."

Engle and Yoo's third step deals with getting better estimates of $\hat{a}$ and $\hat{b}$ by ensuring that standard errors are safe – considering they came from a regression between two non-stationary variables.

Take the absolute value of the ECM and multiply each of the regressor variables in the error correction model by the value of |ECM|.

Saved residuals of Error Correction Model are regressed on new values of independent variables.

Standard errors are the correct standard errors for the ECM.

The only one of interest is the one for ECM.

Johansen Approach

This is a one-step procedure.

Instead of including the errors of our cointegrating regression in a second step, we merge the two steps.

In the two step procedure we used $y_{t-1} - \hat{a} - \hat{b}x_{t-1}$ as the ECM.

In the one step we simply include y_{t-1} and x_{t-1}.

The coefficient on y_{t-1} is our error correction term.

These results should be pretty similar to our two-step estimates except for the constant.

Since the two step ECM includes the constant of the cointegrating regression and the one step doesn't this difference is moved to the constant in the one-step equation.

What if we mistakenly estimate a one-step error correction model and the two variables are not cointegrated?

If both *x* and *y* are integrated or long memoried then the OLS residuals will signal this lack of cointegration by themselves being integrated or long memoried.

That is, if they are cointegrated, the integration should get wiped out by the ECM in the OLS regression.

If – because the variables aren't cointegrated – it doesn't get wiped out, it will appear in the residuals.

Lack of cointegration may also be indicated by any of the parameters in the ECM portion of the equation being estimated as zero.

Pesaran, Shin, and Smith's Bounds Method

A one-step bounds method called the Autoregressive Distributed Lag (ARDL) approach.

A new approach for political scientists – I may be the only one who has used it.
PSS test has three major advantages:

First, it's a one-step method – often preferred (Beck 1992; DeBoef 2001).

Second, the series of interest can be of any level of integration so that stationarity pre-testing can be avoided.

This flexibility makes the test adaptable to fractionally integrated data and in that respect is superior to the Johansen procedure.

Third, the ARDL procedure is powerful for small sample sizes, especially in comparison to Johansen's (PSS 2001, 311).

Begin by estimating the unrestricted error correction model (UECM):

$$\Delta Y_t = \beta_0 + \sum_{i=1}^{k} \lambda_1 \Delta Y_{t-k} + \sum_{i=0}^{k} \lambda_2 \Delta X_{t-k} + \lambda_3 Y_{t-1} + \lambda_4 X_{t-1} + e_t \qquad (9)$$

Where ΔY and ΔX are the differenced versions of *Y* and *X*, $\lambda_1 - \lambda_4$ are vectors of parameters, and *k* is the lag length we choose using the Akaike Information Criterion, Bayesian Information Criterion, and the Residual Mean Squared.

Long-run relationships are tested for using two separate statistics.

First, a Wald test (*F*-statistic) is performed on the joint null hypothesis that the level variables' coefficients are jointly equal to zero, i.e. that $\lambda_3 = \lambda_4 = 0$.

The test statistic bounds can be found on page 300 of PSS.

The second step follows Banerjee, *et al.* (1998), and tests the (non-standard) *t* value of the lagged dependent variable to assess error correction in the cointegrated series (PSS 2001, 303).

We can also pay attention to causal ordering.

Switching the dependent and independent variables will give us a different story.

It may be that one series is "chasing another – the error correction may be in one direction, or both.

Summary of Error Correction Models

Error correction models are attractive because they address the threat of spurious regressions while simultaneously enabling the analyst to study long-term relationships among a set of variables.

However, theses models are not a panacea.

As in traditional time series regression models – including those with lagged endogenous variables – the effects of all of the X variables are specified such that they have a common dynamic captured by the ECM coefficient.

This restriction may be theoretically implausible.

Also, we need to address questions regarding the exogeneity of the X's to have confidence in our parameter estimates.

This is not a problem unique to error correction models but they are naturally worth exploring when we look at cointegrating relationships.

That is, the cointegrating relationship is such a close one, we'd like to be able to conclude which variable is causing which.

Exogeneity

With rare exceptions, the time series models estimated by political scientists are single-equation specifications.

This implies the theory that causality is unidirectional.

The concept of exogeneity is crucial for evaluating inferences based on these models.

By weak exogeneity we mean that $x \rightarrow y$.

y may $\rightarrow x$ but only with a lag.

If not even with a lag, we can say that x is strongly exogenous to y.

From Beck, 1991:

Think of the Error Correction Model as a system of equations with 2 equations.

Since we say that x and y are in an equilibrium relationship, we might assume that this means the following:

When x diverges from y, the disequilibria are reequilibriated by adjustments of y **and** x.

That is, disruptions in the equilibrium are met by adjustments from both y and x.

Or, put another way, x may respond to changes in y.

This system looks like this:

$$\Delta y_t = \sum_{j=1}^{k} \beta_j \Delta y_{t-j} + \sum_{j=1}^{k} \gamma_j \Delta x_{t-j} + \rho_y (y_{t-1} - \omega x_{t-1}) + C_y + \varepsilon_t \qquad (10a)$$

$$\Delta x_t = \sum_{j=1}^{k} \theta_j \Delta x_{t-j} + \sum_{j=1}^{k} \kappa_j \Delta y_{t-j} + \rho_x (y_{t-1} - \omega x_{t-1}) + C_x + v_t \qquad (10b)$$

The two constant terms C_x and C_y allow *x* and *y* to drift.

If *x* is weakly exogenous to *y* and *y* is endogenous, then all of the disequilibria are reequilibriated by adjustments of *y*, not *x*.

This implies that $\rho_x = 0$ in Equation 10b. So there is no need to estimate equation 10b.

We need only estimate 10a.

Thus, the test that $\rho_x = 0$ is one test of weak exogeneity.

Tests for weak exogeneity are crucial for establishing the credibility of single equation models such as the error correction models above.

Charemza and Deadman (1997) have a three stage test for testing weak exogeneity.

Use this for a single right-hand-side variable.

Again thinking of *y* and *x*.

First, specify a model for *x*.

Second, add the error correction mechanism to this model.

If *x* is weakly exogenous to *y* the error correction mechanism should be insignificant in the model for *x*.

Third, the residuals of the model for *x*, (absent the ECM) should be insignificant when added as a predictor in the error correction model for *y*.

This is a generalization of the Houseman-Woo Exogeneity test.

This is a powerful tool – it tells us which variable drives which.

With this accomplished we can be sure that we have correctly identified the dependent variable and the independent variable.

If there is more than one *x* in the model we need to do this test for each of the *x*’s.

If we find that *x* is strongly exogenous to *y* and that this is true no matter what other independent variables are in the model, we call this **super exogeneity**.

Near Integrated

Let’s get away from the “knife-edged” decision of whether a series is I(0) or I(1).

DeBoef and Granato say that political scientists’ attempt to characterize a series into one of these two is arbitrary.

There are two types of long memories series.

These series are in some senses technically stationary but they exhibit long memory.

That is, as opposed to a short-memoried ARMA series or a perfect memoried series that is I(1).

For the near integrated case:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t$$

β_1 may be as low as 0.8 and we will have huge problems of spurious regressions.

Thus, a near integrated series can be dangerous to analyze in level form.

DeBoef and Granato explain this as:

$$y_t = \rho y_{t-1} + \mu_t$$

Where $\rho = 1$ the series is integrated.

Where $|\rho| < 1$ the series is stationary.

$|\rho| = 1 + \varepsilon$ is the near integrated case where ε is small and negative.

What can we say about a series like this?

Asymptotically, it is stationary. $|\rho| = 1 + \varepsilon$

It may remember a great deal of previous period, but not all of it so that over a sufficient period, shocks fade away.

It may look like a random walk but shocks don't persist and it will eventually return to its mean level.

This is especially true for a series that is bounded.

The ACF of a near integrated series will show strong correlations at long lags.

These correlations should die down at a geometric rate.

For example: How can we say that a series such as presidential approval **isn't** mean and variance stationary when we know that presidential approval can never be lower than 0 or above 100.

D & G say that only when the series approaches an infinite length will it behave as a stationary series.

In finite samples, it may be impossible to distinguish the near integrated series from the integrated series.

Figure 2 of D & G show two series that are very similar. The two series experience the same shocks but the integrated one remembers 100% of the previous period and the near-integrated remembers 95%.

The ADF test would lead us to conclude non-stationarity for both of them. Again, the DF tests have low power, especially in the case of near- or fractional- alternatives.

A near integrated series will have time dependent variance which is one of the sources of the spurious regression problem.

What should be done?

Two motivations here:

We don't live in Asymptopia. We will never have very long samples so deal with the data as you have them.

Also, this is reinforced by the pragmatic nature of time series analysis – we want to model our series in a way that makes good forecasts or tells us the relationships between variables without threats to inference. Again, deal with the data as you have them.

Differencing the variables removes the threats to inference.

There is a threat of "over-differencing."

D & G show that doing it for NI series builds autocorrelation into the differenced series and I have found the same thing for FI series.

Essentially, an over-differenced series has an artificial moving average parameter built into it.

What exactly qualifies as near-integration is unclear.

Depends on the length of the series.

Fractionally Integrated Series

Now lets look at relaxing the assumption that *d* in a (*p,d,q*) model can only hold integer values.

This generalizes the ARIMA model to the ARFIMA model – autoregressive, fractionally integrated, moving average.

Look again at the simplified model:

$$(1-L)^d y_t = \frac{(1-\theta L)}{(1-\phi L)}\varepsilon_t$$

y_t is our variable of interest.

L is the lag operator.

θ is for moving average parameters.

ϕ is for autoregressive parameters.

ε_t is the error term.

d is the differencing parameter – it tells us how many times y_t needs to be differenced to be rendered stationary.

When *d* = 0, the left hand side becomes y_t.

When *d*=1, the left hand side becomes differenced y_t.

These are just special cases of *d* which we now allow to hold any value – though for our purposes the range -0.5 < *d* < 1.5 covers all the ground we'll need.

When 0< *d* < 0.5 the series is mean reverting and has finite variance.

When 0.5 < *d* < 1.0 the series is mean reverting with infinite variance.

Fractionally integrated series have long, but not perfect memory.

They will exhibit a mix of traits from the stationary and non-stationary case.

We will see long significant correlations in the ACF that do not die down at a geometric rate.

Where do fractionally integrated series come from?

There are 3 possibilities but the one that is most important to political scientists is that a fractionally integrated series will arise when a series is created by aggregating units that have different autoregressive processes.

Granger (1980) examines independent series x_{jt} for *j*= 1,2,…*n* individuals.

Each has its own AR parameter α_j.

$$x_{j,t} = \alpha_j x_{j,t-1} + \varepsilon_t$$

Granger shows that this type of heterogeneity will create a series that is fractionally integrated.

For example if we create a series of public opinion and some of the people exhibit perfect memory and some have only short memory.

Aggregating these individuals will create a series that shows strong correlations as it remembers the perfect memoried, but this will not be the same as the unit root case because a portion of the aggregated series has only short memory.

In public opinion this can be expected based on Key's Switchers and Stand-Patters, or Converse's Black and White model.

Zaller talks about information being normally distributed so we can expect people to react differently to shocks.

This extends well beyond public opinion and I have found that most series political scientists use are aggregations of something.

Examples:

White House Staff growth over time. Staffing is in many areas (policies, campaigns, press, etc.) and staffing in each area follows its own autoregressive process.

This is true of budgeting and government growth in general – the growth of budgets follow different autoregressive patterns.

Lebo and Moore 2003 discusses action-reaction models of international politics.

How do countries treat each other. There are many ways that countries interact with each other – trade, military alliances or feuds, etc. – and these may have different autoregressive processes.

Supreme Court liberalism over time – 9 justices aggregated.

Its pretty hard to think of a series that political scientists use where this wouldn't be a possibility worth exploring.

So given this we should expect fractional levels of *d*.

The more memory we expect a series to have, the closer *d* should be to 1.

We can have short-run components here as well.

AR parameters and MA parameters.

In most cases, using *d* will solve most of the autocorrelation and (0,*d*,0) models are appropriate.

Testing for FI.

1st – do you have theoretical reasons for expecting heterogeneity at the individual level?

Then, use unit root tests before moving on to direct estimators for *d*.

NOTE: Personally, I have less and less faith with this step over time. As the literature becomes more full of fractional integration techniques, going straight to estimation of *d* – provided you are careful about it – may be the best course.

Use Dickey-Fuller, Augmented Dickey Fuller, Variance Ratio, KPSS, Phillips Perron or whatever tests you have available.

Do they all lead to the same conclusion?

DF/ADF and Variance Ratio test will tell you if you should reject the hypothesis that d=1.

KPSS (η_τ) will tell you if there is a trend.

KPSS (η_μ) will tell you if you can reject the hypothesis that d=0.

From these you may be rejecting both *d*=1 and *d*=0 at the same time → this will lead you to want to estimate *d*.

Using various lag lengths for ADF and VR, or lag truncation parameters for KPSS will give you a lot of results.

If they are all in agreement that *d* is an integer, you may not want to estimate *d*.

Otherwise, there are several estimators.

Sowell's (1992) Fully Informed Maximum Likelihood Estimator is applied to first differenced data.

In RATS this procedure is called ARF500.src

This will give an estimate of *d* for the differenced version of the variable.

For the *d* value of the original variable add 1 to this estimate.

Sowell's procedure in RATS allows simultaneous estimation of AR and MA parameters.

But, its unlikely to work for anything beyond the simplest *p* and *q* values.

Also, we are given the AIC and SBC.

We can use these and our parameter estimates to choose the most appropriate ARFIMA (*p*,*d*,*q*) model – just as we did for ARIMA modeling.

In many or most cases a (0,*d*,0) model will suffice.

Robinson's Gaussian Semiparametric Estimator is also applied to first differenced data.

In RATS this procedure is called RGSER.src. (The last R is for "revised" – beware the earlier version RGSE.src).

Again, for the *d* value of the original variable add 1 to this estimate.

Once we have an estimate of *d* for a series we can proceed to "fractional differencing."

Remember that in ARIMA modeling we applied the filter $(1-L)^d$ to make our series stationary.

If *d* = 0 this meant we left the series in level form.

If *d*=1 we differenced it once.

For ARFIMA modeling we are applying the same filter, $(1-L)^d$, but now we are allowing any real number for the value of *d*.

We'll now call $(1-L)^d$ the fractional differencing operator.

$$(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)L^k}{\Gamma(k+1)\Gamma(-d)}$$

Where $\Gamma(.)$ is the gamma function and *k* is the lag length.

Applying this filter takes a series integrated of order *d*, that is I(*d*) and differences it by exactly that amount, *d*.

Doing so creates a series that is stationary.

For this, we use the FIF.src procedure in RATS.

We use the differenced form of the variable (the same one we used to estimate *d* using either Sowell or Robinson's estimator) and we tell the procedure the degree to which we want the series differenced (the value given by Sowell's or Robinson's estimate).

This should be done for each variable in the model we are ultimately interested in estimating.

That is, if we collect data on $y_t, x_{1t}, x_{2t}, x_{3t}$ and we want to estimate a model of y_t.

We need to estimate d_y and apply the filter fractional integration filter to y_t: $(1-L)^{d_y} y_t$.

Then we need to estimate d_{X1t} and calculate: $(1-L)^{d_{x1}}$

And do the same for x_2 and x_3.

We can then go on to estimate:

$$F\Delta y_t = \beta_0 + \beta_1 F\Delta X_{1t} + \beta_2 F\Delta X_{2t} + \beta_3 F\Delta X_{3t}$$

Doing so eliminates threats to inference – spurious regression results, autocorrelation as evidenced by Durbin-Watson statistics, R^2 statistics biased upwards.

Fractional Cointegration

Studies have shown that fractional integration may be an extremely common phenomenon for political time series. (Box-Steffensmeier and Smith 1996; Lebo, Walker, and Clarke 2000; Byers, Davidson, and Peel 2000; Lebo and Moore 2003).

Also, remember that cointegration says in order to test if two series are cointegrated we take two unit root series and test the level of integration of their combination.

By that definition, if a series is FI, it is not a candidate for cointegration.

What we need to do is generalize the concept of cointegration and study the possibility of fractional cointegration.

Another way of stating the qualification for cointegration is:

If the combination of two variables is of a lower order of integration than are either of the original series, those series are cointegrated.

This definition works for fractional integration.

To test whether x_t and y_t are fractionally cointegrated:

Estimate d for each.

That is, what is d_{x_t} and d_{y_t} ?

Next, run a cointegrating regression of y_t on x_t (remember to keep them in level form).

From this cointegrating regression we get a set of residuals, our error correction mechanism.

Next, we need to test the level of integration for them.

We estimate d for the ECM, giving us: d_{ECM} .

If $d_{ECM} < d_{x_t}$ and $d_{ECM} < d_{y_t}$ we can conclude that the series are fractionally cointegrated.

That is, d_{ECM} need not be zero for us to find fractional cointegration.

Finding this means that the series are in a long-run equilibrium relationship.

If $d_{ECM} > 0$, this tells us that the ECM is itself a long memory process.

That is, disruptions to the long-run equilibrium readjusted but at a rate much slower than predicted by a traditional ECM.

We can then model $F\Delta y_t$ using $F\Delta x_t$, other variables, and the ECM.*

*However, there is a potential problem: if $d_{ECM} > 0$, the ECM is not level stationary and should *not* be used as an independent variable in a regression model.

Using a variable that is not I(0) leads to threats to inference.

So, we must take account of this and difference the ECM by its value of *d*.

That is, use $(1-L)^{d_{ECM}} ECM_{t-1}$ as an independent variable.

As with traditional cointegration, we can use the *t* statistic of the ECM as a further check for cointegration.

This is the Fractional Error Correction Mechanism (FECM). See Clarke and Lebo (2003) for an explanation.

Also, this can all be done in one step – but we cannot use an FECM in the one-step procedure.

So the process of when to use an error correction mechanism is as follows.

From: Lebo and Moore 2003.

Figure 1: When to use an (F)ECM