Lecture Notes 2: Testing for Unit Roots

Carl Bonham, August, 2005

Department of Economics, University of Hawaii
and University of Hawaii Economic Research Organization

1 Testing for Orders of Integration

Economic theory implies that certain variables should not diverge from each other by too great an extent in the long run. The variables may drift apart in the short run, but in the long run, economic forces such as arbitrage, the profit motive, and maximizing life-time utility will bring them together again. For instance, consumer theory tells us that consumption and permanent income should move closely together, covered interest parity tells us that interest rates on assets denominated in different currencies will move together when augmented by information about forward exchange rates. Thus, economic equilibria act as attractors and force different variables to move together in the LR even if they do not in the SR, and even if they are individually non-stationary. The econometric theory of cointegration describes how individually non-stationary processes may be tied together by a statistical equilibrium, and since most macroeconomic time series tend to be dominated by long run trend behavior, it is important to determine the source of this trending behavior.

Suppose we wish to test the hypothesis that y_t is generated by:

$$y_t = y_{t-1} + \epsilon_t, \qquad \epsilon_t \sim IN[0, \sigma^2] \tag{1}$$

We might want to try and estimate the following regression and test for $\rho = 1$.

$$y_t = \rho y_{t-1} + \epsilon_t, \tag{2}$$

If ϵ_t is white noise, then for $\rho = 1$, equation (2) represents a nonstationary random walk process. Of course if $|\rho| < 1$, Then we have a stationary first order autoregressive process.

The problem with using equation (2) and a t-test for $\rho = 1$ is that the distribution of the t-statistic under the null hypothesis of a random walk is non standard.[1] The source of the problem is that the variance of y_t grows with t . That is, $var(y_t) = t\sigma_\epsilon^2$, and $var(y_{t-1}\epsilon_t) = E[y_{t-1}^2]E[\epsilon_t^2] = t\sigma_\epsilon^4$. As a result, $\hat{p}$ converges to its true value at rate T (as opposed to the conventional rate $\sqrt{T}$) and the distribution for $T(\hat{\rho} - \rho)$ is non-normal. The t-stat for $\hat{\rho}$, $t_\rho = \hat{\rho}/SE_\rho = \hat{\rho}\frac{\sum y_{t-1}}{1/T\sum \epsilon_t} = T\hat{\rho}\frac{\sum y_{t-1}}{\sum \epsilon_t}$ has a Dickey Fuller distribution that is skewed to the left and looks more like a Chi-squared distribution. The solution is to work with a model which is stationary under the null hypothesis, and to use critical values for t-tests calculated by monte-carlo simulation.

Dickey and Fuller (1979) suggested what is known as a DF test. Estimate the following regression:

$$\Delta y_t = \delta y_{t-1} + \epsilon_t, \tag{3}$$

and test the null hypothesis $H_0 : \delta = 0$. Notice that by adding y_{t-1} to both sides of equation (3), you can show that $(1 + \delta) = \rho$. Therefore, testing the null $\delta = 0$ is equivalent to testing $\rho = 1$. Under the null hypothesis, we have a stationary model, i.e. $\Delta y_t = \epsilon_t$.

The DF test involves testing for the negativity of δ.

- The null is $H_0 : \delta = 0$, and
- the alternative hypothesis is $H_a : \delta < 0$.
- A rejection of the null then implies $\rho < 1$ and hence $y_t \sim I(0)$.
- If we are unable to reject the null, then we conclude that $y_t \sim I(d)$, where $d \geq 1$.

[1] Recall that Granger and Newbold illustrated that the t-statistic produced excessive rejections of the null of no relationship when regressing a single I(1) series on an unrelated I(1) series.

In other words, finding that $\delta = 0$ could occur when a series is integrated of order greater than 1. For instance, prices (and other nominal series) are sometimes found to be I(2) so that the rate of inflation is I(1). So we cannot really stop with the first set of unit root tests. If $\delta = 0$ cannot be rejected, you should difference your data and test again. If $y_t \sim I(1)$, then $\Delta y_t \sim I(0)$, so we can test for whether $y_t \sim I(2)$ by estimating the following regression:

$$\Delta^2 y_t = \delta \Delta y_{t-1} + \epsilon_t, \tag{4}$$

and testing the null that $\delta = 0$.

If the null is rejected in favor of the alternative, $\delta < 0$, then the series Δy_t is I(0), and $y_t \sim I(1)$. If we cannot reject the null hypothesis, then we may go on and conduct another DF test using $\Delta^3 y_t$ to determine that $y_t \sim I(2)$.

Overdifferencing If a series is already stationary, then forming Δy_t for the dependent variable is overdifferencing. This will tend to produce large positive values of the DF test statistic.

1.1 What Test Statistic

Although, under the null, the model in equation (3) is stationary, the t-statistic still does not have a standard distribution. Rather it possesses a DF distribution and is sometimes referred to as a DF-t-test. The reason for this result is that, under the null, equation (3) represents a regression of an I(0) variable on an I(1) variable. The distribution of the DF-statistic is skewed to the left and has a mean which is significantly negative. More importantly, the distribution of the test depends on the parameters estimated. That is, is the model estimated with a drift or higher order deterministic terms?

PcGive and many other stat packages will automatically produce critical values for DF tests based on the simulations of MacKinnon (1991). For example, the subroutine (URADF.src in rats will conduct DF tests and report critical values). I have included tables from DF later.

jump to givewin or RATS to demonstrate simple unit root tests

1.2 Deterministic elements

So far, we have only illustrated testing for unit roots in isolation. But you probably have noticed that PcGive allows for the possibility of constants, trends, and even seasonal dummies in the unit root tests. This is because it is common to be interested in whether a series has not only a unit root, but also whether it exhibits nozero drift, or deterministic behavior as well. For example, we may want to test for whether α or β are significantly different from zero in:

$$\Delta y_t = \alpha + \beta t + \delta y_{t-1} + \epsilon_t. \tag{5}$$

Equation (5) can be used to test the joint null that y_t contains a stochastic trend, but no deterministic trend, i.e. $H_0 : \delta = 0, \beta = 0$. To do this we need to use an F-test. We can compare the results of the F-test with the critical values reported in Dickey and Fuller's (1981, Table VI, p. 1063) paper.

1.3 Augmented DF Tests

As with any modeling exercise, the ability to conduct valid inference requires that the residuals in our model not be serially correlated. Thus Dickey and Fuller suggested augmenting equation (5) with lagged values of the dependent variable Δy_t to mop up any possible serial correlation. The ADF test is then conducted by estimating the following regression:

$$\Delta y_t = \alpha + \beta t + \delta y_{t-1} + \sum_{i=1}^{k} \gamma_i \Delta y_{t-i} + \epsilon_t, \tag{6}$$

and testing the null hypothesis of a unit root, $H_0 : \delta = 0$ against the alternative hypothesis of trend stationarity. The critical values for such tests are reported by PcGive and are available in most graduate metrics texts: Hamilton, Greene ....

1.3.1 ADF tests in practice

In practice, choosing the lag length for the lagged dependent variable can be crucial. PcGive is designed to implement a search procedure suggested by Robert Hall.

- You begin with a relatively large value for the lag length, k (such as 1.5 years of lags).
- Check to see whether the last lagged dependent variable is significant at the 5 or 10 percent level. Also check the residuals from (6) for serial correlation. If $\hat{\gamma}_k$ is significantly different from zero, and the errors appear iid, then **you stop**.
- If the last lag is insignificantly different from zero, and the errors appear iid, reduce the lag length by one and reestimate the model.
- The lag length is reduced by one until the last lag is significant and the residuals are serially uncorrelated.

Unfortunately, PcGive does not automatically produce F-tests for serial correlation of the residuals from equation (6). However, PcGive does provide other useful information in the form of information criterion. PcGive unit root tests are discussed in Chapter 16 (http://www.pcgive.com/index.html?content=/volume1.html#unitroot). Also, I have written a simple algebra file to force PcGive to follow a Hall type procedure. Check out unitroot.alg in http://www2.hawaii.edu/ bonham/664/materials/.

Before moving on to discuss seasonal unit root tests, lets spend a few moments discussing the role of the deterministic factors in the ADF regressions. It turns out that these elements can be crucial to the success or failure of tests for unit roots. The main issues surrounding the deterministic components in (6) concern whether or not those components are both in the equation estimated as well as the DGP. Unfortunately, the DGP is unknowable, so there is no clear cut roadmap that leads to correct results. However a number of rules can be suggested as a

Theoretical Tool kit for conducting unit root tests. **(see Campbell and Perron (1991))**

1. When the deterministic regressors include all the deterministic components of the DGP, the DF statistic (t_δ) and its critical values vary with the set of included deterministic regressors;

2. The true size of the test (as compared to the critical value (also known as the *nominal size* of the test) for t_δ increases as you increase the number of included deterministic regressors. In other words, the probability of a *Type-I error (the probability of an incorrect rejection of a true null hypothesis)* increases.

3. If you omit a deterministic component that is growing at a rate at least as fast as the included deterministic regressors, then the *power* of the DF-test t_δ goes to zero as the sample size grows. Recall that the *power of a test is equal to the probability of rejecting a false hypothesis.*

4. If you include all of the deterministic components that exist in the DGP, the power of a test of the unit root hypothesis against the TS alternative decreases as additional deterministic regressors are added.

5. The power of ADF tests depends very little on the number of observations per se but is rather influenced by the span of the data. For a given span, additional observations obtained using data sampled more frequently lead only to a marginal increase in power.

6. A non rejection of the unit root hypothesis may be due to misspecification of the deterministic components included as regressors.

Practical Matters Charemza and Deadman (1997, 114) suggest that the problem of including in equation (6) a deterministic regressor that does not exist in the DGP is the more serious problem because researchers pick the size of their tests and expect it to remain constant! However, the most serious problem with unit root tests is their power against alternative hypotheses. Unit root tests will too often indicate that a series contains a unit root (especially when the true root is close to 1). Thus, it is often quite difficult to distinguish between the null of a unit root and the alternative TS process. To deal with these problems, I generally follow a fairly complicated sequential approach to testing for unit roots.

1. begin by graphing your data!

2. begin with the most general model (usually includes a constant and trend, in some cases even a polynomial in time) and choose the lag length, k, sequentially as described above.

3. having chosen the number of lags of the dependent variable to include (to ensure iid residuals), test the unit root hypothesis $H_0 : \delta = 0$ using the critical values for t_δ in DF or table 1.

4. If the unit root hypothesis is rejected, **STOP**.

5. Otherwise, test the hypothesis that the time trend does not belong, $H_0 : \beta = 0$ using the critical values for the $\tau_{\beta\tau}$ test in Table III of Dickey and Fuller (1981, p. 1062) (see Table 1 below).

6. If the trend is significantly different from zero, STOP (you already know the results of your unit root test). If the trend is insignificantly different from zero, **remove it** from the model and reestimate.

7. Test the unit root hypothesis $H_0 : \delta = 0$ using the critical values for t_δ in DF or table 1.

8. If the unit root hypothesis is rejected, STOP (recall the low power of the test)

9. Otherwise, test the hypothesis that the constant does not belong, $H_0 : \alpha = 0$ using the critical values for the $\tau_{\alpha\mu}$ test in Table I of Dickey and Fuller (1981, p. 1062) (or see Table 1 below).

10. If the intercept is significantly different from zero, **STOP** (you already know the results of the unit root test). If the intercept is insignificantly different from zero, remove it from the model.

11. Test the unit root hypothesis $H_0 : \delta = 0$ using the critical values for t_δ in DF or table 1.

12. When you are uncertain of your results, also check the correlogram of the series in question, its first differences etc.

Table 1: Critical values for ADF test [a]

		Fractiles								
		t_α			t_β			t_δ		
Model	T	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
Trend, Drift	25	4.05	3.20	2.77	4.05	3.20	2.77	-4.38	-3.6	-3.24
	50	3.87	3.14	2.75	3.60	2.81	2.38	-4.15	-3.5	-3.18
	100	3.78	3.11	2.73	3.53	2.79	2.38	-4.04	-3.45	-3.15
	250	3.74	3.09	2.73	3.49	2.79	2.38	-3.99	-3.43	-3.13
Drift but no Trend	25	3.41	2.61	2.20				-3.75	-3.00	-2.63
	50	3.28	2.56	2.18				-3.58	-2.93	-2.60
	100	3.22	2.54	2.17				-3.51	-2.89	-2.58
	250	3.18	2.52	2.16				-3.46	-2.88	-2.57

[a] From Dickey and Fuller (1981, p. 1062)

1.4 Seasonal Unit Roots

Seasonal unit roots and testing for seasonal integration is discussed in Charemza and Deadman (1997, 105-9). Bonham and Gangnes (1995) (hereafter BG) applied the S. Hyllberg and Yoo (1990) procedure to monthly data, but here we will focus only on quarterly series. The main advantage of seasonal unit root tests is where you need to make use of data that cannot be seasonally adjusted (say for an intervention study such as BG), or even as a pretest before seasonal adjustment. The reason BG(95) became interested in seasonal integration is that the seasonality in some tourism statistics is so strong that it masked the zero frequency unit roots. In fact, if a series has seasonal unit roots, then standard ADF test statistic do not have the same distribution as for nonseasonal series. Furthermore, seasonally adjusting series which contain seasonal unit roots can alias the seasonal roots to the zero frequency, so there are a number of reasons why economists are interested in seasonal unit roots.

The notation used in Charemza and Deadman is slightly different from that used in Bonham and Gangnes (1995). For convenience I will repeat the notation here from Charemza and Deadman (1997, 107-9). The HEGY test for seasonal integration is conducted by estimating the following regression:

$$\Delta^4 y_t = \alpha + \beta t + \sum_{j=2}^{4} b_j Q_{jt} + \sum_{i=1}^{4} \pi_i Y_{it-1} + \sum_{l=1}^{k} \gamma_l \Delta^4 y_{t-l} + \epsilon_t, \tag{7}$$

where Q_{jt} is a seasonal dummy, and the Y_{it} are given below.

$$\begin{aligned}
Y_{1t} = &\quad (1+L)(1+L^2)y_t &&= y_t + y_{t-1} + y_{t-2} + y_{t-3};\\
Y_{2t} = &\quad -(1-L)(1+L^2)y_t &&= -y_t + y_{t-1} - y_{t-2} + y_{t-3};\\
Y_{3t} = &\quad -(1-L)(1+L)y_t &&= -y_t + y_{t-2};
\end{aligned} \tag{8}$$

$$Y_{4t} = \quad -(L)(1-L)(1+L)y_t \quad = Y_{3t-1} = -y_{t-1} + y_{t-3}.$$

Equation (7) is estimated by OLS and tests are conducted for $\pi_1 = 0$, for $\pi_2 = 0$ and a joint test of the hypothesis $\pi_3 = \pi_4 = 0$. The HEGY test is really a joint test for LR (or zero frequency) unit roots and seasonal unit roots. If none of the π_i are equal to zero, then the series is stationary (both at seasonal and nonseasonal frequencies). Interpretation of the different π_i is as follows:

1. If $\pi_1 < 0$, then there is no long-run (nonseasonal) unit root. π_1 is on $Y_{1t} = S(B)y_t$ which has had all of the seasonal roots removed.
2. If $\pi_2 < 0$, then there is no *semi-annual* unit root.
3. If π_3 and $\pi_4 < 0$, then there is no unit root in the annual cycle.

Just as in the ADF tests, it is important to ensure that the residuals from estimating (7) are white noise. Thus, in testing for seasonal unit roots, it is important to follow the sequential procedures detailed above. Again, begin by testing for the appropriate lag length for the dependent variable (to ensure serially uncorrelated residuals), and then test whether deterministic components belong in the model. Table 2 below provides critical values from HEGY for conducting seasonal unit root tests.

Table 2:
Critical values for seasonal unit roots[a]

		Fractiles								
			't'π_1			't'π_2			'F':$\pi_3 n \pi_4$	
Model	T	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
No Inter	48	-2.72	-1.95	-1.59	-2.67	-1.95	-1.60	5.02	3.26	2.45
No S dum.	100	-2.60	-1.97	-1.61	-2.61	-1.92	-1.57	4.89	3.12	2.39
No trend	200	-2.62	-1.94	-1.62	-2.60	-1.95	-1.61	4.76	3.12	2.37
Inter.	48	-3.66	-2.96	-2.62	-2.68	-1.95	-1.60	4.78	3.04	2.32
No S. dum.	100	-3.47	-2.88	-2.58	-2.68	-1.95	-1.60	4.77	3.08	2.35
No Trend	200	-3.48	-2.87	-2.57	-2.58	-1.92	-1.59	4.76	3.12	2.37
Inter.	48	-3.77	-3.08	-2.72	-3.75	-3.04	-2.69	9.22	6.60	5.50
S. dum.	100	-3.55	-2.95	-2.63	-3.60	-2.94	-2.63	8.74	6.57	5.56
No. Trend	200	-3.51	-2.91	-2.59	-3.50	-2.89	-2.60	8.93	6.61	5.56
Inter.	48	-4.23	-3.56	-3.21	-2.65	-1.91	-1.57	4.64	2.95	2.23
No. S. dum.	100	-4.07	-3.47	-3.16	-2.58	-1.94	-1.60	4.70	2.98	2.31
No. Trend	200	-4.05	-3.44	-3.15	-2.59	-1.95	-1.62	4.66	3.07	2.34
Inter.	48	-4.46	-3.71	-3.37	-3.80	-3.08	-2.73	9.27	6.55	5.37
S. dum.	100	-4.09	-3.71	-3.37	-3.80	-3.08	-2.73	8.79	6.60	5.52
Trend	200	-4.05	-3.49	-3.18	-3.52	-2.91	-2.60	8.96	6.57	5.56

[a]From Hylleberg et. al (1990 pp 226-227.)

1.5 Structural Change

Recall that the random walk model may be written in moving average form as

$$y_t = y_0 + \alpha t + \sum_{j=1}^{t} \epsilon_j \tag{9}$$

The constant term, α in the random walk with drift model represents a deterministic trend. If the same process is subject to a structural break at some time, say $t = \tau$, then we might model that break by using an intervention dummy. Such an intervention dummy could be a step or a pulse dummy, or some function of these two.

Step vs Pulse Dummys

A step dummy is equal to zero up to time $t = \tau$, and is equal to 1 for the remainder of the sample, $t = \tau, \tau + 1, \ldots T$. A pulse dummy is the first difference of a step dummy, i.e. it is zero up to $t = \tau$, equal to 1 at time $t = \tau$, and equal to zero again from $t = \tau + 1$ to T.

Usually pulse dummys are used to reflect temporary events and steps are used for permanent changes, but in an integrated process, a pulse is like a one time jump in the shocks ϵ_t, and hence is permanent. See Charemza and Deadman (1997, 115-18).

To see the differential effects of a pulse vs step dummy in both stationary and non-stationary models, consider the following stochastic process, with mean μ_1 over the period $t = 1, \ldots, \tau$ but mean $\mu_1 + \mu_2$ over the period $t = \tau + 1, \ldots, T$.

$$
\begin{aligned}
y_t - \mu_1 &= \phi(y_{t-1} - \mu_1) + \epsilon_t && \forall\, t \leq \tau \\
y_t - (\mu_1 + \mu_2) &= \phi(y_{t-1} - \{\mu_1 + \mu_2\}) + \epsilon_t && \forall\, t > \tau
\end{aligned} \tag{10}
$$

Using the step dummy,

$$
\begin{aligned}
D_t &= 0 \quad \text{if } t \leq \tau \\
&= 1 \quad \text{if } t > \tau,
\end{aligned}
$$

rewrite equation (10) as

$$y_t - (\mu_1 + \mu_2 D_t) = \phi(y_{t-1} - \{\mu_1 + \mu_2 D_{t-1}\}) + \epsilon_t \tag{11}$$

Notice that for $|\phi| < 1$, the mean μ_1 acts as an attractor for y_t for $t \leq \tau$, while the shifted mean $\mu_1 + \mu_2$ acts as an attractor for $t > \tau$. What about the nonstationary case? Let $\phi = 1$.

$$y_t = y_{t-1} + \mu_2(D_t - D_{t-1}) + \epsilon_t \tag{12}$$

When y_t is $I(1)$, μ_1 is not identified, but μ_2 (the shift in the mean) is. Now suppose you wanted to test for a unit root in y_t. Subtract y_{t-1} from both sides of (11).

$$\begin{aligned} y_t - y_{t-1} &= (\phi - 1)y_{t-1} + (1 - \phi)\mu_1 + \mu_2(D_t - \phi D_{t-1}) + \epsilon_t \\ \Delta y_t &= (\phi - 1)y_{t-1} + (1 - \phi)\{\mu_1 + \mu_2 D_{t-1}\} + \mu_2 \Delta D_t + \epsilon_t \\ \Delta y_t &= \rho y_{t-1} - \rho\{\mu_1 + \mu_2 D_{t-1}\} + \mu_2 \Delta D_t + \epsilon_t \qquad (13) \end{aligned}$$

So if we wanted to include a step dummy to allow for a mean shift while testing the null of a unit root, we would need to include both the lagged dummy variable D_{t-1} and the pulse dummy ΔD_t.

The same type of analysis can be used to examine the effect of a change in deterministic trend, or a combination of mean shift and shift in trend. It can also be extended to multiple breaks. Consider a stochastic process that experiences a shift in trend at time τ. Using the same dummy variable as above,

$$\begin{aligned} y_t &- (\delta_1 + \delta_2 D_t)t = \phi\{y_{t-1} - (\delta_1 + \delta_2 D_{t-1})(t-1)\} + \epsilon_t && (14) \\ &\text{or subtracting } y_{t-1} \text{ from both sides} \\ \Delta y_t &= (\phi - 1)y_{t-1} + (\delta_1 + \delta_2 D_t)t - \phi\{(\delta_1 + \delta_2 D_{t-1})\}(t-1) + \epsilon_t \\ \Delta y_t &= \rho y_{t-1} + \phi(\delta_1 + \delta_2 D_{t-1}) - [\delta_2 \Delta D_t - \rho(\delta_1 + \delta_2 D_{t-1})]t + \epsilon_t \\ & && (15) \end{aligned}$$

So once again, we find that we need both a pulse dummy and a shift dummy to capture the intended effects.

1.5.1 Additive Outlier Test

The problem with structural breaks for unit root tests is that series which have a structural change in their mean will behave as if they are I(1) even if they are stationary around a broken mean. In other words, a series that is stationary but undergoes a single permanent mean shift will appear to unit root tests as if it were a random walk.

A number of tests have been proposed to deal with the effects of "known" structural breaks. These tests effectively estimate the impact of the structural break and then remove it from the series before testing for a unit root. The only problem with such tests is that they require previous knowledge of when the break occurred. An alternative approach is to use rolling–recursive–unit root tests as for instance described in Benerjee et al. (1993).

The modified ADF test as described in Perron (1989) involves identifying the point of structural change (possibly by using recursive parameter estimates from DF tests, or recursive Chow tests), followed by a detrending of the original series using dummy variables to capture the change in either α, β, or both. Next the residuals from the "auxiliary" regression are then tested for unit roots using an ADF test.

To conduct an additive outlier test (i.e., assuming a one time change in the drift term α) on the series y_t, you first estimate the regression:

$$y_t = \alpha + \phi D_t + \epsilon_t, \tag{16}$$

where $D_t = 1$ if $t \geq \tau$, and zero otherwise. Next, the residuals from equation (16) are tested for a unit root by estimating:

$$\Delta\hat{\epsilon}_t = \omega P_t + \delta\hat{\epsilon}_{t-1} + \sum_{l=1}^{k} \gamma_l \Delta\hat{\epsilon}_{t-l} + v_t. \tag{17}$$

In (17), $P_t = \Delta D_t = 1$ if $t = \tau$, and zero otherwise. The non-standard critical values for this test of $H_0 : \delta = 0$ are provided in Charemza and Deadman (1997, 301-3) for the additive outlier test, or in Perron (1989) for more complicated variants. To use the tables, you also need to know the value of λ, the ratio of the subsample up to and including the break point to the entire sample.

EXAMPLE with yen/dollar exch rate

1.5.2 Recursive parameter constancy test

Given the need to identify break points before conducting Perron type ADF tests, it is useful to spend some time discussing recursive methods for testing parameter constancy. One method of testing parameter constancy is to reserve a portion of your sample for out of sample forecast analysis. PcGive makes it very simple to conduct Forecast Chow tests and a forecast χ^2 test. These tests are described in Hendry and Doornik (1996, pp. 46, and 222-223). The null hypothesis of the forecast χ^2 test is 'no structural change in any parameter between the sample and the forecast periods' (denoted s and f respectively), $H_0 : \beta_s = \beta_f, \sigma_s^2 = \sigma_f^2$. The chow test has the same null hypothesis as the χ^2 test, but the latter will always be larger than the chow test because it drops the asymptotically negligible term $(V[\hat{\beta}])$ in its denominator. Thus, the difference between the chow and chi-square forms indicates the relative increase in prediction uncertainty that occurs because we are estimating β (as opposed to knowing the true parameter).

In contrast to the forecast tests, recursive methods allow us to study the constancy of parameters within the sample period. Thus we can use parameter constancy tests to aid us in identifying structural breaks within the sample. There are a wide variety of test for parameter constancy. PcGive facilitates three tests: Recursive LS parameter estimates, one-step-residuals, and scaled recursive Chow test statistics.

Recursive LS parameters

In simple terms, recursive methods are used to efficiently re-estimate the parameters of a model each time a new observation becomes available. The efficiency arises because of updating formulae that do not require the inversion of the cross product matrix of the independent variables. Beginning with a small sub sample of $t = 1, 2, \ldots, n$, where $n \geq k$, and k is the number of estimated parameters in the model (not the same as the number of lagged dependent variables in (17)), recursive methods obtain OLS estimates for the k parameters of the model. Then the sample period is increased by one observation, (i.e. $t = 1, 2, \ldots, n+1$), and new OLS estimates are obtained. This process continues until the complete sample has been used, and the final recursive parameter estimates obtained using data through time T, are the same as the OLS estimates over the full sample. (See Hendry and Doornik (1996, pp. 178-79.)

By estimating a model recursively, and then choosing recursive graphics ($Alt + r$) command, you can graph the coefficients on any of the independent variables plus ± 2 standard errors . (See Hendry and Doornik(1996, pp 58-59, 148-49, and 232-33).) When a parameter estimate lies outside of its confidence interval from previous sub samples, then it indicates a potential structural break at that point.

One-step Residuals

At each recursion, a single residual can be calculated for the last time period in the sub sample, $\tilde{v_\tau} = y_\tau - x\prime_\tau \hat{\beta}_\tau$. Using the set of residuals from $t = 1, \ldots, \tau$, the standard error of the regression can be calculated. The graph of 1-step-residuals bordered by $0 \pm 2\hat{\sigma}_\tau^2$ over $\tau = n, \ldots, T$ are also used to identify structural breaks.

Suppose that your estimated model has relatively stable parameters up to a point in your sample where a structural break occurs. At that point, the parameters change abruptly. Then the residuals calculated for each sample end point should be of similar magnitude as the ones preceding them until the break point is reached. When the break point is reached, the residuals should appear abnormal, because the model (which assumes parameter constancy) will not fit the sample well due to the break. Thus, residuals that lie outside the 2-standard-error bands are indicative of parameter change at that point.

Scaled Recursive Chow Tests

There are three variants of a scaled Chow test available in PcGive: *1-Step Chow Tests*, *Break-point F-tests* ($N \downarrow$ Chow-Tests,) and *Forecast F-tests* ($N \uparrow$, Chow-Tests). Each of the tests is described in Hendry and Doornik (1996, pp. 232-33).

A 1-step Chow test compares the residual sums of squares (RSS_τ) for the model fitted up to and including period $\tau - 1$ and the residual sums of squares up to and including period τ (RSS_τ). The test statistic is given by

$$\frac{(RSS_\tau - RSS_{\tau-1})(\tau-k-1)}{RSS_{\tau-1}}$$

The statistic is calculated for each sample period ending in $\tau, \tau+1 \ldots, T$ and is graphed over time. Because the critical values for F-tests will change as the sample size changes, all of the Chow-tests are scaled by their critical values before they are plotted by PcGive. This means that the significant critical value in the graphs is a horizontal line at unity.

Breakpoint($N \downarrow$ Chow-Tests) compare the residual sums of squares from the *full sample* estimates to the residual sums of squares from a decreasing number of forecasts (t = T - n + 1 to 1 forecasts). A typical statistic is

$$\frac{(RSS_T - RSS_t)(t-k-1)}{RSS_{t-1}(T-t+1)},$$

where RSS_T is the full sample residual sums of squares and RSS_t is the forecast error sums of squares for t forecasts.

Forecast($N \uparrow$ Chow-Tests) compare the residual sums of squares from the initial sub sample estimation to the residual sums of squares from an increasing number of forecasts (from 1 to T-n+1). This tests the model estimated over the period 1 to n-1 against an alternative which allows any form of change over the period n to T.

Examples of the use of recursive parameter constancy tests are given in Charemza and Deadman (1997, 50-1), as well as Hendry and Doornik (2001, 65-67).

2 Panel Tests

Single equation unit root tests can be plagued by low power when confronted with the typical sample size available for research in international monetary economics/finance. For instance, Mark (2001, 39-40) conducts MonteCarlo experiments to calculate the power of DF tests. By generating data for an AR(1) model with $\rho = 0.96$ using a standard DF regression and t-test, Mark calculate the percentage of rejections of the unti root null. He concluded that the false null hypothesis is not consistently rejected until the sample size grows to as many as 1,000 observations. With a sample size of only 100 observations, the false null is rejected a little less than 10% of the time. In other words, the power of the test is close to its nominal size!

So how do you get 1,000 observations without waiting another 150 years? How about 100 observations on 100 countries? In the single equation ADF regression, $\hat{\rho}$ converges at rate T, but in a panel ADF regression, $\hat{\rho}$ converges at rate $T\sqrt{N}$. So a panel unit root test has the potential to significantly increase the power of unit root tests—even without 100 cross-section units. Here we will briefly discuss the Levin et al. (2002), and Im et al.

(2003) panel unit root tests and then cover Pesaran (2005) simple panel unit root test in the presence of cross-sectional dependance.

2.1 Levin and Lin

Let y_{it} be a balanced panel of $i = 1, \ldots N$ time series with $t = 1, \ldots T$ observations each. Suppose that $y_{i,t}$ is generated by the following DGP,

$$\Delta y_{i,t} = \alpha_i + \delta_i t + \theta_t + \beta y_{i,t-1} + \epsilon_{i,t}, \tag{18}$$

where $\epsilon_{i,t} = \sum_{j=0}^{\infty} \phi_{ij} \epsilon_{i,t-j} + \eta_{it}$. Note that if $\beta_i = 0$, and $\delta_i = 0$, then y_{it} is a random walk with drift, where each cross-sectional unit has its own drift term. Because the distribution of the panel DF tests depend on the maintained assumption that the individual errors η_{it} are *iid* both over time and cross-sectionally, it is common to include θ_t in the panel regression as a crude method of modeling cross-sectional dependence. Writing (18) in error-component form, note that θ_t is a common factor removed from each error process at each point in time.

Specifically, including the common time effect, θ_t, can be achieved by removing the cross-sectional mean from each y_{it}, i.e. rewriting equation (18) in deviations from group means $\tilde{y}_{it} = y_{it} - (1/N) \sum_{i=1}^{N} y_{it}$

The Levin and Lin (LL) test is for the null hypothesis

$$H_O : \beta_1 = \beta_2 = \cdots = \beta_N = \beta = 0,$$

against the alternative,

$$H_1 : \beta_1 = \beta_2 = \cdots = \beta_N = \beta < 0.$$

In other words, the LL test imposes homogeneity of the β_i across individual cross-sectional units under both the null and the alternative hypotheses.

2.1.1 LL in practice

The LL procedure is widely available as a subroutine for RATS, STATA, GAUSS etc.

To allow for heterogeneity of the parameters on the lagged differences (used to mop up the serial correlation), individual trends, and interecepts, these are removed from the individual series by first stage regressions.

$$\begin{aligned}
\Delta\tilde{y}_{it} &= \alpha_i + \delta_i t + \sum_{j=1}^{k_i} \phi_{ij} \Delta\tilde{y}_{i,t-j} + e_{it} && (19) \\
\tilde{y}_{it} &= \gamma_i + \omega_i t + \sum_{j=1}^{k_i} \mu_{ij} \Delta\tilde{y}_{i,t-j} + v_{it} && (20)
\end{aligned}$$

The residuals $\hat{e}_{it}$ and $\hat{v}_{it}$ from these first stage regressions are the differences and levels of y_{it} with the heterogeneous dynamics, trends, and individual fixed effects removed. Next, an additional auxilliary regression is used to obtain estimates of the long-run variance of $\Delta\tilde{y}_{it}$ used to standardize the residuals from equation (19 and 20) before performing the pooled UR tests. Specifically, estimate

$$\hat{e}_{it} = \delta_i \hat{v}_{it-1} + u_{it}. \qquad (21)$$

Form the variance $var(\hat{e}_{it})$, $\hat{\sigma}^2_{e_i} = \frac{1}{T-k_i-1}\sum_{t=k+2}^{T}\hat{u}^2_{it}$, and normalize the residuals as,

$$\begin{aligned}\tilde{e}_{it} &= \frac{\hat{e}_{it}}{\hat{\sigma}^2_{e_i}}, \\ \tilde{v}_{it} &= \frac{\hat{v}_{it}}{\hat{\sigma}^2_{e_i}}.\end{aligned} \tag{22}$$

Finally, the normalized residuals are then used to run the pooled ADF regression:

$$\tilde{e}_{it} = \beta\tilde{v}_{i,t-1} + w_{it} \tag{23}$$

The studentized coefficient,

$$\tau = \hat{\beta}\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{v}_{i,t-1}}{\hat{\sigma}^2_w} \tag{24}$$

is not asymptotically standard normal, and diverges as $NT \to \infty$. The LL test is a complicated adjustment to the standard t-test,

$$\tau_{ll} = \frac{\tau - NT \cdot S_N \tau \mu^*_T \hat{\sigma}_w \hat{\beta}^{-1}}{\sigma^*_T} \tag{25}$$

where $S_N = \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{\sigma}_{y_i}}{\hat{\sigma}_{e_i}}$, and $\hat{\sigma}_{y_i}$ is the long-run variance of $\Delta\tilde{y}_{it}$ estimated using the standard Newey and West (1987) formula,

$$\hat{\sigma}^2_{y_i} = \hat{\gamma}_{i0} + 2\sum_{j=1}^{\bar{k}}(1 - \frac{j}{\bar{k}+1})\hat{\gamma}_{ij}, \tag{26}$$

where $\hat{\gamma}_{ij} = \frac{1}{T-1}\sum_{t=2+j}^{T}\Delta\tilde{y}_{it}\Delta\tilde{y}_{it-j}$. The μ^*_T and σ^*_T are adjustment factors calculated by LL (see table 2.2 in Mark). In general, the LL test seems to perform well (see table 2.4 of

Mark). For a sample of $T = 100, N = 20$, the test rejects a false null 94% of the time at the 5% level when 80% of the cross-section units are stationary processes.

2.2 Im Pesaran and Shin

The LL test is messy. In contrast, the IPS test is a relatively straightforward combination of "independent" individual ADF test statistics. For each i, consider the following standard ADF type regression

$$\Delta y_{i,t} = \alpha_i + \delta_i t + \theta_t + \beta_i y_{i,t-1} + \sum_{j=1}^{p_i} \rho_{ij} \Delta y_{i,t-j} + \epsilon_{i,t}, \tag{27}$$

where the null hypothesis of unit roots is

$$H_O : \beta_i = 0 \quad \text{for all i,}$$

against the alternatives,

$$\begin{aligned} H_1 : \quad & \beta_i < 0, \quad i = 1, 2, \ldots, N_1, \\ & \beta_i = 0, \quad i = N_1 + 1, \ldots, N. \end{aligned}$$

The IPS test, therefore, allows for individual fixed effects, individual time trends, and allows β_i to differ over the cross section under the alternative hypothesis. In fact, the IPS test allows for some fraction (N_1/N) of the individual series to have unit roots.

The IPS test is based on a combination of the individual unit root tests for the N cross-sectional units. Let t_i, $i = 1, \ldots, N$ be the t-test for the null of a unit root in (27). For $E[t_i] = \mu$, and $var(t_i) = \sigma^2$, the IPS t-bar test is $\bar{t} = 1/N(\sum_{j=1}^{N} t_i)$, and $\sqrt{N}\frac{\bar{t}-\mu}{\sigma^2} \rightarrow N(0, 1)$. IPS compute μ and σ^2 by Monte Carlo simulations and report critical values for various values of N and T.

A potential problem with both the IPS and LL test is the requirement that the residuals in the individual ADF regressions be cross-sectionally independent, and the attempt to model this cross-sectional dependence with a single (i.e. pooled) common time dummy. In effect this practice assumes that there is no heterogeneity in the cross-sectional correlation of the errors. Suppose for example that we are interested in testing for the stationarity of the real exchange rate. There are many reasons to believe that shocks to real exchange rates might be more highly correlated among close trading partners than among distant trading partners. To the extent that the crude common time-dummy approach is indadequate and therefore leaves behind significant cross-sectional correlation in the errors, it is likely that the IPS and LL tests will over reject the null hypothesis of a unit root. Below we cover a very simple adjustment to the IPS t-bar test to remove the effects of cross-sectional correlation.

2.3 Cross-Sectionally Augmented Panel ADF tests

Let y_{it}, the observation on the ith cross-section unit at time t, be generated as

$$y_{it} = (1 - \phi)a_i + \phi_i y_{it-1} + u_{it} \qquad i = 1, \ldots, N; \quad t = 1, \ldots, T, \quad (28)$$

where initial value, $y_{i,0}$ are given, and the error term u_{it} has the one-factor structure with f_t the unobserved common effect.

$$u_{it} = \gamma_i f_t + \epsilon_{it}. \tag{29}$$

In (29), ϵ_{it} is the individual specific (idiosyncratic) error. Now rewrite (28) as an ADF type equation,

$$\Delta y_{it} = \alpha_i + \beta_i y_{it-1} + \gamma_i f_t + \epsilon_{it}, \tag{30}$$

where $\alpha_i = (1 - \phi_i)$, and $\beta_i = -(1 - \phi_i)$.

The unit root hypothesis is:

$$H_O : \beta_i = 0 \quad \text{for all i},$$

against the possibly heterogeneous alternatives,

$$\begin{aligned} H_1 : \quad & \beta_i < 0, \quad i = 1, 2, \ldots, N_1, \\ & \beta_i = 0, \quad i = N_1 + 1, \ldots, N. \end{aligned}$$

Pesaran makes the following assumptions:

- N_1/N, the fraction of individual $y_{i,t}$ that are $I(0)$ is non-zero and tends to the fixed value $0 < \delta < 1$ as $N \to \infty$.

- ϵ_{it}, $i = 1, \ldots, N$, $t = 1, \ldots, T$ are *iid* across both i & t, mean zero, and variance σ_i^2.

- The common factor f_t is serially uncorrelated with mean zero and constant variance σ_f^2

- ϵ_{it}, f_t, and γ_i are *iid* $\forall\ i$.

- $\bar{\gamma} = N^{-1} \sum_{i=1}^{N} \gamma_i \neq 0$.

Using the last assumption above, note that you can write the cross-sectional mean of y_{it} as

$$\bar{y}_t = \bar{y}_{t-1} + \bar{\gamma} f_t + 0, \tag{31}$$

or $f_t = (\bar{y}_t - \bar{y}_{t-1})/\bar{\gamma}$. In the simple case where u_{it} is serially uncorrelated, $\bar{y}_t$ and $\bar{y}_{t-1}$ or equivalently $\bar{y}_{t-1}$ and $\Delta\bar{y}_t$ are sufficient for filtering out the effects of the unobserved common factor f_t. So Pesaran bases his test of the unit root null on the t-ratio of the OLS estimate of b_i in the "Cross Sectional Augmented Dickey Fuller" (CADF) regression,

$$\Delta y_{it} = a_i + b_i y_{it-1} + c_i \bar{y}_{t-1} + d_i \Delta\bar{y}_t + e_{it}. \tag{32}$$

The exact null distribution of the t-ratio, $t_i(N, T)$ will depend on nuisance parameters (although not asymptotically), and Pesaran conducts simulations to derive the critical values. He finds that the distribution of the t-ratio in the CADF regression is skewed to the left even more than the standard DF distribution. The CADF distribution has a substantially negative mean and its standard deviation is less than unity.

2.3.1 CADF Panel Unit Root Tests

Pesaran considers the following cross-sectionally augmented version of the IPS test,

$$CIPS(N,T) = N^{-1}\sum_{i=1}^{T} t_i(N,T), \tag{33}$$

where $t_i(N,T)$ is the CADF statistic for the i_{th} cross-section unit given by the t-ratio of the coefficient on y_{it-1} in the CADF regression (32). Pesaran also considers an average of the truncated version of the CADF to help deal with the problems created by a lack of independence of the individual CADF statistics. He finds that the finite sample distribution of $CIPS(N,T)$ and the truncated version $CIPS^{*}(N,T)$ differ only for very small values of T and are indistinguishable for $T > 20$. See Tables 3a-3c in Pesaran (2005) for critical values.

Serially Correlated Errors

Pesaran also considers several extensions of the CIPS test procedure to deal with individual specific error terms that are serially correlated. Each of the three models he considers results in the same specification for the CADF regression but with different error specifications and parameter heterogeneity. Consider the case of an $AR(p)$ error specification. The relevant individual CADF statistic is the t-ratio of b_i in

$$\Delta y_{it} = a_i + b_i y_{it-1} + c_i \bar{y}_{t-1} + \sum_{j=0}^{p} d_{ij} \Delta \bar{y}_{t-j} + \sum_{j=1}^{p} \delta_{ij} \bar{y}_{t-j} + e_{it}. \tag{34}$$

The CIPS test is again conducted as a simple average of the $i = 1, \ldots, N$ $CADF_i$, and the critical values are the same as reported in tables 3a-c.

Small Sample Monte Carlo Results

In his MC studies, Pesaran allows the residual serial correlation to differ over i and finds:

- When the degree of cross sectional dependence is low, the IPS test tends to be correctly sized, but when the CD is high it tends to over-reject by a substantial amount.
- In the case of serially correlated errors, there are serious size distortions of the CADF if they are not augmented to account for the serial correlation.
- When the tests are augmented with Δy_{it-j}, the size of the test stabilizes near 5%, but for $T \geq 20$ the CIPS appears to be oversized although the truncated version is correctly sized for T as small as 10.
- The tests tend to have low power for $T < 20$, but with $T = 20$ or higher, the power of the tests rise quite rapidly with N.

References

Anindya Benerjee, Juan Dolado, John W. Galbraith, and David F. Hendry. *Co-Integration, Error-Correction, and the Econometric Analysis of Non-Stationary Data*. Advanced Texts in Econometrics. Oxford University Press, Oxford, 1993.

John Y. Campbell and Pierre Perron. Pitfalls and opportunities: What macroeconomists should know about unit roots. In Olivier Jean Blanchard and Stanley Fischer, editors, *NBER Macroeconomics Annual*, pages 141–201. Cambridge: The MIT Press, 1991.

Woiciech. W. Charemza and Derek F. Deadman. *New Directions in Econometric Practice: General to Specific Modelling, Cointegration and Vector Autoregression*. Edward Elgar, 2nd edition, 1997.

David F. Hendry and Jurgen A. Doornik. *Empirical Econometric Modelling Using PcGive*, volume I. Timberlake Consultants, LTD., London, 2001.

Kyung So Im, M. Hashem Pesaran, and Yongcheol Shin. Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, 115:53–74, 2003.

Andrew Levin, Chien-Fu Lin, and Chia-Shang James Chu ;. Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of Econometrics*, pages 1–24, 2002.

J. G. MacKinnon. Critical values for cointegration tests. In R. F. Engle and C. W. J. Granger, editors, *Long Run Economic Relationships*, pages 267–276. Oxford: Oxford Universtiy Press, 1991.

Nelson C. Mark. *International Macroeconomics and Finance: Theory and Econometric Methods*. Blackwell Publishing, Oxford, 2001.

W. Newey and K. West. A simple positive semi-definite heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708, May 1987.

Pierre Perron. The great crash, the oil shock, and the unit root hypothesis. *Econometrica*, 57:1361– 1402, 1989.

Hashem M. Pesaran. A simple panel unit root test in the presence of cross section dependence. Technical report, University of Cambridge Department of Economics Working Paper, January 2005. URL `http://www.econ.cam.ac.uk/faculty/pesaran/PesaranCADFJan05.pdf`.

C. Granger S. Hyllberg, R. F. Engle and B. S. Yoo. Seasonal integration and co-integration. *Journal of Econometrics*, 44:215–28, 1990.