2.5 Panel Unit-Root Tests

Univariate/single-equation econometric methods for testing unit roots can have low power and can give imprecise point estimates when working with small sample sizes. Consider the popular Dickey–Fuller test for a unit root in a time-series $\{q_t\}$ and assume that the time-series are generated by

$$\Delta q_t = \alpha_0 + \alpha_1 t + (\rho - 1)q_{t-1} + \epsilon_t, \tag{2.67}$$

where $\epsilon_t \overset{iid}{\sim} N(0, \sigma^2)$. If $\rho = 1, \alpha_1 = \alpha_0 = 0$, q_t follows a driftless unit root process. If $\rho = 1, \alpha_1 = 0, \alpha_0 \neq 0$, q_t follows a unit root process with drift If $|\rho| < 1$, y_t is stationary. It is mean reverting if $\alpha_1 = 0$, and is stationary around a trend if $\alpha_1 \neq 0$.

To do the Dickey–Fuller test for a unit root in q_t, run the regression (2.67) and compare the studentized coefficient for the slope to the Dickey–Fuller distribution critical values. Table 2.1 shows the power of the Dickey–Fuller test when the truth is $\rho = 0.96$.[20] With 100 observations, the test with 5 percent size rejects the unit root only 9.6 percent of the time when the truth is a mean reverting process.

100 quarterly observations is about what is available for exchange rate studies over the post Bretton-Woods floating period, so low power is a potential pitfall in unit-root tests for international economists. But again, from Table 2.1, if you had 1000 observations, you are almost guaranteed to reject the unit root when the truth is that q_t is stationary with $\rho = 0.96$. How do you get 1000 observations without having to wait 250 years? How about combining the 100 time-series observations from 10 roughly similar countries.[21] This is the motivation for recently proposed panel unit-root tests have by Levin and Lin [91], Im, Pesaran and Shin [78], and Maddala and Wu [99]. We begin with the popular Levin–Lin test.

[20] Power is the probability that the test correctly rejects the null hypothesis because the null happens to be false.

[21] It turns out that the 1000 cross-section–time-series observations contain less information than 1000 observations from a single time-series. In the time-series, $\hat{\rho}$ converges at rate T, but in the panel, $\hat{\rho}$ converges at rate $T\sqrt{N}$ where N is the number of cross-section units, so in terms of convergence toward the asymptotic distribution, it's better to get more time-series observations.

Table 2.1: Finite Sample Power of Dickey–Fuller test, $\rho = 0.96$.

	T	5 percent	10 percent
Test	25	5.885	11.895
equation	50	6.330	12.975
includes	75	7.300	14.460
constant	100	9.570	18.715
	1000	99.995	100.000
Test	25	5.715	10.720
equation	50	5.420	10.455
includes	75	5.690	11.405
trend	100	7.650	14.665
	1000	99.960	100.000

Notes: Table reports percentage of rejections at 5 percent or 10 percent critical value when the alternative hypothesis is true with $\rho = 0.96$. 20000 replications. Critical values are from Hamilton (1994) Table B.6.

The Levin–Lin Test

Let $\{q_{it}\}$ be a balanced panel[22] of N time-series with T observations which are generated by

$$\Delta q_{it} = \delta_i t + \beta_i q_{it-1} + u_{it}, \tag{2.68}$$

where $-2 < \beta_i \leq 0$, and u_{it} has the error-components representation

$$u_{it} = \alpha_i + \theta_t + \epsilon_{it}. \tag{2.69}$$

α_i is an individual–specific effect, θ_t is a single factor common time effect, and ϵ_{it} is a stationary but possibly serially correlated idiosyncratic effect that is independent across individuals. For each individual i, ϵ_{it} has the Wold moving-average representation

$$\epsilon_{it} = \sum_{j=0}^{\infty} \theta_{ij}\epsilon_{it-j} + u_{it}. \tag{2.70}$$

[22] A panel is balanced if every individual has the same number of T observations.

q_{it} is a unit root process if $\beta_i = 0$ and $\delta_i = 0$. If there is no drift in the unit root process, then $\alpha_i = 0$. The common time effect θ_t is a crude model of cross-sectional dependence.

Levin–Lin propose to test the null hypothesis that all individuals have a unit root

$$H_0 : \beta_1 = \cdots = \beta_N = \beta = 0,$$

against the alternative hypothesis that all individuals are stationary

$$H_A : \beta_1 = \cdots = \beta_N = \beta < 0.$$

The test imposes the homogeneity restrictions that β_i are identical across individuals under both the null and under the alternative hypothesis.

The test proceeds as follows. First, you need to decide if you want to control for the common time effect θ_t. If you do, you subtract off the cross-sectional mean and the basic unit of analysis is

$$\tilde{q}_{it} = q_{it} - \frac{1}{N}\sum_{j=1}^{N} q_{jt}. \tag{2.71}$$

Potential pitfalls of including common-time effect. Doing so however involves a potential pitfall. θ_t, as part of the error-components model, is assumed to be *iid*. The problem is that there is no way to impose independence. Specifically, if it is the case that each q_{it} is driven in part by *common* unit root factor, θ_t is a unit root process. Then $\tilde{q}_{it} = q_{it} - \frac{1}{N}\sum_{i=1}^{N} q_{it}$ will be stationary. The transformation renders all the deviations from the cross-sectional mean stationary. This might cause you to reject the unit root hypothesis when it is true. Subtracting off the cross-sectional average is not necessarily a fatal flaw in the procedure, however, because you are subtracting off only one potential unit root from each of the N time-series. It is possible that the N individuals are driven by N distinct and independent unit roots. The adjustment will cause all originally nonstationary observations to be stationary only if all N individuals are driven by the *same* unit root. An alternative strategy for modeling cross-sectional dependence is to do a bootstrap, which is discussed below. For now, we will proceed with the transformed observations. The resulting test equations are

$$\Delta\tilde{q}_{it} = \alpha_i + \delta_i t + \beta_i \tilde{q}_{it-1} + \sum_{j=1}^{k_i} \phi_{ij} \Delta\tilde{q}_{it-j} + \epsilon_{it}. \tag{2.72}$$

The slope coefficient on $\tilde{q}_{it-1}$ is constrained to be equal across individuals, but no such homogeneity is imposed on the coefficients on the lagged differences nor on the number of lags k_i. To allow for this specification in estimation, regress $\Delta\tilde{q}_{it}$ and $\tilde{q}_{it-1}$ on a constant (and possibly trend) and k_i lags of $\Delta\tilde{q}_{it}$.[23]

$$\Delta\tilde{q}_{it} = a_i + b_i t + \sum_{j=1}^{k_i} c_{ij} \Delta\tilde{q}_{it-j} + \hat{e}_{it}, \tag{2.73}$$

$$\tilde{q}_{it-1} = a_i' + b_i' t + \sum_{j=1}^{k_i} c_{ij}' \Delta\tilde{q}_{it-j} + \hat{v}_{it}, \tag{2.74}$$

where $\hat{e}_{it}$ and $\hat{v}_{it}$ are OLS residuals. Now run the regression

$$\hat{e}_{it} = \delta_i \hat{v}_{it-1} + \hat{u}_{it}, \tag{2.75}$$

set $\hat{\sigma}_{ei}^2 = \frac{1}{T-k_i-1} \sum_{t=k_i+2}^{T} \hat{u}_{it}^2$, and form the normalized observations

$$\tilde{e}_{it} = \frac{\hat{e}_{it}}{\hat{\sigma}_{ei}}, \qquad \tilde{v}_{it} = \frac{\hat{v}_{it}}{\hat{\sigma}_{ei}}. \tag{2.76}$$

Denote the long run variance of Δq_{it} by $\sigma_{qi}^2 = \gamma_0^i + 2\sum_{j=0}^{\infty} \gamma_j^i$, where $\gamma_0^i = E(\Delta q_{it}^2)$ and $\gamma_j^i = E(\Delta q_{it} \Delta q_{it-j})$. Let $\bar{k} = \frac{1}{N}\sum_{i=1}^{N} k_i$ and estimate σ_{qi}^2 by Newey and West [114]

$$\hat{\sigma}_{qi}^2 = \hat{\gamma}_0^i + 2\sum_{j=1}^{\bar{k}} \left(1 - \frac{j}{\bar{k}+1}\right) \hat{\gamma}_j^i, \tag{2.77}$$

[23] To choose k_i, one option is to use AIC or BIC. Another option is to use Hall's [69] general-to-specific method recommended by Campbell and Perron [19]. Start with some maximal lag order ℓ and estimate the regression. If the absolute value of the t-ratio for $\hat{c}_{i\ell}$ is less than some appropriate critical value, c^*, reset k_i to $\ell - 1$ and repeat the process until the t-ratio of the estimated coefficient with the longest lag exceeds the critical value c^*.

Table 2.2: Mean and Standard Deviation Adjustments for Levin–Lin τ Statistic, reproduced from Levin and Lin [91]

		τ^*_{NC}		τ^*_C		τ^*_{CT}	
$\tilde{T}$	$\overline{K}$	$\mu^*_{\tilde{T}}$	$\sigma^*_{\tilde{T}}$	$\mu^*_{\tilde{T}}$	$\sigma^*_{\tilde{T}}$	$\mu^*_{\tilde{T}}$	$\sigma^*_{\tilde{T}}$
25	9	0.004	1.049	-0.554	0.919	-0.703	1.003
30	10	0.003	1.035	-0.546	0.889	-0.674	0.949
35	11	0.002	1.027	-0.541	0.867	-0.653	0.906
40	11	0.002	1.021	-0.537	0.850	-0.637	0.871
45	11	0.001	1.017	-0.533	0.837	-0.624	0.842
50	12	0.001	1.014	-0.531	0.826	-0.614	0.818
60	13	0.001	1.011	-0.527	0.810	-0.598	0.780
70	13	0.000	1.008	-0.524	0.798	-0.587	0.751
80	14	0.000	1.007	-0.521	0.789	-0.578	0.728
90	14	0.000	1.006	-0.520	0.782	-0.571	0.710
100	15	0.000	1.005	-0.518	0.776	-0.566	0.695
250	20	0.000	1.001	-0.509	0.742	-0.533	0.603
∞	–	0.000	1.000	-0.500	0.707	-0.500	0.500

where $\hat{\gamma}^i_j = \frac{1}{T-1}\sum_{t=2+j}^{T} \Delta\tilde{q}_{it}\Delta\tilde{q}_{it-j}$. Let $s_i = \frac{\hat{\sigma}_{qi}}{\hat{\sigma}_{ei}}$, $S_N = \frac{1}{N}\sum_{i=1}^{N} s_i$ and run the pooled cross-section time-series regression

$$\tilde{e}_{it} = \beta\tilde{v}_{it-1} + \tilde{\epsilon}_{it}. \tag{2.78}$$

The *studentized coefficient* is $\tau = \hat{\beta}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{v}_{it-1}/\hat{\sigma}_{\tilde{\epsilon}}$ where $\hat{\sigma}_{\tilde{\epsilon}} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\epsilon}_{it}$. As in the univariate case, τ is not asymptotically standard normally distributed. In fact, τ diverges as the number of observations NT gets large, but Levin and Lin show that the adjusted statistic

$$\tau^* = \frac{\tau - N\tilde{T}S_N\tau\mu^*_{\tilde{T}}\hat{\sigma}_{\epsilon}^{-2}\hat{\beta}^{-1}}{\sigma^*_{\tilde{T}}} \xrightarrow{D} N(0,1), \tag{2.79}$$

as $\tilde{T}\to\infty, N\to\infty$ where $\tilde{T} = T - \bar{k} - 1$, and $\mu^*_{\tilde{T}}$ and $\sigma^*_{\tilde{T}}$ are adjustment factors reproduced from Levin and Lin's paper in Table 2.2.

Performance of Levin and Lin's adjustment factors in a controlled en-

Table 2.3: How Well do Levin–Lin adjustments work? Percentiles from a Monte Carlo Experiment.

Statistic	N	T	trend	2.5%	5%	50%	95%	97.5%
τ	20	100	no	-7.282	-6.995	-5.474	-3.862	-3.543
	20	500	no	-7.202	-6.924	-5.405	-3.869	-3.560
τ^*	20	100	no	-2.029	-1.732	-0.092	1.613	1.965
	20	500	no	-1.879	-1.557	0.012	1.595	1.894
τ	20	100	yes	-10.337	-10.038	-8.642	-7.160	-6.896
	20	500	yes	-10.126	-9.864	-8.480	-7.030	-6.752
τ^*	20	100	yes	-1.171	-0.825	0.906	2.997	3.503
	20	500	yes	-1.028	-0.746	0.702	2.236	2.571

vironment. Suppose the data generating process (the truth) is, that each individual is the unit root process

$$\Delta q_{it} = \alpha_i + \sum_{j=1}^{2} \phi_{ij} \Delta q_{it-j} + \epsilon_{it}, \tag{2.80}$$

where $\epsilon_{it} \overset{iid}{\sim} N(0, \sigma_i)$, and each of the σ_i is drawn from a uniform distribution over the range 0.1 to 1.1. That is, $\sigma_i \sim U[0.1, 1.1]$. Also, $\phi_{ij} \sim U[-0.3, 0.3]$, and $\alpha_i \sim N(0, 1)$ if a drift is included, (otherwise $\alpha = 0$).[24] Table 2.3 shows the Monte Carlo distribution of Levin and Lin's τ and τ^* generated from this process. Here are some things to note from the table. First, the median value of τ is very far from 0. It would get bigger (in absolute value) if we let N get bigger. Second, τ^* looks like a standard normal variate when there is no drift in the DGP (and no trend in the test equation). Third, the Monte Carlo distribution for τ^* looks quite different from the asymptotic distribution when there is drift in the DGP and a trend is included in the test equation. This is what we call finite sample size distortion of the test. When there is known size distortion, you might want to control for it by doing a bootstrap, which is covered below.

[24] Instead of me arbitrarily choosing values of these parameters for each of the individual units, I let the computer pick out some numbers at random.

Table 2.4: Size adjusted power of Levin–Lin test with $T = 100, N = 20$

Proportion	Constant		Trend	
stationary[a/]	5 %	10%	5 %	10%
0.2	0.141	0.275	0.124	0.218
0.4	0.329	0.439	0.272	0.397
0.6	0.678	0.761	0.577	0.687
0.8	0.942	0.967	0.906	0.944
1.0	1.000	1.000	1.000	1.000

Notes: [a/]Proportion of individuals in the panel that are stationary. Stationary components have root equal to 0.96. Source: Choi [26].

Another option is to try to correct for the size distortion. The question here is, if you correct for size distortion, does the Levin–Lin test have good power? That is, will it reject the null hypothesis when it is false with high probability? The answer suggested in Table 2.4 is yes. It should be noted, that even though the Levin-Lin test is motivated in terms of a homogeneous panel, it has moderate ability to reject the null when the truth is a mixed panel in which some of the individuals are unit root process and others are stationary.

Bias Adjustment

The OLS estimator $\hat{\rho}$ is biased downward in small samples. Kendall [85] showed that the bias of the least squares estimator is $E(\hat{\rho}) - \rho \simeq -(1 + 3\rho)/T$. A bias-adjusted estimate of ρ is

$$\hat{\rho^*} = \frac{T\hat{\rho} + 1}{T - 3}. \tag{2.81}$$

The panel estimator of the serial correlation coefficient is also biased downwards in small samples. A first-order bias-adjustment of the panel estimate of ρ can be done using a result by Nickell [116] who showed that

$$(\hat{\rho} - \rho) \to \frac{A_T B_T}{C_T}, \tag{2.82}$$

as $T \to \infty, N \to \infty$ where $A_T = \frac{-(1+\rho)}{T-1}$, $B_T = 1 - \frac{1}{T}\frac{(1-\rho^T)}{(1-\rho)}$, and $C_T = 1 - \frac{2\rho(1-B_T)}{[(1-\rho)(T-1)]}$.

Bootstrapping τ^*

The fact that τ diverges can be distressing. Rather than to rely on the asymptotic adjustment factors that may not work well in some regions of the parameter space, researchers often choose to test the unit root hypothesis using a bootstrap distribution of τ.[25] Furthermore, the bootstrap provides an alternative way to model cross-sectional dependence in the error terms, as discussed above. The method discussed here is called the *residual* bootstrap because we will be resampling from the residuals.

To build a bootstrap distribution under the null hypothesis that all individuals follow a unit-root process, begin with the data generating process (DGP)

$$\Delta q_{it} = \mu_i + \sum_{j=1}^{k_i} \phi_{ij} \Delta q_{i,t-j} + \epsilon_{it}. \tag{2.83}$$

Since each q_{it} is a unit root process, its first difference follows an autoregression. While you may prefer to specify the DGP as an unrestricted vector autoregression for all N individuals, the estimation such a system turns out not to be feasible for even moderately sized N.

The individual equations of the DGP can be fitted by least squares. If a linear trend is included in the test equation a constant must be included in (2.83). To account for dependence across cross-sectional units, estimate the joint error covariance matrix $\mathbf{\Sigma} = \mathrm{E}(\underline{\epsilon}_t \underline{\epsilon}_t')$ by $\hat{\mathbf{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T} \hat{\underline{\epsilon}}_t \hat{\underline{\epsilon}}_t'$ where $\hat{\underline{\epsilon}}_t = (\hat{\epsilon}_{1t}, \ldots, \hat{\epsilon}_{Nt})$ is the vector of OLS residuals.

The parametric bootstrap distribution for τ is built as follows.

1. Draw a sequence of length $T + R$ innovation vectors from $\tilde{\underline{\epsilon}}_t \sim N(0, \hat{\mathbf{\Sigma}})$.

2. Recursively build up pseudo–observations $\{\hat{q}_{it}\}, i = 1, \ldots, N$, $t = 1, \ldots, T + R$ according to (2.83) with the $\tilde{\underline{\epsilon}}_t$ and estimated values of the coefficients $\hat{\underline{\mu}}_i$ and $\hat{\phi}_{ij}$.

[25] For example, Wu [135] and Papell [118].

3. Drop the first R pseudo-observations, then run the Levin–Lin test on the pseudo-data. *Do not* transform the data by subtracting off the cross-sectional mean and do not make the τ^* adjustments. This yields a realization of τ generated in the presence of cross-sectional dependent errors.

4. Repeat a large number (2000 or 5000) times and the collection of τ and $\bar{t}$ statistics form the bootstrap distribution of these statistics under the null hypothesis.

This is called a parametric bootstrap because the error terms are drawn from the parametric normal distribution. An alternative is to do a nonparametric bootstrap. Here, you resample the estimated residuals, which are in a sense, the data. To do a nonparametric bootstrap, do the following. Estimate (2.83) using the data. Denote the OLS residuals by

$$\begin{array}{cc} (\hat{\epsilon}_{11}, \hat{\epsilon}_{21}, \ldots, \hat{\epsilon}_{N1}) & \leftarrow \text{obs. } 1 \\ (\hat{\epsilon}_{12}, \hat{\epsilon}_{22}, \ldots, \hat{\epsilon}_{N2}) & \leftarrow \text{obs. } 2 \\ \vdots & \vdots \\ (\hat{\epsilon}_{1T}, \hat{\epsilon}_{2T}, \ldots, \hat{\epsilon}_{NT}) & \leftarrow \text{obs. T} \end{array}$$

Now resample the residual *vectors* with replacement. For each observation $t = 1, \ldots, T$, draw one of the T possible residual vectors with probability $\frac{1}{T}$. Because the entire vector is being resampled, the cross-sectional correlation observed in the data is preserved. Let the resampled vectors be

$$\begin{array}{cc} (\epsilon^*_{11}, \epsilon^*_{21}, \ldots, \epsilon^*_{N1}) & \leftarrow \text{obs. } 1 \\ (\epsilon^*_{12}, \epsilon^*_{22}, \ldots, \epsilon^*_{N2}) & \leftarrow \text{obs. } 2 \\ \vdots & \vdots \\ (\epsilon^*_{1T}, \epsilon^*_{2T}, \ldots, \epsilon^*_{NT}) & \leftarrow \text{obs. T} \end{array}$$

and use these resampled residuals to build up values of Δq_{it} recursively using (2.83) with $\hat{\mu}_i$ and $\hat{\phi}_{ij}$, and run the Levin-Lin test on these observations but do not subtract off the cross-sectional mean, and do not make the τ^* adjustments. This gives a realization of τ. Now repeat a large number of times to get the nonparametric bootstrap distribution of τ.

The Im, Pesaran and Shin Test

Im, Pesaran and Shin suggest a very simple panel unit root test. They begin with the ADF representation (2.72) for individual i (reproduced here for convenience)

$$\Delta\tilde{q}_{it} = \alpha_i + \gamma_i t + \beta_i \tilde{q}_{it-1} + \sum_{j=1}^{k_i} \phi_{ij} \Delta\tilde{q}_{it-j} + \epsilon_{it}, \tag{2.84}$$

where $\mathrm{E}(\epsilon_{it}\epsilon_{js}) = 0, i \neq j$ for all t, s. A common time effect may be removed in which case $\tilde{q}_{it} = q_{it} - (1/N)\sum_{i=1}^{N} q_{it}$ is the deviation from the cross-sectional average as the basic unit of analysis.

Let τ_i be the studentized coefficient from the ith ADF regression. Since the ϵ_{it} are assumed to be independent across individuals, the τ_i are also independent, and by the central limit theorem, $\overline{\tau}_{NT} = \frac{1}{N}\sum_{i=1}^{N} \tau_i$ converges to a normal distribution first as $T \to \infty$ then as $N \to \infty$. That is

$$\frac{\sqrt{N}[\bar{\tau}_{NT} - \mathrm{E}(\tau_{it}|\beta_i = 0)]}{\sqrt{\mathrm{Var}(\tau_{it}|\beta = 0)}} \xrightarrow{D} \mathrm{N}(0, 1), \tag{2.85}$$

as $T \to \infty, N \to \infty$. IPS report selected critical values for $\bar{\tau}_{NT}$ with the conditional mean and variance adjustments of the distribution. A selected set of these critical values are reproduced in Table 2.5. An alternative to relying on the asymptotic distribution is to do a residual bootstrap of the $\overline{\tau}_{NT}$ statistic. As before, when doing the bootstrap, do not subtract off the cross-sectional mean.

The Im, Pesaran and Shin test as well as the Maddala–Wu test (discussed below) relax the homogeneity restrictions under the alternative hypothesis. Here, the null hypothesis

$$H_0 : \beta_1 = \cdots = \beta_N = \beta = 0,$$

is tested against the alternative

$$H_A : \beta_1 < 0 \cup \beta_2 < 0 \cdots \cup \beta\beta_N < 0.$$

The alternative hypothesis is *not* H_0, which is less restrictive than the Levin–Lin alternative hypothesis.

Table 2.5: Selected Exact Critical Values for the IPS $\bar{\tau}_{NT}$ Statistic

		Constant			Trend		
	T	20	40	100	20	40	100
A. 5 percent							
N	5	-2.19	-2.16	-2.15	-2.82	-2.77	-2.75
	10	-1.99	-1.98	-1.97	-2.63	-2.60	-2.58
	15	-1.91	-1.90	-1.89	-2.55	-2.52	-2.51
	20	-1.86	-1.85	-1.84	-2.49	-2.48	-2.46
	25	-1.82	-1.81	-1.81	-2.46	-2.44	-2.43
B. 10 percent							
N	5	-2.04	-2.02	-2.01	-2.67	-2.63	-2.62
	10	-1.89	-1.88	-1.88	-2.52	-2.50	-2.49
	15	-1.82	-1.81	-1.81	-2.46	-2.44	-2.44
	20	-1.78	-1.78	-1.77	-2.42	-2.41	-2.40
	25	-1.75	-1.75	-1.75	-2.39	-2.38	-2.38

Source: Im, Pesaran and Shin [78].

The Maddala and Wu Test

Maddala and Wu [99] point out that the IPS strategy of combining independent tests to construct a joint test is an idea suggested by R.A. Fisher [53]. Maddala and Wu follow Fisher's suggestion and propose following test. Let the p-value of τ_i from the augmented Dickey–Fuller test for a unit root be $p_i = \text{Prob}(\tau < \tau_i) = \int_{-\infty}^{\tau_i} f(x)dx$ be the p-value of τ_i from the ADF test on (2.72), where $f(\tau)$ is the probability density function of τ. Solve for $g(p)$, the density of p_i by the method of transformations, $g(p_i) = f(\tau_i)|J|$ where $J = d\tau_i/dp_i$ is the Jacobian of the transformation, and $|J|$ is its absolute value. Since $dp_i/d\tau_i = f(\tau_i)$, the Jacobian is $1/f(\tau_i)$ and $g(p_i) = 1$ for $0 \le p_i \le 1$. That is, p_i is uniformly distributed on the interval $[0,1]$ ($p_i \sim U[0,1]$).

Next, let $y_i = -2\ln(p_i)$. Again, using the method of transformations, the probability density function of y_i is $h(y_i) = g(p_i)|dp_i/dy_i|$. But $g(p_i) = 1$ and $|dp_i/dy_i| = p_i/2 = (1/2)e^{-y_i/2}$, so it follows that $h(y_i) = (1/2)e^{-y_i/2}$ which is the chi-square distribution with 2 degrees of freedom. Under cross-sectional independence of the error terms ϵ_{it}, the joint test statistic also has a chi-square distribution

$$\lambda = -2\sum_{i=1}^{N} \ln(p_i) \sim \chi^2_{2N}. \tag{2.86}$$

The asymptotic distribution of the IPS test statistic was established by sequential $T \to \infty, N \to \infty$ asymptotics, which some econometricians view as being too restrictive.[26] Levin and Lin derive the asymptotic distribution of their test statistic by allowing both N and T simultaneously to go to infinity. A remarkable feature of the Maddala–Wu–Fisher test is that it avoids issues of sequential or joint N, T asymptotics. (2.86) gives the exact distribution of the test statistic.

The IPS test is based on the sum of τ_i, whereas the Maddala–Wu test is based on the sum of the log p-values of τ_i. Asymptotically, the two tests should be equivalent, but can differ in finite samples. Another advantage of Maddala–Wu is that the test statistic distribution does not depend on nuisance parameters, as does IPS and LL. The disadvantage is that p-values need to be calculated numerically.

Potential Pitfalls of Panel Unit-Root Tests

Panel unit-root tests need to be applied with care. One potential pitfall with panel tests is that the rejection of the null hypothesis does not mean that all series are stationary. It is possible that out of N time-series, only 1 is stationary and (N-1) are unit root processes. This is an example of a mixed panel. Whether we want the rejection of the unit root process to be driven by a single outlier or not depends on the purpose the researcher uses the test.[27]

[26] That is, they deduce the limiting behavior of the test statistic first by letting $T \to \infty$ holding N fixed, then letting $N \to \infty$ and invoking the central limit theorem.

[27] Bowman [17] shows that both the LL and IPS tests have low power against outlier driven alternatives. He proposes a test that has maximal power. Taylor and Sarno [131] propose a test based on Johansen's [80] maximum likelihood approach that can test for the number of unit-root series in the panel. Computational considerations, however, generally limit the number of time-series that can be analyzed to 5 or less.

A second potential pitfall is that cross-sectional independence is a regularity condition for these tests. Transforming the observations by subtracting off the cross-sectional means will leave some residual dependence across individuals if common time effects are generated by a multi-factor process. This residual cross-sectional dependence can potentially generate errors in inference.

A third potential pitfall concerns potential small sample size distortion of the tests. While most of the attention has been aimed at improving the power of unit root tests, Schwert [126] shows that there are regions of the parameter space under which the size of the augmented Dickey–Fuller test is wrong in small samples. Since the panel tests are based on the augmented Dickey–Fuller test in some way or another, it is probably the case that this size distortion will get impounded into the panel test. To the extent that size distortion is an issue, however, it is not a problem that is specific to the panel tests.

2.6 Cointegration

The unit root processes $\{q_t\}$ and $\{f_t\}$ will be *cointegrated* if there exists a linear combination of the two time-series that is stationary. To understand the implications of cointegration, let's first look at what happens when the observations are not cointegrated.

No cointegration. Let $\xi_{qt} = \xi_{qt-1} + \epsilon_{qt}$ and $\xi_{ft} = \xi_{ft-1} + \epsilon_{ft}$ be two independent random walk processes where $\epsilon_{qt} \overset{iid}{\sim} \mathrm{N}(0, \sigma_q^2)$ and $\epsilon_{ft} \overset{iid}{\sim} \mathrm{N}(0, \sigma_f^2)$. Let $\underline{z}_t = (z_{qt}, z_{ft})'$ follow a stationary bivariate process such as a VAR. The exact process for $\underline{z}_t$ doesn't need to explicitly modeled at this point. Now consider the two unit root series built up from these components

$$\begin{aligned} q_t &= \xi_{qt} + z_{qt}, \\ f_t &= \xi_{ft} + z_{ft}. \end{aligned} \tag{2.87}$$

Since q_t and f_t are driven by *independent* random walks, they will drift arbitrarily far apart from each other over time. If you try to find a value of β to form a stationary linear combination of q_t and f_t, you will