

Nonstationary Panels

Baltagi: Econometric Analysis of Panel Data (Wiley, 2005)
Chapter 12

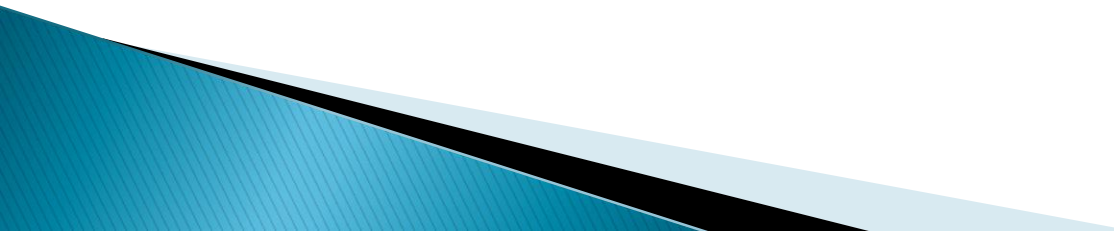
Prof. Dr. Robert Kunst
UK 40871

Ziad Al Hajeb
Harun Akbas
9/6/2010

Outline

- ▶ 12.1 Introduction
- ▶ 12.2 Panel Unit Roots Tests assuming Cross-sectional Independence
 - 12.2.1 Levin, Lin and Chu Test
 - 12.2.2 Im, Pesaran and Shin Test
 - 12.2.3 Breitung's Test
 - 12.2.4 Combining p -Value Tests
 - 12.2.5 Residual-Based LM Test
- ▶ 12.3 Panel Unit Roots Tests allowing for Cross-sectional Dependence
 - Moon and Perron Test

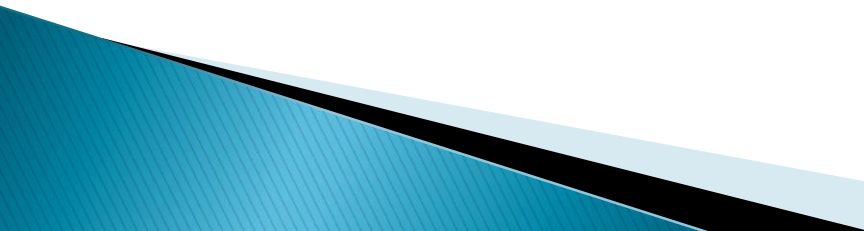
Outline

- ▶ 12.4 Spurious Regression in Panel Data
 - ▶ 12.5 Panel Cointegration Tests
 - ▶ 12.6 Estimation and Inference in Panel
- 

Introduction

- ▶ The growing use of cross-country data over time to study purchasing power parity, growth convergence and International R&D spillovers
- ▶ The focus of panel data econometrics shifted towards studying the asymptotics of macro panels with large N (number of countries) and large T (length of the time series) rather than the usual asymptotics of micro panels with large N and small T

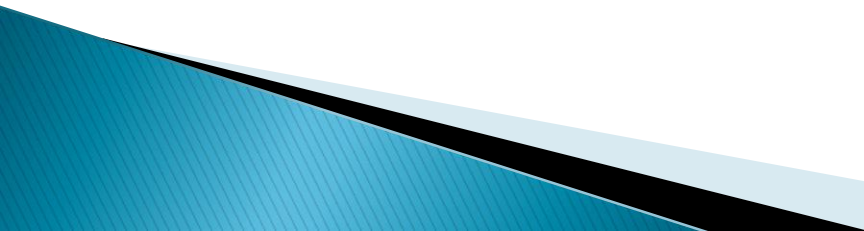
Introduction

- ▶ The fact that T is allowed to increase to infinity in macro panel data generated two strands of ideas:
 1. rejected the homogeneity of the regression parameters implicit in the use of a pooled regression model in favor of heterogeneous regressions
 2. applied time series procedures to panels, worrying about nonstationarity, spurious regressions and cointegration
- 

Panel Unit Roots Tests assuming Cross-sectional Independence

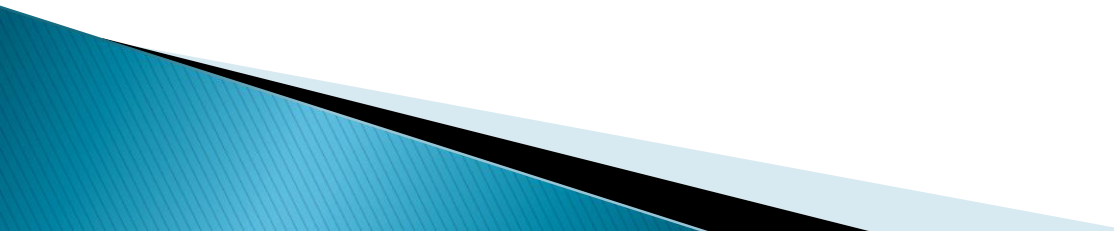
- ▶ Quah (1994) suggested a test for unit root in a panel data model without fixed effects where both N and T go to infinity at the same rate such that N/T is constant.
- ▶ Levin et al. (2002), LLC, generalized this model to allow for fixed effects, individual deterministic trends and heterogeneous serially correlated errors

Panel Unit Roots Tests assuming Cross-sectional Independence

- ▶ LLC argue that individual unit root tests have limited power against alternative hypotheses with highly persistent deviations from equilibrium, particularly in small samples
 - ▶ Therefore, suggest a more powerful panel unit root test than performing individual unit root tests for each cross-section using an increased sample size of panel data
- 

Panel Unit Roots Tests assuming Cross-sectional Independence

Levin, Lin, and Chu (LLC) test

- ▶ Null hypothesis: each individual time series contains a unit root
 - ▶ Alternative hypothesis: each individual time series is stationary
 - ▶ Maintained hypothesis:
- 

Panel Unit Roots Tests assuming Cross-sectional Independence

$$\Delta y_{it} = \rho y_{i,t-1} + \sum_{L=1}^{p_i} \theta_{iL} \Delta y_{it-L} + \alpha_{mi} d_{mt} + \varepsilon_{it} \quad m = 1, 2, 3$$

- ▶ d_{mt} : the vector of deterministic variables
- ▶ α_{mi} : the corresponding vector of coefficients for model $m = 1, 2, 3$.
- ▶ $d_{1t} = \{\text{empty set}\}$, $d_{2t} = \{1\}$ and $d_{3t} = \{1, t\}$.
- ▶ Since the lag order p_i is unknown, LLC suggest a three-step procedure to implement their test

Panel Unit Roots Tests assuming Cross-sectional Independence

1. Perform separate augmented Dickey-Fuller (ADF) regressions for each cross-section:

$$\Delta y_{it} = \rho_i y_{i,t-1} + \sum_{L=1}^{p_i} \theta_{iL} \Delta y_{i,t-L} + \alpha_{mi} d_{mt} + \varepsilon_{it} \quad m = 1, 2, 3$$

- i. choose a maximum lag order p_{max} and use t -statistic of $\hat{\theta}_{iL}$ to determine if a smaller lag order is preferred (t -statistics are distributed $N(0, 1)$ under the $H_0(\theta_{iL} = 0)$)

Panel Unit Roots Tests assuming Cross-sectional Independence

ii. In order to get orthogonalized residuals

Run Δy_{it} on $\Delta y_{i,t-L}(L=1, \dots, p_i)$ and d_{mt} to get residuals \hat{e}_{it}
Run $y_{i,t-1}$ on $\Delta y_{i,t-L}(L=1, \dots, p_i)$ and d_{mt} to get residuals $\hat{v}_{i,t-1}$

iii. Standardize these residuals to control for different variances across i

$$\tilde{e}_{it} = \hat{e}_{it} / \hat{\sigma}_{\varepsilon i} \quad \text{and} \quad \tilde{v}_{i,t-1} = \hat{v}_{it} / \hat{\sigma}_{\varepsilon i}$$

where $\hat{\sigma}_{\varepsilon i}$ is the standard error from each ADF regression, for $i = 1, \dots, N$

Panel Unit Roots Tests assuming Cross-sectional Independence

2. Estimate the ratio of long-run to short-run standard deviations. The long-run variance under the null hypothesis of a unit root:

$$\hat{\sigma}_{yi}^2 = \frac{1}{T-1} \sum_{t=2}^T \Delta y_{it}^2 + 2 \sum_{L=1}^{\bar{K}} w_{\bar{K}L} \left[\frac{1}{T-1} \sum_{t=2+L}^T \Delta y_{it} \Delta y_{i,t-L} \right]$$

where $w_{\bar{K}L} = 1 - (L/(\bar{K} + 1))$ for a Bartlett kernel

Panel Unit Roots Tests assuming Cross-sectional Independence

3. Compute the panel test statistics by
 - i. running the pooled regression:

$$\tilde{e}_{it} = \rho \tilde{v}_{i,t-1} + \tilde{\varepsilon}_{it}$$

based on $N\tilde{T}$ observations where $\tilde{T} = T - \bar{p} - 1$. \tilde{T} is the average number of observations per individual where \bar{p} is the average lag order

Panel Unit Roots Tests assuming Cross-sectional Independence

- ii. calculate the conventional t-statistic under $H_0: \rho = 0$

$$\hat{\rho} = \frac{\sum_{i=1}^N \sum_{t=2+p_i}^T \tilde{v}_{i,t-1} \tilde{e}_{it}}{\sum_{i=1}^N \sum_{t=2+p_i}^T \tilde{v}_{i,t-1}^2}$$
$$\hat{\sigma}(\hat{\rho}) = \hat{\sigma}_{\tilde{\varepsilon}} / \left[\sum_{i=1}^N \sum_{t=2+p_i}^T \tilde{v}_{i,t-1}^2 \right]^{1/2}$$

$$\hat{\sigma}_{\tilde{\varepsilon}}^2 = \frac{1}{NT} \sum_{i=1}^N \sum_{t=2+p_i}^T (\tilde{e}_{it} - \hat{\rho} \tilde{v}_{i,t-1})^2$$

is the estimated variance of $\tilde{\varepsilon}_{it}$

Panel Unit Roots Tests assuming Cross-sectional Independence

iii. calculate the adjusted t-statistic

$$t_{\rho}^* = \frac{t_{\rho} - N\tilde{T}\hat{S}_N\hat{\sigma}_{\tilde{\varepsilon}}^{-2}\hat{\sigma}(\hat{\rho})\mu_{m\tilde{T}}^*}{\sigma_{m\tilde{T}}^*}$$

where $\mu_{m\tilde{T}}^*$ and $\sigma_{m\tilde{T}}^*$ are the mean and standard deviation adjusted and tabulated by LLC

Panel Unit Roots Tests assuming Cross-sectional Independence

- ▶ t_{ρ}^* is asymptotically distributed:
 - as $N(0,1)$
 - requires $\sqrt{N_T}/T \rightarrow 0$ where N is an arbitrary monotonically increasing function of T
- ▶ Limitations
 - depends upon the *independence* assumption across cross-sections and is not applicable if cross-sectional correlation is present
 - the assumption that *all* cross-sections have or do not have a unit root is restrictive

Panel Unit Roots Tests assuming Cross-sectional Independence

- ▶ Suggestions for using the LLC test:
 - for panels of moderate size ($10 < N < 250$, $25 < T < 250$)
 - for very large T , individual unit root time series tests are sufficiently powerful
 - For very large N and very small T , usual panel data procedures

Panel Unit Roots Tests assuming Cross-sectional Independence

Im, Pesaran and Shin (IPS) test

- ▶ LLC test is restrictive because it requires ρ to be homogeneous across i
- ▶ IPS test allow ρ to be heterogeneous across i
- ▶ Null Hypothesis:

$$H_0: \rho_i = 0 \text{ for all } i$$

Panel Unit Roots Tests assuming Cross-sectional Independence

- ▶ Alternative hypothesis:

$$H_1: \begin{cases} \rho_i < 0 & \text{for } i = 1, 2, \dots, N_1 \\ \rho_i = 0 & \text{for } i = N_1 + 1, \dots, N \end{cases}$$

it requires the fraction of the individual time series that are stationary to be nonzero, otherwise the panel unit root test will be inconsistent

Panel Unit Roots Tests assuming Cross-sectional Independence

- ▶ Test statistic:

$$\bar{t} = \frac{1}{N} \sum_{i=1}^N t_{\rho_i}$$

where t_{ρ_i} is the individual t –statistic for testing $H_0 : \rho_i = 0$ for all i

- ▶ “the average of the individual ADF statistics”

Panel Unit Roots Tests assuming Cross-sectional Independence

- ▶ for fixed N and $T \rightarrow \infty$, the standard result is:

$$t_{\rho_i} \Rightarrow \frac{\int_0^1 W_{iZ} dW_{iZ}}{\left[\int_0^1 W_{iZ}^2 \right]^{1/2}} = t_{iT}$$

- ▶ t_{iT} is assumed to be IID with finite mean and variance

Panel Unit Roots Tests assuming Cross-sectional Independence

- ▶ by the Lindeberg–Levy central limit theorem, as $T \rightarrow \infty$ followed by $N \rightarrow \infty$ sequentially, the Asymptotic distribution is:

$$t_{\text{IPS}} = \frac{\sqrt{N} \left(\bar{t} - \frac{1}{N} \sum_{i=1}^N E[t_{iT} | \rho_i = 0] \right)}{\sqrt{\frac{1}{N} \sum_{i=1}^N \text{var}[t_{iT} | \rho_i = 0]}} \Rightarrow N(0, 1)$$

- ▶ the values of $E[t_{iT} | \rho_i = 0]$ and $\text{var}[t_{iT} | \rho_i = 0]$ are computed by IPS via simulations for different values of T and ρ_i 's

Panel Unit Roots Tests assuming Cross-sectional Independence

- ▶ if a large enough lag order is selected, IPS is generally better than LLC

Panel Unit Roots Tests assuming Cross-sectional Independence

Breitung's test

- ▶ LLC and IPS tests suffer from a dramatic loss of power if individual-specific trends are included due to the bias correction
- ▶ Breitung's test statistic without bias adjustment is obtained in a 3 step process

Panel Unit Roots Tests assuming Cross-sectional Independence

1. Same as in LLC, but $\Delta y_{i,t-L}$ is used in obtaining the residuals \hat{e}_{it} and $\hat{v}_{i,t-1}$, basically separate augmented ADF regressions for each cross section i
2. Transform \hat{e}_{it} using the forward orthogonalization transformation by Arellano and Bover (1995)

$$e_{it}^* = \sqrt{\frac{T-t}{(T-t+1)}} \left(\tilde{e}_{it} - \frac{\tilde{e}_{i,t+1} + \dots + \tilde{e}_{i,T}}{T-t} \right)$$

Panel Unit Roots Tests assuming Cross-sectional Independence

$$v_{i,t-1}^* = \begin{cases} \tilde{v}_{i,t-1} & \text{without intercept or trend} \\ \tilde{v}_{i,t-1} - \tilde{v}_{i,1} & \text{with intercept} \\ \tilde{v}_{i,t-1} - \tilde{v}_{i,1} - \frac{t-1}{T} \tilde{v}_{i,T} & \text{with intercept and trend} \end{cases}$$

- ▶ Run the pooled regression

$$e_{it}^* = \rho v_{i,t-1}^* + \varepsilon_{it}^*$$

t-statistic for $H_0 : \rho = 0$, which has in the limit a standard $N(0, 1)$ distribution

Panel Unit Roots Tests assuming Cross-sectional Independence

- ▶ Let G_{iT_i} be a unit root test statistic for the i th group
- ▶ Assume $G_{iT_i} \Rightarrow G_i$ where G_i is a nondegenerate random Variable as $T_i \rightarrow \infty$
- ▶ Let p_i be the corresponding asymptotic p -value
- ▶ Fisher type test with P-value from unit root tests for each cross section i

$$P = -2 \sum_{i=1}^N \ln p_i$$

Panel Unit Roots Tests assuming Cross-sectional Independence

- ▶ When N is large, a modified P test is proposed:

$$P_m = \frac{1}{2\sqrt{N}} \sum_{i=1}^N (-2 \ln p_i - 2)$$

- ▶ By applying the Lindberg–Levy central limit theorem we get:

$$P_m \Rightarrow N(0, 1) \text{ as } T_i \rightarrow \infty \text{ followed by } N \rightarrow \infty$$

Panel Unit Roots Tests assuming Cross-sectional Independence

- ▶ Advantages:
 - different lag orders may be used
 - other unit root tests may be applied
- ▶ Disadvantages:
 - p-values have to be derived by Monte Carlo simulations
- ▶ Recommendations:
 - combined p-value tests outperform the IPS test, the Z test performs best

Panel Unit Roots Tests assuming Cross-sectional Independence

Residual-based LM test

- ▶ Null Hypothesis: No unit root in any of the series
- ▶ Alternative Hypothesis: Unit root in the Panel
- ▶ Hadri (2000) considers the following two models:

$$y_{it} = r_{it} + \varepsilon_{it} \quad i = 1, \dots, N; \quad t = 1, \dots, T$$

$$y_{it} = r_{it} + \beta_i t + \varepsilon_{it}$$

Panel Unit Roots Tests assuming Cross-sectional Independence

- ▶ Where $\varepsilon_{it} \sim \text{IIN}(0, \sigma_\varepsilon^2)$ and $u_{it} \sim \text{IIN}(0, \sigma_\mu^2)$ are mutually independent normals that are IID across i
- ▶ Using back substitution:

$$y_{it} = r_{io} + \beta_{it} + \sum_{s=1}^t u_{is} + \varepsilon_{it} = r_{io} + \beta_{it} + v_{it}$$

- ▶ The stationary hypothesis:

$$H_0 : \sigma_u^2 = 0$$

Panel Unit Roots Tests assuming Cross-sectional Independence

- ▶ The LM statistic is given by:

$$LM_1 = \frac{1}{N} \left(\sum_{i=1}^N \frac{1}{T^2} \sum_{t=1}^T S_{it}^2 \right) / \hat{\sigma}_\varepsilon^2$$

- ▶ Where $S_{it} = \sum_{s=1}^t \hat{\varepsilon}_{is}$ are the partial sum of OLS residuals and $\hat{\sigma}_\varepsilon^2 = \frac{1}{NT} \sum_{i=1}^N$ is a consistent estimate of σ_ε^2 under the null Hypotheses

Panel Unit Roots Tests assuming Cross-sectional Independence

- ▶ The alternative LM test that allows for heteroskedasticity across i

$$LM_2 = \frac{1}{N} \left(\sum_{i=1}^N \left(\frac{1}{T^2} \sum_{t=1}^T s_{it}^2 / \hat{\sigma}_{\varepsilon i}^2 \right) \right)$$

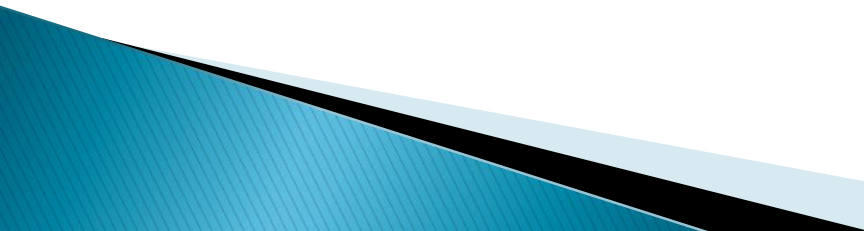
- ▶ The test statistic is given by $Z = \sqrt{N}(LM - \xi_1)/\zeta$ and is asymptotically distributed as $N(0, 1)$

Panel Unit Roots Tests assuming Cross-sectional Independence

General Remarks:

1. The empirical sizes of the IPS and the Fisher test are reasonably close to their nominal size 0.05 when N is small. But the Fisher test shows mild size distortions at $N = 100$, which is expected from the asymptotic theory. Overall, the IPS t -bar test has the most stable size.

Panel Unit Roots Tests assuming Cross-sectional Independence

2. In terms of the size-adjusted power, the Fisher test seems to be superior to the IPS t -bar test.
 3. When a linear time trend is included in the model, the power of all tests decreases considerably.
- 

Panel Unit Roots Tests allowing for Cross-sectional Dependence

Moon and Perron test

- ▶ Null Hypothesis: $H_0: \rho_i = 0$ for all i
- ▶ Alternative Hypothesis: $H_1: \rho_i < 0$ for some i
- ▶ The model is used to capture cross-section correlation, consider the following model:

$$y_{it} = \alpha_i + y_{it}^0$$
$$y_{it}^0 = \rho_i y_{i,t-1}^0 + \epsilon_{it}$$

Panel Unit Roots Tests allowing for Cross-sectional Dependence

- ▶ ϵ_{it} is generated by M unobservable random factors f_t and idiosyncratic shocks e_{it}

$$\epsilon_{it} = \Lambda_i' f_t + e_{it}$$

- ▶ Λ_i are nonrandom factor loading coefficient vectors and the number of factors M is unknown

Panel Unit Roots Tests allowing for Cross-sectional Dependence

- ▶ Test statistic

$$t_a = \frac{\sqrt{NT}(\hat{\rho}_{pool}^+ - 1)}{\sqrt{\frac{2\phi_e^4}{w_e^4}}}$$

- ▶ The pooled bias-correlated estimate of ρ is

$$\hat{\rho}_{pool}^+ = \frac{\text{tr}(Y_{-1} Q_{\Lambda} Y') - NT \lambda_e^N}{\text{tr}(Y_{-1} Q_{\Lambda} Y'_{-1})}$$

- ▶ where Y is a $T \times N$ matrix of the data, Y_{-1} contains lagged values

Panel Unit Roots Tests allowing for Cross-sectional Dependence

- ▶ Asymptotic distribution:

$$t_a \Rightarrow N(0, 1)$$

- ▶ where $N \rightarrow \infty$ and $T \rightarrow \infty$ such that $N/T \rightarrow 0$