Nonstationary Panels


Ulrich Gunter, University of Vienna

UK 40871: Econometric Methods for Panel Data

June 3, 2009
Agenda

1 Spurious Regressions in Panel Data
Agenda

1. Spurious Regressions in Panel Data

2. Panel Cointegration Tests
   - Residual-based DF and ADF Tests
   - Finite Sample Properties
Agenda

1. Spurious Regressions in Panel Data

2. Panel Cointegration Tests
   - Residual-based DF and ADF Tests
   - Finite Sample Properties

3. Estimation and Inference in Panel Cointegration Models
Spurious Regressions in Panel Data

• **ENTORF (1997):** For $T \to \infty$ and $N$ finite, nonsense regression phenomenon holds for spurious fixed effects

⇒ This implies seemingly significant $t$-statistics and high $R^2$ in case of FE estimation

• **PHILLIPS/MOON (1999):** Long-run variance matrix of two unit-root nonstationary variables $y_t, X_t$:

$$
\Omega = \begin{pmatrix}
\Omega_{yy} & \Omega_{yx} \\
\Omega_{xy} & \Omega_{xx}
\end{pmatrix}
$$
• When $\Omega$ is rank-deficient, long-run regression coefficient $\beta = \Omega_{yx}\Omega_{xx}^{-1}$ can be interpreted as cointegrating vector since linear combination $y_t - \beta X_t$ is stationary.

• **Phillips/Moon (1999):** Extend above concept to panel regressions with nonstationary data.

• Heterogeneity across individuals $i$ can be characterized by heterogeneous long-run covariance matrices $\Omega_i$, randomly drawn from population with mean $E[\Omega_i] = \Omega$.

\[
\Rightarrow \quad \beta = E[\Omega_{yi}x_i]E[\Omega_{xi}x_i]^{-1} = \Omega_{yx}\Omega_{xx}^{-1}\]
Hence, we get a fundamental framework for studying sequential and joint limit theories in nonstationary panel data, which allows for four cases:

1. Panel spurious regression

2. Heterogeneous panel cointegration

3. Homogeneous panel cointegration

4. Near-homogeneous panel cointegration
For all four cases, **Phillips/Moon (1999)** find that pooled OLS estimator is consistent and has normal limiting distribution:

- \( \hat{\beta} \) is \( \sqrt{N} \)-consistent for \( \beta \) and has a normal limiting distribution for spurious panel regressions and cross-section regressions with time-averaged data under quite weak regularity conditions.

- This is different to OLS in pure time-series analysis, where \( \hat{\beta} \) has a functional of Brownian motions as limiting distribution and is therefore not consistent for \( \beta \).

\( \Rightarrow \) Idea in **Phillips/Moon (1999)**: Independent cross-section data in panels add information compared to pure time-series data.
Panel Cointegration Tests

Economists pool data on similar countries such as G7, OECD, or EU to increase power of unit-root or cointegration tests in case they want to test for issues such as convergence of growth or purchasing power parity

Tests with two opposing null hypotheses:

1. **Null of no cointegration**: e.g. residual-based Dickey-Fuller (DF) and Augmented Dickey-Fuller (ADF) tests (see Kao 1999)

2. **Null of cointegration**: e.g. residual-based LM tests (see McCoskey/Kao 1998), Pedroni tests (see Pedroni 2000, 2004), or likelihood-based cointegration tests (see Larsson et al. 2001)
Residual-based DF and ADF Tests

• Panel regression model:

\[ y_{it} = x_{it}'\beta + z_{it}'\gamma + e_{it} \]

where \( y_{it}, x_{it} \) are \( I(1) \) and non-cointegrated

• For \( z_{it} = \{\mu_i\} \), KAO (1999) proposes DF- and ADF-type tests under null of no cointegration, which can be calculated from FE residuals:

\[ \hat{e}_{it} = \rho\hat{e}_{it-1} + \nu_{it} \]

where \( \hat{e}_{it} = \tilde{y}_{it} - \tilde{x}_{it}'\beta, \tilde{y}_{it} = y_{it} - \bar{y}_i \).
• $H_0$ of no cointegration corresponds to $\rho = 1$

• OLS estimate of $\rho$ and corresponding $t$-statistic $t_{\hat{\rho}}$:

$$\hat{\rho} = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} \hat{e}_{it} \hat{e}_{it-1}}{\sum_{i=1}^{N} \sum_{t=2}^{T} \hat{e}_{it}^2}$$

$$t_{\hat{\rho}} = \frac{(\hat{\rho} - 1) \sqrt{\sum_{i=1}^{N} \sum_{t=2}^{T} \hat{e}_{it-1}^2}}{s_e}$$

where $s_e^2 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} (\hat{e}_{it} - \hat{\rho} \hat{e}_{it-1})^2$
Kao (1999) proposes four DF-type tests based on $\hat{\rho}$ or $t\hat{\rho}$:

- $DF_\rho = \frac{\sqrt{NT}(\hat{\rho} - 1) + 3\sqrt{N}}{\sqrt{10.2}}$
- $DF_t = \sqrt{1.25}t\hat{\rho} + \sqrt{1.875}N$
- $DF^*_\rho = \frac{\sqrt{NT}(\hat{\rho} - 1) + \frac{3\sqrt{N}\hat{\sigma}^2_\nu}{\hat{\sigma}^2_{0\nu}}}{\sqrt{3 + \frac{36\hat{\sigma}^4_\nu}{5\hat{\sigma}^4_{0\nu}}}}$
- $DF^*_t = \frac{t\hat{\rho} + \frac{\sqrt{6NT\hat{\sigma}_\nu}}{2\hat{\sigma}_{0\nu}}}{\sqrt{\frac{\hat{\sigma}^2_{0\nu}}{2\hat{\sigma}^2_\nu} + \frac{3\hat{\sigma}^2_\nu}{10\hat{\sigma}^2_{0\nu}}}}$

where $\hat{\sigma}_\nu^2 = \hat{\Sigma}_{yy} - \hat{\Sigma}_{yx} \hat{\Sigma}_{xx}^{-1}$, $\hat{\sigma}^2_{0\nu} = \hat{\Omega}_{yy} - \hat{\Omega}_{yx} \hat{\Omega}_{xx}^{-1}$
Residual-based DF and ADF Tests

- \(DF_\rho, DF_t\) are based on strongly exogenous regressors and errors, where \(DF^*_\rho, DF^*_t\) are based on an endogenous relationship between regressors and errors.

- ADF-type test based on following regression and null of no cointegration:
  \[
  \hat{e}_{it} = \rho \hat{e}_{it-1} + \sum_{j=1}^{p} \theta_j \Delta \hat{e}_{it-j} + \nu_{itp}
  \]
  \[
  \Rightarrow ADF = t_{ADF} + \frac{\sqrt{6N\hat{\sigma}_\nu}}{2\hat{\sigma}_{0\nu}}
  \]
  \[
  \sqrt{\frac{\hat{\sigma}_{0\nu}^2}{2\hat{\sigma}_\nu^2} + \frac{3\hat{\sigma}_\nu^2}{10\hat{\sigma}_{0\nu}^2}}
  \]
  
  - O.c.s. that asymptotic distributions of \(DF_\rho, DF_t, DF^*_\rho, DF^*_t\), and \(ADF\) converge to \(N(0, 1)\).
Finite Sample Properties

- **McCoskey/Kao (1999)** conduct Monte Carlo experiments to compare statistical power of different residual-based cointegration tests.

- Reassessing null of no cointegration proposed because of low statistical power of tests associated with null of no cointegration, especially in time-series cases (near observation equivalence problem).

⇒ Authors find that in cases, where economic theory predicts long-run steady-state relationships between two $I(1)$ variables, null of cointegration is more appropriate than null of no cointegration.
Estimation and Inference in Panel Cointegration Models

- **Kao/Chiang** (2000) consider the following panel regression model:

\[ y_{it} = x_{it}' \beta + z_{it}' \gamma + u_{it} \]

where \( x_{it} = x_{it-1} + \varepsilon_t \) are cross-sectionally independent \( I(1) \) processes.

\[ \Rightarrow y_{it} \text{ and } x_{it} \text{ are cointegrated} \]

- Corresponding OLS estimator:

\[ \hat{\beta}_{OLS} = \left( \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{x}_{it}' \right)^{-1} \left( \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{y}_{it} \right) \]
• O.c.s. that $\hat{\beta}_{OLS}$ is inconsistent using panel data, which is different to pure time-series cases

• Choi (2002) uses subsequent panel regression model:

$$y_{it} = \alpha + \beta x_{it} + u_{it}$$

where $x_{it}$ is nearly nonstationary, $u_{it}$ is $I(0)$ and $z_{it}$ is instrumental variable

• Corresponding panel IV estimator:

$$\hat{\beta}_{IV} = \left[ \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_{i.})(z_{it} - \bar{z}_{i.}) \right]^{-1} \left[ \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \bar{y}_{i.})(z_{it} - \bar{z}_{i.}) \right]$$
O.c.s. that $\hat{\beta}_{IV}$ is asymptotically normally distributed, which is different to pure time-series cases.

References