Panel Data Methods using Stata
4B: Panels - Cointegration

© A. Colin Cameron
Univ. of Calif. Davis

November 2013

1. Introduction

- Now consider panel models with dynamics.
- Especially panel regression of an $I(1)$ y_t on an $I(1)$ x_t when they are cointegrated
  - i.e. there is a long-run relationship between them.
- For time series the standard way is systems estimation
  - joint estimation of systems model for y_t and x_t
  - use Johansen's method based on a vector error correction model.
- For panel data a systems approach is generally not feasible
  - instead single-equation methods
  - there are several different methods, most notably dynamic OLS and fully-modified OLS.
- Recent concern is correlation across panels.
- Eviews 8 has panel cointegration tests and estimators
  - Stata has none - there are a few addons.

Outline

2. Cointegration in Time Series

- A process is integrated of order d, denoted $I(d)$, if when differenced d times it yields a stationary series
  - a white noise process is I(0)
  - an AR(1) process with $|\rho| < 1$ is I(0), as is MA(q)
  - a random walk is I(1) since $y_t = y_{t-1} + \varepsilon_t$ implies $\Delta y_t = \varepsilon_t$.
- The I(1) processes x_t and y_t are cointegrated if a linear combination of them is I(0)
  - i.e. $y_t - \alpha - \beta x_t$ is I(0).
- We consider case where x_t and y_t are I(1) and we OLS regress

$$y_t = \alpha + \beta x_t + u_t.$$

  - Complications exists if x_t and y_t are not cointegrated
  - Complications exists if x_t and y_t are cointegrated.

Spurious Regressions

- Suppose x_t and y_t are I(1) and y_t and x_t are not cointegrated $(\beta = 0)$
- Then the OLS estimator is inconsistent
  - $\widehat{\beta}$ does not converge to zero
  - conventional methods will reject $H_0 : \beta = 0$ much too often and find "spurious" regressions
  - this extends to dynamic models with lags
- Analogous to unrelated deterministic trends
  - if $y_t = \delta t + u_t$ and $x_t = \gamma t + v_t$ where $\mathrm{Cor}[u_t, v_t] = 0$
  - then OLS $y_t = \alpha + \beta x_t + u_t$ (i.e. without t as regressor) has $\widehat{\beta} \overset{p}{\nrightarrow} 0$.

Cointegrated Regressions

- Suppose x_t and y_t are I(1) and y_t and x_t are cointegrated $(\beta \neq 0)$
- Then the OLS estimator is consistent even though x_t is endogenous!
  - OLS converges to β but at rate T not $\sqrt{T}$
  - in finite samples, however, $\widehat{\beta}$ can be very biased
  - and it can be inefficient because u_t is serially correlated.
- There are several better ways to estimate cointegrated relationships (and to test for cointegration)
  - these come from different representations of the model
  - most notably Granger-Representation Theorem and Stock and Watson common trends
- Big distinction between single equation and systems methods.

Systems Testing and Estimation

- Let $\mathbf{y}_t$ denote an $m \times 1$ vector process (before $m = 2$ and $\mathbf{y}_t = [y_t \; x_t]'$)
- Consider a VAR (vector autoregression) with p lags

$$\underset{m\times 1}{\mathbf{y}_t} = \underset{k\times 1}{\mathbf{C}}\underset{m\times k}{\mathbf{D}_t'} + \sum_{j=1}^{p} \underset{m\times m}{\mathbf{A}_j}\;\underset{m\times 1}{\mathbf{y}_{t-j}} + \underset{m\times 1}{\mathbf{u}_t}$$

  - where $\mathbf{D}_t$ are exogenous (and could include deterministic trends)
  - $\mathbf{u}_t$ is serially uncorrelated error

- Rewrite equivalently as

$$\begin{aligned} \Delta\mathbf{y}_t &= \mathbf{C}\mathbf{D}_t' + \mathbf{\Pi}\mathbf{y}_{t-1} + \sum_{j=1}^{p-1} \mathbf{\Gamma}_j \Delta\mathbf{y}_{t-j} + \mathbf{u}_t \\ \mathbf{\Pi} &= \sum_{j=1}^{p} \mathbf{A}_j - \mathbf{I}; \quad \mathbf{\Gamma}_j = -\sum_{l=j+1}^{p} \mathbf{A}_l \end{aligned}$$

Cointegration in Systems

- Now bring in cointegration
- If there are r cointegrating relationships then
  - $\text{rank}(\mathbf{\Pi}) = r$ so $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $m \times r$ with rank r
  - so model as VAR in first difference with $\mathbf{y}_{t-1}$ also regressor
- Johansen (1988) method
  - determine number of cointegrating relationships via rank of $\mathbf{\Pi}$
  - then estimate subject to restriction $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$.
- Other related systems methods
  - Ahn and Reinsel (1988) similar to Johansen
  - Stock and Watson (1988) common trends
  - Park and Phillips (1988)

Single-equation Methods

- Model is $y_t = \mathbf{d}_t'\delta + \mathbf{x}_t'\beta + u_t$
  - y_t is I(1)
  - $\mathbf{d}_t$ are deterministic I(0) variables
  - $\mathbf{x}_t$ are m I(1) variables and are not cointegrated - often $m = 1$
  - the innovations $\Delta\mathbf{x}_t$ may be correlated with u_t - endogenous regressors
  - there is at most one cointegrating relationship between y_t and $\mathbf{x}_t$
- The direct regression is the levels estimator
  - given cointegration $\widehat{\beta}$ is superconsistent (at rate T not $\sqrt{T}$)
  - despite $\mathbf{x}_t$ endogenous and u_t serially correlated
  - but $\widehat{\beta}$ is biased for finite T.

Error Correction Model

- A better estimator uses the error correction model (ECM) form
  - for case $m = 1$ with $y_t = \delta + \alpha y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + u_t$
  - then in long run $y = \delta + \lambda x$ where $\lambda = \frac{\beta_0 + \beta_1}{1 - \alpha_1}$
  - rewrite as ECM: $\Delta y_t = \delta + (\alpha - 1)(y_{t-1} - \lambda x_{t-1}) + \beta_1 \Delta x_t + u_t$
  - like model in first differences except add "error correction term" $y_{t-1} - \lambda x_{t-1}$
  - And estimate above by OLS or equivalently Δy_t on $y_{t-1}, x_{t-1}, \Delta x_t$
  - More generally OLS of Δy_t on $\mathbf{d}_t, y_{t-1}, \mathbf{x}_{t-1}, \boldsymbol{\Delta} x_t$
  - Better than levels but not best

Fully Modified and Dynamic OLS Tests

- Fully modified OLS (Phillips and Hansen (1990)
  - adjusts the levels estimator for regressor endogeneity and error serial correlation
- Dynamic OLS (Saikkonen (1991))
  - can show $u_t = \sum_{k=-\infty}^{\infty} \boldsymbol{\Delta} x'_{t+k}\gamma_k + v_t$ where v_t is orthogonal to all leads and lags of $\boldsymbol{\Delta} x_t$
  - then can do OLS in $y_t = \mathbf{d}'_t\delta + \mathbf{x}'_t\beta + \sum_{k=-\infty}^{\infty} \boldsymbol{\Delta} x'_{t+k}\gamma_k + v_t$
  - in practice need to truncate and to e.g. $k = -2$ to 2.

Residual-based Cointegration Tests

- If y_t and $\mathbf{x}_t$ are cointegrated then $u_t = y_t - \mathbf{d}_t'\delta - \mathbf{x}_t'\boldsymbol{\beta} \sim I(0)$
  - so get residuals $\widehat{u}_t$ from OLS of $y_t = \mathbf{d}_t'\delta + \mathbf{x}_t'\boldsymbol{\beta} + u_t$
  - perform the usual unit root tests on these residuals
  - will need to adjust the statistics / critical values due to estimation of $\boldsymbol{\beta}$
  - for most tests $H_0 : u_t \sim I(1)$ so no cointegration vs H_a: cointegration.
- Engle and Granger (1987) do ADF tests on the residuals $\widehat{u}_t$
- Phillips and Ouliaris (1990) do Phillips-Perron adjustment on the residuals $\widehat{u}_t$.
- Hamilton (1994) Tables B.8 and B.9 give the critical values.

3. Panel Cointegration Tests

- Tests are usually single-equation tests, mostly residual-based
  - usually H_0 : unit root residual (no cointegration)
- Pedroni (1994, 2004) residual-based ADF tests
  - in most general case $y_{it} = \mathbf{d}'_{it}\delta_i + \mathbf{x}'_{it}\boldsymbol{\beta}_i + u_{it}$
  - OLS gives $\widehat{u}_{it}$
  - ADF test OLS regresses $\widehat{u}_{it} = \rho_i \widehat{u}_{it} + \sum_{k=1}^{K} \gamma_{ik} \Delta \widehat{u}_{i,t-k}$
  - asymptotics are $N \rightarrow \infty$ and then $T \rightarrow \infty$
  - get various tests depending on $\mathbf{d}_{it}$, heterogeneity of ρ_i, ...
  - since $N \rightarrow \infty$ in each case the statistic $\overset{d}{\rightarrow} N[0,1]$ upon appropriate recentering and scaling.
- Kao (1998) residual-based ADF tests
  - similar to Pedroni except constrain $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ at the first step.

Panel Cointegration Tests

- McCoskey and Kao (1998) residual-based KPSS (LM) tests
  - similar to Hadri (2000) for panel unit roots
  - here H_0 : stationary residual (cointegration)
- Maddala and Wu (1999) Fisher/Johanson combined systems tests
  - do separate Johanson systems tests for each cross-section
  - combine p-values as in Fisher-type test.
- Westerlund (2007, 2008) tests based on error correction model (ECM)
  - $\Delta y_{it} = \mathbf{d}_t'\delta_i + \alpha_i(y_{i,t-1} - \mathbf{x}_{i,t-1}'\boldsymbol{\beta}_i) + \sum_{k=1}^{K_i} \alpha_{ij}\Delta y_{i,t-k}$ $+ \sum_{k=-J_i}^{K_i} \Delta\mathbf{x}_{i,t-k}'\gamma_{ij} + \varepsilon_{it}$, where iid
  - $\alpha_i(y_{i,t-1} - \mathbf{x}_{i,t-1}'\boldsymbol{\beta}_i) = \alpha_i y_{i,t-1} - \mathbf{x}_{i,t-1}'\alpha_i\boldsymbol{\beta}_i = \alpha_i y_{i,t-1} - \mathbf{x}_{i,t-1}'\boldsymbol{\lambda}_i$
  - test $H_0 : \alpha_i = 0$ no cointegration as then y_{it} does not respond to the shock $y_{i,t-1} - \mathbf{x}_{i,t-1}'\boldsymbol{\beta}$
  - group mean tests allow $\alpha_i' s$ to differ and panel tests have $\alpha_i = \alpha$
- Eviews 8 does Pedroni tests, Kao tests and Maddala-Wu.
- Stata 12 does not do panel cointegration tests
  - Stata addon `xtwest` does Westerlund tests.

4. Panel Cointegration Estimation

- Use single equations estimators
  - Panel fully modified estimators - Phillips and Moon (1999), Pedroni (2000)
  - Panel dynamic OLS estimators - Kao and Chiang (2000)
  - Baltagi chapter 12.6 has short discussion
  - Kao, Chiang and Chen (2001) is good application
  - Eviews 8 does both panel dynamic and fully modified.
- Easiest is panel dynamic OLS
  - Estimate $y_{it} = \mathbf{d}_t'\delta_i + \mathbf{x}_{it}'\beta + \sum_{k=-J_i}^{K_i} \Delta\mathbf{x}_{i,t+k}'\gamma_{i,k} + v_{it}$
  - the coefficients β are homogeneous while others are heterogeneous
  - Stata add-on `xtdolshm` does panel dynamic OLS
- Markus Eberhardt has references to Stata code on his website
  - concerned with both heterogeneity and stationarity

Panel Dynamic OLS Example

- Kao, Chiang and Chen (2001) data 22 countries 1970-89
  - ltfp - Log of Total Factor Productivity
  - lrd - Log of domestic R&D capital stock (logSd)
  - lfrd - Log of import-share weighted foreign R&D capital stock (logSf)
  - $k = -2, -1, 0, 1$

```
.
. * Dynamic OLS
. xtdolshm ltfp lrd lfrd, nlag(2) nlead(1)


DOLS Hom. Panel data Coint. Estimation results   Number of obs      =       352
Group variable: id                               Number of groups   =        22
Wald chi2(2)   =      56.49                      Obs per group: min =        20
Prob > chi2    =      0.000                                     avg =        20
                                                                max =        20
                                                 R-squared          =    0.5114
                                                 Adj R-squared      =   0.3054

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lrd |   .1068545   .0228706     4.67   0.000      .062029      .15168
        lfrd |   .0558467    .038506     1.45   0.147    -.0196237    .1313171
------------------------------------------------------------------------------
```

5. Summary

- For panel cointegration the methods are not well-established.
- Eviews 8 has several single-equation commands for cointegration testing and estimation.
- Stata has some add-ons
  - `xtwest` for cointegration test
  - `xtdolshm` for estimation when cointegrated.
  - Marcus Eberhardt's website https://sites.google.com/site/medevecon/home has many useful links to papers, data and Stata code, including Stata add-ons, for both cross-sectional dependence and unit roots.

6. References

- Panel Unit Roots and Cointegration Surveys

  - Banerjee, A. (1999), "Panel Data Unit Roots and Cointegration: An Overview," *Oxford Bulletin of Statistics*, 607-629. Very nice exposition of the literature to 1999. http://onlinelibrary.wiley.com/doi/10.1111/1468-0084.0610s1607/pdf
  - Breitung, J., and M.H. Pesaran (2008), "Unit Roots and Cointegration in Panels," in L. Matyas and P. Sevestre (Eds.), *The Econometrics of Panel Data*, 279–322. Kluwer Academic Publishers. More recent and concerned about cross-section correlation. http://www.econ.cam.ac.uk/dae/repec/cam/pdf/cwpe0535.pdf
  - Smith R.P., and A-M. Fuertes (2010), "Panel Time-Series." Lengthy survey used for a course taught at cemmap. http://www.ems.bbk.ac.uk/faculty/smith/RSpanel.pdf
  - Baltagi, B. (2008), Econometric Analysis of Panel Data, fourth edition, chapter 12.

- Panel Cointegration Tests

  - McCoskey, S. and C. Kao (1998), "A residual-based test of the null of cointegration in panel data," Econometric Reviews, 17, 57-84.
  - Pedroni, P. (2004), "Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis," Econometric Theory, 20, 597-265.
  - Westerlund, J. (2007), "Testing for Error Correction in Panel Data," Oxford Bulletin of Economics and Statistics, 69, 709-748.
  - Westerlund, J. and D. Persyn (2008), "Error Correction Based Cointegration Tests for Panel Data," Stata Journal, 8, 232-241.

- Panel Cointegration Estimation

  - Kao, C.-H., M.-H, Chiang, and B. Chen (2002), "International R&D Spillovers: An Application of Estimation and Inference in Panel Cointegration," Oxford Bulletin of Economics and Statistics, 61, 691-709.
  - Kao, C., and M.-H. Chiang (2001), "On the Estimation and Inference of a Cointegrated Regression in Panel Data," Advances in Econometrics, 15, 179-222.
  - Phillips, P.C.B. and H. Moon (1999), "Linear Regression Limit Theory for Nonstationary Panel Data," Econometrica, 67, 1057-1111.

- Cross-Sectional Dependence in Panels with Cointegration

  - Bai, J. (2009), "Panel data models with interactive fixed effects," *Econometrica*, 1229–1279.
    http://onlinelibrary.wiley.com/doi/10.3982/ECTA6135/pdf
  - Eberhardt, M., C. Helmers and H. Strauss (2012), "Do Spillovers Matter When Estimating Private Returns to R&D?," *Review of Economics and Statistics*, forthcoming. Applies CCE and MG estimators.
    http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1966020
  - Marcus Eberhardt's website
    https://sites.google.com/site/medevecon/home This has many useful links to papers, data and Stata code, including Stata add-ons, for both cross-sectional dependence and unit roots.
  - Pedroni, P. (2001), "Fully modified OLS for heterogeneous cointegrated panels," Advances in Econometrics, 15, 93-130.