Large Dimensional Data in Econometrics: The Curses and Blessings

Chihwa Kao

Syracuse University

June 2009

Introduction

- Century of large dimension data.

Introduction

- Century of large dimension data.
- Multivariate statistics: assumes that $\frac{n}{T}$ is small, e.g., assume n is fixed. n could be the number of subjects and T could be the number of observations. Anderson (2003).

Introduction

- Century of large dimension data.
- Multivariate statistics: assumes that $\frac{n}{T}$ is small, e.g., assume n is fixed. n could be the number of subjects and T could be the number of observations. Anderson (2003).
- Micro panel models: n is large and T is small. Hsiao (2003).

Introduction

- Century of large dimension data.
- Multivariate statistics: assumes that $\frac{n}{T}$ is small, e.g., assume n is fixed. n could be the number of subjects and T could be the number of observations. Anderson (2003).
- Micro panel models: n is large and T is small. Hsiao (2003).
- Large panel models $(n, T) \to \infty$ and $\frac{n}{T} \to 0$ e.g., my contribution

Introduction

- Century of large dimension data.
- Multivariate statistics: assumes that $\frac{n}{T}$ is small, e.g., assume n is fixed. n could be the number of subjects and T could be the number of observations. Anderson (2003).
- Micro panel models: n is large and T is small. Hsiao (2003).
- Large panel models $(n, T) \to \infty$ and $\frac{n}{T} \to 0$ e.g., my contribution
  - Kao (1999), *Journal of Econometrics*. Panel spurious regression

Introduction

- Century of large dimension data.
- Multivariate statistics: assumes that $\frac{n}{T}$ is small, e.g., assume n is fixed. n could be the number of subjects and T could be the number of observations. Anderson (2003).
- Micro panel models: n is large and T is small. Hsiao (2003).
- Large panel models $(n, T) \to \infty$ and $\frac{n}{T} \to 0$ e.g., my contribution
  - Kao (1999), *Journal of Econometrics*. Panel spurious regression
  - Hong and Kao (2004), *Econometrica*: Testing serial correlation using wavelet in panel

Introduction

- Century of large dimension data.
- Multivariate statistics: assumes that $\frac{n}{T}$ is small, e.g., assume n is fixed. n could be the number of subjects and T could be the number of observations. Anderson (2003).
- Micro panel models: n is large and T is small. Hsiao (2003).
- Large panel models $(n, T) \to \infty$ and $\frac{n}{T} \to 0$ e.g., my contribution
  - Kao (1999), *Journal of Econometrics*. Panel spurious regression
  - Hong and Kao (2004), *Econometrica*: Testing serial correlation using wavelet in panel
  - Bai, Kao and Ng (2009), *Journal of Econometrics*. Panel cointegration with common factors

Introduction

- Century of large dimension data.
- Multivariate statistics: assumes that $\frac{n}{T}$ is small, e.g., assume n is fixed. n could be the number of subjects and T could be the number of observations. Anderson (2003).
- Micro panel models: n is large and T is small. Hsiao (2003).
- Large panel models $(n, T) \to \infty$ and $\frac{n}{T} \to 0$ e.g., my contribution
  - Kao (1999), *Journal of Econometrics*. Panel spurious regression
  - Hong and Kao (2004), *Econometrica*: Testing serial correlation using wavelet in panel
  - Bai, Kao and Ng (2009), *Journal of Econometrics*. Panel cointegration with common factors
- Today, it is common for n to be large or even huge, so that $\frac{n}{T} \to c$ where $c \in [0, \infty]$.

Introduction

- Century of large dimension data.
- Multivariate statistics: assumes that $\frac{n}{T}$ is small, e.g., assume n is fixed. n could be the number of subjects and T could be the number of observations. Anderson (2003).
- Micro panel models: n is large and T is small. Hsiao (2003).
- Large panel models $(n, T) \to \infty$ and $\frac{n}{T} \to 0$ e.g., my contribution
  - Kao (1999), *Journal of Econometrics*. Panel spurious regression
  - Hong and Kao (2004), *Econometrica*: Testing serial correlation using wavelet in panel
  - Bai, Kao and Ng (2009), *Journal of Econometrics*. Panel cointegration with common factors
- Today, it is common for n to be large or even huge, so that $\frac{n}{T} \to c$ where $c \in [0, \infty]$.
- New challenges and opportunities!

- Today, it is common for n to be large or even huge, so that $\frac{n}{T} \longrightarrow c$ where $c \in [0, \infty]$.

- Today, it is common for n to be large or even huge, so that $\frac{n}{T} \longrightarrow c$ where $c \in [0, \infty]$.
- New challenges and opportunities!

Signal vs Noise

- Data is “too” noisy comparing to the signal

Signal vs Noise

- Data is “too” noisy comparing to the signal
- Model is not identifiable

Signal vs Noise

- Data is “too” noisy comparing to the signal
- Model is not identifiable
- Estimator (of parameter) is inconsistent

Signal vs Noise

- Data is “too” noisy comparing to the signal

- Model is not identifiable
- Estimator (of parameter) is inconsistent
- Test statistic

Signal vs Noise

- Data is "too" noisy comparing to the signal
- Model is not identifiable
- Estimator (of parameter) is inconsistent
- Test statistic
  - may have incorrect size

Signal vs Noise

- Data is “too” noisy comparing to the signal
- Model is not identifiable
- Estimator (of parameter) is inconsistent
- Test statistic
  - may have incorrect size
  - Test statistic is not consistent, i.e., the power will not go to ∞ under the alternative

Examples of Large Dimensional Data

Example

Climate studies: T might be the number of time points and n the number of observation stations. $\frac{n}{T}$ is moderate.

Example

Financial data: large covariance and correlation matrices, with $n \approx 400$ financial indicators are publicly posted daily and used for value-at-risk calculations.

Example

Information Retrieval/search engines: A common search engine strategy forms huge term by document incidence matrices (n and T at least in the thousands).

Challenges and Opportunities

- With the rapid development of computing power, econometricians/statisticians are beginning to discuss/use large dimensional data.

Challenges and Opportunities

- With the rapid development of computing power, econometricians/statisticians are beginning to discuss/use large dimensional data.
- New research areas, theory and applications, in economics, finance, and econometrics!

Orientation

- Classical panel econometrics is a long-established field:

Orientation

- Classical panel econometrics is a long-established field:
  - E.g., Textbooks by Hsiao (2003) and Baltagi (2005).

Orientation

- Classical panel econometrics is a long-established field:
  - E.g., Textbooks by Hsiao (2003) and Baltagi (2005).
  - Large n (cross-sectional dimension) and small/fixed T (time series dimension).

Orientation

- Classical panel econometrics is a long-established field:
  - E.g., Textbooks by Hsiao (2003) and Baltagi (2005).
  - Large n (cross-sectional dimension) and small/fixed T (time series dimension).
  - No cross-sectional dependence.

Orientation

- Classical panel econometrics is a long-established field:
  - E.g., Textbooks by Hsiao (2003) and Baltagi (2005).
  - Large n (cross-sectional dimension) and small/fixed T (time series dimension).
  - No cross-sectional dependence.
- Random matrix theory (RMT)

Orientation

- Classical panel econometrics is a long-established field:
  - E.g., Textbooks by Hsiao (2003) and Baltagi (2005).
  - Large n (cross-sectional dimension) and small/fixed T (time series dimension).
  - No cross-sectional dependence.
- Random matrix theory (RMT)
  - nuclear physics 1950's, now many areas of mathematics and probability.

Orientation

- Classical panel econometrics is a long-established field:
  - E.g., Textbooks by Hsiao (2003) and Baltagi (2005).
  - Large n (cross-sectional dimension) and small/fixed T (time series dimension).
  - No cross-sectional dependence.
- Random matrix theory (RMT)
  - nuclear physics 1950's, now many areas of mathematics and probability.
  - Mehta, M. L. (1990), *Random Matrices*, Academic Press, New York.

Orientation

- Classical panel econometrics is a long-established field:
  - E.g., Textbooks by Hsiao (2003) and Baltagi (2005).
  - Large n (cross-sectional dimension) and small/fixed T (time series dimension).
  - No cross-sectional dependence.
- Random matrix theory (RMT)
  - nuclear physics 1950's, now many areas of mathematics and probability.
  - Mehta, M. L. (1990), *Random Matrices*, Academic Press, New York.
  - Bai, Z. D. (1999). "Methodologies in Spectral Analysis of Large Dimensional Random Matrices, A Review," *Statistical Sinica*, 9, 611-677.

Orientation

- Classical panel econometrics is a long-established field:
  - E.g., Textbooks by Hsiao (2003) and Baltagi (2005).
  - Large n (cross-sectional dimension) and small/fixed T (time series dimension).
  - No cross-sectional dependence.
- Random matrix theory (RMT)
  - nuclear physics 1950's, now many areas of mathematics and probability.
  - Mehta, M. L. (1990), *Random Matrices*, Academic Press, New York.
  - Bai, Z. D. (1999). "Methodologies in Spectral Analysis of Large Dimensional Random Matrices, A Review," *Statistical Sinica*, 9, 611-677.
- Large panels in Econometrics - large n and large T: Requires $\frac{n}{T} \to 0$ (This means T should be much larger than n) as $(n, T) \to 0$.

Orientation

- Classical panel econometrics is a long-established field:
  - E.g., Textbooks by Hsiao (2003) and Baltagi (2005).
  - Large n (cross-sectional dimension) and small/fixed T (time series dimension).
  - No cross-sectional dependence.
- Random matrix theory (RMT)
  - nuclear physics 1950's, now many areas of mathematics and probability.
  - Mehta, M. L. (1990), *Random Matrices*, Academic Press, New York.
  - Bai, Z. D. (1999). "Methodologies in Spectral Analysis of Large Dimensional Random Matrices, A Review," *Statistical Sinica*, 9, 611-677.
- Large panels in Econometrics - large n and large T: Requires $\frac{n}{T} \to 0$ (This means T should be much larger than n) as $(n, T) \to 0$.
  - Nonstationary panel: Kao (1999), Phillips and Moon (1999).

Orientation

- Classical panel econometrics is a long-established field:
  - E.g., Textbooks by Hsiao (2003) and Baltagi (2005).
  - Large n (cross-sectional dimension) and small/fixed T (time series dimension).
  - No cross-sectional dependence.
- Random matrix theory (RMT)
  - nuclear physics 1950's, now many areas of mathematics and probability.
  - Mehta, M. L. (1990), *Random Matrices*, Academic Press, New York.
  - Bai, Z. D. (1999). "Methodologies in Spectral Analysis of Large Dimensional Random Matrices, A Review," *Statistical Sinica*, 9, 611-677.
- Large panels in Econometrics - large n and large T: Requires $\frac{n}{T} \to 0$ (This means T should be much larger than n) as $(n, T) \to 0$.
  - Nonstationary panel: Kao (1999), Phillips and Moon (1999).
  - Large factor model:

Orientation

- Classical panel econometrics is a long-established field:
  - E.g., Textbooks by Hsiao (2003) and Baltagi (2005).
  - Large n (cross-sectional dimension) and small/fixed T (time series dimension).
  - No cross-sectional dependence.
- Random matrix theory (RMT)
  - nuclear physics 1950's, now many areas of mathematics and probability.
  - Mehta, M. L. (1990), *Random Matrices*, Academic Press, New York.
  - Bai, Z. D. (1999). "Methodologies in Spectral Analysis of Large Dimensional Random Matrices, A Review," *Statistical Sinica*, 9, 611-677.
- Large panels in Econometrics - large n and large T: Requires $\frac{n}{T} \longrightarrow 0$ (This means T should be much larger than n) as $(n, T) \longrightarrow 0$.
  - Nonstationary panel: Kao (1999), Phillips and Moon (1999).
  - Large factor model:
    - Bai, J. (2003), "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135-171.

Orientation

- Classical panel econometrics is a long-established field:
  - E.g., Textbooks by Hsiao (2003) and Baltagi (2005).
  - Large n (cross-sectional dimension) and small/fixed T (time series dimension).
  - No cross-sectional dependence.
- Random matrix theory (RMT)
  - nuclear physics 1950's, now many areas of mathematics and probability.
  - Mehta, M. L. (1990), *Random Matrices*, Academic Press, New York.
  - Bai, Z. D. (1999). "Methodologies in Spectral Analysis of Large Dimensional Random Matrices, A Review," *Statistical Sinica*, 9, 611-677.
- Large panels in Econometrics - large n and large T: Requires $\frac{n}{T} \to 0$ (This means T should be much larger than n) as $(n, T) \to 0$.
  - Nonstationary panel: Kao (1999), Phillips and Moon (1999).
  - Large factor model:
    - Bai, J. (2003), "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135-171.
    - Bai, J., Kao, C., and Serena Ng, S. (2009), "Panel Cointegration with

Large Covariance Matrices

- Sample covariance matrix is fundamental to econometrics/multivariate statistics.

Large Covariance Matrices

- Sample covariance matrix is fundamental to econometrics/multivariate statistics.
- Classical Results: When the number of subjects, say n, is fixed, as the the number of samples (say T) tends to infinity, the sample covariance matrix is a good approximate of the population covariance matrix.

Large Covariance Matrices

- Sample covariance matrix is fundamental to econometrics/multivariate statistics.
- Classical Results: When the number of subjects, say n, is fixed, as the the number of samples (say T) tends to infinity, the sample covariance matrix is a good approximate of the population covariance matrix.
- Still OK, when $\frac{n}{T} \to 0$.

Large Covariance Matrices

- Sample covariance matrix is fundamental to econometrics/multivariate statistics.
- Classical Results: When the number of subjects, say n, is fixed, as the the number of samples (say T) tends to infinity, the sample covariance matrix is a good approximate of the population covariance matrix.
- Still OK, when $\frac{n}{T} \longrightarrow 0$.
- However, when n is large and comparable with T, $\frac{n}{T} \longrightarrow c$, it is known from RMT that the sample covariance matrix is no longer a good approximation to the population covariance matrix.

Identification

- What is Consistency? Inconsistency?

Identification

- What is Consistency? Inconsistency?
- Spurious Regression: Phillips (1986). OLS of a spurious regression is inconsistent

$$y_t = \beta x_t + u_t$$

where $y_t \sim I(1)$, $x_t \sim I(1)$, and $u_t \sim I(1)$. y_t and x_t are independent. $\beta = 0$.

Identification

- What is Consistency? Inconsistency?
- Spurious Regression: Phillips (1986). OLS of a spurious regression is inconsistent

$$y_t = \beta x_t + u_t$$

  where $y_t \sim I(1)$, $x_t \sim I(1)$, and $u_t \sim I(1)$. y_t and x_t are independent. $\beta = 0$.

- Not identifiable.

Identification

- What is Consistency? Inconsistency?
- Spurious Regression: Phillips (1986). OLS of a spurious regression is inconsistent

$$y_t = \beta x_t + u_t$$

where $y_t \sim I(1)$, $x_t \sim I(1)$, and $u_t \sim I(1)$. y_t and x_t are independent. $\beta = 0$.

- Not identifiable.
  - $\widehat{\beta} = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2}$.

Identification

- What is Consistency? Inconsistency?
- Spurious Regression: Phillips (1986). OLS of a spurious regression is inconsistent

$$y_t = \beta x_t + u_t$$

where $y_t \sim I(1)$, $x_t \sim I(1)$, and $u_t \sim I(1)$. y_t and x_t are independent. $\beta = 0$.

- Not identifiable.
  - $\widehat{\beta} = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2}$.
  - $[rT] = t$, $\frac{x_t}{\sqrt{T}} \xrightarrow{d} B_x(r)$, $\frac{y_t}{\sqrt{T}} \xrightarrow{d} B_y(r)$

Identification

- What is Consistency? Inconsistency?
- Spurious Regression: Phillips (1986). OLS of a spurious regression is inconsistent

$$y_t = \beta x_t + u_t$$

where $y_t \sim I(1)$, $x_t \sim I(1)$, and $u_t \sim I(1)$. y_t and x_t are independent. $\beta = 0$.

- Not identifiable.
  - $\widehat{\beta} = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2}$.
  - $[rT] = t$, $\frac{x_t}{\sqrt{T}} \xrightarrow{d} B_x(r)$, $\frac{y_t}{\sqrt{T}} \xrightarrow{d} B_y(r)$
  - $\frac{1}{T^2} \sum_{t=1}^{T} x_t y_t = \sum_{t=1}^{T} \frac{x_t}{\sqrt{T}} \frac{y_t}{\sqrt{T}} \frac{1}{T} \xrightarrow{d} \int_0^1 B_x(r) B_y(r)\, dr$

Identification

- What is Consistency? Inconsistency?
- Spurious Regression: Phillips (1986). OLS of a spurious regression is inconsistent

$$y_t = \beta x_t + u_t$$

where $y_t \sim I(1)$, $x_t \sim I(1)$, and $u_t \sim I(1)$. y_t and x_t are independent. $\beta = 0$.

- Not identifiable.
  - $\widehat{\beta} = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2}$.
  - $[rT] = t$, $\frac{x_t}{\sqrt{T}} \xrightarrow{d} B_x(r)$, $\frac{y_t}{\sqrt{T}} \xrightarrow{d} B_y(r)$
  - $\frac{1}{T^2} \sum_{t=1}^{T} x_t y_t = \sum_{t=1}^{T} \frac{x_t}{\sqrt{T}} \frac{y_t}{\sqrt{T}} \frac{1}{T} \xrightarrow{d} \int_0^1 B_x(r) B_y(r) dr$
  - $\frac{1}{T^2} \sum_{t=1}^{T} x_t^2 = \sum_{t=1}^{T} \left(\frac{x_t}{\sqrt{T}} \right)^2 \frac{1}{T} \xrightarrow{d} \int_0^1 B_x^2(r) dr$

Identification

- What is Consistency? Inconsistency?
- Spurious Regression: Phillips (1986). OLS of a spurious regression is inconsistent

$$y_t = \beta x_t + u_t$$

where $y_t \sim I(1)$, $x_t \sim I(1)$, and $u_t \sim I(1)$. y_t and x_t are independent. $\beta = 0$.

- Not identifiable.
  - $\widehat{\beta} = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2}$.
  - $[rT] = t$, $\frac{x_t}{\sqrt{T}} \xrightarrow{d} B_x(r)$, $\frac{y_t}{\sqrt{T}} \xrightarrow{d} B_y(r)$
  - $\frac{1}{T^2} \sum_{t=1}^{T} x_t y_t = \sum_{t=1}^{T} \frac{x_t}{\sqrt{T}} \frac{y_t}{\sqrt{T}} \frac{1}{T} \xrightarrow{d} \int_0^1 B_x(r) B_y(r)\, dr$
  - $\frac{1}{T^2} \sum_{t=1}^{T} x_t^2 = \sum_{t=1}^{T} \left(\frac{x_t}{\sqrt{T}} \right)^2 \frac{1}{T} \xrightarrow{d} \int_0^1 B_x^2(r)\, dr$
  - $\widehat{\beta} \xrightarrow{d} \frac{\int_0^1 B_x(r) B_y(r) dr}{\int_0^1 B_x^2(r) dr} = \beta^* \neq \beta$, β^* is a random variable

Identification

- What is Consistency? Inconsistency?
- Spurious Regression: Phillips (1986). OLS of a spurious regression is inconsistent

$$y_t = \beta x_t + u_t$$

  where $y_t \sim I(1)$, $x_t \sim I(1)$, and $u_t \sim I(1)$. y_t and x_t are independent. $\beta = 0$.

- Not identifiable.
  - $\widehat{\beta} = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2}$.
  - $[rT] = t$, $\frac{x_t}{\sqrt{T}} \xrightarrow{d} B_x(r)$, $\frac{y_t}{\sqrt{T}} \xrightarrow{d} B_y(r)$
  - $\frac{1}{T^2} \sum_{t=1}^{T} x_t y_t = \sum_{t=1}^{T} \frac{x_t}{\sqrt{T}} \frac{y_t}{\sqrt{T}} \frac{1}{T} \xrightarrow{d} \int_0^1 B_x(r) B_y(r)\, dr$
  - $\frac{1}{T^2} \sum_{t=1}^{T} x_t^2 = \sum_{t=1}^{T} \left(\frac{x_t}{\sqrt{T}} \right)^2 \frac{1}{T} \xrightarrow{d} \int_0^1 B_x^2(r)\, dr$
  - $\widehat{\beta} \xrightarrow{d} \frac{\int_0^1 B_x(r) B_y(r) dr}{\int_0^1 B_x^2(r) dr} = \beta^* \neq \beta$, β^* is a random variable
  - Time series spurious regression is not identifiable.

Identification:

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

Identification:

- $y_t = \alpha + \beta x_t + u_t$
- $\widehat{\beta} = \frac{\sum_{t=1}^{T}(x_t - \overline{x})y_t}{\sum_{t=1}^{T}(x_t - \overline{x})^2}$.

Identification:

- $y_t = \alpha + \beta x_t + u_t$
- $\widehat{\beta} = \frac{\sum_{t=1}^{T}(x_t - \overline{x})y_t}{\sum_{t=1}^{T}(x_t - \overline{x})^2}$.
- $[rT] = t$, $\frac{x_t}{\sqrt{T}} \xrightarrow{d} B_x(r)$, $\frac{y_t}{\sqrt{T}} \xrightarrow{d} B_y(r)$

Identification:

- $y_t = \alpha + \beta x_t + u_t$
- $\widehat{\beta} = \frac{\sum_{t=1}^{T}(x_t - \overline{x})y_t}{\sum_{t=1}^{T}(x_t - \overline{x})^2}$.
- $[rT] = t$, $\frac{x_t}{\sqrt{T}} \xrightarrow{d} B_x(r)$, $\frac{y_t}{\sqrt{T}} \xrightarrow{d} B_y(r)$
- $\frac{x_t - \overline{x}}{\sqrt{T}} \xrightarrow{d} \underline{B}_x(r)$, $\frac{y_t - \overline{y}}{\sqrt{T}} \xrightarrow{d} \underline{B}_x(r)$, $\underline{B}_x(r) = B_x(r) - \int_0^1 B_x(r)\, dr$

Identification:

- $y_t = \alpha + \beta x_t + u_t$
- $\widehat{\beta} = \frac{\sum_{t=1}^{T}(x_t - \overline{x})y_t}{\sum_{t=1}^{T}(x_t - \overline{x})^2}$.
- $[rT] = t$, $\frac{x_t}{\sqrt{T}} \xrightarrow{d} B_x(r)$, $\frac{y_t}{\sqrt{T}} \xrightarrow{d} B_y(r)$
- $\frac{x_t - \overline{x}}{\sqrt{T}} \xrightarrow{d} \underline{B}_x(r)$, $\frac{y_t - \overline{y}}{\sqrt{T}} \xrightarrow{d} \underline{B}_x(r)$, $\underline{B}_x(r) = B_x(r) - \int_0^1 B_x(r)\,dr$
- $\frac{1}{T^2}\sum_{t=1}^{T}(x_t - \overline{x})\,y_t \xrightarrow{d} \int_0^1 \underline{B}_x(r)\,B_y(r)\,dr = \int_0^1 \underline{B}_x(r)\,\underline{B}_y(r)\,dr$

Identification:

- $y_t = \alpha + \beta x_t + u_t$
- $\widehat{\beta} = \frac{\sum_{t=1}^{T}(x_t - \overline{x})y_t}{\sum_{t=1}^{T}(x_t - \overline{x})^2}$.
- $[rT] = t$, $\frac{x_t}{\sqrt{T}} \xrightarrow{d} B_x(r)$, $\frac{y_t}{\sqrt{T}} \xrightarrow{d} B_y(r)$
- $\frac{x_t - \overline{x}}{\sqrt{T}} \xrightarrow{d} \underline{B}_x(r)$, $\frac{y_t - \overline{y}}{\sqrt{T}} \xrightarrow{d} \underline{B}_x(r)$, $\underline{B}_x(r) = B_x(r) - \int_0^1 B_x(r)\,dr$
- $\frac{1}{T^2}\sum_{t=1}^{T}(x_t - \overline{x})\,y_t \xrightarrow{d} \int_0^1 \underline{B}_x(r)\,B_y(r)\,dr = \int_0^1 \underline{B}_x(r)\,\underline{B}_y(r)\,dr$
- $\frac{1}{T^2}\sum_{t=1}^{T} x_t^2 = \sum_{t=1}^{T}\left(\frac{x_t}{\sqrt{T}}\right)^2 \frac{1}{T} \xrightarrow{d} \int_0^1 \underline{B}_x^2(r)\,dr$

Identification:

- $y_t = \alpha + \beta x_t + u_t$
- $\widehat{\beta} = \frac{\sum_{t=1}^{T}(x_t - \overline{x})y_t}{\sum_{t=1}^{T}(x_t - \overline{x})^2}$.
- $[rT] = t$, $\frac{x_t}{\sqrt{T}} \xrightarrow{d} B_x(r)$, $\frac{y_t}{\sqrt{T}} \xrightarrow{d} B_y(r)$
- $\frac{x_t - \overline{x}}{\sqrt{T}} \xrightarrow{d} \underline{B}_x(r)$, $\frac{y_t - \overline{y}}{\sqrt{T}} \xrightarrow{d} \underline{B}_x(r)$, $\underline{B}_x(r) = B_x(r) - \int_0^1 B_x(r)\, dr$
- $\frac{1}{T^2}\sum_{t=1}^{T}(x_t - \overline{x})\, y_t \xrightarrow{d} \int_0^1 \underline{B}_x(r)\, B_y(r)\, dr = \int_0^1 \underline{B}_x(r)\, \underline{B}_y(r)\, dr$
- $\frac{1}{T^2}\sum_{t=1}^{T} x_t^2 = \sum_{t=1}^{T}\left(\frac{x_t}{\sqrt{T}}\right)^2 \frac{1}{T} \xrightarrow{d} \int_0^1 \underline{B}_x^2(r)\, dr$
- $\widehat{\beta} \xrightarrow{d} \frac{\int_0^1 \underline{B}_x(r)\underline{B}_y(r)dr}{\int_0^1 \underline{B}_x^2(r)dr}$

Panel Spurious Regression

- Identification:

Panel Spurious Regression

- Identification:
  - Panel Spurious Regression: Kao (1999). OLS of a panel spurious regression is consistent, n-consistent.

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

where $y_{it} \sim I(1)$, $x_{it} \sim I(1)$, and $u_{it} \sim I(1)$.

Panel Spurious Regression

- Identification:
  - Panel Spurious Regression: Kao (1999). OLS of a panel spurious regression is consistent, n-consistent.

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

where $y_{it} \sim I(1)$, $x_{it} \sim I(1)$, and $u_{it} \sim I(1)$.

  - $\widehat{\beta} \xrightarrow{p} 0$, Identification can be achieved in panel model.

Panel Spurious Regression

- Identification:
  - Panel Spurious Regression: Kao (1999). OLS of a panel spurious regression is consistent, n-consistent.

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

    where $y_{it} \sim I(1)$, $x_{it} \sim I(1)$, and $u_{it} \sim I(1)$.
  - $\widehat{\beta} \xrightarrow{p} 0$, Identification can be achieved in panel model.
  - Spuriousness in time series can be removed in panel model with large n and large T.

Panel Spurious Regression

- $\widehat{\beta} = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} (x_t - \overline{x}) y_t}{\sum_{i=1}^{n} \sum_{t=1}^{T} (x_t - \overline{x})^2}.$

Panel Spurious Regression

- $\widehat{\beta} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(x_t - \overline{x})y_t}{\sum_{i=1}^{n}\sum_{t=1}^{T}(x_t - \overline{x})^2}$.
- $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\left(x_t - \overline{x}\right)y_t \xrightarrow{d} \frac{1}{n}\sum_{i=1}^{n}\int_0^1 \underline{B}_x\left(r\right)\underline{B}_y\left(r\right)dr$ as $T \longrightarrow \infty$

Panel Spurious Regression

- $\widehat{\beta} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(x_t - \overline{x})y_t}{\sum_{i=1}^{n}\sum_{t=1}^{T}(x_t - \overline{x})^2}$.
- $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}(x_t - \overline{x})\, y_t \xrightarrow{d} \frac{1}{n}\sum_{i=1}^{n}\int_0^1 \underline{B}_x(r)\,\underline{B}_y(r)\,dr$ as $T \to \infty$
- $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}(x_t - \overline{x})^2 \xrightarrow{d} \frac{1}{n}\sum_{i=1}^{n}\int_0^1 \underline{B}_x^2(r)\,dr$ as $T \to \infty$

Panel Spurious Regression

- $\widehat{\beta} = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} (x_t - \overline{x}) y_t}{\sum_{i=1}^{n} \sum_{t=1}^{T} (x_t - \overline{x})^2}$.
- $\frac{1}{n} \sum_{i=1}^{n} \frac{1}{T^2} \sum_{t=1}^{T} (x_t - \overline{x}) \, y_t \xrightarrow{d} \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \underline{B}_x(r) \, \underline{B}_y(r) \, dr$ as $T \to \infty$
- $\frac{1}{n} \sum_{i=1}^{n} \frac{1}{T^2} \sum_{t=1}^{T} (x_t - \overline{x})^2 \xrightarrow{d} \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \underline{B}_x^2(r) \, dr$ as $T \to \infty$
- $\frac{1}{n} \sum_{i=1}^{n} \int_0^1 \underline{B}_x(r) \, \underline{B}_y(r) \, dr \xrightarrow{p} E\left[\int_0^1 \underline{B}_x(r) \, \underline{B}_y(r) \, dr \right] = 0$ as $n \to \infty$

Panel Spurious Regression

- $\widehat{\beta} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(x_t - \overline{x})y_t}{\sum_{i=1}^{n}\sum_{t=1}^{T}(x_t - \overline{x})^2}$.
- $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}(x_t - \overline{x})\,y_t \xrightarrow{d} \frac{1}{n}\sum_{i=1}^{n}\int_0^1 \underline{B}_x(r)\,\underline{B}_y(r)\,dr$ as $T \to \infty$
- $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}(x_t - \overline{x})^2 \xrightarrow{d} \frac{1}{n}\sum_{i=1}^{n}\int_0^1 \underline{B}_x^2(r)\,dr$ as $T \to \infty$
- $\frac{1}{n}\sum_{i=1}^{n}\int_0^1 \underline{B}_x(r)\,\underline{B}_y(r)\,dr \xrightarrow{p} E\left[\int_0^1 \underline{B}_x(r)\,\underline{B}_y(r)\,dr\right] = 0$ as $n \to \infty$
- $\frac{1}{n}\sum_{i=1}^{n}\int_0^1 \underline{B}_x^2(r)\,dr \xrightarrow{p} E\left[\int_0^1 \underline{B}_x^2(r)\,dr\right] = \sigma_x^2\left[\int_0^1 E\underline{W}_x^2(r)\,dr\right]$

Panel Spurious Regression

- $\widehat{\beta} = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} (x_t - \overline{x}) y_t}{\sum_{i=1}^{n} \sum_{t=1}^{T} (x_t - \overline{x})^2}$.
- $\frac{1}{n} \sum_{i=1}^{n} \frac{1}{T^2} \sum_{t=1}^{T} \left(x_t - \overline{x} \right) y_t \xrightarrow{d} \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \underline{B}_x \left(r \right) \underline{B}_y \left(r \right) dr$ as $T \to \infty$
- $\frac{1}{n} \sum_{i=1}^{n} \frac{1}{T^2} \sum_{t=1}^{T} \left(x_t - \overline{x} \right)^2 \xrightarrow{d} \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \underline{B}_x^2 \left(r \right) dr$ as $T \to \infty$
- $\frac{1}{n} \sum_{i=1}^{n} \int_0^1 \underline{B}_x \left(r \right) \underline{B}_y \left(r \right) dr \xrightarrow{p} E \left[\int_0^1 \underline{B}_x \left(r \right) \underline{B}_y \left(r \right) dr \right] = 0$ as $n \to \infty$
- $\frac{1}{n} \sum_{i=1}^{n} \int_0^1 \underline{B}_x^2 \left(r \right) dr \xrightarrow{p} E \left[\int_0^1 \underline{B}_x^2 \left(r \right) dr \right] = \sigma_x^2 \left[\int_0^1 E \underline{W}_x^2 \left(r \right) dr \right]$
- $\int_0^1 E \left[W_x^2 \left(r \right) \right] dr = \int_0^1 r dr = \frac{1}{2}, \int_0^1 E \left[\underline{W}_x^2 \left(r \right) \right] dr = \frac{1}{6}$

Panel Spurious Regression

- $\widehat{\beta} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(x_t - \overline{x})y_t}{\sum_{i=1}^{n}\sum_{t=1}^{T}(x_t-\overline{x})^2}$.
- $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}(x_t - \overline{x})\, y_t \xrightarrow{d} \frac{1}{n}\sum_{i=1}^{n}\int_0^1 \underline{B}_x(r)\,\underline{B}_y(r)\,dr$ as $T \to \infty$
- $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}(x_t - \overline{x})^2 \xrightarrow{d} \frac{1}{n}\sum_{i=1}^{n}\int_0^1 \underline{B}_x^2(r)\,dr$ as $T \to \infty$
- $\frac{1}{n}\sum_{i=1}^{n}\int_0^1 \underline{B}_x(r)\,\underline{B}_y(r)\,dr \xrightarrow{p} E\left[\int_0^1 \underline{B}_x(r)\,\underline{B}_y(r)\,dr\right] = 0$ as $n \to \infty$
- $\frac{1}{n}\sum_{i=1}^{n}\int_0^1 \underline{B}_x^2(r)\,dr \xrightarrow{p} E\left[\int_0^1 \underline{B}_x^2(r)\,dr\right] = \sigma_x^2\left[\int_0^1 E\underline{W}_x^2(r)\,dr\right]$
- $\int_0^1 E\left[W_x^2(r)\right]dr = \int_0^1 r dr = \frac{1}{2}$, $\int_0^1 E\left[\underline{W}_x^2(r)\right]dr = \frac{1}{6}$
- $\widehat{\beta} \xrightarrow{p} 0$ as $T \to \infty$ and follows $n \to \infty$ sequentially.

Panel Spurious Regression

- $\widehat{\beta} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(x_t - \overline{x})y_t}{\sum_{i=1}^{n}\sum_{t=1}^{T}(x_t - \overline{x})^2}$.
- $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}(x_t - \overline{x})\,y_t \xrightarrow{d} \frac{1}{n}\sum_{i=1}^{n}\int_0^1 \underline{B}_x(r)\,\underline{B}_y(r)\,dr$ as $T \to \infty$
- $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}(x_t - \overline{x})^2 \xrightarrow{d} \frac{1}{n}\sum_{i=1}^{n}\int_0^1 \underline{B}_x^2(r)\,dr$ as $T \to \infty$
- $\frac{1}{n}\sum_{i=1}^{n}\int_0^1 \underline{B}_x(r)\,\underline{B}_y(r)\,dr \xrightarrow{p} E\left[\int_0^1 \underline{B}_x(r)\,\underline{B}_y(r)\,dr\right] = 0$ as $n \to \infty$
- $\frac{1}{n}\sum_{i=1}^{n}\int_0^1 \underline{B}_x^2(r)\,dr \xrightarrow{p} E\left[\int_0^1 \underline{B}_x^2(r)\,dr\right] = \sigma_x^2\left[\int_0^1 E\underline{W}_x^2(r)\,dr\right]$
- $\int_0^1 E\left[W_x^2(r)\right]dr = \int_0^1 r dr = \frac{1}{2}$, $\int_0^1 E\left[\underline{W}_x^2(r)\right]dr = \frac{1}{6}$
- $\widehat{\beta} \xrightarrow{p} 0$ as $T \to \infty$ and follows $n \to \infty$ sequentially.
- $\widehat{\beta}$ is consistent and the panel spurious model is identified.

n-consistent

- $\sqrt{n}\widehat{\beta} = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}(x_t - \overline{x})y_t}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}(x_t - \overline{x})^2}.$

n-consistent

- $\sqrt{n}\widehat{\beta} = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}(x_t - \overline{x})y_t}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}(x_t - \overline{x})^2}$.
- $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\left(x_t - \overline{x}\right)^2 \xrightarrow{p} \frac{\sigma_x^2}{6}$ as $(n, T) \to \infty$ sequentially/jointly.

n-consistent

- $\sqrt{n}\widehat{\beta} = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}(x_t-\overline{x})y_t}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}(x_t-\overline{x})^2}$.
- $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\left(x_t-\overline{x}\right)^2 \xrightarrow{p} \frac{\sigma_x^2}{6}$ as $(n, T) \longrightarrow \infty$ sequentially/jointly.
- $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\left(x_t-\overline{x}\right)y_t \xrightarrow{d} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^1 \underline{B}_x(r)\,\underline{B}_y(r)\,dr$ as $T \longrightarrow \infty$

n-consistent

- $\sqrt{n}\widehat{\beta} = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}(x_t - \overline{x})y_t}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}(x_t - \overline{x})^2}$.
- $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\left(x_t - \overline{x}\right)^2 \xrightarrow{p} \frac{\sigma_x^2}{6}$ as $(n, T) \longrightarrow \infty$ sequentially/jointly.
- $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\left(x_t - \overline{x}\right)y_t \xrightarrow{d} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^1 \underline{B}_x(r)\,\underline{B}_y(r)\,dr$ as $T \longrightarrow \infty$
- $E\left[\int_0^1 \underline{B}_x(r)\,\underline{B}_y(r)\,dr\right] = 0$ and $Var\left[\int_0^1 \underline{B}_x(r)\,\underline{B}_y(r)\,dr\right] = \frac{\sigma_x^2\sigma_y^2}{90}$

n-consistent

- $\sqrt{n}\widehat{\beta} = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}(x_t-\overline{x})y_t}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}(x_t-\overline{x})^2}$.
- $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\left(x_t-\overline{x}\right)^2 \xrightarrow{p} \frac{\sigma_x^2}{6}$ as $(n,T)\longrightarrow\infty$ sequentially/jointly.
- $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\left(x_t-\overline{x}\right)y_t \xrightarrow{d} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^1 \underline{B}_x(r)\,\underline{B}_y(r)\,dr$ as $T\longrightarrow\infty$
- $E\left[\int_0^1 \underline{B}_x(r)\,\underline{B}_y(r)\,dr\right]=0$ and $Var\left[\int_0^1 \underline{B}_x(r)\,\underline{B}_y(r)\,dr\right]=\frac{\sigma_x^2\sigma_y^2}{90}$
- $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^1 \underline{B}_x(r)\,\underline{B}_y(r)\,dr \xrightarrow{d} N\left(0,\frac{2}{5}\frac{\sigma_y^2}{\sigma_x^2}\right)$ as $n\longrightarrow\infty$

n-consistent

- $\sqrt{n}\widehat{\beta} = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{T^2} \sum_{t=1}^{T} (x_t - \overline{x}) y_t}{\frac{1}{n} \sum_{i=1}^{n} \frac{1}{T^2} \sum_{t=1}^{T} (x_t - \overline{x})^2}.$
- $\frac{1}{n} \sum_{i=1}^{n} \frac{1}{T^2} \sum_{t=1}^{T} (x_t - \overline{x})^2 \xrightarrow{p} \frac{\sigma_x^2}{6}$ as $(n, T) \longrightarrow \infty$ sequentially/jointly.
- $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{T^2} \sum_{t=1}^{T} (x_t - \overline{x}) y_t \xrightarrow{d} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^1 \underline{B}_x(r) \underline{B}_y(r) dr$ as $T \longrightarrow \infty$
- $E\left[\int_0^1 \underline{B}_x(r) \underline{B}_y(r) dr\right] = 0$ and $Var\left[\int_0^1 \underline{B}_x(r) \underline{B}_y(r) dr\right] = \frac{\sigma_x^2 \sigma_y^2}{90}$
- $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^1 \underline{B}_x(r) \underline{B}_y(r) dr \xrightarrow{d} N\left(0, \frac{2}{5} \frac{\sigma_y^2}{\sigma_x^2}\right)$ as $n \longrightarrow \infty$
- $\sqrt{n}\widehat{\beta} = O_p(1)$.

A Less Developed Theory

- Contemporary large panel - when n is large and comparable with T such that $\frac{n}{T} \longrightarrow c$.

A Less Developed Theory

- Contemporary large panel - when n is large and comparable with T such that $\frac{n}{T} \longrightarrow c$.
- Is there a payoff to econometrics from RMT?

A Less Developed Theory

- Contemporary large panel - when n is large and comparable with T such that $\frac{n}{T} \longrightarrow c$.
- Is there a payoff to econometrics from RMT?
- Baltagi, Feng and Kao (2009): Testing for Cross-sectional Dependence in Fixed Effects Panel Data Models

A Less Developed Theory

- Contemporary large panel - when n is large and comparable with T such that $\frac{n}{T} \longrightarrow c$.
- Is there a payoff to econometrics from RMT?
- Baltagi, Feng and Kao (2009): Testing for Cross-sectional Dependence in Fixed Effects Panel Data Models
- Expand arsenal of econometric tools for thinking about these large panel models.

Factor-Loading Model

$$R_n = \Lambda F^{'} + \varepsilon$$

where R_n is a $n \times T$ matrix of asset returns, F is an $T \times k$ matrix of factors, Λ is a $n \times k$ matrix of loadings and ε is a $n \times T$ matrix of idiosyncratic errors.

- How many dimensions of "significant eigenvalues"? That is how many factors?

Factor-Loading Model

$$R_n = \Lambda F^{'} + \varepsilon$$

where R_n is a $n \times T$ matrix of asset returns, F is an $T \times k$ matrix of factors, Λ is a $n \times k$ matrix of loadings and ε is a $n \times T$ matrix of idiosyncratic errors.

- How many dimensions of "significant eigenvalues"? That is how many factors?
- From RMT we know that sample eigenvalues are biased of the corresponding population eigenvalues if $\frac{n}{T} \to c$.

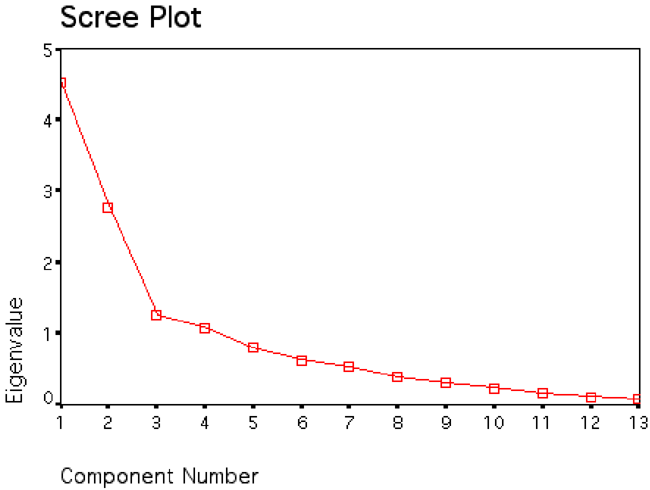
Factor-Loading Model

$$R_n = \Lambda F^{'} + \varepsilon$$

where R_n is a $n \times T$ matrix of asset returns, F is an $T \times k$ matrix of factors, Λ is a $n \times k$ matrix of loadings and ε is a $n \times T$ matrix of idiosyncratic errors.

- How many dimensions of "significant eigenvalues"? That is how many factors?
- From RMT we know that sample eigenvalues are biased of the corresponding population eigenvalues if $\frac{n}{T} \to c$.
- The sample eigenvalues are more spread out than the population eigenvalues.

Scree plot



Scree Plot
5
4
3
2
1
0
Eigenvalue
1
2
3
4
5
6
7
8
9
10
11
12
13
Component Number

Dynamic Conditional Correlation (DCC) model

- What happens if $\frac{n}{T} \longrightarrow c$?

Dynamic Conditional Correlation (DCC) model

- What happens if $\frac{n}{T} \longrightarrow c$?
- Test for constant conditional correlation (CCC) model.

Weak Instruments Models in Panel Set-up

- Two-stage least squares (2SLS) is inconsistent if instruments are weak

Weak Instruments Models in Panel Set-up

- Two-stage least squares (2SLS) is inconsistent if instruments are weak
- Consistency of 2SLS can be recovered using panel

Weak Instruments Models in Panel Set-up

- Two-stage least squares (2SLS) is inconsistent if instruments are weak
- Consistency of 2SLS can be recovered using panel
- Consistent Estimation with Weak Instruments in Panel Data

Change Point estimation in Time Series Set-up

- Change point estimate is inconsistent (not identifiable) in time series, $\widehat{k} = k_0 + O_p\left(\frac{1}{\|\delta_T\|^2}\right)$, $\widehat{k}$ is inconsistent

Change Point estimation in Time Series Set-up

- Change point estimate is inconsistent (not identifiable) in time series, $\widehat{k} = k_0 + O_p\left(\frac{1}{\|\delta_T\|^2}\right)$, $\widehat{k}$ is inconsistent
- k_0 is the true change point date and δ_T is the jump size

Change Point estimation in Time Series Set-up

- Change point estimate is inconsistent (not identifiable) in time series, $\widehat{k} = k_0 + O_p\left(\frac{1}{\|\delta_T\|^2}\right)$, $\widehat{k}$ is inconsistent
- k_0 is the true change point date and δ_T is the jump size
- $\widehat{\tau} = \frac{\widehat{k}}{T}$ $\widehat{\tau} = \tau_0 + O_p\left(\frac{1}{T\|\delta_T\|^2}\right)$ $\widehat{\tau}$ is T-consistent, $T\left(\widehat{\tau} - \tau_0\right) = O_p\left(\frac{1}{\|\delta_T\|^2}\right)$

Change Point estimation in Time Series Set-up

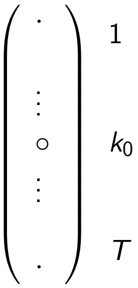
- Change point estimate is inconsistent (not identifiable) in time series, $\widehat{k} = k_0 + O_p\left(\frac{1}{\|\delta_T\|^2}\right)$, $\widehat{k}$ is inconsistent
- k_0 is the true change point date and δ_T is the jump size
- $\widehat{\tau} = \frac{\widehat{k}}{T}$ $\widehat{\tau} = \tau_0 + O_p\left(\frac{1}{T\|\delta_T\|^2}\right)$ $\widehat{\tau}$ is T-consistent, $T\left(\widehat{\tau} - \tau_0\right) = O_p\left(\frac{1}{\|\delta_T\|^2}\right)$
- In time series need to assume: shrinking Break: $\delta_T \to 0$ and $T^{1/2-\alpha}\delta_T \to \infty$ for some $\alpha \in \left(0, \frac{1}{2}\right)$

Change Point estimation in Panel Set-up

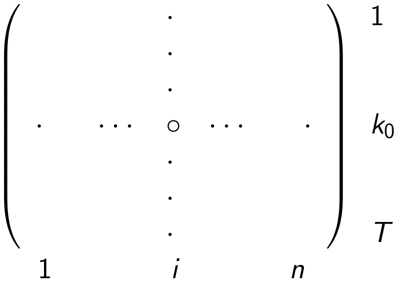
- $\widehat{k} = k_0 + O_p\left(\frac{1}{n\|\delta_T\|^2}\right)$ $\widehat{k}$ is n-consistent $\delta_T = \beta_2 - \beta_1$

Change Point estimation in Panel Set-up

- $\widehat{k} = k_0 + O_p\left(\frac{1}{n\|\delta_T\|^2}\right)$ $\widehat{k}$ is n-consistent $\delta_T = \beta_2 - \beta_1$
- $\widehat{\tau} = \frac{\widehat{k}}{T}$ $\widehat{\tau} = \tau_0 + O_p\left(\frac{1}{nT\|\delta_T\|^2}\right)$ $\widehat{\tau}$ is nT-consistent



$a)$ Time series data



$b)$ Panel data

Test for Cross-Sectional Dependence in Panel Models

- Suppose that $(x_1, x_2, \ldots, x_T)$ are *iid* samples drawn from a n-dimensional multivariate normal population $N(0, \Sigma_n)$.

Test for Cross-Sectional Dependence in Panel Models

- Suppose that $(x_1, x_2, ..., x_T)$ are *iid* samples drawn from a n-dimensional multivariate normal population $N(0, \Sigma_n)$.
- The null hypothesis $H_0 : \Sigma_n = I_n$.

Test for Cross-Sectional Dependence in Panel Models

- Suppose that $(x_1, x_2, ..., x_T)$ are *iid* samples drawn from a n-dimensional multivariate normal population $N(0, \Sigma_n)$.
- The null hypothesis $H_0 : \Sigma_n = I_n$.
- Reject H_0 if $\lambda_1 > 1$.

Test for Cross-Sectional Dependence in Panel Models

- Suppose that $(x_1, x_2, ..., x_T)$ are *iid* samples drawn from a n-dimensional multivariate normal population $N(0, \Sigma_n)$.
- The null hypothesis $H_0 : \Sigma_n = I_n$.
- Reject H_0 if $\lambda_1 > 1$.
- Reject the null hypothesis if $l_1 > u$ where l_1 is the largest eigenvalue of the sample covariance matrix.

Test for Cross-Sectional Dependence in Panel Models

- Suppose that $(x_1, x_2, ..., x_T)$ are *iid* samples drawn from a n-dimensional multivariate normal population $N(0, \Sigma_n)$.
- The null hypothesis $H_0 : \Sigma_n = I_n$.
- Reject H_0 if $\lambda_1 > 1$.
- Reject the null hypothesis if $l_1 > u$ where l_1 is the largest eigenvalue of the sample covariance matrix.
- Any guidance from distribution theory when $\frac{n}{T} \rightarrow c$?

Test for Cross-Sectional Dependence in Panel Models

- Suppose that $(x_1, x_2, ..., x_T)$ are *iid* samples drawn from a n-dimensional multivariate normal population $N(0, \Sigma_n)$.
- The null hypothesis $H_0 : \Sigma_n = I_n$.
- Reject H_0 if $\lambda_1 > 1$.
- Reject the null hypothesis if $l_1 > u$ where l_1 is the largest eigenvalue of the sample covariance matrix.
- Any guidance from distribution theory when $\frac{n}{T} \to c$?
- Johnstone, I. M. (2001), "On the Distribution of the Largest Eigenvalue in Principal Component Analysis," *Annals of Statistics*, 29, 295-327.

Largest Eigenvalue of Random Covariance Matrices

- Suppose that $(x_1, x_2, ..., x_T)$ are *iid* samples drawn from a n-dimensional multivariate normal population $N(0, \Sigma_n)$.

Largest Eigenvalue of Random Covariance Matrices

- Suppose that $(x_1, x_2, ..., x_T)$ are *iid* samples drawn from a n-dimensional multivariate normal population $N(0, \Sigma_n)$.
- Let $X = \left(x_1^{'}, x_2^{'}, ..., x_T^{'}\right)^{'}$ be $T \times n$ data matrix and $\frac{1}{n}X^{'}X$ be the sample covariance matrix.

Largest Eigenvalue of Random Covariance Matrices

- Suppose that $(x_1, x_2, ..., x_T)$ are *iid* samples drawn from a n-dimensional multivariate normal population $N(0, \Sigma_n)$.
- Let $X = \left(x_1', x_2', ..., x_T'\right)'$ be $T \times n$ data matrix and $\frac{1}{n}X'X$ be the sample covariance matrix.
- Principal component analysis (PCA). We are interested in recovering as much of the total variance in the data as possible while reducing the dimensionality of the problem from n to k.

Largest Eigenvalue of Random Covariance Matrices

- Suppose that $(x_1, x_2, ..., x_T)$ are *iid* samples drawn from a n-dimensional multivariate normal population $N(0, \Sigma_n)$.
- Let $X = \left(x_1^{'}, x_2^{'}, ..., x_T^{'}\right)^{'}$ be $T \times n$ data matrix and $\frac{1}{n}X^{'}X$ be the sample covariance matrix.
- Principal component analysis (PCA). We are interested in recovering as much of the total variance in the data as possible while reducing the dimensionality of the problem from n to k.
- We are looking from k vectors $(e_1, e_2, ..., e_k)$ in R^n such that $\sum_{m=1}^{k} var\left(\langle x_i, e_m \rangle\right)$ is maximal.

Largest Eigenvalue of Random Covariance Matrices

- Suppose that $(x_1, x_2, ..., x_T)$ are *iid* samples drawn from a n-dimensional multivariate normal population $N(0, \Sigma_n)$.
- Let $X = \left(x_1^{'}, x_2^{'}, ..., x_T^{'}\right)^{'}$ be $T \times n$ data matrix and $\frac{1}{n}X^{'}X$ be the sample covariance matrix.
- Principal component analysis (PCA). We are interested in recovering as much of the total variance in the data as possible while reducing the dimensionality of the problem from n to k.
- We are looking from k vectors $(e_1, e_2, ..., e_k)$ in R^n such that $\sum_{m=1}^{k} var(\langle x_i, e_m \rangle)$ is maximal.
- One should choose for e the eigenvectors associated with the first k eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_k$ of Σ_n.

Largest Eigenvalue of Random Covariance Matrices

- Assume $\Sigma_n = I_n$. So all the population eigenvalues are equal to 1.

Largest Eigenvalue of Random Covariance Matrices

- Assume $\Sigma_n = I_n$. So all the population eigenvalues are equal to 1.
- Let l_1 is the largest eigenvalue of sample covariance matrix.

Largest Eigenvalue of Random Covariance Matrices

- Assume $\Sigma_n = I_n$. So all the population eigenvalues are equal to 1.
- Let l_1 is the largest eigenvalue of sample covariance matrix.
- Anderson (2003), Theorem 13.5.1: $\sqrt{T}\left(l_1 - \lambda_1\right) \xrightarrow{d} N\left(0, 2\lambda_1^2\right)$ as $T \to \infty$ if n is fixed.

Largest Eigenvalue of Random Covariance Matrices

- Assume $\Sigma_n = I_n$. So all the population eigenvalues are equal to 1.
- Let l_1 is the largest eigenvalue of sample covariance matrix.
- Anderson (2003), Theorem 13.5.1: $\sqrt{T}\left(l_1 - \lambda_1\right) \xrightarrow{d} N\left(0, 2\lambda_1^2\right)$ as $T \to \infty$ if n is fixed.
- l_1 is consistent for λ_1.

Largest Eigenvalue of Random Covariance Matrices

- Assume $\Sigma_n = I_n$. So all the population eigenvalues are equal to 1.
- Let l_1 is the largest eigenvalue of sample covariance matrix.
- Anderson (2003), Theorem 13.5.1: $\sqrt{T}\,(l_1 - \lambda_1) \xrightarrow{d} N\left(0, 2\lambda_1^2\right)$ as $T \to \infty$ if n is fixed.
- l_1 is consistent for λ_1.
- Geman (1980), Yin et al (1988): $l_1 \overset{a.c.}{\to} \left(1 + \sqrt{c}\right)^2$ if $\frac{n}{T} \to c$.

Largest Eigenvalue of Random Covariance Matrices

- Assume $\Sigma_n = I_n$. So all the population eigenvalues are equal to 1.
- Let l_1 is the largest eigenvalue of sample covariance matrix.
- Anderson (2003), Theorem 13.5.1: $\sqrt{T}\left(l_1 - \lambda_1\right) \xrightarrow{d} N\left(0, 2\lambda_1^2\right)$ as $T \to \infty$ if n is fixed.
- l_1 is consistent for λ_1.
- Geman (1980), Yin et al (1988): $l_1 \overset{a.c.}{\to} \left(1 + \sqrt{c}\right)^2$ if $\frac{n}{T} \to c$.
- Choose $n = T$, $l_1 \to 4$ whereas $\lambda_1 = 1$.

Largest Eigenvalue of Random Covariance Matrices

- Assume $\Sigma_n = I_n$. So all the population eigenvalues are equal to 1.
- Let l_1 is the largest eigenvalue of sample covariance matrix.
- Anderson (2003), Theorem 13.5.1: $\sqrt{T}\left(l_1 - \lambda_1\right) \xrightarrow{d} N\left(0, 2\lambda_1^2\right)$ as $T \to \infty$ if n is fixed.
- l_1 is consistent for λ_1.
- Geman (1980), Yin et al (1988): $l_1 \overset{a.c.}{\to} \left(1 + \sqrt{c}\right)^2$ if $\frac{n}{T} \to c$.
- Choose $n = T$, $l_1 \to 4$ whereas $\lambda_1 = 1$.
- l_1 is an inconsistent estimator in the large n and T setting.