Lecture 5B: Cointegration — Using FIML

Vitor Possebom

EESP-FGV

Econometrics 2

Administrative

- Recommended Reading: Hamilton's Chapters 20.2 (pages 635 and 636), 20.3 (pages 645-648) and 20.4
- Problem Set 4 - Deadline: June 13th at 9:00 am

Outline

1. Motivation

2. Full-Information Maximum Likelihood (FIML) Estimation

3. Testing the Null Hyothesis of h Cointegrating Relations

4. Overview of Unit Roots: To difference or Not to Difference?

Motivation

Motivation

In Lecture 5A, we learned how to handle cointegration using OLS.

Motivation

In Lecture 5A, we learned how to handle cointegration using OLS.

Those approaches were imperfect:

Motivation

In Lecture 5A, we learned how to handle cointegration using OLS.

Those approaches were imperfect:

- They rely on arbitrary normalizations. (Which variable is our first variable?)

Motivation

In Lecture 5A, we learned how to handle cointegration using OLS.

Those approaches were imperfect:

- They rely on arbitrary normalizations. (Which variable is our first variable?)
- If $a_{11} = 0$, the previous arbitrary normalization implies a misspecified model.

Motivation

In Lecture 5A, we learned how to handle cointegration using OLS.

Those approaches were imperfect:

- They rely on arbitrary normalizations. (Which variable is our first variable?)
- If $a_{11} = 0$, the previous arbitrary normalization implies a misspecified model.

Hence, we need a method that avoids those problems.

Motivation

In Lecture 5A, we learned how to handle cointegration using OLS.

Those approaches were imperfect:

- They rely on arbitrary normalizations. (Which variable is our first variable?)
- If $a_{11} = 0$, the previous arbitrary normalization implies a misspecified model.

Hence, we need a method that avoids those problems.

Johansen (1988, 1991) proposed a solution based on full-information maximum likelihood (FIML) estimation.

Motivation

In Lecture 5A, we learned how to handle cointegration using OLS.

Those approaches were imperfect:

- They rely on arbitrary normalizations. (Which variable is our first variable?)
- If $a_{11} = 0$, the previous arbitrary normalization implies a misspecified model.

Hence, we need a method that avoids those problems.

Johansen (1988, 1991) proposed a solution based on full-information maximum likelihood (FIML) estimation.

- It also allows us to test for the number of cointegrating relations.

Full-Information Maximum Likelihood (FIML) Estimation

Outline

Full-Information Maximum Likelihood (FIML) Estimation

Our goal is to use FIML to estimate a system characterized by exactly h cointegrating equations.

Full-Information Maximum Likelihood (FIML) Estimation

Our goal is to use FIML to estimate a system characterized by exactly h cointegrating equations.

Let $\{Y_t\}$ denote a vector process with n variables.

Full-Information Maximum Likelihood (FIML) Estimation

Our goal is to use FIML to estimate a system characterized by exactly h cointegrating equations.

Let $\{Y_t\}$ denote a vector process with n variables. The maintained hypothesis is that $\{Y_t\}$ follows a $VAR(p)$ in levels:

$$Y_t = \alpha + \Phi_1 \cdot Y_{t-1} + \Phi_2 \cdot Y_{t-2} + \ldots + \Phi_p \cdot Y_{t-p} + \epsilon_t.$$

Full-Information Maximum Likelihood (FIML) Estimation

We saw how to represent this $VAR(p)$ model as a vector error correction ($VEC(p-1)$) model:

Full-Information Maximum Likelihood (FIML) Estimation

We saw how to represent this $VAR(p)$ model as a vector error correction $(VEC(p-1))$ model:

$$\Delta Y_t = \zeta_1 \cdot \Delta Y_{t-1} + \zeta_2 \cdot \Delta Y_{t-2} + \ldots + \zeta_{p-1} \cdot \Delta Y_{t-p+1} + \alpha + \zeta_0 \cdot Y_{t-1} + \epsilon_t,$$

with

Full-Information Maximum Likelihood (FIML) Estimation

We saw how to represent this $VAR\,(p)$ model as a vector error correction ($VEC\,(p-1)$) model:

$$\Delta Y_t = \zeta_1 \cdot \Delta Y_{t-1} + \zeta_2 \cdot \Delta Y_{t-2} + \ldots + \zeta_{p-1} \cdot \Delta Y_{t-p+1} + \alpha + \zeta_0 \cdot Y_{t-1} + \epsilon_t,$$

with

$$\mathbb{E}\left[\epsilon_t\right] = 0; \quad \mathbb{E}\left[\epsilon_t \epsilon_\tau'\right] = \begin{cases} \Omega & \text{for } t = \tau \\ 0 & \text{otherwise.} \end{cases}$$

Full-Information Maximum Likelihood (FIML) Estimation

We saw how to represent this $VAR(p)$ model as a vector error correction ($VEC(p-1)$) model:

$$\Delta Y_t = \zeta_1 \cdot \Delta Y_{t-1} + \zeta_2 \cdot \Delta Y_{t-2} + \ldots + \zeta_{p-1} \cdot \Delta Y_{t-p+1} + \alpha + \zeta_0 \cdot Y_{t-1} + \epsilon_t,$$

with

$$\mathbb{E}[\epsilon_t] = 0; \quad \mathbb{E}\left[\epsilon_t \epsilon_\tau'\right] = \begin{cases} \Omega & \text{for } t = \tau \\ 0 & \text{otherwise.} \end{cases}$$

Imposing that each individual process $\{Y_{k,t}\}$ is $I(1)$ and that there are h linear combinations of $\{Y_t\}$ that are stationary,

Full-Information Maximum Likelihood (FIML) Estimation

We saw how to represent this $VAR(p)$ model as a vector error correction ($VEC(p-1)$) model:

$$\Delta Y_t = \zeta_1 \cdot \Delta Y_{t-1} + \zeta_2 \cdot \Delta Y_{t-2} + \ldots + \zeta_{p-1} \cdot \Delta Y_{t-p+1} + \alpha + \zeta_0 \cdot Y_{t-1} + \epsilon_t,$$

with

$$\mathbb{E}[\epsilon_t] = 0; \quad \mathbb{E}\left[\epsilon_t \epsilon_\tau'\right] = \begin{cases} \Omega & \text{for } t = \tau \\ 0 & \text{otherwise.} \end{cases}$$

Imposing that each individual process $\{Y_{k,t}\}$ is $I(1)$ and that there are h linear combinations of $\{Y_t\}$ that are stationary, we have that

$$\zeta_0 = -BA'$$

for B and $(n \times h)$ matrix and A' an $(h \times n)$ matrix.

Full-Information Maximum Likelihood (FIML) Estimation

Full-Information Maximum Likelihood (FIML) Estimation

Estimation:

- Consider a sample of $T + p$ observation on $\{Y_t\}$.

Full-Information Maximum Likelihood (FIML) Estimation

Estimation:

- Consider a sample of $T + p$ observation on $\{Y_t\}$.
- Assume that the disturbances ϵ_t are Gaussian.

Full-Information Maximum Likelihood (FIML) Estimation

Estimation:

- Consider a sample of $T + p$ observation on $\{Y_t\}$.
- Assume that the disturbances ϵ_t are Gaussian.
- The log likelihood of $(Y_1, Y_2, \ldots, Y_T)$ conditional on $(Y_{-p+1}, Y_{-p+2}, \ldots, Y_0)$ is given by

Full-Information Maximum Likelihood (FIML) Estimation

Estimation:

- Consider a sample of $T + p$ observation on $\{Y_t\}$.
- Assume that the disturbances ϵ_t are Gaussian.
- The log likelihood of $(Y_1, Y_2, \ldots, Y_T)$ conditional on $(Y_{-p+1}, Y_{-p+2}, \ldots, Y_0)$ is given by

$$\mathcal{L}\left(\Omega, \zeta_1, \ldots, \zeta_{p-1}, \alpha, \zeta_0\right) = -\left(\frac{Tn}{2}\right) \cdot \log\left(2\pi\right) - \left(\frac{T}{2}\right) \cdot \log\left|\Omega\right| - \frac{\sum_{t=1}^{T} X_t' \Omega X_t}{2}$$

where $X_t := \Delta Y_t - \zeta_1 \cdot \Delta Y_{t-1} - \ldots - \zeta_{p-1} \cdot \Delta Y_{t-p+1} - \alpha - \zeta_0 \cdot Y_{t-1}$.

Full-Information Maximum Likelihood (FIML) Estimation

Estimation:

- Consider a sample of $T+p$ observation on $\{Y_t\}$.
- Assume that the disturbances ϵ_t are Gaussian.
- The log likelihood of $(Y_1, Y_2, \ldots, Y_T)$ conditional on $(Y_{-p+1}, Y_{-p+2}, \ldots, Y_0)$ is given by

$$\mathcal{L}(\Omega, \zeta_1, \ldots, \zeta_{p-1}, \alpha, \zeta_0) = -\left(\frac{Tn}{2}\right) \cdot \log(2\pi) - \left(\frac{T}{2}\right) \cdot \log|\Omega| - \frac{\sum_{t=1}^{T} X_t' \Omega X_t}{2}$$

where $X_t := \Delta Y_t - \zeta_1 \cdot \Delta Y_{t-1} - \ldots - \zeta_{p-1} \cdot \Delta Y_{t-p+1} - \alpha - \zeta_0 \cdot Y_{t-1}$.

We choose $(\Omega, \zeta_1, \ldots, \zeta_{p-1}, \alpha, \zeta_0)$ so as to maximize $\mathcal{L}(\Omega, \zeta_1, \ldots, \zeta_{p-1}, \alpha, \zeta_0)$ subject to $\zeta_0 = -BA'$.

Full-Information Maximum Likelihood (FIML) Estimation

Full-Information Maximum Likelihood (FIML) Estimation

Hamilton provides a detailed discussion on how to implement this FIML Estimator.

Full-Information Maximum Likelihood (FIML) Estimation

Hamilton provides a detailed discussion on how to implement this FIML Estimator. If you are interested, check pages 636-644.

Full-Information Maximum Likelihood (FIML) Estimation

Hamilton provides a detailed discussion on how to implement this FIML Estimator. If you are interested, check pages 636-644.

In practice, we use `R` to implement FIML estimation.

Full-Information Maximum Likelihood (FIML) Estimation

Hamilton provides a detailed discussion on how to implement this FIML Estimator. If you are interested, check pages 636-644.

In practice, we use `R` to implement FIML estimation.

Specifically, we use the function `ca.jo` and `vec2var` to analyze cointegrated series.

Full-Information Maximum Likelihood (FIML) Estimation

Hamilton provides a detailed discussion on how to implement this FIML Estimator. If you are interested, check pages 636-644.

In practice, we use `R` to implement FIML estimation.

Specifically, we use the function `ca.jo` and `vec2var` to analyze cointegrated series.

Code `johansen.R` illustrates how to use these function. We will discuss a brief theoretical discussion on how to uncover the number h of cointegrating relations.

Testing the Null Hyothesis of h Cointegrating Relations

Outline

Testing the Null Hyothesis of h Cointegrating Relations

- We want to test the null hypothesis of h cointegrating relations against the alternative of $h+1$ cointegrating relations.

Testing the Null Hyothesis of h Cointegrating Relations

- We want to test the null hypothesis of h cointegrating relations against the alternative of $h + 1$ cointegrating relations.
- To do so, we use a likelihood ratio test.

Testing the Null Hyothesis of h Cointegrating Relations

- We want to test the null hypothesis of h cointegrating relations against the alternative of $h+1$ cointegrating relations.
- To do so, we use a likelihood ratio test.
- The distribution of its test statistic is the same as the distribution of the largest **eigenvalue** of the the following matrix:

$$Q = \left[\int_0^1 W(r)\, dW(r)'\right]' \left[\int_0^1 W(r)\, W(r)'\, dr\right]^{-1} \left[\int_0^1 W(r)\, dW(r)'\right]$$

where $W(r)$ is a $(n-h)$-dimensional standard Brownian motion.

Testing the Null Hyothesis of h Cointegrating Relations

- We want to test the null hypothesis of h cointegrating relations against the alternative of $h+1$ cointegrating relations.
- To do so, we use a likelihood ratio test.
- The distribution of its test statistic is the same as the distribution of the largest **eigenvalue** of the the following matrix:

$$Q = \left[\int_0^1 W(r)\, dW(r)'\right]' \left[\int_0^1 W(r)\, W(r)'\, dr\right]^{-1} \left[\int_0^1 W(r)\, dW(r)'\right]$$

  where $W(r)$ is a $(n-h)$-dimensional standard Brownian motion.
- Critical values are obtained via Monte Carlo simulations.

Testing the Null Hyothesis of h Cointegrating Relations

Testing the Null Hyothesis of h Cointegrating Relations

In practice, we use `R` to implement this test.

Testing the Null Hyothesis of h Cointegrating Relations

In practice, we use `R` to implement this test. Specifically, we use the function `ca.jo`.

Testing the Null Hyothesis of h Cointegrating Relations

In practice, we use `R` to implement this test. Specifically, we use the function `ca.jo`.

Code `johansen.R` illustrates how to use this function.

Testing the Null Hyothesis of h Cointegrating Relations

In practice, we use `R` to implement this test. Specifically, we use the function `ca.jo`.

Code `johansen.R` illustrates how to use this function.

It also explains how to connect $VEC\,(p-1)$ models to their $VAR\,(p)$ representation.

Testing the Null Hyothesis of h Cointegrating Relations

In practice, we use `R` to implement this test. Specifically, we use the function `ca.jo`.

Code `johansen.R` illustrates how to use this function.

It also explains how to connect $VEC\,(p-1)$ models to their $VAR\,(p)$ representation.

This connection is useful for forecasting and analyzing IRFs.

Testing the Null Hyothesis of h Cointegrating Relations

In practice, we use `R` to implement this test. Specifically, we use the function `ca.jo`.

Code `johansen.R` illustrates how to use this function.

It also explains how to connect $VEC\,(p-1)$ models to their $VAR\,(p)$ representation.

This connection is useful for forecasting and analyzing IRFs. (Go over the code here.)

Overview of Unit Roots: To difference or Not to Difference?

Outline

Overview of Unit Roots: To difference or Not to Difference?

Consider a vector $\{Y_t\}$ whose dynamics we would like to describe in terms of a vector autoregression.

Overview of Unit Roots: To difference or Not to Difference?

Consider a vector $\{Y_t\}$ whose dynamics we would like to describe in terms of a vector autoregression. However, some elements in $\{Y_t\}$ may be non-stationary.

Overview of Unit Roots: To difference or Not to Difference?

Consider a vector $\{Y_t\}$ whose dynamics we would like to describe in terms of a vector autoregression. However, some elements in $\{Y_t\}$ may be non-stationary.

We have two options:

Overview of Unit Roots: To difference or Not to Difference?

Consider a vector $\{Y_t\}$ whose dynamics we would like to describe in terms of a vector autoregression. However, some elements in $\{Y_t\}$ may be non-stationary.

We have two options:

1. Difference any apparently nonstationary series.

Overview of Unit Roots: To difference or Not to Difference?

Consider a vector $\{Y_t\}$ whose dynamics we would like to describe in terms of a vector autoregression. However, some elements in $\{Y_t\}$ may be non-stationary.

We have two options:

1. Difference any apparently nonstationary series.
    - *Pros:* If the true process is a *VAR* in differences, then it will eliminate the nonstandard asymptotic distributions.

Overview of Unit Roots: To difference or Not to Difference?

Consider a vector $\{Y_t\}$ whose dynamics we would like to describe in terms of a vector autoregression. However, some elements in $\{Y_t\}$ may be non-stationary.

We have two options:

1. Difference any apparently nonstationary series.
    - *Pros:* If the true process is a *VAR* in differences, then it will eliminate the nonstandard asymptotic distributions.
    - *Cons:* If the true process is a cointegrated *VAR*, then a *VAR* in differences is misspecified.

Overview of Unit Roots: To difference or Not to Difference?

Consider a vector $\{Y_t\}$ whose dynamics we would like to describe in terms of a vector autoregression. However, some elements in $\{Y_t\}$ may be non-stationary.

We have two options:

2. Investigate carefully the nature of the nonstationarity.

Overview of Unit Roots: To difference or Not to Difference?

Consider a vector $\{Y_t\}$ whose dynamics we would like to describe in terms of a vector autoregression. However, some elements in $\{Y_t\}$ may be non-stationary.

We have two options:

2. Investigate carefully the nature of the nonstationarity. Once the nature of the nonstationarity is understood, a stationary representation for the system can be estimated.

Overview of Unit Roots: To difference or Not to Difference?

Consider a vector $\{Y_t\}$ whose dynamics we would like to describe in terms of a vector autoregression. However, some elements in $\{Y_t\}$ may be non-stationary.

We have two options:

2. Investigate carefully the nature of the nonstationarity. Once the nature of the nonstationarity is understood, a stationary representation for the system can be estimated.
    - This is my preferred approach and the one I follow in my lecture notes and the PSets.
    - *Cons:* Despite being a careful approach, the restrictions imposed may still be invalid

Overview of Unit Roots: To difference or Not to Difference?

Consider a vector $\{Y_t\}$ whose dynamics we would like to describe in terms of a vector autoregression. However, some elements in $\{Y_t\}$ may be non-stationary.

We have two options:

2. Investigate carefully the nature of the nonstationarity. Once the nature of the nonstationarity is understood, a stationary representation for the system can be estimated.
    - This is my preferred approach and the one I follow in my lecture notes and the PSets.
    - *Cons:* Despite being a careful approach, the restrictions imposed may still be invalid because we may have falsely concluded that a series is nonstationary or that cointegration is not present.

Overview of Unit Roots: To difference or Not to Difference?

Consider a vector $\{Y_t\}$ whose dynamics we would like to describe in terms of a vector autoregression. However, some elements in $\{Y_t\}$ may be non-stationary.

We have two options:

2. Investigate carefully the nature of the nonstationarity. Once the nature of the nonstationarity is understood, a stationary representation for the system can be estimated.
    - This is my preferred approach and the one I follow in my lecture notes and the PSets.
    - *Cons:* Despite being a careful approach, the restrictions imposed may still be invalid because we may have falsely concluded that a series is nonstationary or that cointegration is not present. Moreover, alternative tests can produce conflicting results.

Thank you!

Contact Information:

Vitor Possebom

E-mail: vitor.possebom@fgv.br

Website: sites.google.com/site/vitorapossebom/

References