Intro to Econometrics: Recitation 9

Gustavo Pereira

December 4, 2019

Roadmap

- Overview of restricted estimation

- Hypothesis testing

Restricted estimation

Big picture

- Setup: linear projection model,

$$y_i = x_i' \beta + u_i$$

where $\mathbf{E} x_i u_i = 0$

Restricted estimation

Big picture

- Setup: linear projection model,

$$y_i = x_i'\beta + u_i$$

  where $\mathbf{E} x_i u_i = 0$

- Suppose the relationship $F(\beta) = 0$ holds for some F.

Restricted estimation

Big picture

- Setup: linear projection model,

$$y_i = x_i'\beta + u_i$$

  where $\mathbf{E}x_i u_i = 0$
- Suppose the relationship $F(\beta) = 0$ holds for some F.
- Two applications:

Restricted estimation

Big picture

- Setup: linear projection model,

$$y_i = x_i'\beta + u_i$$

  where $\mathbf{E}x_i u_i = 0$

- Suppose the relationship $F(\beta) = 0$ holds for some F.
- Two applications:
  1. In the case we know $F(\beta) = 0$ holds in advance, can we use this to improve our estimate of $\hat{\beta}$?

Restricted estimation

Big picture

- Setup: linear projection model,

$$y_i = x_i'\beta + u_i$$

  where $\mathbf{E}x_i u_i = 0$

- Suppose the relationship $F(\beta) = 0$ holds for some F.
- Two applications:
  1. In the case we know $F(\beta) = 0$ holds in advance, can we use this to improve our estimate of $\hat{\beta}$?
  2. If we wish to *test* whether $F(\beta) = 0$, is there something we can do?

Restricted estimation
Big picture

- Consider the restricted least squares estimator,

$$\tilde{\beta} = \arg \max_{F(\beta)=0} (y - X\beta)'(y - X\beta)$$

Restricted estimation

Big picture

- Consider the restricted least squares estimator,

$$\tilde{\beta} = \arg\max_{F(\beta)=0} (y - X\beta)'(y - X\beta)$$

- Comments:

Restricted estimation

Big picture

- Consider the restricted least squares estimator,

$$\tilde{\beta} = \arg \max_{F(\beta)=0} (y - X\beta)'(y - X\beta)$$

- Comments:

1. By construction, $F(\tilde{\beta}) = 0$

Restricted estimation

Big picture

- Consider the restricted least squares estimator,

$$\tilde{\beta} = \arg \max_{F(\beta)=0} (y - X\beta)'(y - X\beta)$$

- Comments:
  1. By construction, $F(\tilde{\beta}) = 0$
  2. Of course, in general $F(\hat{\beta}) \neq 0$ where $\hat{\beta}$ is the unrestricted solution

Restricted estimation

Big picture

- Consider the restricted least squares estimator,

$$\tilde{\beta} = \arg\max_{F(\beta)=0} (y - X\beta)'(y - X\beta)$$

- Comments:
  1. By construction, $F(\tilde{\beta}) = 0$
  2. Of course, in general $F(\hat{\beta}) \neq 0$ where $\hat{\beta}$ is the unrestricted solution
- Results: under regularity conditions, if $F(\beta) = 0$ then

Restricted estimation

Big picture

- Consider the restricted least squares estimator,

$$\tilde{\beta} = \arg \max_{F(\beta)=0} (y - X\beta)'(y - X\beta)$$

- Comments:
  1. By construction, $F(\tilde{\beta}) = 0$
  2. Of course, in general $F(\hat{\beta}) \neq 0$ where $\hat{\beta}$ is the unrestricted solution
- Results: under regularity conditions, if $F(\beta) = 0$ then
  1. $\tilde{\beta}$ is consistent for β

Restricted estimation

Big picture

- Consider the restricted least squares estimator,

$$\tilde{\beta} = \arg \max_{F(\beta)=0} (y - X\beta)'(y - X\beta)$$

- Comments:
  1. By construction, $F(\tilde{\beta}) = 0$
  2. Of course, in general $F(\hat{\beta}) \neq 0$ where $\hat{\beta}$ is the unrestricted solution
- Results: under regularity conditions, if $F(\beta) = 0$ then
  1. $\tilde{\beta}$ is consistent for β
  2. $\sqrt{n}\left(\tilde{\beta} - \beta\right) \xrightarrow{d} N(0, \tilde{V})$

Restricted estimation

Big picture

- What do we gain from all this? Questions:

Restricted estimation

Big picture

- What do we gain from all this? Questions:
- Is $\tilde{\beta}$ necessarily 'better' than $\hat{\beta}$ under $F(\beta) = 0$?

Restricted estimation

Big picture

- What do we gain from all this? Questions:
- Is $\tilde{\beta}$ necessarily 'better' than $\hat{\beta}$ under $F(\beta) = 0$?
  - In general, no. See sec 8.9. However, finite sample results in the homoskedastic linear regression with linear constraint.

Restricted estimation

Big picture

- What do we gain from all this? Questions:
- Is $\tilde{\beta}$ necessarily 'better' than $\hat{\beta}$ under $F(\beta) = 0$?
  - In general, no. See sec 8.9. However, finite sample results in the homoskedastic linear regression with linear constraint.
- Can we use this machinery to tests hypotheses of the type $F(\beta) = 0$?

Restricted estimation

Big picture

- What do we gain from all this? Questions:
- Is $\tilde{\beta}$ necessarily 'better' than $\hat{\beta}$ under $F(\beta) = 0$?
  - In general, no. See sec 8.9. However, finite sample results in the homoskedastic linear regression with linear constraint.
- Can we use this machinery to tests hypotheses of the type $F(\beta) = 0$?
  - Yes. Options:

Restricted estimation

Big picture

- What do we gain from all this? Questions:
- Is $\tilde{\beta}$ necessarily 'better' than $\hat{\beta}$ under $F(\beta) = 0$?
  - In general, no. See sec 8.9. However, finite sample results in the homoskedastic linear regression with linear constraint.
- Can we use this machinery to tests hypotheses of the type $F(\beta) = 0$?
  - Yes. Options:
    1. $F(\hat{\beta})$ has a sufficiently bigger than zero magnitude (Wald test)

Restricted estimation

Big picture

- What do we gain from all this? Questions:
- Is $\tilde{\beta}$ necessarily 'better' than $\hat{\beta}$ under $F(\beta) = 0$?
  - In general, no. See sec 8.9. However, finite sample results in the homoskedastic linear regression with linear constraint.
- Can we use this machinery to tests hypotheses of the type $F(\beta) = 0$?
  - Yes. Options:
    1. $F(\hat{\beta})$ has a sufficiently bigger than zero magnitude (Wald test)
    2. $\hat{\beta}$ is too far from $\tilde{\beta}$ (minimum distance statistic)

Restricted estimation

Big picture

- What do we gain from all this? Questions:
- Is $\tilde{\beta}$ necessarily 'better' than $\hat{\beta}$ under $F(\beta) = 0$?
  - In general, no. See sec 8.9. However, finite sample results in the homoskedastic linear regression with linear constraint.
- Can we use this machinery to tests hypotheses of the type $F(\beta) = 0$?
  - Yes. Options:
    1. $F(\hat{\beta})$ has a sufficiently bigger than zero magnitude (Wald test)
    2. $\hat{\beta}$ is too far from $\tilde{\beta}$ (minimum distance statistic)
    3. The Lagrange multiplier associated with F is large ("score" type test)

Restricted estimation

Minimum distance

- Take OLS objective, add subtract $\hat{\beta}$:

$$J = (y - X\beta)'(y - X\beta) = (\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta) + \text{rest}$$

where "rest" doesn't depend on β

Restricted estimation

Minimum distance

- Take OLS objective, add subtract $\hat{\beta}$:

$$J = (y - X\beta)'(y - X\beta) = (\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta) + \text{rest}$$

  where "rest" doesn't depend on β

- CLS is the solution to a more general problem:

$$\tilde{\beta} = \arg\min_{F(\beta)=0} (\hat{\beta} - \beta)'W(\hat{\beta} - \beta)$$

  with $W = X'X$.

Restricted estimation

Minimum distance: results

- Index $\tilde{\beta}(W)$ by the choice of weighting matrix

Restricted estimation

Minimum distance: results

- Index $\tilde{\beta}(W)$ by the choice of weighting matrix
- Then:

Restricted estimation

Minimum distance: results

- Index $\tilde{\beta}(W)$ by the choice of weighting matrix
- Then:
  1. $\tilde{\beta}(W)$ is consistent

Restricted estimation

Minimum distance: results

- Index $\tilde{\beta}(W)$ by the choice of weighting matrix
- Then:
  1. $\tilde{\beta}(W)$ is consistent
  2. $\sqrt{n}(\tilde{\beta}(W) - \beta) \xrightarrow{d} N(0, \tilde{V}_\beta(W))$ where the expression for $\tilde{V}_\beta(W)$ is very long and not very informative

Restricted estimation

Minimum distance: results

- Index $\tilde{\beta}(W)$ by the choice of weighting matrix
- Then:
    1. $\tilde{\beta}(W)$ is consistent
    2. $\sqrt{n}(\tilde{\beta}(W) - \beta) \xrightarrow{d} N(0, \tilde{V}_\beta(W))$ where the expression for $\tilde{V}_\beta(W)$ is very long and not very informative
- The following holds. For any symmetric positive definite matrix W,

$$\tilde{V}_\beta(W^*) \leq \tilde{V}_\beta(W)$$

where $W^* = V_\beta^{-1}$

Restricted estimation

Minimum distance: results

- Index $\tilde{\beta}(W)$ by the choice of weighting matrix
- Then:
  1. $\tilde{\beta}(W)$ is consistent
  2. $\sqrt{n}(\tilde{\beta}(W) - \beta) \xrightarrow{d} N(0, \tilde{V}_\beta(W))$ where the expression for $\tilde{V}_\beta(W)$ is very long and not very informative
- The following holds. For any symmetric positive definite matrix W,

$$\tilde{V}_\beta(W^*) \leq \tilde{V}_\beta(W)$$

where $W^* = V_\beta^{-1}$

- In the case where F is linear, also

$$\tilde{V}_\beta(W^*) \leq V_\beta$$

Restricted estimation

Minimum distance: results

- Useful exercises:

Restricted estimation

Minimum distance: results

- Useful exercises:
    1. Make sure you can derive the expressions for constrained OLS and MD estimator

Restricted estimation

Minimum distance: results

- Useful exercises:
    1. Make sure you can derive the expressions for constrained OLS and MD estimator
    2. Check out the proof of efficient MD weighting

Statistical tests

Review of first half

- Parameter space decomposed into $B = B_0 \sqcup B_1$

Statistical tests

Review of first half

- Parameter space decomposed into $B = B_0 \sqcup B_1$
- Action space: $\mathcal{A} = \{0, 1\}$ or $\mathcal{A} = [0, 1]$

Statistical tests

Review of first half

- Parameter space decomposed into $B = B_0 \sqcup B_1$
- Action space: $\mathcal{A} = \{0, 1\}$ or $\mathcal{A} = [0, 1]$
- Loss function

$$\mathcal{L}(a, \beta) = \mathbf{1}(a = 0, \beta \in B_1) + \mathbf{1}(a = 1, \beta \in B_0)$$

Statistical tests

Review of first half

- Parameter space decomposed into $B = B_0 \sqcup B_1$
- Action space: $\mathcal{A} = \{0, 1\}$ or $\mathcal{A} = [0, 1]$
- Loss function

$$\mathcal{L}(a, \beta) = \mathbf{1}(a = 0, \beta \in B_1) + \mathbf{1}(a = 1, \beta \in B_0)$$

- Statistical test: let data realizations be denoted by $z \in \mathcal{Z}$

$$\phi : \mathcal{Z} \to \{0, 1\}$$

Statistical tests

Review of first half

- Parameter space decomposed into $B = B_0 \sqcup B_1$
- Action space: $\mathcal{A} = \{0, 1\}$ or $\mathcal{A} = [0, 1]$
- Loss function

$$\mathcal{L}(a, \beta) = \mathbf{1}(a = 0, \beta \in B_1) + \mathbf{1}(a = 1, \beta \in B_0)$$

- Statistical test: let data realizations be denoted by $z \in \mathcal{Z}$

$$\phi : \mathcal{Z} \to \{0, 1\}$$

- Define $\xi(\beta) = \mathbf{E}_\beta[\phi(z)]$

Statistical tests

Review of first half

- Parameter space decomposed into $B = B_0 \sqcup B_1$
- Action space: $\mathcal{A} = \{0, 1\}$ or $\mathcal{A} = [0, 1]$
- Loss function

$$\mathcal{L}(a, \beta) = \mathbf{1}(a = 0, \beta \in B_1) + \mathbf{1}(a = 1, \beta \in B_0)$$

- Statistical test: let data realizations be denoted by $z \in \mathcal{Z}$

$$\phi : \mathcal{Z} \to \{0, 1\}$$

- Define $\xi(\beta) = \mathbf{E}_\beta[\phi(z)]$
  - For $\beta \in B_0$, $\xi(\beta)$ is rate of type I error

    "How often does your test reject a true null?"

Statistical tests

Review of first half

- Parameter space decomposed into $B = B_0 \sqcup B_1$
- Action space: $\mathcal{A} = \{0, 1\}$ or $\mathcal{A} = [0, 1]$
- Loss function

$$\mathcal{L}(a, \beta) = \mathbf{1}(a = 0, \beta \in B_1) + \mathbf{1}(a = 1, \beta \in B_0)$$

- Statistical test: let data realizations be denoted by $z \in \mathcal{Z}$

$$\phi : \mathcal{Z} \to \{0, 1\}$$

- Define $\xi(\beta) = \mathbf{E}_\beta[\phi(z)]$
  - For $\beta \in B_0$, $\xi(\beta)$ is rate of type I error

    "How often does your test reject a true null?"
  - For $\beta \in B_1$, $\xi(\beta)$ is the *power* of a test

    "how capable is your test of rejecting a false null?"

Statistical tests

Review of first half

- In the simple null vs simple alternative case:

Statistical tests

Review of first half

- In the simple null vs simple alternative case:
  - Admissible tests maximize power subject to size

Statistical tests

Review of first half

- In the simple null vs simple alternative case:
  - Admissible tests maximize power subject to size
  - Admissible tests take the LR form, i.e., reject when

$$\frac{f(z, \beta_1)}{f(z, \beta_0)}$$

is large

Statistical tests

Review of first half

- In the simple null vs simple alternative case:
  - Admissible tests maximize power subject to size
  - Admissible tests take the LR form, i.e., reject when

$$\frac{f(z,\beta_1)}{f(z,\beta_0)}$$

    is large
- Constructed LR tests in the composite case, argued that it's approximately the same as Wald and Score tests

Statistical tests: asymptotics

- We move to the linear projection/regression model

Statistical tests: asymptotics

- We move to the linear projection/regression model
- Problem: distribution of data is not fully specified $\implies$ no likelihood function

Statistical tests: asymptotics

- We move to the linear projection/regression model
- Problem: distribution of data is not fully specified $\implies$ no likelihood function
  - E.g., linear projection model

$$y_i = x_i'\beta + u_i \qquad \mathbf{E}[x_i u_i] = 0$$

indexed by β;

Statistical tests: asymptotics

- We move to the linear projection/regression model
- Problem: distribution of data is not fully specified $\implies$ no likelihood function
  - E.g., linear projection model
    $$y_i = x_i'\beta + u_i \qquad \mathbf{E}[x_i u_i] = 0$$
    indexed by β;
- Approach: asympotics

Statistical tests: asymptotics

- We move to the linear projection/regression model
- Problem: distribution of data is not fully specified $\Longrightarrow$ no likelihood function
  - E.g., linear projection model
    $$y_i = x_i'\beta + u_i \qquad \mathbf{E}[x_i u_i] = 0$$
    indexed by β;
- Approach: asympotics
  - We will focus on constructing tests with correct (asymptotic) size

Statistical tests: asymptotics

- We move to the linear projection/regression model
- Problem: distribution of data is not fully specified $\implies$ no likelihood function
  - E.g., linear projection model

$$y_i = x_i'\beta + u_i \qquad \mathbf{E}[x_i u_i] = 0$$

  indexed by β;
- Approach: asympotics
  - We will focus on constructing tests with correct (asymptotic) size
  - Then we will try to think about comparing tests based on some notion of power

Statistical tests: asymptotics

Main example

- In the linear projection model,

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N(0, V_{\beta})$$

Statistical tests: asymptotics

Main example

- In the linear projection model,

$$\sqrt{n}\left(\hat{\beta}-\beta\right) \xrightarrow{d} N(0, V_\beta)$$

- Suppose we want to test $\mathrm{H}_0 : \beta = \beta_0$ vs $\mathrm{H}_1 : \beta \neq \beta_0$. Note that

$$\sqrt{n}\hat{V}_\beta^{-1/2}\left(\hat{\beta}-\beta\right) \xrightarrow{d} N(0, \mathbf{I}_k)$$

Statistical tests: asymptotics

Main example

- In the linear projection model,

$$\sqrt{n}\left(\hat{\beta}-\beta\right) \xrightarrow{d} N(0, V_\beta)$$

- Suppose we want to test $\mathrm{H}_0 : \beta = \beta_0$ vs $\mathrm{H}_1 : \beta \neq \beta_0$. Note that

$$\sqrt{n}\hat{V}_\beta^{-1/2}\left(\hat{\beta}-\beta\right) \xrightarrow{d} N(0, \mathbf{I}_k)$$

- Thus, if the null is correct,

$$W_n = n\left(\hat{\beta}-\beta_0\right)' \hat{V}_\beta^{-1}\left(\hat{\beta}-\beta_0\right) \xrightarrow{d} \chi_k^2$$

Statistical tests: asymptotics

Main example

- Consider the following procedure. Let K be a $1-\alpha$ quantile of χ^2_k

Statistical tests: asymptotics

Main example

- Consider the following procedure. Let K be a $1-\alpha$ quantile of χ^2_k
  - Reject the null if $W_n > K$

Statistical tests: asymptotics

Main example

- Consider the following procedure. Let K be a $1-\alpha$ quantile of χ^2_k
  - Reject the null if $W_n > K$
  - Accept the null otherwise

Statistical tests: asymptotics

Main example

- Consider the following procedure. Let K be a $1-\alpha$ quantile of χ^2_k
  - Reject the null if $W_n > K$
  - Accept the null otherwise
- What happens if the null is correct?

$$\mathbf{P}_0\left(W_n > K\right) = 1 - F_n(K) \to 1 - F(K) = \alpha$$

Statistical tests: asymptotics

Main example

- Consider the following procedure. Let K be a $1-\alpha$ quantile of χ^2_k
  - Reject the null if $W_n > K$
  - Accept the null otherwise
- What happens if the null is correct?

$$\mathbf{P}_0\left(W_n > K\right) = 1 - F_n(K) \to 1 - F(K) = \alpha$$

- What happens if $\beta \neq \beta_0$?

Statistical tests: asymptotics

Main example

- Consider the following procedure. Let K be a $1-\alpha$ quantile of χ^2_k
  - Reject the null if $W_n > K$
  - Accept the null otherwise
- What happens if the null is correct?

$$\mathbf{P}_0\left(W_n > K\right) = 1 - F_n(K) \to 1 - F(K) = \alpha$$

- What happens if $\beta \neq \beta_0$?
  - By CMT,

$$\frac{W_n}{n} = \left(\hat{\beta} - \beta_0\right)' \hat{V}_{\beta}^{-1} \left(\hat{\beta} - \beta_0\right) \xrightarrow{p} (\beta - \beta_0)' V_{\beta}^{-1} (\beta - \beta_0)$$

Statistical tests: asymptotics

Main example

- Consider the following procedure. Let K be a $1-\alpha$ quantile of χ^2_k
  - Reject the null if $W_n > K$
  - Accept the null otherwise
- What happens if the null is correct?

$$\mathbf{P}_0\left(W_n > K\right) = 1 - F_n(K) \to 1 - F(K) = \alpha$$

- What happens if $\beta \neq \beta_0$?
  - By CMT,

$$\frac{W_n}{n} = \left(\hat{\beta} - \beta_0\right)' \hat{V}_\beta^{-1}\left(\hat{\beta} - \beta_0\right) \xrightarrow{p} (\beta - \beta_0)' V_\beta^{-1}(\beta - \beta_0)$$

  - Therefore

$$W_n = n\frac{W_n}{n} \xrightarrow{p} +\infty \implies \mathbf{P}_\beta(W_n > K) \to 1$$

Statistical tests: asymptotics

Summary

- Testing $\mathrm{H}_0 : \beta \in B_0$ vs $\mathrm{H}_1 : \beta \in B_1$

Statistical tests: asymptotics

Summary

- Testing $\mathrm{H}_0 : \beta \in B_0$ vs $\mathrm{H}_1 : \beta \in B_1$
- Find statistic T_n and critical region K_n

Statistical tests: asymptotics

Summary

- Testing $\mathrm{H}_0 : \beta \in B_0$ vs $\mathrm{H}_1 : \beta \in B_1$
- Find statistic T_n and critical region K_n
- Asymptotic level requirement:

$$\lim \mathbf{P}_\beta(T_n \in K_n) \leq \alpha$$

for all $\beta \in B_0$. Typically achieve that by finding *pivotal* T_n

Statistical tests: asymptotics

Summary

- Testing $\mathrm{H}_0 : \beta \in B_0$ vs $\mathrm{H}_1 : \beta \in B_1$
- Find statistic T_n and critical region K_n
- Asymptotic level requirement:

$$\lim \mathbf{P}_\beta(T_n \in K_n) \leq \alpha$$

  for all $\beta \in B_0$. Typically achieve that by finding *pivotal* T_n

- Power: if test is level α, and

$$\lim \mathbf{P}_\beta(T_n \in K_n) \to 1$$

  for every $\beta \in B_1$, we call it *consistent*

Statistical tests: application

- Consider the model

$$y_i = x_i'\beta + u_i$$

Statistical tests: application

- Consider the model

$$y_i = x_i'\beta + u_i$$

- Want to test $r(\beta) = \theta_0$

Statistical tests: application

- Consider the model

$$y_i = x_i'\beta + u_i$$

- Want to test $r(\beta) = \theta_0$
- Two approaches:

Statistical tests: application

- Consider the model

$$y_i = x_i' \beta + u_i$$

- Want to test $r(\beta) = \theta_0$
- Two approaches:
  - Apply delta method:

$$\sqrt{n}\left[r(\hat{\beta}) - r(\beta)\right] \to r'(\beta)N(0, V_\beta)$$

Statistical tests: application

- Consider the model

$$y_i = x_i'\beta + u_i$$

- Want to test $r(\beta) = \theta_0$
- Two approaches:
  - Apply delta method:

$$\sqrt{n}\left[r(\hat{\beta}) - r(\beta)\right] \to r'(\beta)N(0, V_\beta)$$

  - Use the properties of restricted estimators

Statistical tests: application

- Consider the model

$$y_i = x_i' \beta + u_i$$

- Want to test $r(\beta) = \theta_0$
- Two approaches:
  - Apply delta method:

$$\sqrt{n}\left[r(\hat{\beta}) - r(\beta)\right] \to r'(\beta) N(0, V_\beta)$$

  - Use the properties of restricted estimators
    - It is often easier to solve minimization of squared residuals in the restricted model