1 Panel Robust Variance Estimator

The sample covariance matrix becomes

$$V\left(\hat{b}\right)=\left(\sum_{i=1}^{N}\tilde{X}_i'\tilde{X}\right)^{-1}\left(\sum_{i=1}^{N}\tilde{X}_i'\hat{u}_i\hat{u}_i'\tilde{X}_i\right)\left(\sum_{i=1}^{N}\tilde{X}_i'\tilde{X}\right)^{-1} \tag{1}$$

and its associated t-statistic becomes

$$t_{\hat{b}}=\frac{\hat{b}}{\sqrt{\left(\sum_{i=1}^{N}\tilde{X}_i'\tilde{X}\right)^{-1}\left(\sum_{i=1}^{N}\tilde{X}_i'\hat{u}_i\hat{u}_i'\tilde{X}_i\right)\left(\sum_{i=1}^{N}\tilde{X}_i'\tilde{X}\right)^{-1}}} \tag{2}$$

Consider two regressors: First let

$$\xi_i=X_i'\hat{u}_i=[x_{1,i}\hat{u}_i \quad x_{2,i}\hat{u}_i]$$

where

$$x_{k,i}=(x_{k,i1},...,x_{k,iT})'$$

Then calculate $\sum_{i=1}^{N}\xi_i'\xi_i$ which is $T\times T$ matrix.

Read Lecture note in Econometric I and find out the potential issue on this panel robust variance estimator.

2 Monte Carlo Studies

2.1 Why Do We need MC?

1. Verify asymptotic results. If an econometric theory is correct, the asymptotic results should be replicatable by means of Monte Carlo studies.
   (a) Large sample theory: T or N must be very large. At least $T=500$.
   (b) Generalize assumptions. See if a change in an assumption makes any difference in asymptotic results.
2. Examine finite sample performance. In finite sample, asymptotic results are just approximation. We don't know if or not an econometric theory works well in the finite sample.
   (a) Useful to compare with various estimators.
   (b) MSE and Bias become important to the estimation methods.
   (c) Size and Power become issues on various testing procedures & covariance estimation.

2.2 How to do MC

1. Need a data generating process (DGP), and distributional assumption.
   (a) DGP depends on an econometric theory and its assumptions.
   (b) Need to generate pseudo random variables from a certain distribution

2.2.1 Example 1: Verifying asymptotic result of OLSE

DGP:

$$\text{Model: } y_i = a + x_i\beta + u_i$$

Now we take a particular case like

$$u_i \sim iidN\left(0,1\right), \quad x_i \sim iidN\left(0,I_k\right)$$

where $a = \beta = 0$.

Step by Step procedure

1. Find out the parameters of interest. (here we are interested in consistency of OLSE)
2. Generate n pseudo random variables of u, x and y. Since $a = \beta = 0$, $y_i = u_i$.
3. Calculate OLSE for β and a. (plus the estimates of parameters of interest)
4. Repeat 2 and 3 S times. record all $\hat{\beta}$.
5. calculate mean of $\hat{\beta}$ and variance of them. (how do we know the convergence rate?)
6. Repeat 2-5 by changing n.

2.2.2 Example 2: Verifying asymptotic result of OLSE Testing

DGP:

$$\text{Model: } y_i = a + x_i\beta + u_i$$

Now we take a particular case like

$$u_i \sim iidN\left(0,1\right), \quad x_i \sim iidN\left(0,I_k\right)$$

where $a = \beta = 0$.

Step by Step procedure

1. Find out the parameters of interest. ($t-$statistic)
2. Generate n pseudo random variables of u, x and y. And calculate t ratio for β and a.
3. Repeat 2 and 3 S times. record all $t_{\hat{\beta}}$.
4. Sort $t_{\hat{\beta}}$ and find out the lower and upper 2.5% values. Compare them with the asymptotic critical value.
5. Repeat 2-4 by changing n.

2.2.3 Exercise 1: Use NW estimator and calculate t ratio. Compare the size and power of the tests (ordinary and NW t-ratios)

Asymptotic theory: Both of them are consistent. The ordinary t ratio becomes more efficient. Why?

Size of the test Change step 4 in Example 2 as follows:

Let

$$\mathbf{t}^* = \left|\hat{\mathbf{t}}_\beta\right|$$

sort t^*. Find when $t_j^* > 1.96$. And $1 - j^*/S$ becomes the size of the test.

Power of the test Change $\beta = 0.01, 0.05, 0.1, 0.2$.

Repeat the above procedures, and find $1 - j^*/S$. This becomes the power of the test.

2.2.4 Exercise 2: Re-do Bertrand et al.

3 Review Asymptotic Theory

3.1 Most Basic Theory

$$y_i = \beta x_i + u_i$$

where

$$u_i \sim iid\left(0, \sigma_u^2\right)$$

$$\hat{\beta} = \beta + \left(x'x\right)^{-1} x'u = \beta + \frac{\sum_{i=1}^n x_i u_i}{\sum_{i=1}^n x_i^2} = \beta + \frac{\frac{1}{n}\sum_{i=1}^n x_i u_i}{\frac{1}{n}\sum_{i=1}^n x_i^2}$$

First let

$$\frac{1}{n}\sum_{i=1}^n x_i u_i = \frac{1}{n}\sum_{i=1}^n \xi_i \to^d N\left(0, \frac{\sigma_\xi^2}{n}\right)$$

Hence we have

$$\sqrt{n}\frac{1}{n}\sum_{i=1}^n \xi_i \to^d N\left(0, n\frac{\sigma_\xi^2}{n}\right)$$

or

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i \to^d N\left(0, \sigma_\xi^2\right)$$

Next,

$$\text{plim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^n x_i^2 = Q_x, \text{ let say.}$$

Then

$$\hat{\beta} - \beta = \frac{\frac{1}{n}\sum_{i=1}^n x_i u_i}{\frac{1}{n}\sum_{i=1}^n x_i^2}$$

or

$$\sqrt{n}\left(\hat{\beta} - \beta\right) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i u_i}{\frac{1}{n}\sum_{i=1}^n x_i^2} \to^d N\left(0, Q_x^{-1}\sigma_\xi^2 Q_x^{-1}\right)$$

3.2 Addition Constant term

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

where

$$x_i = a_x + x_i^o, \quad y_i = a_y + y_i^o.$$

$$u_i \sim iid\left(0, \sigma_u^2\right)$$

$$\hat{\beta} = \beta + \left(\tilde{x}'\tilde{x}\right)^{-1} \tilde{x}'\tilde{u} = \beta + \frac{\sum_{i=1}^n \tilde{x}_i \tilde{u}_i}{\sum_{i=1}^n \tilde{x}_i^2} = \beta + \frac{\frac{1}{n}\sum_{i=1}^n \tilde{x}_i \tilde{u}_i}{\frac{1}{n}\sum_{i=1}^n \tilde{x}_i^2}$$

First let

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i\tilde{u}_i &= \frac{1}{n}\sum_{i=1}^{n}\left(x_i-\frac{1}{n}\sum_{i=1}^{n}x_i\right)\left(u_i-\frac{1}{n}\sum_{i=1}^{n}u_i\right)=\frac{1}{n}\sum_{i=1}^{n}x_i^o u_i^o-\frac{1}{n^2}\left(\sum x_i^o\right)\left(\sum u_i^o\right)\\ &= \frac{1}{n}\sum_{i=1}^{n}x_i^o u_i^o-\left(\frac{1}{n}\sum x_i^o\right)\left(\frac{1}{n}\sum u_i^o\right)=\frac{1}{n}\sum_{i=1}^{n}\xi_i+O_p\left(\frac{1}{\sqrt{n}}\right)O_p\left(\frac{1}{\sqrt{n}}\right)\\ &= \frac{1}{n}\sum_{i=1}^{n}\xi_i+O_p\left(\frac{1}{n}\right)\end{aligned}$$

Hence we have

$$\begin{aligned}\sqrt{n}\frac{1}{n}\sum_{i=1}^{n}\tilde{\xi}_i &= \sqrt{n}\frac{1}{n}\sum_{i=1}^{n}\xi_i+\sqrt{n}\left(\frac{1}{n}\sum x_i^o\right)\left(\frac{1}{n}\sum u_i^o\right)\\ &\to {}^{d}N\left(0,n\frac{\sigma_\xi^2}{n}\right)+O_p\left(\frac{1}{\sqrt{n}}\right)=N\left(0,\sigma_\xi^2\right).\end{aligned}$$

Next,

$$\text{plim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i^2=Q_x,\ \text{let say.}$$

Then

$$\sqrt{n}\left(\hat{\beta}-\beta\right)=\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{x}_i\tilde{u}_i}{\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i^2}\to^{d}N\left(0,Q_x^{-1}\sigma_\xi^2Q_x^{-1}\right).$$

4 Power of the Test (Local Alternative Approach)

Consider the model

$$y_i = \beta x_i + u_i$$

and under the null hypothesis, we have

$$\beta = \beta_o$$

Now we want to analyze the power of the test asymptotically. Under the alternative, we have

$$\beta = \beta_o + c$$

where $c \neq 0$.

Suppose that we are interested in comparing two estimates, let say OLSE and FGLSE ($\hat{\beta}_1$ and $\hat{\beta}_2$). Then we have

$$\frac{\sqrt{n}\left(\hat{\beta}_1 - \beta\right)}{\sqrt{V\left(\hat{\beta}_1\right)}} \to^d N(0,1) + O_p\left(N^{-1/2}\right)$$

or

$$\frac{\sqrt{n}\left(\hat{\beta}_1 - \beta_o\right)}{\sqrt{V\left(\hat{\beta}_1\right)}} \to^d N(0,1) + \sqrt{n}c + O_p\left(N^{-1/2}\right)$$

Hence as long as $c \neq 0$, the power of the test goes to one. In other words, the dominant term becomes the second term ($\sqrt{n}c$)

Similary, we have

$$\frac{\sqrt{n}\left(\hat{\beta}_2 - \beta_o\right)}{\sqrt{V\left(\hat{\beta}_2\right)}} \to^d N(0,1) + \sqrt{n}c + O_p\left(N^{-1/2}\right)$$

Hence we can't compare two tests.

Now, to avoid this, let

$$\beta = \beta_o + \frac{c}{\sqrt{n}}$$

so that $\beta \to \beta$ as $n \to \infty$. Then we have

$$\frac{\sqrt{n}\left(\hat{\beta}_\kappa - \beta\right)}{\sqrt{V\left(\hat{\beta}_\kappa\right)}} \to^d N(c,1) + O_p\left(N^{-1/2}\right).$$

Hence depending on the value of c, we can compare the power of the test (across different estimates).

5 Panel Regression

5.1 Regression Types

1. Pooled OLS estimator (POLS)

$$y_{it} = a + \beta x_{it} + \gamma z_{it} + u_{it}$$

2. Least squares dummy variables (LSDV) or Withing group (WG) or Fixed effects (FE) estimator

$$y_{it} = a_i + \beta x_{it} + \gamma z_{it} + u_{it}$$

3. Random Effect (RE) or PFGLS estimator

$$y_{it} = a + \beta x_{it} + \gamma z_{it} + e_{it}, \quad e_{it} = a_i - a + u_{it}$$

Let $X = (x_{11}, x_{12}, ..., x_{1T}, x_{21}, ..., x_{NT})'$, $\mathbf{x}_i = (x_{i1}, ..., x_{iT})'$, $\mathbf{x}_t = (x_{1t}, ..., x_{Nt})'$. Define Z, $\mathbf{z}_i$ and $\mathbf{z}_t$ in the similar way. Let $W = (X\ Z)'$. Then

5.2 Covariance estimators:

1. Ordinary estimator: $\hat{\sigma}_u^2 (W'W)^{-1}$

2. White estimator

   (a) Cross sectional heteroskedasticity: $NT (W'W)^{-1} \left(\frac{1}{N} \sum_{i=1}^{n} \hat{\mathbf{u}}_i^2 \mathbf{w}_i' \mathbf{w}_i \right) (W'W)^{-1}$

   (b) Time series heteroskedasticity: $NT (W'W)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} \hat{\mathbf{u}}_t^2 \mathbf{w}_t' \mathbf{w}_t \right) (W'W)^{-1}$

   (c) Cross and Time heteroskedasticity: $NT (W'W)^{-1} \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{it}^2 \mathbf{w}_{it} \mathbf{w}_{it}' \right) (W'W)^{-1}$

3. Panel Robust Covariane estimator: $N (W'W)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{w}_i' \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i' \mathbf{w}_i \right) (W'W)^{-1}$

4. LRV estimator ? Why not?

5.3 Pooled GLS Estimators

$$\hat{\delta} = \left[W' \left(\Omega^{-1} \otimes I \right) W \right]^{-1} \left[W' \left(\Omega^{-1} \otimes I \right) y \right]$$

5.3.1 How to estimate Ω :

1. Time Series Correlation:

   (a) AR1: easy to extend. $\Omega = \begin{bmatrix} 1 & & \rho^{T-1} \\ \vdots & \ddots & \vdots \\ \rho^{T-1} & & 1 \end{bmatrix}$

   (b) Unknown. $\hat{\Omega}_{sh} = \frac{1}{N}\sum_{i=1}^{N} \hat{u}_{is}\hat{u}_{ih}$. Required small T and large N.

2. Cross sectional correlation

   (a) Spatial: Easy.

   (b) Unknown. $\hat{\Omega}_{sh} = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_{st}\hat{u}_{ht}$

5.4 Seemingly Unrelated Regression

$$\hat{\delta} = \left[W'\left(I \otimes \Omega^{-1}\right)W\right]^{-1}\left[W'\left(I \otimes \Omega^{-1}\right)y\right]$$

6 Bootstrap

Reference: "The BOOTSTRAP" by Joel L. Horowitz (Chapter 52 in Handbook of Econometrics Vol 5)

6.1 What is the bootstrap

It is a method for estimating the distribution of an estimator or test statistics by resampling the data.

Example 1 (Bias correction) Model

$$y_t = a + \rho y_{t-1} + e_t,$$

where e_t is a white noise process. It is well known that $\mathrm{E}(\hat{\rho} - \rho) = -\dfrac{1+3\rho}{T} + O\left(T^{-2}\right)$. Here I am explaining how to reduce Kendall bias (not eliminating) by using the following bootstrap procedure.

1. Estimate OLSE for a and ρ, denote them as $\hat{a}$ and $\hat{\rho}$. Get OLS residual $\hat{e}_t = y_t - \hat{a} - \hat{\rho} y_{t-1}$.
2. generate $T + K$ random variables from the uniform distribution of $U(1, T-1)$. Make them as integers.

```
ind = rand(t+k,1)*(t-1); % generate from U(0,T-1).
ind = 1+floor(ind); % make integers. 0.1 => 1.
```

3. Draw $(T+K) \times 1$ vector of e_t^* from $\hat{e}_t$.

```
esta = e(ind,:);
```

4. Recentering e_t^* to make its mean be zero. Generate pseudo y_t^* from e_t^*, and discard the first K obs.

```
esta = esta - mean(esta); ysta = esta;
       for i=2:t+k;
       ysta(i,:) = ahat+rhohat*ysta(i-1,:) + esta(i,:);
       end;
ysta =ysta(k+1:t+k,:);
```

5. Estimate $\hat{a}^*$ and $\hat{\rho}^*$ with y_t^*.

6. Repeat step 2 and 5 M times.

7. Calculate the sample mean of $\hat{\rho}^*$. Calculate the bootstrap bias, $B = \frac{1}{M}\sum_{m=1}^{M}\hat{\rho}_m^* - \hat{\rho}$ where $\hat{\rho}_m^*$ is the mth time bootstrapped point estimate of ρ. Subtract B from $\hat{\rho}$.

$$\hat{\rho}_{\text{mue}} = \hat{\rho} - B$$

where mue stands from mean unbiased estimator. Note that

$$E\left(\hat{\rho}_{\text{mue}} - \rho\right) = O\left(T^{-2}\right).$$

6.2 How the bootstrap works

First let the estimates be a function of T. For example, $\hat{\rho}$ be $\hat{\rho}_T$. Now define

$$\hat{\rho}_T = \frac{\sum \tilde{y}_{it}\tilde{y}_{it-1}}{\sum \tilde{y}_{it-1}^2} = g\left(z\right), \text{ let say}$$

where z is a 2×1 vector. That is, $z = (z_1, z_2)$ and $z_1 = \frac{1}{T}\sum \tilde{y}_{it}\tilde{y}_{it-1}$ and $z_2 = \frac{1}{T}\sum \tilde{y}_{it-1}^2$.

From A Tyalor expansion (or Delta method), we have

$$\hat{\rho}_T = \rho + \frac{\partial g}{\partial z}\left(z - z_o\right) + \frac{1}{2}\left(z - z_o\right)'\left(\frac{\partial^2 g}{\partial z \partial z'}\right)\left(z - z_o\right) + O_p\left(T^{-2}\right)$$

Now taking expectations yields

$$\begin{aligned} E\left(\hat{\rho}_T - \rho\right) &= E\frac{\partial g}{\partial z}\left(z - z_o\right) + \frac{1}{2}E\left(z - z_o\right)'\left(\frac{\partial^2 g}{\partial z \partial z'}\right)\left(z - z_o\right) + O\left(T^{-2}\right) \\ &= \frac{1}{2}E\left(z - z_o\right)'\left(\frac{\partial^2 g}{\partial z \partial z'}\right)\left(z - z_o\right) + O\left(T^{-2}\right) \end{aligned}$$

since $E\left(z - z_o\right) = 0$ always.

The first term in the above becomes $O\left(T^{-1}\right)$, that is $-\dfrac{1+3\rho}{T}$. We want to eliminate this part (not reduce it). The bootstrapped $\hat{\rho}_T^*$ becomes

$$\hat{\rho}_T^* = \hat{\rho}_T + \frac{\partial g}{\partial z}\left(z^* - z_o\right) + \frac{1}{2}\left(z^* - z_o\right)'\left(\frac{\partial^2 g}{\partial z \partial z'}\right)\left(z^* - z_o\right) + O_p\left(T^{-2}\right)$$

where $z^* = \left(z_1^*,\ z_2^*\right)$, and $z_1^* = \frac{1}{T}\sum \tilde{y}_{it}^*\tilde{y}_{it-1}^*$, etc. Note that we generate y_{it}^* from $\hat{\rho}_T$, $\hat{\rho}_T^*$ can be expanded around $\hat{\rho}_T$ not around the true value of ρ. Now taking expectation E^* in the sense that

$$E^* \to E \text{ as } M, T \to \infty.$$

Then we have

$$\begin{aligned} E^*\left(\hat{\rho}_T^* - \hat{\rho}_T\right) &= \frac{1}{2} E^*\left(z^* - z_o\right)' \left(\frac{\partial^2 g}{\partial z \partial z'}\right)\left(z^* - z_o\right) + O\left(T^{-2}\right) \\ &= B^* \end{aligned}$$

Note that in general

$$B^* = B + O\left(T^{-2}\right)$$

hence we have

$$\hat{\rho}_{\text{mue}} = \hat{\rho}_T - B^* = \hat{\rho}_T - E^*\left(\hat{\rho}_T^* - \hat{\rho}_T\right)$$

6.3 Bootstrapping Critical Value

Example 2. (Using the same example 1) Generate t-ratio for $\hat{\rho}_m^*$ M times. Sort them, and find 95% critical value from the bootstrapped t-ratio. Compare it with the actual t-ratio.

Asymptotic Refinement Notation:

F_0 is the true cumulative density function. For an example, cdf of normal distribution.
t_β is the t-statistic of β.
$t_{n,\beta}$ is the sample t-statistic of $\hat{\beta}$ where n is the sample size.
$G(\tau, F_0) = P(t_\beta \leq \tau)$. That is the function G is the true CDF of t_β.
$G_n(\tau, F_0) = P(t_{n,\beta} \leq \tau)$. The function G_n is the exact finite sample CDF of $t_{n,\beta}$

Asymptotically $G_n \to G$ as $n \to \infty$. Denote that $G_n(\tau, F_n)$ is the bootstrapped function for $t_{n,\beta}^*$ where F_n is the finite sample CDF.

Definition: Pivotal statistics If $G_n(\tau, F_0)$ does not depend on F_0, then $t_{n,\beta}$ is said to be pivotal.

Example 3 (exact finite sample CDF for AR(1) with a unknown constant) From Tanaka (1983, Econometrica), the exact finite sample CDF for $t_{\hat{\rho}}$ is given by

$$P\left(t_{T,\hat{\rho}} \leq x\right) = \Phi(x) + \frac{\phi(x)}{\sqrt{T}} \frac{2\rho + 1}{\sqrt{1 - \rho^2}} + O\left(T^{-1}\right)$$

where Φ is the CDF of normal distribution and ϕ is PDF of normal. Here Tanaka assumes F_0 is normal. That is, y_t is distributed as normal. Of course, if y_t has a different distribution, the exact finite sample PDF is unknown. However, $t_{T,\hat{\rho}}$ is pivotal since as $T \to \infty$, its limiting distribution goes to $\Phi(x)$.

Now under some regularity conditions (see Theorem 3.1 Horowitz), we have

$$G_n(\tau, F_0) = G(\tau, F_0) + \frac{1}{\sqrt{n}} g_1(\tau, F_0) + \frac{1}{n} g_2(\tau, F_0) + \frac{1}{n^{3/2}} g_3(\tau, F_0) + O\left(n^{-2}\right)$$

uniformly over τ.

Meanwhile the bootstrapped $t_{\hat{\beta}}$ has the following properties

$$G_n(\tau, F_n) = G(\tau, F_n) + \frac{1}{\sqrt{n}} g_1(\tau, F_n) + \frac{1}{n} g_2(\tau, F_n) + \frac{1}{n^{3/2}} g_3(\tau, F_n) + O\left(n^{-2}\right)$$

When $t_{n,\hat{\beta}}$ is not a pivotal statistic In this case, we have

$$G_n(\tau, F_0) - G_n(\tau, F_n) = \left[G(\tau, F_0) - G(\tau, F_n)\right] + \frac{1}{\sqrt{n}}\left[g_1(\tau, F_0) - g_1(\tau, F_n)\right] + O\left(n^{-1}\right)$$

Note that $G(\tau, F_0) - G(\tau, F_n) = O\left(n^{-1/2}\right)$. Hence the bootstrap makes an error of size $O\left(n^{-1/2}\right)$. Also note that $G_n(\tau, F_0)$ also makes an error of size $O\left(n^{-1/2}\right)$, so that the bootstrap does not reduce (neither increase) the size of the error.

When $t_{n,\hat{\beta}}$ is a pivotal In this case, we have

$$G(\tau, F_0) - G(\tau, F_n) = 0$$

by definition. Then we have

$$G_n(\tau, F_0) - G_n(\tau, F_n) = \frac{1}{\sqrt{n}}\left[g_1(\tau, F_0) - g_1(\tau, F_n)\right] + O\left(n^{-1}\right)$$

and $g_1(\tau, F_0) - g_1(\tau, F_n) = O\left(n^{-1/2}\right)$. Hence we have

$$G_n(\tau, F_0) - G_n(\tau, F_n) = O\left(n^{-1}\right),$$

which implies that the bootstrap reduces the size of an error.

6.4 Exercise: Sieve Bootstrap

(Read Li and Maddala, 1997)

Consider the following cross sectional regression

$$y_{it} = a + \beta x_{it} + u_{it} \tag{3}$$

We want to test the null hypothesis of $\beta = 0$. We suspect that x_{it} and u_{it} are serially correlated, but not cross correlated. Consider the following sieve bootstrap procedure

1. Run (3) and get $\hat{a}, \hat{\beta}$, and $\hat{u}_{it}$.

2. Run the following regression

$$\begin{bmatrix} x_{it} \\ u_{it} \end{bmatrix} = \begin{bmatrix} \mu_x \\ 0 \end{bmatrix} + \begin{bmatrix} \rho_x & 0 \\ 0 & \rho_u \end{bmatrix} \begin{bmatrix} x_{it} \\ u_{it} \end{bmatrix} + \begin{bmatrix} e_{it} \\ \varepsilon_{it} \end{bmatrix}$$

   and get $\hat{\mu}_x, \hat{\rho}_x, \hat{\rho}_u$ and their residuals of $\hat{e}_{it}$ and $\hat{\varepsilon}_{it}$. Recentering them.

3. Generate pseudo x^*_{it} and u^*_{it}.

   ind = rand(t+k,1)*(t-1); % generate from U(0,T-1).
   ind = 1+floor(ind); % make integers. 0.1 => 1.
   F = [ehat espi]; % $\hat{e}_{it}$ and $\hat{\varepsilon}_{it}$
   Fsta = F(ind,:); % use the same ind. Important!
   repeat what you learnt before....

4. Generate y^*_{it} under the null,

$$y^*_{it} = \hat{a} + u^*_{it}.$$

5. Run (3) with y^*_{it} and x^*_{it}, and get the bootstrapped critical value.

7 Maximum Likelihood Estimation

7.1 The likelihood function

Let $y_1, ..., y_n$, $\{y_i\}$, be a sequence of random variables which has iid$N\left(\mu, \sigma^2\right)$. Its probability density function $f\left(y|\mu, \sigma^2\right)$ can be written as

$$f\left(y = y_i|\mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left(y_i - \mu\right)^2\right].$$

Note that this pdf states that with given μ and σ^2, the probability of $y_i = y$. Now let think about the joint density of y such that

$$\begin{aligned} f\left(y_1, ..., y_n|\mu, \sigma^2\right) &= f\left(y_1|\mu, \sigma^2\right) \times \cdots \times f\left(y_n|\mu, \sigma^2\right) \\ &= \prod_{i=1}^{n} f\left(y_i|\mu, \sigma^2\right) \text{ due to independence} \end{aligned}$$

That is, with given μ and σ^2, the joint pdf states that the probability of a sequence of y to be $\{y_i\}$. This concept is very useful when we do both/or MC and bootstrap.

Now consider the mirror image case. Given $\{y_i\}$, what are the most probable estimates for μ and σ^2? To answer this question, we consider the likelihood (probability) of μ and σ^2. Let $\theta = \left(\mu, \sigma^2\right)$. Then we can re-interpret the joint pdf as the likelihood function. That is,

$$f\left(y|\theta\right) = L\left(\theta|y\right).$$

And then we maximize the likelihood with given $\{y_i\}$.

$$\arg\max_{\theta} L\left(\theta|y\right).$$

However it is often difficult to maximize directly L function due to nonlinearity. Hence alternatively we maximize the log likelihood

$$\arg\max_{\theta} \ln L\left(\theta|y\right).$$

In practice (computer programming) it is much easier to minimize the negative log likelihood such that

$$\arg\min_{\theta} -\ln L\left(\theta|y\right)$$

Of course, we have to get the first order conditions with respect to θ, and find the optimal values of θ.

Example 1 Normal random variables. $\{y_i\}$, $i = 1, ..., n$. Want to estimate μ and σ^2.

$$\begin{aligned} L\left(\mu, \sigma^2 | y\right) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left(y_i - \mu\right)^2\right] \\ &= \left[\frac{1}{\sqrt{2\pi\sigma^2}}\right]^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - \mu\right)^2\right] \end{aligned}$$

since

$$\exp(a)\exp(b) = \exp(a+b).$$

Hence

$$\ln L = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2}\sum_{i=1}^{n}\left[\frac{\left(y_i - \mu\right)^2}{\sigma^2}\right]$$

Note that

$$\begin{aligned} \frac{\partial \ln L}{\partial \mu} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(y_i - \mu\right) = 0, \\ \frac{\partial \ln L}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}\left(y_i - \mu\right)^2 = 0 \end{aligned}$$

From this, we have

$$\hat{\mu}_{mle} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

and

$$\hat{\sigma}^2_{mle} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{\mu}_{mle}\right)^2$$

Properties of an MLE (Theorem 16.1 Green)

1. Consistency: $\hat{\theta}_{mle} \to^p \theta$

2. Asymptotic normality

   $$\hat{\theta}_{mle} \to^d N\left(\theta, \left[I(\theta)^{-1}\right]\right)$$

   where

   $$I(\theta) = -E\left[\frac{\partial^2 \ln L}{\partial\theta\partial\theta'}\right] = -E(H)$$

   where H is the Hessian matrix.

3. Asymptotic Efficiency: $\hat{\theta}_{mle}$ is asymptotically efficient and achieves the Cramer-Rao lower bound for consistent estimators.

8 Sample Midterm Exam

Model

$$y_{it} = a_i + \beta x_{it} + u_{it}, \quad t = 1, ..., T; i = 1, ..., N \tag{4}$$

$$u_{it} = \rho u_{it-1} + v_{it}, \quad x_{it} = \rho x_{it-1} + e_{it} \text{ for time series and panel cases} \tag{5}$$

where v_{it},e_{it} are independent each other.

8.1 Matlab Exercise:

1. Estimators
   (a) Cross section: Let $t = 1$, $N = n$. Provide matlab codes for OLS, WLS (weighted least squares)
   (b) Time series: Let $N = 1, T = T$. Provide matlab codes for OLS.
   (c) Panel data: Provide matlab codes for POLS, LSDV, PGLS (infeasible GLS)
2. t-statistics
   (a) Cross section: provide t ratios for ordinary and white.
   (b) Time series: provide t ratios for ordinary and NW.
   (c) Panel Data: provide t ratios for ordinary and panel robust.
3. Monte Carlo Study. Assume all innovations are iidN(0,1). (Don't need to write up matlab codes)
   (a) want to show that $\hat{\beta}_{LSDV}$ is inconsistent. Write down how you can do by means of MC.
   (b) want to show that $t_{\hat{\beta}} = \hat{\beta}_{LSDV} / \sqrt{\hat{\sigma}_u^2 / \left\{ \sum_{i=1}^N \sum_{t=1}^T \left(x_{it} - \frac{1}{T} \sum_{t=1}^T x_{it} \right)^2 \right\}}$ suffers from size distortion. Write down step by step procedure how you can show it by means of MC.
4. Bootstrap.
   (a) write up the bootstrap procedure (step by step) how to construct the bootstrapped critical value for $t_{\hat{\beta}}$ in 3.b. under the null hypothesis $\beta = 0$.

8.2 Theory

1. Basic: Derive the limiting distribution of $\hat{\beta}_{LSDV}$ in (1) and (2)

2. Suppose that

$$u_{it} = \theta_t + \varepsilon_{it} \tag{6}$$

where θ_t is independent from x_{it}.

   (a) You run eq. (1). (y on a_i and x_{it}). Show $\hat{\beta}_{LSDV}$ is consistent.

   (b) Further assume that ε_{it} is a white noise but θ_t follows an AR(1) process. How can you obtain more efficient estimator by using a simple transformation. (Don't think about MLE)

3. Now we have

$$u_{it} = \lambda_i \theta_t + \varepsilon_{it}$$

   (a) Show that $\hat{\beta}_{LSDV}$ is still consistent as long as u_{it} is independent from x_{it}.

   (b) Can you eliminate $\lambda_i \theta_t$? If so, how?

4. DGP is given by

$$y_{it} = a + \beta x_{it} + \omega_{it}, \quad \omega_{it} = (a_i - a) + u_{it}, \quad u_{it} = \rho u_{it-1} + \varepsilon_{it}, \quad \varepsilon_{it} \sim iidN\left(0, \sigma_\varepsilon^2\right). \tag{7}$$

   (a) you want to estimate the set of parameters by maximizing log likelihood function. Write down the set of parameters.

   (b) Write down the log likelihood function and its F.O.C.

   (c) Derive MLEs when $\rho = 0$ and this information is given to you.