MLE and MOM

1 Introduction

Maximum Likelihood Estimation (MLE) and Method of Moments (MOM) estimation are two different estimation strategies. Sometimes they result in different estimators, sometimes not. Sometimes MLE provides direction for moments to focus on in MOM.

2 MLE

2.1 Examples

1.

$$\begin{aligned} X_i &\sim iidBin\left(N,p\right), \quad i=1,2,..,n \\ &\Rightarrow f\left(x_i\right)=\binom{N}{x_i}p^{x_i}\left(1-p\right)^{N-x_i} \\ &\Rightarrow L\left(x\mid p\right)=\prod_{i=1}^{n}\binom{N}{x_i}p^{x_i}\left(1-p\right)^{N-x_i} \\ &\Rightarrow \log L\left(x\mid p\right)=\sum_{i=1}^{n}\log\binom{N}{x_i}+x_i\log p \\ &\quad +\left(N-x_i\right)\log\left(1-p\right) \\ &\Rightarrow \frac{\partial\log L\left(x\mid p\right)}{\partial p}=\sum_{i=1}^{n}\left[\frac{x_i}{\widehat{p}}-\frac{N-x_i}{1-\widehat{p}}\right]=0 \\ &\Rightarrow \widehat{p}=\frac{1}{nN}\sum_{i=1}^{n}x_i. \end{aligned}$$

2.

$$\begin{aligned} X_i &\sim iidU\left(-\theta,\theta\right), \quad i=1,2,..,n \\ &\Rightarrow L\left(x\mid\theta\right)=\prod_{i=1}^{n}\frac{1}{2\theta}1\left(-\theta\leq x_i\leq\theta\right). \end{aligned}$$

Consider $L_{\theta}\left(x\mid\theta\right)$ (draw a picture) Where is the maximum? Why did this happen?

3.

$$\begin{aligned} y_t &= X_t\beta+u_t, \quad t=1,2,..,T; \\ u_t &\sim iidN\left(0,\sigma^2\right) \end{aligned}$$

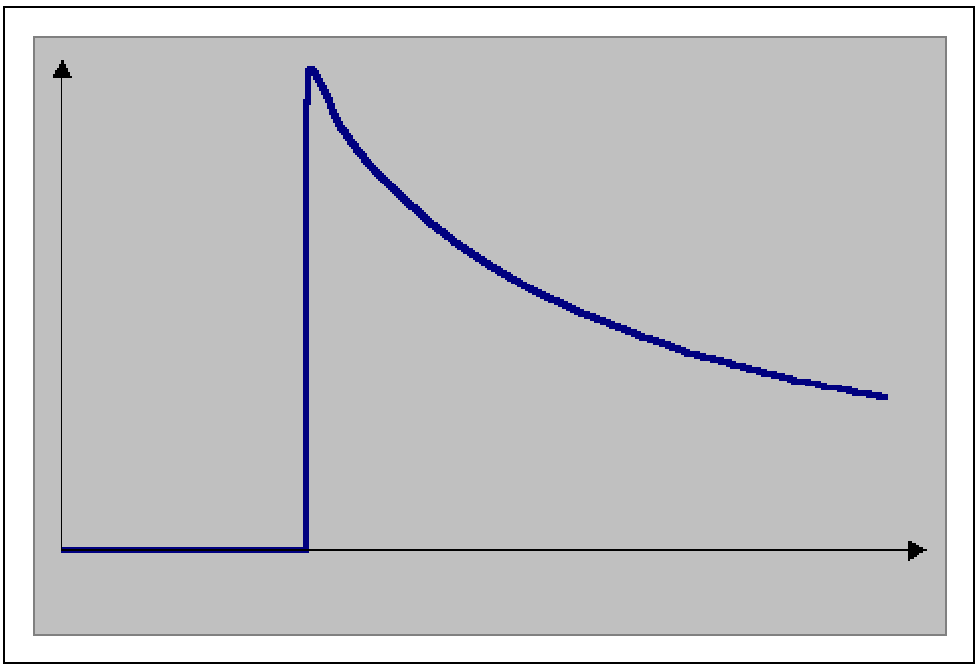

Figure 1:

$$\begin{aligned}
&\Rightarrow \quad L(y \mid \beta, \sigma) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2\sigma} (y_t - X_t \beta)^2 \right\} \\
&\Rightarrow \quad \log L(y \mid \beta, \sigma) = \sum_{t=1}^{T} -\frac{1}{2} \log 2\pi - \log \sigma - \frac{1}{2\sigma} (y_t - X_t \beta)^2 .
\end{aligned}$$

Maximizing $\log L(y \mid \beta, \sigma)$ is equivalent to minimizing

$$\sum_{t=1}^{T} (y_t - X_t \beta)^2$$

which is the OLS problem. So the MLE of β is

$$\widehat{\beta} = (X'X)^{-1} X'y.$$

4.

$$\begin{aligned}
y &= X\beta + u; \\
u &\sim N(0, \Omega).
\end{aligned}$$

$$\begin{aligned}
&\Rightarrow \quad L(y \mid \beta, \Omega) = (2\pi)^{-T/2} |\Omega|^{-1/2} \exp\left\{ -\frac{1}{2} (y - X\beta)' \Omega^{-1} (y - X\beta) \right\} \\
&\Rightarrow \quad \log L(y \mid \beta, \Omega) = -\frac{T}{2} \log 2\pi - \frac{1}{2} \log |\Omega| - \frac{1}{2} (y - X\beta)' \Omega^{-1} (y - X\beta) \\
&\Rightarrow \quad \frac{\partial \log L(y \mid \beta, \Omega)}{\partial \beta} = \frac{1}{2} \left[X'\Omega^{-1} (y - X\beta) + (y - X\beta)' \Omega^{-1} X \right] \\
&\Rightarrow \quad \widehat{\beta} = \left(X'\Omega^{-1} X \right)^{-1} X'\Omega^{-1} y
\end{aligned}$$

which is the GLS estimator.

5. Full Information Maximum Likelihood (FIML):

$$\underset{n\times n}{B} \underset{n\times T}{Y'} = \underset{n\times m}{C} \underset{m\times T}{X'} + \underset{n\times T}{U'}$$

or

$$\underset{n\times n}{B} \underset{n\times 1}{y_t} = \underset{n\times m}{C} \underset{m\times 1}{x_t} + \underset{n\times 1}{u_t}$$

with

$$u_t \sim iidN(0, \Omega).$$

$$\Rightarrow L = (2\pi)^{-nT/2} |\Omega|^{-T/2} |B|^{T} \prod_{t=1}^{T} \exp\left\{ -\frac{1}{2} (By_t - Cx_t)' \Omega^{-1} (By_t - Cx_t) \right\}.$$

The term $|B|^T$ is the Jacobian. Let

$$\underset{(n+m)\times 1}{Z_t} = \begin{pmatrix} y_t \\ x_t \end{pmatrix}; \quad \underset{n\times(n+m)}{P} = (B \mid -C).$$

Then the log likelihood function can be written as

$$\begin{aligned}
\log L &= -\frac{nT}{2}\log 2\pi - \frac{T}{2}|\Omega| + T\log B - \frac{1}{2}\sum_{t=1}^{T}(PZ_t)'\,\Omega^{-1}(PZ_t) \\
&= -\frac{nT}{2}\log 2\pi - \frac{T}{2}|\Omega| + T\log B - \frac{1}{2}trZP'\Omega^{-1}PZ' \\
&= -\frac{nT}{2}\log 2\pi - \frac{T}{2}|\Omega| + T\log B - \frac{1}{2}tr\Omega^{-1}PZ'ZP'
\end{aligned}$$

$\Rightarrow$

$$\begin{aligned}
d\log L &= -\frac{T}{2}tr\left(\Omega^{-1}d\Omega\right) + \frac{1}{2}tr\left(\Omega^{-1}d\Omega\Omega^{-1}PZ'ZP'\right) \\
&= -\frac{T}{2}tr\left[d\Omega\left(\Omega^{-1} - \Omega^{-1}P\frac{Z'Z}{T}P'\Omega^{-1}\right)\right] = 0 \\
&\Rightarrow\ \widehat{\Omega}^{-1} = \widehat{\Omega}^{-1}P\frac{Z'Z}{T}P'\widehat{\Omega}^{-1} \\
&\Rightarrow\ \widehat{\Omega} = P\frac{Z'Z}{T}P' = \frac{1}{T}\sum_{t=1}^{T}(By_t - Cx_t)(By_t - Cx_t)' \\
&= \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t\widehat{u}_t'.
\end{aligned}$$

We can now plug $\widehat{\Omega}$ into the log likelihood function (concentrate) to get

$$\begin{aligned}
= \ & -\frac{nT}{2}\log 2\pi - \frac{T}{2}\left|P\frac{Z'Z}{T}P'\right| + T\log B - \frac{1}{2}tr\left(P\frac{Z'Z}{T}P'\right)^{-1}PZ'ZP' \\
& -\frac{nT}{2}\log 2\pi - \frac{T}{2}\left|P\frac{Z'Z}{T}P'\right| + T\log B - \frac{1}{2}(n+m).
\end{aligned}$$

We can now maximize over P (subject to any restrictions on P).

2.2 Asymptotics of MLE

Let

$$X_i \sim iidf(x \mid \theta).$$

Let $\widehat{\theta}$ be the MLE of θ.

Theorem 1 *plim* $\widehat{\theta} = \theta$.

Proof.

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f\left(x_i \mid \widehat{\theta}\right) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f\left(x_i \mid \theta\right) + \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta\partial\theta'}\log f\left(x_i \mid \theta\right)\left(\widehat{\theta}-\theta\right)+o_p\left(\frac{1}{n}\right) \quad (1)$$

is a Taylor series approximation of the FOC for the log likelihood function (assuming that there is an interior solution for the maximiming θ). Note that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f\left(x_i \mid \widehat{\theta}\right)=0 \quad (2)$$

because $\widehat{\theta}$ is chosen so that equation (2) is true. So equation (1) becomes

$$0=\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f\left(x_i \mid \theta\right)+\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta\partial\theta'}\log f\left(x_i \mid \theta\right)\left(\widehat{\theta}-\theta\right)+o_p\left(\frac{1}{n}\right). \quad (3)$$

Consider the plim of this equation, and first consider

$$plim\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f\left(x_i \mid \theta\right).$$

Note that

$$\begin{aligned}
E\frac{\partial}{\partial\theta}\log f\left(x_i \mid \theta\right) &= \int\frac{\partial}{\partial\theta}\log f\left(x_i \mid \theta\right)f\left(x_i \mid \theta\right)dx_i \\
&= \int\frac{f_\theta\left(x_i \mid \theta\right)}{f\left(x_i \mid \theta\right)}f\left(x_i \mid \theta\right)dx_i \\
&= \int f_\theta\left(x_i \mid \theta\right)dx_i \\
&= \frac{\partial}{\partial\theta}\int f\left(x_i \mid \theta\right)dx_i \\
&= \frac{\partial}{\partial\theta}1=0.
\end{aligned}$$

$$\Rightarrow plim\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f\left(x_i \mid \theta\right)=0.$$

Next consider

$$plim\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta\partial\theta'}\log f\left(x_i \mid \theta\right).$$

Note that

$$
\begin{aligned}
E\frac{\partial^2}{\partial\theta\partial\theta'}\log f(x_i \mid \theta) &= \int \frac{\partial^2}{\partial\theta\partial\theta'}\log f(x_i \mid \theta)\, f(x_i \mid \theta)\, dx_i \\
&= \int \frac{\partial}{\partial\theta'}\left[\frac{\partial}{\partial\theta}\log f(x_i \mid \theta)\right] f(x_i \mid \theta)\, dx_i \\
&= \int \frac{\partial}{\partial\theta'}\frac{f_\theta(x_i \mid \theta)}{f(x_i \mid \theta)} f(x_i \mid \theta)\, dx_i \\
&= \int \left[\frac{f_{\theta\theta'}(x_i \mid \theta)}{f(x_i \mid \theta)} - \frac{f_\theta(x_i \mid \theta)\, f_{\theta'}(x_i \mid \theta)}{f(x_i \mid \theta)^2}\right] f(x_i \mid \theta)\, dx_i \\
&\quad -\int \frac{f_\theta(x_i \mid \theta)\, f_{\theta'}(x_i \mid \theta)}{f(x_i \mid \theta)^2} f(x_i \mid \theta)\, dx_i \\
&\quad -\int \frac{\partial \log f(x_i \mid \theta)}{\partial\theta}\frac{\partial \log f(x_i \mid \theta)}{\partial\theta'} f(x_i \mid \theta)\, dx_i \\
&\quad -E\left[\frac{\partial \log f(x_i \mid \theta)}{\partial\theta}\frac{\partial \log f(x_i \mid \theta)}{\partial\theta'}\right].
\end{aligned}
$$

$\Rightarrow$

$$
plim\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta\partial\theta'}\log f(x_i \mid \theta) = -E\left[\frac{\partial \log f(x_i \mid \theta)}{\partial\theta}\frac{\partial \log f(x_i \mid \theta)}{\partial\theta'}\right].
$$

$\Rightarrow$ equation (1) becomes

$$
\begin{aligned}
0 &= plim\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta\partial\theta'}\log f(x_i \mid \theta)\left(\widehat{\theta}-\theta\right) \qquad (4) \\
&= plim\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta\partial\theta'}\log f(x_i \mid \theta)\, plim\left(\widehat{\theta}-\theta\right).
\end{aligned}
$$

As long as

$$
plim\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta\partial\theta'}\log f(x_i \mid \theta)
$$

has full rank (this is an identification condition), equation (4) implies that

$$
plim\left(\widehat{\theta}-\theta\right) = 0.
$$

■

Theorem 2 $\sqrt{n}\left(\widehat{\theta}-\theta\right) \sim N(0,V)$ *with*

$$
V = E\left[\frac{\partial \log f(x_i \mid \theta)}{\partial\theta}\frac{\partial \log f(x_i \mid \theta)}{\partial\theta'}\right].
$$

Proof. From equation (3)

$$\sqrt{n}\left(\widehat{\theta}-\theta\right)=-\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^{2}}{\partial\theta\partial\theta'}\log f\left(x_{i}\mid\theta\right)\right]^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f\left(x_{i}\mid\theta\right)\right]$$

$\Rightarrow$

$$\begin{aligned}
& plim\ n\left(\widehat{\theta}-\theta\right)\left(\widehat{\theta}-\theta\right)' \\
= & -plim\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^{2}}{\partial\theta\partial\theta'}\log f\left(x_{i}\mid\theta\right)\right]^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f\left(x_{i}\mid\theta\right)\right]\cdot \\
& \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\theta'}\log f\left(x_{i}\mid\theta\right)\right]\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^{2}}{\partial\theta\partial\theta'}\log f\left(x_{i}\mid\theta\right)\right]^{-1}
\end{aligned}$$

$$\begin{aligned}
= & -plim\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^{2}}{\partial\theta\partial\theta'}\log f\left(x_{i}\mid\theta\right)\right]^{-1}\cdot \\
& \left[\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial}{\partial\theta}\log f\left(x_{i}\mid\theta\right)\frac{\partial}{\partial\theta'}\log f\left(x_{j}\mid\theta\right)\right]\cdot \\
& \left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^{2}}{\partial\theta\partial\theta'}\log f\left(x_{i}\mid\theta\right)\right]^{-1}
\end{aligned}$$

$$\begin{aligned}
= & -\left[plim\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^{2}}{\partial\theta\partial\theta'}\log f\left(x_{i}\mid\theta\right)\right]^{-1}\cdot \\
& \left[plim\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial}{\partial\theta}\log f\left(x_{i}\mid\theta\right)\frac{\partial}{\partial\theta'}\log f\left(x_{j}\mid\theta\right)\right]\cdot \\
& \left[plim\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^{2}}{\partial\theta\partial\theta'}\log f\left(x_{i}\mid\theta\right)\right]^{-1}
\end{aligned}$$

$$\begin{aligned}
= & -\left[plim\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^{2}}{\partial\theta\partial\theta'}\log f\left(x_{i}\mid\theta\right)\right]^{-1}\cdot \\
& \left[plim\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f\left(x_{i}\mid\theta\right)\frac{\partial}{\partial\theta'}\log f\left(x_{i}\mid\theta\right)\right]\cdot \\
& \left[plim\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^{2}}{\partial\theta\partial\theta'}\log f\left(x_{i}\mid\theta\right)\right]^{-1}
\end{aligned}$$

$$
\begin{aligned}
&= \left[plim\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta\partial\theta'}\log f\left(x_i \mid \theta\right)\right]^{-1}\cdot \\
&\quad \left[plim\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta\partial\theta'}\log f\left(x_i \mid \theta\right)\right]\cdot \\
&\quad \left[plim\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta\partial\theta'}\log f\left(x_i \mid \theta\right)\right]^{-1} \\
&= \left[plim\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta\partial\theta'}\log f\left(x_i \mid \theta\right)\right]^{-1}.
\end{aligned}
$$

The result follows from a Central Limit Theorem. ■

3 MOM

3.1 Examples

1.

$$
\begin{aligned}
X_i &\sim iidBin\left(N,p\right), \quad i=1,2,..,n \\
&\Rightarrow EX_i = Np.
\end{aligned}
$$

Set

$$\overline{X} = Np$$

and solve for p to get

$$\widehat{p} = \frac{\overline{X}}{N}.$$

Note that the MLE is equal to the MOM estimator.

2.

$$X_i \sim iidU\left(-\theta,\theta\right), \quad i=1,2,..,n.$$

$$EX_i = 0$$

which does not depend on θ. So the first moment does not help.

$$VarX_i = \frac{\theta^2}{3}.$$

So set

$$s^2 = \frac{\theta^2}{3}$$

to get

$$\widehat{\theta} = \sqrt{3s^2}.$$

Note that the MLE is not equal to the MOM estimator.

3.

$$\begin{aligned} y_t &= X_t\beta + u_t, \quad t = 1, 2, .., T; \\ u_t &\sim iid\left(0, \sigma^2\right). \end{aligned}$$

$$EX'(y - X\beta) = 0$$

$$\Rightarrow \widehat{\beta} = (X'X)^{-1} X'y.$$

4.

$$\begin{aligned} y_t^* &= X_t\beta + u_t, \quad t = 1, 2, .., T; \\ u_t &\sim iidN\left(0, \sigma^2\right); \\ y_t &= 1\left(y_t^* > 0\right) \end{aligned}$$

with $\{y_t, X_t\}_{t=1}^T$ observed. For MLE,

$$\begin{aligned} \Pr\left[y_t = 1 \mid X_t\right] &= \Pr\left[y_t^* > 0 \mid X_t\right] \\ &= \Pr\left[X_t\beta + u_t > 0\right] \\ &= \Pr\left[u_t > -X_t\beta\right] \\ &= 1 - \Phi\left(\frac{-X_t\beta}{\sigma^2}\right) \\ &= \Phi\left(\frac{X_t\beta}{\sigma^2}\right) \end{aligned}$$

$$\Rightarrow \Pr\left[y_t = 0 \mid X_t\right] = 1 - \Phi\left(\frac{X_t\beta}{\sigma^2}\right).$$

$$\Rightarrow L(y \mid \beta, \sigma) = \prod_{t=1}^{T} \Phi\left(\frac{X_t\beta}{\sigma}\right)^{y_t} \left[1 - \Phi\left(\frac{X_t\beta}{\sigma}\right)\right]^{1-y_t}.$$

Note that we can identify only

$$\beta^* = \frac{\beta}{\sigma}.$$

So, without loss of generality, set $\sigma = 1$; the likelihood function becomes

$$\begin{aligned} L(y \mid \beta) &= \prod_{t=1}^{T} \Phi(X_t\beta)^{y_t} \left[1 - \Phi(X_t\beta)\right]^{1-y_t} \\ \log L(y \mid \beta) &= \sum_{t=1}^{T} y_t \log \Phi(X_t\beta) + (1 - y_t) \log \left[1 - \Phi(X_t\beta)\right]. \end{aligned}$$

The value of β that maximizes the log likelihood function is the MLE of β. In general, this will require numerical optimization.

However, consider the FOC of the log likelihood function:

$$\begin{aligned} \frac{\partial \log L(y \mid \beta)}{\partial \beta} &= \sum_{t=1}^{T} y_t \frac{\phi(X_t\beta)}{\Phi(X_t\beta)} X_t' - (1 - y_t) \frac{\phi(X_t\beta)}{1 - \Phi(X_t\beta)} X_t' \qquad (5) \\ &= \sum_{t=1}^{T} \frac{\phi(X_t\beta) X_t'}{\Phi(X_t\beta)\left[1 - \Phi(X_t\beta)\right]} \left[y_t \left(1 - \Phi(X_t\beta)\right) - (1 - y_t) \Phi(X_t\beta)\right] \\ &= \sum_{t=1}^{T} \frac{\phi(X_t\beta) X_t'}{\Phi(X_t\beta)\left[1 - \Phi(X_t\beta)\right]} \left[y_t - \Phi(X_t\beta)\right] = 0. \end{aligned}$$

Note that

$$\begin{aligned} E\left[y_t \mid X_t\right] &= 1 \Pr\left[y_t = 1 \mid X_t\right] + 0 \Pr\left[y_t = 0 \mid X_t\right] \\ &= \Pr\left[y_t = 1 \mid X_t\right] \\ &= \Phi(X_t\beta). \end{aligned}$$

Think of

$$\frac{\phi(X_t\beta) X_t}{\Phi(X_t\beta)\left[1 - \Phi(X_t\beta)\right]} = Z_t,$$

and think of equation (5) as

$$\sum_{t=1}^{T} Z_t' \left[y_t - E(y_t \mid X_t)\right] = 0. \qquad (6)$$

This is a moment condition which can be used to estimate β. Note that any moment (orthogonality) condition of the form in equation (6) would provide a consistent estimate of β; MLE just tells us what the best instruments Z_t would be. How do we know they are best?

3.2 Asymptotics of MOM Estimators

Assume the model is

$$\underset{n\times 1}{g}\left(X_t, \underset{m\times 1}{\theta}\right) = u_t \qquad (7)$$

where u_t is a set of errors uncorrelated with a set of instruments $\underset{n\times k}{Z_t}$. Give some examples.

Then define the MOM estimator of θ as

$$\underset{k\times nT}{Z'}\underset{Tn\times 1}{g\left(X,\widehat{\theta}\right)}=0.$$

Theorem 3 *plim* $\widehat{\theta}=\theta$.

Proof. Write a Taylor series approximation of equation (7):

$$\frac{1}{T}\sum_{t=1}^{T}g\left(X_t,\widehat{\theta}\right)=\frac{1}{T}\sum_{t=1}^{T}g\left(X_t,\theta\right)+\frac{1}{T}\sum_{t=1}^{T}g_\theta\left(X_t,\theta\right)\left(\widehat{\theta}-\theta\right)$$

and premultiply by Z_t to get

$$\frac{1}{T}\sum_{t=1}^{T}Z_t'g\left(X_t,\widehat{\theta}\right)=\frac{1}{T}\sum_{t=1}^{T}Z_t'g\left(X_t,\theta\right)+\frac{1}{T}\sum_{t=1}^{T}Z_t'g_\theta\left(X_t,\theta\right)\left(\widehat{\theta}-\theta\right).$$

The term on the left is identically zero given the definition of $\widehat{\theta}$. So we get

$$0=\frac{1}{T}\sum_{t=1}^{T}Z_t'g\left(X_t,\theta\right)+\frac{1}{T}\sum_{t=1}^{T}Z_t'g_\theta\left(X_t,\theta\right)\left(\widehat{\theta}-\theta\right).$$

Now consider *plim*s. Note that

$$EZ_t'g\left(X_t,\theta\right)=0$$

$$\Rightarrow plim\frac{1}{T}\sum_{t=1}^{T}Z_t'g\left(X_t,\theta\right)=0.$$

Thus, as long as

$$plim\frac{1}{T}\sum_{t=1}^{T}Z_t'g_\theta\left(X_t,\theta\right)$$

has full rank (identification condition),

$$\begin{aligned}0 &= plim\frac{1}{T}\sum_{t=1}^{T}Z_t'g_\theta\left(X_t,\theta\right)\left(\widehat{\theta}-\theta\right)\\ &= plim\frac{1}{T}\sum_{t=1}^{T}Z_t'g_\theta\left(X_t,\theta\right)plim\left(\widehat{\theta}-\theta\right)\\ &\Rightarrow plim\left(\widehat{\theta}-\theta\right)=0.\end{aligned}$$

■

Theorem 4 *$\sqrt{T}\left(\widehat{\theta}-\theta\right) \sim N(0,V)$ with*

$$
\begin{aligned}
V &= \left[plim\frac{1}{T}\sum_{t=1}^{T} Z_t' g_\theta\left(X_t,\theta\right)\right]^{-1}\left[plim\frac{1}{T}\sum_{t=1}^{T} Z_t'\Omega Z_s\right]\cdot \\
&\quad \left[plim\frac{1}{T}\sum_{t=1}^{T} g_\theta'\left(X_t,\theta\right)Z_t\right]^{-1}
\end{aligned}
$$

where

$$\Omega = Eu_t u_t'$$

Proof.

$$\widehat{\theta}-\theta = \left[\frac{1}{T}\sum_{t=1}^{T} Z_t' g_\theta\left(X_t,\theta\right)\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T} Z_t' g\left(X_t,\theta\right)\right]$$

$$\Rightarrow \sqrt{T}\left(\widehat{\theta}-\theta\right) = \left[\frac{1}{T}\sum_{t=1}^{T} Z_t' g_\theta\left(X_t,\theta\right)\right]^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t' g\left(X_t,\theta\right)\right].$$

Note that

$$
\begin{aligned}
& plimT\left(\widehat{\theta}-\theta\right)\left(\widehat{\theta}-\theta\right)' \\
=\ & plim\left[\frac{1}{T}\sum_{t=1}^{T} Z_t' g_\theta\left(X_t,\theta\right)\right]^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t' g\left(X_t,\theta\right)\right]\cdot \\
& \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g'\left(X_t,\theta\right)Z_t\right]\left[\frac{1}{T}\sum_{t=1}^{T} g_\theta'\left(X_t,\theta\right)Z_t\right]^{-1} \\
=\ & plim\left[\frac{1}{T}\sum_{t=1}^{T} Z_t' g_\theta\left(X_t,\theta\right)\right]^{-1}\cdot \\
& \left[\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} Z_t' g\left(X_t,\theta\right)g'\left(X_s,\theta\right)Z_s\right]\left[\frac{1}{T}\sum_{t=1}^{T} g_\theta'\left(X_t,\theta\right)Z_t\right]^{-1}
\end{aligned}
$$

$$
\begin{aligned}
= \quad & \left[plim\frac{1}{T}\sum_{t=1}^{T} Z_t' g_\theta\left(X_t,\theta\right)\right]^{-1} \cdot \\
& \left[plim\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} Z_t' g\left(X_t,\theta\right) g'\left(X_s,\theta\right) Z_s\right] \cdot \\
& \left[plim\frac{1}{T}\sum_{t=1}^{T} g_\theta'\left(X_t,\theta\right) Z_t\right]^{-1} \\
& \left[plim\frac{1}{T}\sum_{t=1}^{T} Z_t' g_\theta\left(X_t,\theta\right)\right]^{-1} \left[plim\frac{1}{T}\sum_{t=1}^{T} Z_t' \Omega Z_s\right] \cdot \\
& \left[plim\frac{1}{T}\sum_{t=1}^{T} g_\theta'\left(X_t,\theta\right) Z_t\right]^{-1} .
\end{aligned}
$$

A Central Limit Theorem implies the result. ■

4 Testing

4.1 Wald Tests

Assume

$$\sqrt{T}\left(\widehat{\theta}-\theta\right) \sim N\left(0,V\right), \tag{8}$$

and consider the test

$$H_0 : R\theta = c \ \text{ vs } H_A : R\theta \neq c.$$

Equation (8) implies that

$$\sqrt{T}\left(R\widehat{\theta}-c\right) \sim N\left(0,RVR'\right)$$

$$\Rightarrow W = \sqrt{T}\left(R\widehat{\theta}-c\right)'\left[RVR'\right]^{-1}\left(R\widehat{\theta}-c\right) \sim \chi_k^2$$

where $k = Rank\left(R\right)$.

4.2 Likelihood Ratio Tests

Assume $\widehat{\theta}$ is the MLE of θ, and consider the test

$$H_0 : R\theta = c \ \text{ vs } H_A : R\theta \neq c.$$

Let $\widehat{\theta}_R$ be the restricted MLE and $\widehat{\theta}_U$ be the unrestricted MLE. The likelihood ratio statistic is

$$LR = \frac{L\left(X \mid \widehat{\theta}_U\right)}{L\left(X \mid \widehat{\theta}_r\right)}$$

with

$$2 \log LR \sim \chi^2_k.$$

Example 5

$$\begin{aligned} X_i &\sim iidBin\left(N, p\right), \quad i = 1, 2, .., n; \\ H_0 &: p = .5 \text{ } vs \text{ } H_0 : p \neq .5. \end{aligned}$$

$$\widehat{p}_U = \frac{\overline{X}}{N}; \quad \widehat{p}_R = .5$$

$$\begin{aligned} \Rightarrow \quad \log LR &= \sum_{i=1}^{n} x_i \log \widehat{p}_U + (N - x_i) \log (1 - \widehat{p}_U) \\ &- \sum_{i=1}^{n} x_i \log \widehat{p}_R + (N - x_i) \log (1 - \widehat{p}_R) \\ &= \sum_{i=1}^{n} x_i \log \frac{\widehat{p}_U}{\widehat{p}_R} + (N - x_i) \log \frac{1 - \widehat{p}_U}{1 - \widehat{p}_R} \\ &= n\overline{X} \log \frac{\widehat{p}_U}{\widehat{p}_R} + n\left(N - \overline{X}\right) \log \frac{1 - \widehat{p}_U}{1 - \widehat{p}_R} \end{aligned}$$

$\Rightarrow$

$$\frac{\partial \log LR}{\partial \widehat{p}_R} = -\frac{n\overline{X}}{\widehat{p}_R} + \frac{n\left(N - \overline{X}\right)}{1 - \widehat{p}_R}$$

(which equals 0 *if* $\widehat{p}_R = \widehat{p}_U$*) and*

$$\frac{\partial^2 \log LR}{\partial \widehat{p}_R^2} = \frac{n\overline{X}}{\widehat{p}_R^2} + \frac{n\left(N - \overline{X}\right)}{\left(1 - \widehat{p}_R\right)^2} > 0.$$

We can pick a critical value c *such that we reject* H_0 *iff* $\log LR > c$ *From the picture, it is clear that this is equivalent to picking critical values* (a, b) *such that we do not reject* H_0 *iff* $a < \widehat{p}_R < b$*. The goal is to pick critical values that imply the right size and power. We know that*

$$nN\widehat{p}_U = \sum_{i=1}^{n} x_i \sim Bin\left(nN, p\right).$$

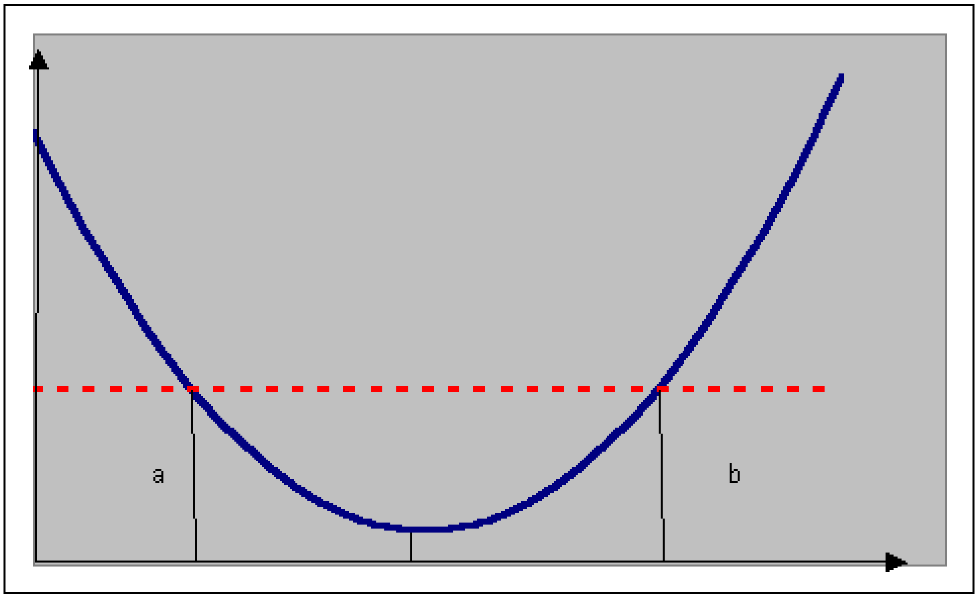

Figure 2:

We can choose a size α*,*

$$\Pr\left[\textit{Reject } H_0 \mid H_0\right] = \alpha.$$

Then we can approximate

$$\sqrt{n}\left(\widehat{p}_U - p\right) \sim N\left(0, \frac{p\left(1-p\right)}{N}\right).$$

So we can set

$$\begin{aligned}
\delta_1 &= \widehat{p}_R - z_\alpha \sqrt{\frac{p_R\left(1-p_R\right)}{N}}; \\
\delta_2 &= \widehat{p}_R + z_\alpha \sqrt{\frac{p_R\left(1-p_R\right)}{N}}
\end{aligned}$$

and not reject iff $\delta_1 < \widehat{p}_U < \delta_2$*.*

Example 6

$$\begin{aligned}
y &= X\beta + Z\gamma + u \\
H_0 &: \gamma = 0 \;\textit{vs}\; H_A : \gamma \neq 0.
\end{aligned}$$

$$\log LR = \frac{1}{2\sigma^2}\widehat{u}_R'\widehat{u}_R - \frac{1}{2\sigma^2}\widehat{u}_U'\widehat{u}_U$$

which is monotone with

$$\frac{\left(R_U^2 - R_R^2\right)/k}{\left(1 - R_U^2\right)/\left(T - n - k\right)} \sim F_{k,T-n-k}.$$

4.3 Lagrange Multiplier Tests

Let $\widehat{\theta}_R$ be the restricted MLE of θ. If H_0 is true, then

$$E\frac{\partial}{\partial\theta}\log L\left(X \mid \widehat{\theta}_R\right) = 0.$$

The Lagrange Multiplier (LM) test is

$$\left[\frac{\partial}{\partial\theta}\log L\left(X \mid \widehat{\theta}_R\right)\right]' D^{-1}\left[\frac{\partial}{\partial\theta}\log L\left(X \mid \widehat{\theta}_R\right)\right]\left[\frac{\partial}{\partial\theta}\log L\left(X \mid \widehat{\theta}_R\right)\right] \sim \chi_k^2.$$

Example 7

$$\begin{aligned} f\left(x_i\right) &= \lambda e^{-\lambda x_i}, \quad i = 1, 2, .., n \\ H_0 &: \lambda = 2 \ \textit{vs.}\ H_A : \lambda \neq 2. \end{aligned}$$

$$\Rightarrow \log L = n\log\lambda - \lambda\sum_{i=1}^{n} x_i$$

$$\begin{aligned} \Rightarrow \quad \frac{\partial\log L}{\partial\lambda} &= \frac{n}{\lambda} - \sum_{i=1}^{n} x_i; \\ \Rightarrow \quad \frac{\partial\log L\left(2\right)}{\partial\lambda} &= \frac{n}{2} - n\overline{x}. \end{aligned}$$

Note that

$$\widehat{\lambda} = \overline{x}^{-1}.$$

We could derive the exact small sample test. Instead, note that

$$\sqrt{n}\left(\overline{x} - \frac{1}{2}\right) \sim N\left(0, \frac{1}{4}\right)$$

$$\Rightarrow 4n\left(\overline{x} - \frac{1}{2}\right)^2 \sim \chi_1^2.$$

Example 8

$$\begin{aligned} y &= X\beta + Z\gamma + u; \\ u &\sim N\left(0, \sigma^2 I\right); \\ H_0 &: \gamma = 0 \ vs. \ H_A : \gamma \neq 0. \end{aligned}$$

$$\log L = -\frac{1}{2\sigma^2}\left(y - X\beta - Z\gamma\right)'\left(y - X\beta - Z\gamma\right)$$

$$\begin{aligned} &\Rightarrow \quad \frac{\partial \log L}{\partial \gamma} \propto Z'\left(y - X\beta - Z\gamma\right) \\ &\Rightarrow \quad \frac{\partial \log L\left(0\right)}{\partial \gamma} \propto Z'\left(y - X\beta\right) = Z'\widehat{u}_R. \end{aligned}$$

$$Z'\left(y - X\widehat{\beta}\right) = Z'\left(I - P_X\right)u$$

$$\begin{aligned} & EZ'\left(I - P_X\right)uu'\left(I - P_X\right)Z \\ = \; & Z'\left(I - P_X\right)Euu'\left(I - P_X\right)Z \\ = \; & \sigma^2 Z'\left(I - P_X\right)\left(I - P_X\right)Z \\ = \; & \sigma^2 Z'\left(I - P_X\right)Z. \end{aligned}$$

$$\Rightarrow LM = \left(y - X\widehat{\beta}\right)' Z\left[\sigma^2 Z'\left(I - P_X\right)Z\right]^{-1} Z'\left(y - X\widehat{\beta}\right) \sim \chi^2_k.$$

4.4 Advantages and Disadvantages of Each Test Statistic

1. LM requires only estimation of $\widehat{\theta}_R$.
2. LM is the only test statistic with nice properties when H_0 is on the boundary of the parameter space.
3. W requires only estimation of $\widehat{\theta}_U$.
4. LR is always most powerful but requires estimation of both $\widehat{\theta}_R$ and $\widehat{\theta}_U$.
5. Frequently all three test statistics are the same.

5 Computation

In general, we have a need to optimize in estimation.

5.1 Newton routines for MLE

$$\widehat{\theta} - \theta = -\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta\partial\theta'}\log f\left(x_i \mid \theta\right)\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f\left(x_i \mid \theta\right)\right].$$

BHHH:

$$\widehat{\theta} - \theta = \left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f\left(x_i \mid \theta\right)\frac{\partial}{\partial\theta'}\log f\left(x_i \mid \theta\right)\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f\left(x_i \mid \theta\right)\right].$$

Other bells and whistles:

1. Line search
2. Dampening

5.2 Newton routines for MOM

$$\widehat{\theta} - \theta = \left[\frac{1}{T}Z'g_\theta\right]^{-1}Z'g.$$

5.3 Simplex Routines

Discuss.