

Econ 583 Homework 7

Suggested Solutions: Wald, LM and LR based on GMM and MLE

Eric Zivot

Winter 2013

1 Wald, LR and LM statistics based on generalized method of moments estimation

Let X_1, \dots, X_N be an iid sample of Bernoulli random variables; that is, each X_i has density $f(x; \theta_0) = \theta_0^x(1 - \theta_0)^{1-x}$. Instead of using ML estimation, consider instead estimation of θ_0 using the generalized method of moments (GMM).

1. Consider GMM estimation of θ using the single moment condition

$$g(x_t, \theta) = x_t - \theta$$

Show that $E[g(x_t, \theta)] = 0$ for $\theta = \theta_0$, and $E[g(x_t, \theta)] \neq 0$ for $\theta \neq \theta_0$. Additionally, show that $S = E[g(x_t, \theta_0)^2] = \theta_0(1 - \theta_0)$.

Here

$$E[g(x_t, \theta)] = E[x_t] - \theta = \theta_0 - \theta$$

Clearly, $E[g(x_t, \theta)] = 0$ for $\theta = \theta_0$, and $E[g(x_t, \theta)] \neq 0$ for $\theta \neq \theta_0$. Next,

$$S = E[g(x_t, \theta_0)^2] = E[(x_t - \theta_0)^2] = \text{var}(x_t) = \theta_0(1 - \theta_0)$$

since x_t has a Bernoulli distribution with parameter θ_0 . Note that $S = I(\theta_0|x_i)^{-1}$.

2. Using the sample moment $g_n(\theta) = n^{-1} \sum_{t=1}^n g(x_t, \theta)$, show that the GMM estimator of θ is $\hat{\theta}_{gmm} = n^{-1} \sum_{t=1}^n x_t$. What is the value of the GMM objective function $J(\hat{S}^{-1}, \hat{\theta}_{gmm}) = n g_n(\hat{\theta}_{gmm})' \hat{S}^{-1} g_n(\hat{\theta}_{gmm}) = n g_n(\hat{\theta}_{gmm})^2 \hat{S}^{-1}$? How would you consistently estimate S ?

Equating the sample moment with the population moment gives the estimating equation

$$\begin{aligned} g_n(\hat{\theta}_{gmm}) &= n^{-1} \sum_{t=1}^n g(x_t, \hat{\theta}_{gmm}) = n^{-1} \sum_{t=1}^n x_t - \hat{\theta}_{gmm} = 0 \\ \Rightarrow \hat{\theta}_{gmm} &= n^{-1} \sum_{t=1}^n x_t = \bar{x} \end{aligned}$$

Hence, the GMM estimator is the same as the ML estimator. The GMM objective function evaluated at $\hat{\theta}_{gmm}$ is

$$J(\hat{S}^{-1}, \hat{\theta}_{gmm}) = n g_n(\hat{\theta}_{gmm})^2 \hat{S}^{-1} = n \left(n^{-1} \sum_{t=1}^n x_t - \hat{\theta}_{gmm} \right) \hat{S}^{-1} = 0$$

regardless of the value of \hat{S} since $\hat{\theta}_{gmm} = n^{-1} \sum_{t=1}^n x_t$. This is to be expected since with one moment condition and one parameter to estimate the model is exactly identified.

Since $S = \theta_0(1 - \theta_0)$ a nature estimate is $\hat{S} = \hat{\theta}_{gmm}(1 - \hat{\theta}_{gmm})$. Alternatively, one could use the the usual non-parametric estimator of S :

$$\hat{S} = n^{-1} \sum_{t=1}^n g(x_t, \hat{\theta}_{gmm})^2 = n^{-1} \sum_{t=1}^n (x_t - \hat{\theta}_{gmm})^2$$

3. Determine the asymptotic normal distribution for $\hat{\theta}_{gmm}$. Is this distribution the same as the asymptotic normal distribution for $\hat{\theta}_{mle}$?

Using the asymptotic normality result for GMM estimators we have

$$\begin{aligned} \hat{\theta}_{gmm} &\overset{A}{\approx} N(\theta_0, n^{-1}(G' S^{-1} G)^{-1}) \\ G &= E \left[\frac{dg(x_t, \theta_0)}{d\theta_0} \right], \quad S = E[g(x_t, \theta_0)^2] \end{aligned}$$

Now,

$$\begin{aligned} \frac{dg(x_t, \theta_0)}{d\theta_0} &= \frac{d}{d\theta_0}(x_t - \theta_0) = -1 \\ S &= \theta_0(1 - \theta_0) \end{aligned}$$

so that

$$(G' S^{-1} G)^{-1} = S = \theta_0(1 - \theta_0)$$

and

$$\hat{\theta}_{gmm} \overset{A}{\approx} N(\theta_0, n^{-1}\theta_0(1 - \theta_0))$$

which is equivalent to the asymptotic distribution for the MLE.

Next, consider testing the hypotheses $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$ using a 5 significance level.

4. and 5. Derive the Wald, LR and LM statistics for testing the above hypothesis. Because θ is a scalar, these statistics have the form

$$\begin{aligned} Wald &= n(\hat{\theta}_{gmm} - \theta_0)^2 (G_n(\hat{\theta}_{gmm})^2 \hat{S}^{-1}) \\ LM &= n(G_n(\theta_0) \hat{S}^{-1} g_n(\theta_0))^2 (G_n(\theta_0)^2 \hat{S}^{-1})^{-1} \\ LR &= J(\hat{S}^{-1}, \theta_0) - J(\hat{S}^{-1}, \hat{\theta}_{gmm}) \end{aligned}$$

where $G_n(\theta) = \frac{d}{d\theta} g_n(\theta)$. For the Wald and LR statistic it is useful to estimate S using $\hat{\theta}_{gmm}$. For the LM statistic, it is useful to an estimate of S imposing $\theta = \theta_0$. Are the GMM Wald, LM and LR statistics numerically equivalent? Compare the GMM Wald, LR and LM statistics with the ML equivalents. Do you see any similarities or differences?

The Wald statistic has the form

$$\begin{aligned} Wald &= (\hat{\theta}_{gmm} - \theta_0)^2 (G_n(\hat{\theta}_{gmm})^2 \hat{S}^{-1}) \\ &= \frac{n(\hat{\theta}_{gmm} - \theta_0)^2}{\hat{\theta}_{gmm}(1 - \hat{\theta}_{gmm})} \end{aligned}$$

which is equivalent to the ML Wald statistic.

For the LM statistic, first note that

$$\begin{aligned} G_n(\theta_0) &= 1, \\ g_n(\theta_0) &= n^{-1} \sum_{i=1}^n x_i - \theta_0 = \hat{\theta}_{gmm} - \theta_0 \\ \hat{S} &= \theta_0(1 - \theta_0) \end{aligned}$$

Then, the LM statistic has the form

$$\begin{aligned} LM &= n(G_n(\theta_0) \hat{S}^{-1} g_n(\theta_0))^2 (G_n(\theta_0)^2 \hat{S}^{-1})^{-1} \\ &= \frac{n(\hat{\theta}_{gmm} - \theta_0)^2}{\theta_0(1 - \hat{\theta}_0)} \end{aligned}$$

which is identical to the ML LM statistic.

For the LR statistic, recall that $J(\hat{\theta}_{gmm}, \hat{S}^{-1}) = 0$ since the model is just identified. Therefore, the LR statistic has the form

$$\begin{aligned} LR &= J(\hat{S}^{-1}, \theta_0) - J(\hat{S}^{-1}, \hat{\theta}_{gmm}) \\ &= J(\hat{S}^{-1}, \theta_0) = n g_n(\theta_0)' \hat{S}^{-1} g_n(\theta_0) = n g_n(\theta_0)^2 \hat{S}^{-1} \\ &= \frac{n(\hat{\theta}_{gmm} - \theta_0)^2}{\hat{\theta}_{gmm}(1 - \hat{\theta}_{gmm})} \end{aligned}$$

which is not equivalent to the ML LR statistic. However, the GMM LR statistic is equivalent to the GMM Wald statistic. If you use $\hat{S} = \theta_0(1 - \theta_0)$ then the GMM LR statistics is equivalent to the the GMM LM statistic.

Note: for GMM, $Wald = LM = LR$ if the same \hat{S} is used for all statistics. If $\hat{S} = n^{-1} \sum_{t=1}^n g(x_t, \hat{\theta}_{gmm})^2 = n^{-1} \sum_{t=1}^n (x_t - \hat{\theta}_{gmm})^2$ is used then the GMM statistics will not be equal to the ML statistics in finite samples, but will be as the sample size gets large.

2 Wald, LR and LM statistics based on maximum likelihood estimation

Let X_1, \dots, X_n be an iid sample of Bernoulli random variables; that is, each X_i has density $f(x; \theta_0) = \theta_0^x(1 - \theta_0)^{1-x}$. First, consider estimation of θ_0 using maximum likelihood (ML). Let $x = (x_1, \dots, x_n)$.

1. Derive the log-likelihood function, $\ln L(\theta|\mathbf{x})$, score function, $S(\theta|\mathbf{x}) = \frac{d}{d\theta} \ln L(\theta|\mathbf{x})$, Hessian function, $H(\theta|\mathbf{x}) = \frac{d^2}{d\theta^2} \ln L(\theta|\mathbf{x})$, and information matrix, $I(\theta_0|\mathbf{x}) = -E[H(\theta_0|\mathbf{x})] = var(S(\theta_0|\mathbf{x}))$. Verify that $E[S(\theta_0|\mathbf{x})] = 0$, and show that the ML estimate for θ is $\hat{\theta}_{mle} = n^{-1} \sum_{i=1}^n x_i$.

The joint density/likelihood function for the sample is given by

$$f(\mathbf{x}; \theta) = L(\theta|\mathbf{x}) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i}$$

The log-likelihood function is

$$\begin{aligned} \ln L(\theta|\mathbf{x}) &= \ln \left(\theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i} \right) \\ &= \sum_{i=1}^n x_i \ln(\theta) + \left(n - \sum_{i=1}^n x_i \right) \ln(1 - \theta) \end{aligned}$$

The score function for the Bernoulli log-likelihood is

$$\begin{aligned} S(\theta|\mathbf{x}) &= \frac{\partial \ln L(\theta|\mathbf{x})}{\partial \theta} = \frac{1}{\theta} \sum_{i=1}^n x_i - \frac{1}{1 - \theta} \left(n - \sum_{i=1}^n x_i \right) \\ &= \frac{\sum_{i=1}^n (x_i - \theta)}{\theta(1 - \theta)} \end{aligned}$$

Using the chain rule it can be shown that

$$\begin{aligned} H(\theta|\mathbf{x}) &= \frac{d}{d\theta} S(\theta|\mathbf{x}) = \frac{-n\bar{x}}{\theta} - \frac{n(1 - \bar{x})}{(1 - \theta)^2} \\ &= - \left(\frac{n + S(\theta|\mathbf{x}) - 2\theta S(\theta|\mathbf{x})}{\theta(1 - \theta)} \right) \end{aligned}$$

where $\bar{x} = n^{-1} \sum_{i=1}^n x_i$. Therefore, the information matrix is

$$\begin{aligned} I(\theta_0|\mathbf{x}) &= -E[H(\theta_0|\mathbf{x})] = \frac{n + E[S(\theta_0|\mathbf{x})] - 2\theta_0 E[S(\theta_0|\mathbf{x})]}{\theta_0(1 - \theta_0)} \\ &= \frac{n}{\theta_0(1 - \theta_0)} \end{aligned}$$

since

$$E[S(\theta_0|\mathbf{x})] = \frac{\sum_{i=1}^n (E[x_i] - \theta_0)}{\theta_0(1 - \theta_0)} = 0$$

since $E[x_i] = \theta_0$.

The MLE solves

$$\begin{aligned} S(\hat{\theta}_{mle}|\mathbf{x}) &= \frac{\sum_{i=1}^n (x_i - \hat{\theta}_{mle})}{\hat{\theta}_{mle}(1 - \hat{\theta}_{mle})} = 0 \\ \Rightarrow \hat{\theta}_{mle} &= \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} \end{aligned}$$

2. Determine the asymptotic normal distribution for $\hat{\theta}_{mle}$.

Using the asymptotic properties of the MLE for a regular model,

$$\begin{aligned} \hat{\theta}_{mle} &\overset{A}{\sim} N(\theta_0, I(\theta_0|\mathbf{x})^{-1}) \\ I(\theta_0|\mathbf{x})^{-1} &= \frac{\theta_0(1 - \theta_0)}{n} \end{aligned}$$

Since θ_0 is unknown, a practically useful result replaces $I(\theta_0|\mathbf{x})^{-1}$ with $I(\hat{\theta}_{mle}|\mathbf{x})^{-1}$. Next, consider testing the hypotheses $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$ using a 5% significance level.

3. Derive the Wald, LR and LM statistics for testing the above hypothesis. These statistics have the form

$$\begin{aligned} Wald &= (\hat{\theta}_{mle} - \theta_0)^2 I(\hat{\theta}_{mle}|\mathbf{x}) \\ LM &= S(\theta_0|\mathbf{x})^2 I(\theta_0|\mathbf{x})^{-1} \\ LR &= -2[\ln L(\theta_0|\mathbf{x}) - \ln L(\hat{\theta}_{mle}|\mathbf{x})] \end{aligned}$$

Are these statistics numerically equivalent? For a 5% test, what is the decision rule to reject H_0 for each statistic?

The Wald statistic has the form

$$\begin{aligned}
Wald &= (\hat{\theta}_{mle} - \theta_0)^2 I(\hat{\theta}_{mle} | \mathbf{x}) \\
&= (\hat{\theta}_{mle} - \theta_0)^2 \frac{n}{\hat{\theta}_{mle} (1 - \hat{\theta}_{mle})} \\
&= \frac{n(\hat{\theta}_{mle} - \theta_0)^2}{\hat{\theta}_{mle} (1 - \hat{\theta}_{mle})}
\end{aligned}$$

For the LM statistic, note that

$$\begin{aligned}
S(\theta_0 | \mathbf{x}) &= \frac{\sum_{i=1}^n (x_i - \theta)}{\theta(1 - \theta)} = \frac{n(\hat{\theta}_{mle} - \theta_0)}{\theta_0(1 - \theta_0)} \\
I(\theta_0 | \mathbf{x})^{-1} &= \frac{\theta_0(1 - \theta_0)}{n}
\end{aligned}$$

so that

$$\begin{aligned}
LM &= S(\theta_0 | \mathbf{x})^2 I(\theta_0 | \mathbf{x})^{-1} \\
&= \left(\frac{n(\hat{\theta}_{mle} - \theta_0)}{\theta_0(1 - \theta_0)} \right)^2 \frac{\theta_0(1 - \theta_0)}{n} \\
&= \frac{n(\hat{\theta}_{mle} - \theta_0)^2}{\theta_0(1 - \theta_0)}
\end{aligned}$$

Notice that the LM statistic has the same form as the Wald statistic. The difference between the two statistics is based on how $I(\theta | \mathbf{x})$ is estimated. With the Wald statistic, $I(\theta | \mathbf{x})$ is estimated using $I(\hat{\theta}_{mle} | \mathbf{x})$ and with the LM statistic, $I(\theta | \mathbf{x})$ is estimated using $I(\theta_0 | \mathbf{x})$

For the statistic, note that

$$\begin{aligned}
\ln L(\hat{\theta}_{mle} | \mathbf{x}) &= \sum_{i=1}^n x_i \ln(\hat{\theta}_{mle}) + \left(n - \sum_{i=1}^n x_i \right) \ln(1 - \hat{\theta}_{mle}) \\
\ln L(\theta_0 | \mathbf{x}) &= \sum_{i=1}^n x_i \ln(\theta_0) + \left(n - \sum_{i=1}^n x_i \right) \ln(1 - \theta_0)
\end{aligned}$$

Therefore

$$\begin{aligned}
LR &= -2[\ln L(\theta_0 | \mathbf{x}) - \ln L(\hat{\theta}_{mle} | \mathbf{x})] \\
&= -2 \left[\sum_{i=1}^n (\ln \theta_0 - \ln \hat{\theta}_{mle}) + \left(n - \sum_{i=1}^n x_i \right) (\ln(1 - \theta_0) - \ln(1 - \hat{\theta}_{mle})) \right] \\
&= -2n \left[\hat{\theta}_{mle} \ln \left(\frac{\theta_0}{\hat{\theta}_{mle}} \right) + (1 - \hat{\theta}_{mle}) \ln \left(\frac{1 - \theta_0}{1 - \hat{\theta}_{mle}} \right) \right]
\end{aligned}$$

4. Prove that the Wald and LM statistics have asymptotic $\chi^2(1)$ distributions.

First, consider the Wald statistic:

$$\begin{aligned} Wald &= \frac{n(\hat{\theta}_{mle} - \theta_0)^2}{\hat{\theta}_{mle}(1 - \hat{\theta}_{mle})} \\ &= \left(\frac{\sqrt{n}(\hat{\theta}_{mle} - \theta_0)}{\sqrt{\hat{\theta}_{mle}(1 - \hat{\theta}_{mle})}} \right)^2 \end{aligned}$$

From the asymptotic normality of the MLE, we have that

$$\sqrt{n}(\hat{\theta}_{mle} - \theta_0) \xrightarrow{d} N(0, \theta_0(1 - \theta_0))$$

and from the consistency of the MLE we have that

$$\hat{\theta}_{mle} \xrightarrow{p} \theta_0$$

Slutsky's theorem then gives

$$\sqrt{\hat{\theta}_{mle}(1 - \hat{\theta}_{mle})} \xrightarrow{p} \sqrt{\theta_0(1 - \theta_0)}$$

Finally, by a combination of Slutsky's theorem and the continuous mapping theorem (CMT) we have

$$\begin{aligned} &\left(\frac{\sqrt{n}(\hat{\theta}_{mle} - \theta_0)}{\sqrt{\hat{\theta}_{mle}(1 - \hat{\theta}_{mle})}} \right)^2 \xrightarrow{d} \left(\frac{N(0, \theta_0(1 - \theta_0))}{\sqrt{\theta_0(1 - \theta_0)}} \right)^2 \\ &= N(0, 1)^2 = \chi^2(1) \end{aligned}$$

where $\chi^2(1)$ denotes a chi-square random variable with 1 degree of freedom.

A similar proof works for the LM statistic:

$$\begin{aligned} LM &= \frac{n(\hat{\theta}_{mle} - \theta_0)^2}{\theta_0(1 - \theta_0)} \\ &= \left(\frac{\sqrt{n}(\hat{\theta}_{mle} - \theta_0)}{\sqrt{\theta_0(1 - \theta_0)}} \right)^2 \\ &\xrightarrow{d} \left(\frac{N(0, \theta_0(1 - \theta_0))}{\sqrt{\theta_0(1 - \theta_0)}} \right)^2 = \chi^2(1) \end{aligned}$$