FINC 9311-21 Financial Econometrics Handout

Jialin Yu

1 MLE (continued)

1.1 Cramer-Rao Inequality

Let $\widehat{\theta} = g(X)$ denote an unbiased estimator of θ.

$$\theta = \int g(x) f(x, \theta) \, dx$$

differentiating both sides with respect to θ

$$\begin{aligned} 1 &= \int g(x) \frac{\partial \log f(x, \theta)}{\partial \theta} f(x, \theta) \, dx \\ &= \int g(x) S(\theta) f(x, \theta) \, dx \end{aligned}$$

This, with $\widehat{\theta} = g(X)$, implies

$$E\left[\widehat{\theta}, S(\theta)\right] = Cov\left[\widehat{\theta}, S(\theta)\right] = 1$$

Because

$$Cov\left[\widehat{\theta}, S(\theta)\right] \leq Var\left(\widehat{\theta}\right) Var(S(\theta))$$

therefore

$$Var\left(\widehat{\theta}\right) \geq \frac{1}{Var(S(\theta))} = I(\theta)^{-1}$$

which is the Cramer-Rao inequality.

1.2 Issues with estimators of drawdowns

Suppose you want to estimate the drawdown per day (i.e., the maximum loss of your portfolio per day). Assume you observe your return $r_i \in [\theta, 0]$ and estimate drawdown θ using MLE. The probability density function of

r_i is

$$f(r_i) = \begin{cases} -\frac{1}{\theta} & \text{if } r_i \in [\theta, 0] \\ 0 & \text{otherwise} \end{cases}$$

Maximize

$$\max \sum_i \log f(r_i)$$

implies what ML estimate of θ? Why is this not a good estimator?

The problem is because the possible region of θ (i.e., Θ) depends observations of r_i.

2 GMM Estimator

Suppose θ_0 satisfies $E\left[g\left(x,\theta_0\right)\right]=0$, where g is a "moment function" vector. A GMM estimator is obtained by

$$\begin{aligned}\widehat{\theta} &= \arg\min_{\theta\in\Theta} Q_T\left(\theta\right) \\ &= \arg\min_{\theta\in\Theta}\left[T^{-1}\sum_{i=1}^{T} g\left(x_i,\theta\right)\right]^T \widehat{W}\left[T^{-1}\sum_{i=1}^{T} g\left(x_i,\theta\right)\right]\end{aligned}$$

for some weighting function $\widehat{W}$.

Examples: if you observe stock return r_i where $i=1,2,...,n$ and think expected stock return $E\left(r_i\right)=\mu$ and $E\left(r_i^2\right)=\sigma^2$. What would be your moment condition and your objective function to minimize? Would $\widehat{W}$ matter when there is only one moment condition? How about when there are two moment conditions?

What is the right moment condition that delivers OLS estimates?

What is the moment condition of MLE estimator?

2.1 Consistency of GMM Estimator

Theorem 1 *Suppose the observations are i.i.d.,* $\widehat{W}\xrightarrow{p}W$. *(i)* W *is positive semi-definite and* $WE\left[g\left(x,\theta\right)\right]=0$ *only if* $\theta=\theta_0$ *(ii)* $\theta_0\in\Theta$ *compact (iii)* $g\left(x,\theta\right)$ *is continuous at each* θ *with probability one (iv)* $E\left[\sup_{\theta\in\Theta}\left|g\left(x,\theta\right)\right|\right]<\infty$. *Then* $\widehat{\theta}\xrightarrow{p}\theta$.

Proof: By verifying the conditions of theorem on extremum estimators.

2.2 Asymptotic Normality of GMM Estimator

Theorem 2 *Assume the conditions in theorem 1 are satisfied and (i)* $\theta_0\in Interior\left(\Theta\right)$ *(ii)* $g\left(x,\theta\right)$ *is continuously differentiable in a neighborhood* N *of* θ_0, *with probability approaching one (iii)* $E\left[g\left(x,\theta_0\right)\right]=0$ *and* $E\left[\left\|g\left(x,\theta_0\right)\right\|^2\right]<\infty$ *(iv)* $E\left[\sup_{\theta\in N}\left\|\frac{\partial}{\partial\theta}g\left(x,\theta\right)\right\|\right]<\infty$ *(v)* G^TWG *is nonsingular for* $G=$

$E\left[\frac{\partial}{\partial\theta}g\left(x,\theta_0\right)\right]$. *Let* $\Omega = E\left[g\left(x,\theta_0\right)g\left(x,\theta_0\right)^T\right]$, *then*

$$\sqrt{n}\left(\widehat{\theta}-\theta\right)\xrightarrow{d}N\left(0,\left(G^TWG\right)^{-1}G^TW\Omega WG\left(G^TWG\right)^{-1}\right)$$

See [1] for additional details on consistency/asymptotic normality of GMM/MLE estimators.

When $W = \Omega^{-1}$, an optimal estimator results in the sense that it minimizes the asymptotic variance of the GMM estimator and

$$\sqrt{n}\left(\widehat{\theta}-\theta\right)\xrightarrow{d}N\left(0,\left(G^T\Omega^{-1}G\right)^{-1}\right)$$

To see the intuition of the optimal weighting matrix, consider the previous example of stock return with two moment conditions.

3 GMM Estimation

Moment condition

$$Eh\left(z_t,\theta_0\right)=0$$

where the dimension of h is $M\times 1$. θ_0 is k dimensional.

Conditional GMM

$$E\left[h\left(z_t,\theta_0\right)|I_{t-1}\right]=0$$

But we only learned moment conditions expressed as unconditional expectations... However, conditional GMM can be transformed into unconditional GMM by

$$E\left[A_{t-1}h\left(z_t,\theta_0\right)\right]=0$$

where A_{t-1} is known at time $t-1$.

Therefore, we focus on the unconditional GMM estimation with moment condition

$$Eh\left(z_t,\theta_0\right)=0$$

The parameter estimate θ_T is obtained from

$$\min H_T(\theta_T)' W H_T(\theta_T)$$

where

$$H_T(\theta_T) = \frac{1}{T}\Sigma h(z_t, \theta_T)$$

The weighting matrix W is assumed to be the optimal weighting matrix that we discussed earlier this semester. θ_T is characterized by

$$\frac{\partial}{\partial\theta} H_T(\theta_T)' W H_T(\theta_T) = 0$$

For simplicity of illustration, we use H' to denote $\frac{\partial}{\partial\theta} H_T(\theta_T)$ and we do not distinguish between matrix transpositions. I.e. $\frac{\partial}{\partial\theta} H_T(\theta_T)'$ is also denoted by H' (instead of H'^T). Sometimes, to highlight the dependence on θ_0, we may write H' as $H'(\theta_0)$.

$$\sqrt{T} H_T(\theta_T) = \sqrt{T} H_T(\theta_0) + H'(\theta_0)\sqrt{T}(\theta_T - \theta_0) + o_p(1) \tag{1}$$

for some θ^* between θ_0 and θ_T.

$$H'(\theta_T) W H_T(\theta_T) = 0$$

implies

$$H'(\theta_T) W \left[\sqrt{T} H_T(\theta_0) + H'(\theta_0)\sqrt{T}(\theta_T - \theta_0) + o_p(1)\right] = 0$$

$$\begin{aligned} \sqrt{T}(\theta_T - \theta_0) &= -\left[H'(\theta_T) W H'(\theta_0)\right]^{-1} H'(\theta_T) W \sqrt{T} H_T(\theta_0) + o_p(1) \\ &= -\left[H' W H'\right]^{-1} H' W \sqrt{T} H_T(\theta_0) + o_p(1) \end{aligned}$$

i.e.,

$$\sqrt{T}(\theta_T - \theta_0) \xrightarrow{d} N\left(0, \left[H' W H'\right]^{-1} H' W \Sigma W H' \left[H' W H'\right]^{-1}\right)$$

where Σ is the asymptotic variance of $\sqrt{T}H_T(\theta_0)$. The optimal weighting matrix $W = \Sigma^{-1}$. And with this choice of W

$$\sqrt{T}(\theta_T - \theta_0) \xrightarrow{d} N\left(0, \left[H'WH'\right]^{-1}\right)$$

Substituting this into (1),

$$\sqrt{T}H_T(\theta_T) = \left(I - H'\left[H'WH'\right]^{-1}H'W\right)\sqrt{T}H_T(\theta_0)$$

$$\sqrt{T}W^{1/2}H_T(\theta_T) = \left(I - W^{1/2}H'\left[H'WH'\right]^{-1}H'W^{1/2}\right)\sqrt{T}W^{1/2}H_T(\theta_0)$$

By the choice of W,

$$\sqrt{T}W^{1/2}H_T(\theta_0) \xrightarrow{d} N(0, I)$$

And the matrix $\left(I - W^{1/2}H'\left[H'WH'\right]^{-1}H'W^{1/2}\right)$ is idempotent since

$$\begin{aligned}&\left(I - W^{1/2}H'\left[H'WH'\right]^{-1}H'W^{1/2}\right)^2 \\ &= I - W^{1/2}H'\left[H'WH'\right]^{-1}H'W^{1/2}\end{aligned}$$

The rank of an idempotent matrix is its trace, therefore

$$\begin{aligned}&Rank\left(I - W^{1/2}H\left[H'WH'\right]^{-1}H'W^{1/2}\right) \\ &= trace\left(I - W^{1/2}H\left[H'WH'\right]^{-1}H'W^{1/2}\right) \\ &= M - trace\left(W^{1/2}H\left[H'WH'\right]^{-1}H'W^{1/2}\right) \\ &= M - trace\left(\left[H'WH\right]^{-1}H'W^{1/2}W^{1/2}H\right) \\ &= M - K\end{aligned}$$

Lemma 3 *If an M-dimensional vector $x \sim N(0, I)$ and A is an idempotent matrix with rank K, then*

$$x'Ax \sim \chi^2_K$$

Example, if x is univariate/bi-variate normal, and A is identity matrix.

Proposition 4 *Therefore,*

$$TH_T(\theta_T)' WH_T(\theta_T) \xrightarrow{d} \chi^2_{M-K}$$

This is GMM specification test. What's the intuition of this test? What does the CDF of χ^2 look like?

References

[1] Handbook of Econometrics, Volume 4, Chapter 36, "Large sample estimation and hypothesis testing". It will be posted on Angel course website. Chapters of this handbook can also be downloaded through Columbia University Library website.

[2] Econometrics, by Hayashi, F., Princeton University Press. This book discusses various estimation methods under the unified theme of extremum estimators.