Generalized Method of Moments

Ruming Liu

February 13, 2025

1 Overview

This is a note of the Generalized Method of Moments by Hansen (1982).

2 Framework of GMM

- Step 1. Put forward a model and express this model by population moment conditions.
- Step 2. Use sample data to calculate the sample moments and estimate the estimators.
- Step 3. In this framework, not only the variance of estimators will be calculated and used as significance test. The variance of the sample moments will also be calculated as a test of the correctness of the model. If the variances of sample moments are quite large, we may reject the hypothesis that the sample fits the model.

3 Population Moment Conditions

- Assume the data are x_t and the true parameters b_0 satisfy the conditions

$$E[f(\vec{x_t}, b_0)] = \vec{0}, \tag{1}$$

we call this population moment conditions.

- An example is OLS, where the model should satisfy

$$E[x_t(y_t - b_0 x_t)] = E[x_t \epsilon_t] = 0. \tag{2}$$

- Another example is CCAPM, where the model should satisfy

$$\begin{bmatrix} E[m(b_0)R_f] - 1 \\ E[m(b_0)R_1^e] \\ E[m(b_0)R_2^e] \\ \vdots \\ E[m(b_0)R_K^e] \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3}$$

4 Sample Moment Conditions

- In equation (1), we give the population moment conditions under the assumptions of model. But we only can observe the samples, we denote

$$g_T(b) := \frac{1}{T}\sum_{t=1}^{T} f(x_t, b)^1 \tag{4}$$

as the sample moments after observing sample $\{x_t, t = 1, 2, .., T\}$.

- Now, our goal is to find a appropriate estimators $\hat{b}$ such that

$$g_T(\hat{b}) \sim= 0 \tag{5}$$

- If the model is just identification (number of moment conditions n is equal to the number of estimators p), the $=$ holds. If the model is over-identification (number of moment conditions is larger than number of estimators), the $=$ may not holds for all moment conditions, instead our goal will change to

$$a_{[p\times n]} g_T(\hat{b}) = 0 \tag{6}$$

where a is a $[p \times n]$ matrix, we will talk about how to choose this matrix in the later section.

5 Variance of Estimators and Moments

- In equation (6), we find a way to appropriately estimate the $\hat{b}$, now it is time to test whether the estimators and model are reasonable.

- To test the estimators and model, we need two statistics, the variance of estimators and moments.

[1]And here b can be either true parameters b_0 or estimators $\hat{b}$, we will discuss later.

- Hansen (1982) proves that

$$\sqrt{T}(\hat{b} - b_0) \to N(0, (ad)^{-1} a S a' (ad)^{-1\prime})_{[p \times p]} \tag{7}$$

$$\sqrt{T} g_T(\hat{b}) \to N(0, (I - d(ad)^{-1} a) S (I - d(ad)^{-1} a)')_{[n \times n]} \tag{8}$$

where $S = \sum_{j=-\infty}^{\infty} E[f(x_t, b_0) f'(x_{t-j}, b_0)]$ and $d = \frac{\partial g_T(b_0)}{\partial b'}$, but in reality, if S is unknown we can use data to estimate $\hat{S}$. Similarly, if d is unknown, we can use $\hat{b}$ to estimate $\hat{d}$.[2]

- Now the only issue left is the choice of a, our target is to chose an optimal a such that the estimator $\hat{\beta}$ is the most efficient, Hansen (1982) proves the optimal a should be:

$$a = d' S^{-1} \tag{9}$$

- We can easily find that the equation (6) is equivalent to another form of the GMM:

$$\begin{array}{rrl}
 & \text{GMM estimator } W = S^{-1}: & \hat{b} = \arg\min g_T(b)' S^{-1} g_T(b) \\
\Rightarrow & \text{first order condition:} & \left(\frac{\partial g_T'}{\partial b} S^{-1} \right) g_T(\hat{b}) = 0 \\
\Rightarrow & & \left(\left[\frac{\partial g_T}{\partial b'} \right]' S^{-1} \right) g_T(\hat{b}) = 0 \\
\Rightarrow & & \left(d' S^{-1} \right) g_T(\hat{b}) = 0 \\
\Rightarrow & \text{GMM estimator } a = d' S^{-1}: & a g_T(\hat{b}) = 0
\end{array}$$

- The economic meaning of choosing the weight matrix $W = S^{-1}$: for those moment conditions which are more volatile, we assign less weight to make $\hat{\beta}$ more efficient.

- In this optimal choice, the variances are simplified to:

$$\begin{aligned}
\text{var}(\hat{b}) &= \frac{1}{T} \left(d' S^{-1} d \right)^{-1} \\
\text{var}(g_T(\hat{b})) &= \frac{1}{T} \left(S - d(d' S^{-1} d)^{-1} d' \right) \\
T g_T(\hat{b})' S^{-1} g_T(\hat{b}) &\sim \chi^2_{n-p}
\end{aligned}$$

[2] Recall that in OLS: $var(\beta_{OLS}) = \sigma_\epsilon (X'X)^{-1}$, but σ_ϵ is unknown, so we use $\hat{\sigma_\epsilon}(X'X)^{-1}$ in significance test instead.

- Another technical issue is that the optimal a replies the value of d and S, which usually is unknown. In this case, we will use two steps GMM. The first step is to estimate $\hat{b_{old}}$ by setting $W = I$, use $\hat{b_{old}}$ estimate $\hat{d}$ and $\hat{S}$. Then in second step, we use optimal $\hat{a} = \hat{d}'\hat{S}^{-1}$ to update the $\hat{b_{new}}$.

6 GMM is All You Need

In this section, we will discuss that most of the regressions can be regarded as a special case of GMM framework

6.1 OLS

- The population moment condition of OLS is

$$E[x_t(y_t - x_t'b_0)] = 0 \tag{10}$$

$$\Rightarrow d = \frac{\partial g_T(b_0)}{\partial b} = -E[x_t x_t'] \tag{11}$$

$$\Rightarrow S = \sum_{j=-\infty}^{\infty} E[x_t \epsilon_t x_{t-j} \epsilon_{t-j}] = E[x_t \epsilon_t x_t' \epsilon_t] = \sigma_\epsilon^2 E[x_t x_t'] \tag{12}$$

$$\Rightarrow a = d'S^{-1} = -(\sigma_\epsilon^2)^{-1} I \tag{13}$$

- The sample moment condition is

$$a g_T(\hat{b}) = \vec{0} \Longleftrightarrow g_T(\hat{b}) = (\frac{1}{T}\sum x_t y_t) - (\frac{1}{T}\sum x_t x_t')\hat{b} = \vec{0} \tag{14}$$

$$\Rightarrow \hat{b} = (\frac{1}{T}\sum x_t x_t')^{-1}(\frac{1}{T}\sum x_t y_t) = (XX')^{-1}X'Y \tag{15}$$

- We get the equation of OLS estimator. Below is the variance of OLS estimator

$$Var(\hat{b}) = \frac{1}{T}(d'S^{-1}d)^{-1} = \frac{1}{T}[(-\sigma_\epsilon^{-1}I)(-E[x_t x_t'])]^{-1} = \frac{\sigma_\epsilon^2}{T}E[x_t x_t']^{-1} \tag{16}$$

$$\Rightarrow Var(\hat{b})|X = \sigma_\epsilon^2(X'X)^{-1} \tag{17}$$

6.2 OLS with Heteroscedasticity

- If the residuals are heteroscedasiticity, equation (12) will change to

$$\Rightarrow S = \sum_{j=-\infty}^{\infty} E[x_t\epsilon_t x_{t-j}\epsilon_{t-j}] = E[x_t\epsilon_t x_t'\epsilon_t] = E[\epsilon_t^2 x_t x_t'] \tag{18}$$

$$Var(\hat{b}) = \frac{1}{T}(d'S^{-1}d)^{-1} = \frac{1}{T}[(-E[x_t x_t'])E[\epsilon_t^2 x_t x_t']^{-1}(-E[x_t x_t'])]^{-1} \tag{19}$$

$$\Rightarrow Var(\hat{b})|X = T(X'X)^{-1}(X'DX)(X'X)^{-1} \tag{20}$$

where

$$D = \begin{bmatrix} \epsilon_1^2 & & & \\ & \epsilon_2^2 & & \\ & & \ddots & \\ & & & \epsilon_T^2 \end{bmatrix} \tag{21}$$

- Result is still consistent with White (1980) heteroscedasticity consistent estimator.

6.3 GLS and Autocorrelation

- We can also prove that the scenarios of GLS and Autocorrelation, are still under this framework, they are just with different forms of S.

6.4 2SLS and IV

- Interestingly, 2SLS and IV are still under this framework.

- The population moment conditions are:

$$E[z_t(y_t - x_t'b_0)] = 0 \tag{22}$$

$$g_T(\hat{b}) = Z'(Y - X\hat{b}) \tag{23}$$

$$\Rightarrow S = \sum_{j=-\infty}^{\infty} E[z_t\epsilon_t z_{t-j}\epsilon_{t-j}] = E[z_t\epsilon_t z_t'\epsilon_t] = E[\epsilon_t^2 z_t z_t'] = \sigma_\epsilon^2 E[z_t z_t'] \tag{24}$$

$$\Rightarrow d = \frac{\partial g_T(b_0)}{\partial b} = -E[z_t x_t'] \tag{25}$$

$$\Rightarrow a = d'S^{-1} = -(X'Z)(Z'Z)^{-1} \tag{26}$$

$$ag_T(\hat{b}) = -(X'Z)(Z'Z)^{-1}Z'(Y - X\hat{b}) = \vec{0} \tag{27}$$

$$\hat{b} = [(X'Z)(Z'Z)^{-1}Z'X]^{-1}(X'Z)(Z'Z)^{-1}Z'Y \tag{28}$$

- This is exactly the result in 2SLS.
- If we further assume that the number of instruments and the number of endogenous variables, then it is just identification, $Z'X$ is invertible

$$\hat{b} = (Z'X)^{-1}(Z'Z)(X'Z)^{-1}(X'Z)(Z'Z)^{-1}Z'Y = (Z'X)^{-1}Z'Y \tag{29}$$

7 Reference

- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. Econometrica, Vol. 50(4), 1029 – 1054.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica, Vol. 48(4), 817 – 838.
- https://zhuanlan.zhihu.com/p/92490417
- https://www.zhihu.com/question/358270088/answer/922799892
- https://zhuanlan.zhihu.com/p/672482646