Econometrics II TA Session #11*

Kenta KUDO †

January 14th, 2020

Contents

* All comments welcome!

† E-mail: vge011kk@student.econ.osaka-u.ac.jp

1 Preliminary

In this class, we review how to derive a GMM estimator in general cases. GMM method is widely used in the fields of panel data and finance, so we introduce how to use GMM in the dynamic panel data analysis.

2.1-2.2 How to derive the GMM estimator in the general case
2.3 How to decide a weight matrix?
2.4 Asymptotic distribution of GMM estimator
3.1 Testing hypothesis
4.1 Empirical example
5.1 Notification

2 GMM: non-linear case

2.1 Model Setting

In this section, we review how to derive a GMM estimator of the general case. For instance, suppose the regression model as follows:

$$f(y_i, x_i, \beta) = \epsilon_i, \tag{1}$$

where y_i and x_i represents the observed data and ϵ_i indicates the disturbance term. Here, the exogeneous variable vector is given as $z_i \in \mathbb{R}^{r\times 1}$ and the orthogonality condition is $\mathbb{E}(z_i'\epsilon_i) = 0$. The regressor is $\beta \in \mathbb{R}^{k\times 1}$. If **the order condition** ($r \geq k$) is satisfied, we can apply GMM.

2.2 GMM Estimator

Consider the case that we estimate Eq. (1) and assume that we represent $h(\theta : w_i) := z_i' f(y_i, x_i, \beta)$. Note that θ is a parameter vector. This vector corresponds to β in Eq. (1) and $w_i = (y_i, x_i)$ is the i th observed data. Then, the orthogonality condition is

$$\mathbb{E}[h(\theta : w_i)] = 0,$$

As in the case of linear models, we can derive GMM estimator by solving the following minimization problem:

$$\min_{\theta} q \equiv \bar{g}_n' S^{-1} \bar{g}_n,$$

where $\bar{g}_n(\theta : W) := \frac{1}{n}\sum_{i=1}^{n} h(\theta : w_i)$ and S is a positive definite (and symmetric) matrix. The estimator is obtained from the following first order condition:

$$\frac{\partial q}{\partial \theta} = 2\frac{\partial \bar{g}_n'(\theta : W)}{\partial \theta} S^{-1}\bar{g}_n = \mathbf{0}.$$

To obtain $\widehat{\theta}$, we linearize the first-order condition around $\theta = \widehat{\theta}$,

$$\begin{aligned} 0 &= \frac{\partial \bar{g}_n'(\theta : W)}{\partial \theta} S^{-1}\bar{g}_n(\theta : W) \\ &\approx \frac{\partial \bar{g}_n'(\widehat{\theta} : W)}{\partial \theta} S^{-1}\left(\bar{g}_n(\widehat{\theta} : W) + \frac{\partial \bar{g}_n(\widehat{\theta} : W)}{\partial \theta'}(\theta - \widehat{\theta})\right). \end{aligned}$$

(Note that the second derivative is omitted.) Let $\widehat{D}$ is the first derivative of $\bar{g}_n(\widehat{\theta} : W)$ with respect to θ', then we have

$$\widehat{D}' S^{-1}\bar{g}_n(\theta : W) + \widehat{D}' S^{-1}\widehat{D}(\theta - \widehat{\theta}) = 0. \tag{2}$$

Rewriting Eq. (2), we can estimate the regressor by the iterative procedure for $i = 1, 2, 3, \ldots$:

$$\widehat{\theta}^{i+1} = \widehat{\theta}^{i} - (\widehat{D}_i S^{-1}\widehat{D}_i')^{-1}\widehat{D}_i S^{-1}\bar{g}_n(\widehat{\theta}^{i} : W),$$

where

$$\widehat{D}_i \equiv \frac{\partial \bar{g}_n(\widehat{\theta^i} : W)}{\partial \theta'}.$$

2.3 How to estimate $\widehat{S}$?

Suppose that $h(\theta : w_i)$ has a stationarity and $\Gamma_\tau := \mathbb{E}(h(\theta : w_i)h(\theta : w_{i-\tau})') < \infty$.

Assumption 2.1 (Stationarity) If $h(\theta : w_i)$ has a stationarity, we have

1. $\mathbb{E}(h(\theta : w_i))$ doen not depend on i.
2. $\mathbb{E}(h(\theta : w_i)h(\theta : w_{i-\tau})')$ depends on the time difference γ.

In this case, S is the variance of $\sqrt{n}\bar{g}_n(\theta : W)$:

$$\begin{aligned}
S = \mathrm{V}ar[\sqrt{n}\bar{g}_n(\theta : W)] &= \frac{1}{n}\mathrm{V}ar\left[\sum_{i=1}^{n} h(\theta : w_i)\right] \\
&= \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n} h(\theta : w_i)\sum_{i=1}^{n} h(\theta : w_i)'\right] \\
&= \frac{1}{n}\{n\Gamma_0 + (n-1)(\Gamma_1 + \Gamma_1') + (n-2)(\Gamma_2 + \Gamma_2')\cdots + (\Gamma_{n-1} + \Gamma_{n-1})\} \\
&= \Gamma_0 + \sum_{i=1}^{n-1}\left(1 - \frac{i}{n}\right)(\Gamma_i + \Gamma_i'),
\end{aligned}$$

where $\Gamma_\tau' = \mathbb{E}[h(\theta : w_i)h(\theta : w_{i-\tau})'] = \Gamma_{-\tau}$. The estimator of S is

$$\widehat{S} = \widehat{\Gamma}_0 + \sum_{i=1}^{q-1}\left(1 - \frac{i}{q+1}\right)(\widehat{\Gamma}_i + \widehat{\Gamma}_i').$$

This is the **Newey-West estimator**. Remind that n is replaced by $q+1$, and therefore $q \leq n$. We need to estimate $\widehat{\Gamma}_\tau$ as

$$\widehat{\Gamma}_\tau = \frac{1}{n}\sum_{i=\tau+1}^{n} h(\widehat{\theta} : w_i)h(\widehat{\theta} : w_{i-\tau})'.$$

Note that $\widehat{S} \to S$, because $\widehat{\Gamma_\tau} \to \Gamma_\tau$ as $n \to +\infty$.

2.4 Asymptotic Distribution of GMM Estimator

In this subsection, we assume that the GMM estimator has the following properties. *1

Assumption 2.2 (Assumptions for the Asymptotic Normality of the GMM Estimator)

1. $\widehat{\theta} \to \theta$
2. $\sqrt{n}\bar{g}_n(\theta : W) \to N(0, S),\ S = \lim_{n\to\infty} V[\sqrt{n}\bar{g}_n(\theta : W)]$.

Then, the GMM estimator has **the asymptotic normality** stated as follows.

Theorem 2.3 (Asymptotic Normality of the GMM Estimator) $\widehat{\theta}_{GMM}$ satisfies

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, (DS^{-1}D)^{-1})$$

where D is the first derivative of $\bar{g}_n(\theta : W)$ with respect to θ'.

Proof. The approximation through the linearization of $\bar{g}_n(\widehat{\theta} : W)$ around $\widehat{\theta} = \theta$ yields

$$\bar{g}_n(\widehat{\theta}) = \bar{g}_n(\theta : W) + \frac{\partial \bar{g}_n(\bar{\theta} : W)}{\partial \theta'}(\widehat{\theta} - \theta), \tag{3}$$

where $\bar{\theta} \in (\widehat{\theta}, \theta)$. Substituting Eq. (3) into Eq. (2) at $\widehat{\theta} = \theta$, we have

$$0 = \widehat{D}'\widehat{S}^{-1}(\bar{g}_n(\theta : W) + \bar{D}(\widehat{\theta} - \theta)),$$

where

$$\bar{D} := \frac{\partial \bar{g}_n(\bar{\theta} : W)}{\partial \theta'} \in \mathbb{R}^{r\times k}.$$

By using **(Assumption 1.2)**, the asymptotic distribution is

$$\begin{aligned}\sqrt{n}(\widehat{\theta} - \theta) &= (\widehat{D}'\widehat{S}^{-1}\bar{D})^{-1}\widehat{D}'\widehat{S}^{-1} \cdot \sqrt{n}\bar{g}_n(\theta : W) \\ &\xrightarrow{d} N(0, (DS^{-1}D)^{-1})\end{aligned}$$

where $\widehat{D} \to D, \bar{D} \to D, \widehat{S} \to S$ since $\widehat{\theta} \to \theta,\ \bar{\theta} \to \theta$. □

---

*1 GMM estimator has the consisitency under some general conditions. The formal statement is explained in Chapter 14 of Wooldridge (2010).

3 Testing Hypothesis

In this section, we consider the following hypothesis:

- $H_0 : R(\theta) = 0 \in \mathbb{R}^{p\times 1}$;
- $H_1 : R(\theta) \neq 0$.

Note that $p \geq k$ is the number of restrictions. By the **delta method**, $R(\widehat{\theta})$ is linearlized as

$$R(\widehat{\theta}) = R(\theta) + R_{\bar{\theta}}(\widehat{\theta} - \theta),$$

where

$$R_{\bar{\theta}} := \frac{\partial R(\bar{\theta})}{\partial \theta'} \in \mathbb{R}^{p\times k}.$$

Remind that $\bar{\theta}$ is between θ and $\widehat{\theta}_{\text{GMM}}$. Under the null hypothesis, we have $R(\widehat{\theta}) = R_{\bar{\theta}}(\widehat{\theta}) - R(\theta))$, which implies that the distribution of $R(\widehat{\theta})$ is equivalent to that of $R_{\bar{\theta}}(\widehat{\theta} - \theta)$. The asymptotic distribution of $\sqrt{n}(R(\widehat{\theta}) - R(\theta))$ is

$$\begin{aligned}\sqrt{n}(R(\widehat{\theta}) - R(\theta)) &= \sqrt{n}R_{\bar{\theta}}(\widehat{\theta} - \theta) \\ &\to N(0, R_\theta(D'S^{-1}D)^{-1}R'_\theta),\end{aligned}$$

since $R_{\bar{\theta}} \to R_\theta$ as $\widehat{\theta} \to \theta$. Thus, we have the following distribution:

$$nR(\widehat{\theta})\left(R_\theta(D'S^{-1}D)^{-1}\right)R'(\widehat{\theta}) \to \chi^2(p).$$

Practically, replacing θ by $\widehat{\theta}$, we can derive the test statistic under $H_0 : R(\theta) = 0$,

$$n \cdot (R(\widehat{\theta}_{\text{GMM}}))(R_{\widehat{\theta}}(\widehat{D}'\widehat{S}^{-1}\widehat{D})^{-1})(R'(\widehat{\theta})) \to \chi^2(p).$$

This is a kind of **Wald type tests**.

4 Empirical Example

Today, we estimate the dynamic panel data model by using the plm package. We can estimate the dynamic panel data model by using GMM. In this class, we use EmplUK data, which is related to the number of workers in 140 firms of the United Kingdom in 1976-1984.*2 The model which we estimate is

$$\Delta \log(\text{emp}_{it}) = \alpha_i + \beta_1 \log(\text{emp}_{it-1}) + \beta_2 \log(\text{emp}_{it-2}) + \gamma_1 \log(\text{wage}_{it}) + \gamma_2 \log(\text{wage}_{it-1}) + \delta_1 \log(\text{capital}_{it-1}) + \theta_1 \log(\text{output}_{it}) + \theta_2 \log(\text{output}_{it-1}) + \Delta\lambda_t + \Delta\epsilon_{it}.$$

The variables contained in the above equation are

- emp_{it}: the number of workers,
- wage_{it}: the real wage,
- capital_{it}: the total capital,
- output_{it}: the quantity of output,

where $i = 1, \cdots, 140$ and $t = 4, \cdots, 9$. R code and the result are given as follows.

```
rm(list=ls(all=TRUE))
library(plm)

data("EmplUK", package = "plm")
## Arellano and Bond (1991), table 4 col. b
z1 <- pgmm(log(emp) ~ lag(log(emp), 1:2) + lag(log(wage), 0:1)
           + log(capital) + lag(log(output), 0:1) | lag(log(emp), 2:99),
           data = EmplUK, effect = "twoways", model = "twosteps")
summary(z1, robust = TRUE)

##results
Twoways effects Two steps model

Call:
pgmm(formula = log(emp) ~ lag(log(emp), 1:2) + lag(log(wage),
    0:1) + log(capital) + lag(log(output), 0:1) | lag(log(emp),
    2:99), data = EmplUK, effect = "twoways", model = "twosteps")

Unbalanced Panel: n = 140, T = 7-9, N = 1031

Number of Observations Used: 611

Residuals:
      Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
-0.6190677 -0.0255683  0.0000000 -0.0001339  0.0332013  0.6410272

Coefficients:
                        Estimate Std. Error z-value  Pr(>|z|)
lag(log(emp), 1:2)1     0.474151   0.185398  2.5575 0.0105437 *
```

*2 This data set is provided in plm package. The detail of estimation is explained in Arellano and Bond(1991).

```
lag(log(emp), 1:2)2     -0.052967   0.051749 -1.0235 0.3060506
lag(log(wage), 0:1)0    -0.513205   0.145565 -3.5256 0.0004225 ***
lag(log(wage), 0:1)1     0.224640   0.141950  1.5825 0.1135279
log(capital)             0.292723   0.062627  4.6741 2.953e-06 ***
lag(log(output), 0:1)0  0.609775   0.156263  3.9022 9.530e-05 ***
lag(log(output), 0:1)1 -0.446373   0.217302 -2.0542 0.0399605 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sargan test: chisq(25) = 30.11247 (p-value = 0.22011)
Autocorrelation test (1): normal = -1.53845 (p-value = 0.12394)
Autocorrelation test (2): normal = -0.2796829 (p-value = 0.77972)
Wald test for coefficients: chisq(7) = 142.0353 (p-value = < 2.22e-16)
Wald test for time dummies: chisq(6) = 16.97046 (p-value = 0.0093924)
>
```

5 Notification

The solution to the question (10) of the assignment #01 must be modified. The following second derivative is true:

$$\frac{\partial^2 l(\beta, \sigma^2)}{\partial\beta\partial\sigma^2} = \sum_{i=1}^{n} \left(-\frac{1}{\sigma^4}(y_i - X_i\beta) - \frac{X_i\beta}{2\sigma^4}\frac{\phi_i'}{\Phi_i} + \frac{1}{2\sigma^3}\frac{\phi_i}{\Phi_i} - \frac{X_i\beta}{2\sigma^4}\frac{\phi_i^2}{\Phi_i^2} \right) X_i'.$$

References

[1] Manuel Arellano and Stephen Bond (1991) *"Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations"*, Review of Economic Studies, 58, 277-297.

[2] W. H. Greene (2012) *"Econometric analysis Seventh Edition"*, Pearson.

[3] J. M. Wooldridge (2010) *"Econometric Analysis of Cross Section and Panel Data (2nd Eddition)"*, The MIT Press.