

GMM

Sangmyung Ha*

March 14, 2024

References

1. Bruce E. Hansen (2023), "Econometrics", Chapter 13
2. A. W. van der Vaart (2000), "Asymptotic Statistics", Chapter 5.2-5.3

1 Moment Conditions

Let $(w_i)_{i=1}^n \in \mathbb{R}^r$ be a set of random vectors that have common distribution, characterized by the parameters $\theta_0 \in \Theta \subset \mathbb{R}^p$. Suppose the true parameter θ_0 satisfies a set of moment conditions characterized by $m : \mathbb{R}^r \times \Theta \rightarrow \mathbb{R}^q$. In other words, θ_0 satisfies

$$\mathbb{E}[m(w_i, \theta_0)] = 0_{q \times 1}$$

Then, we may estimate θ_0 by finding $\hat{\theta} \in \Theta$ such that

$$\mathbb{E}_n[m(w_i, \hat{\theta})] := \frac{1}{n} \sum_{i=1}^n m(w_i, \hat{\theta}) = 0_{q \times 1}$$

This is called the sample analog estimator or the method of moments.

Recall that p is the number of unknown parameters and q is the number of moment conditions.

1. $p > q$: Under-identified. There exists an infinite number of $\theta \in \Theta$ such that

$$\mathbb{E}_n[m(w_i, \theta)] = \frac{1}{n} \sum_{i=1}^n [m(w_i, \theta)] = 0_{q \times 1} \tag{1}$$

2. $p = q$: Just-identified. There exists a unique $\theta \in \Theta$ such that (1) is satisfied.
3. $p < q$: Over-identified. There does not exist any $\theta \in \Theta$ such that (1) is satisfied.

*Special thanks to Naoya.

We define the generalized method of moments (GMM) estimator $\hat{\theta}_{GMM}$ of θ as the minimizer of

$$Q_n(\theta) = \left(\frac{1}{n} \sum_{i=1}^n m(w_i, \theta) \right)' W_n \left(\frac{1}{n} \sum_{i=1}^n m(w_i, \theta) \right)$$

for some $q \times q$ dimensional sequence of positive definite weighting matrices W_n such that $W_n \rightarrow_p W$ where W is also positive definite. If we have just-identified case, using any positive definite W_n gives us the same minimizer $\hat{\theta}_{GMM}$ because the solution is given by $\hat{\theta}_{GMM}$ that gives

$$\frac{1}{n} \sum_{i=1}^n m(w_i, \hat{\theta}_{GMM}) = 0_{q \times 1}$$

so that $Q_n(\hat{\theta}_{GMM}) = 0$. It can be shown that the weighting matrix defined by the positive definite matrix $W = V^{-1}$

$$V = \lim_{n \rightarrow \infty} \text{var} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n m(w_i, \theta_0) \right)$$

is optimal, i.e., yields estimator with the smallest variance. The efficient GMM estimator is defined as

$$\hat{\theta}_{eff\ GMM} = \arg \min_{\theta \in \Theta} \left(\frac{1}{n} \sum_{i=1}^n m(w_i, \theta) \right)' V^{-1} \left(\frac{1}{n} \sum_{i=1}^n m(w_i, \theta) \right)$$

In practice, we replace V with a feasible estimator $\hat{V}$, which is why we allow the weight matrix W_n to depend on n . Since V depends on the true parameter θ_0 , we need two step or iterative method in order to calculate the efficient GMM estimator.

1.1 Examples: OLS

Suppose i.i.d. samples (x_i, y_i) can be written as

$$y_i = x_i' \beta + e_i$$

where $\mathbb{E}[x_i e_i] = 0$. Let $w_i = (x_i', y_i)'$, $m(w_i, \beta) = x_i e_i = x_i(y_i - x_i' \beta)$.

1. Moment condition:

$$\mathbb{E}[m(w_i, \beta)] = \mathbb{E}[x_i(y_i - x_i' \beta)] = 0$$

2. Sample moment condition:

$$\mathbb{E}_n[m(w_i, \beta)] = \mathbb{E}_n[x_i(y_i - x_i' \beta)] = \frac{1}{n} \sum_{i=1}^n x_i(y_i - x_i' \beta) = 0$$

For any positive definite weighting matrix W_n , the minimizer

$$\begin{aligned}\hat{\beta}_{GMM} &= \arg \min_{\beta \in \mathbb{R}^k} \left(\frac{1}{n} \sum_{i=1}^n m(w_i, \beta) \right)' W_n \left(\frac{1}{n} \sum_{i=1}^n m(w_i, \beta) \right) \\ &= \arg \min_{\beta \in \mathbb{R}^k} \left(\frac{1}{n} \sum_{i=1}^n x_i(y_i - x_i' \beta) \right)' W_n \left(\frac{1}{n} \sum_{i=1}^n x_i(y_i - x_i' \beta) \right)\end{aligned}$$

can be obtained by solving for β in

$$\frac{1}{n} \sum_{i=1}^n x_i(y_i - x_i' \beta) = 0$$

In other words, use

$$\hat{\beta}_{GMM} = (X'X)^{-1}X'y$$

1.2 Examples: IV

Suppose i.i.d. samples (x_i, z_i, y_i) can be written as

$$\begin{aligned}y_i &= x_i' \beta + e_i \\ z_i &= \gamma x_i + u_i\end{aligned}$$

where $y_i, e_i \in \mathbb{R}, x_i \in \mathbb{R}^p, z_i, u_i \in \mathbb{R}^q$, and $\gamma \in \mathbb{R}^{q \times p}$. Here, β is the parameter of interest, and z_i 's are called instruments. Suppose

1. $\mathbb{E}[x_i e_i] \neq 0$
2. $\mathbb{E}[x_i z_i']$ is full rank.
3. $\mathbb{E}[z_i e_i] = 0$

Let $w_i = (x_i', z_i', y_i)'$ and $m(w_i, \beta) = z_i e_i = z_i(y_i - x_i' \beta)$. The moment condition is given by

$$\mathbb{E}[m(w_i, \beta)] = \mathbb{E}[z_i(y_i - x_i' \beta)] = 0$$

The sample moment condition is given by

$$\frac{1}{n} \sum_{i=1}^n z_i(y_i - x_i' \beta) = 0$$

If $p = q$, the number of parameters and the number of instruments are the same; We have just-identified case and may directly find $\hat{\beta}_{GMM}$ so that the above equality is satisfied. Writing the equation in matrix form,

$$\frac{1}{n} Z'(y - X\beta) = 0$$

The solution is given by

$$\hat{\beta}_{GMM} = (Z'X)^{-1}Z'y$$

If we have $p < q$, we have over-identification, and the GMM estimator is defined by

$$\begin{aligned}\hat{\beta}_{GMM} &= \arg \min_{\beta} \left(\frac{1}{n} \sum_{i=1}^n z_i(y_i - x_i'\beta) \right)' W_n \left(\frac{1}{n} \sum_{i=1}^n z_i(y_i - x_i'\beta) \right) \\ &= \arg \min_{\beta} \left(\frac{1}{n} Z'(y - X\beta) \right)' W_n \left(\frac{1}{n} Z'(y - X\beta) \right) \\ &= \arg \min_{\beta} \left(\frac{1}{n^2} (y - X\beta)' ZW_n Z' (y - X\beta) \right)\end{aligned}$$

The FOC is given by

$$-2X'ZW_nZ'(y - X\beta) = 0$$

which gives

$$\hat{\beta}_{GMM} = (X'ZW_nZ'X)^{-1}X'ZW_nZ'y$$

The optimal weight $W = V^{-1}$ is defined by

$$\begin{aligned}V &= \lim_{n \rightarrow \infty} \text{var} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n m(w_i, \beta) \right) \\ &= \lim_{n \rightarrow \infty} \text{var} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n z_i e_i \right) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbb{E}[z_i z_i' e_i^2] \\ &= \mathbb{E}[z_i z_i' e_i^2] \\ &= \mathbb{E}[z_i z_i'] \sigma_e^2\end{aligned}$$

where we've used independence in third, and identically distributed in the fourth equality. Final equality only holds if we have conditional homoskedasticity, i.e., $\mathbb{E}[e_i|z_i] = \sigma_e^2$. Suppose we estimate V using

$$\begin{aligned}\hat{V} &= \frac{\hat{\sigma}_e^2}{n} \sum_{i=1}^n z_i z_i' \\ &= \frac{\hat{\sigma}_e^2}{n} Z'Z\end{aligned}$$

Then, the optimal weight is given by $W_n = \frac{n}{\hat{\sigma}_e^2}(Z'Z)^{-1}$ and the efficient GMM is given by

$$\begin{aligned}\hat{\beta}_{eff\ GMM} &= (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'y \\ &= (X'P_ZX)^{-1}X'P_Zy \\ &= (\hat{X}'\hat{X})^{-1}\hat{X}'y, \quad \hat{X} = P_ZX\end{aligned}$$

where the scaling term $\frac{\hat{\sigma}_e^2}{n}$ have canceled out. Thus, efficient GMM is equivalent to 2-step least squares if we have conditional homoskedasticity on e_i (conditional on z_i).

2 Consistency

Notations

1. Population moment:

$$m(\theta) = \mathbb{E}[m(w_i, \theta)]$$

2. Population objective function:

$$Q(\theta) = m(\theta)'Wm(\theta)$$

3. Sample moments:

$$m_n(\theta) = \frac{1}{n} \sum_{i=1}^n m(w_i, \theta)$$

4. Sample objective function:

$$Q_n(\theta) = m_n(\theta)'Wm_n(\theta)$$

5. GMM estimator

$$\hat{\theta}_{GMM} = \arg \min_{\theta \in \Theta} Q_n(\theta)$$

We show consistency of the GMM estimator by showing consistency the M-estimators. M-estimator is an estimator that is defined as either a minimizer or a maximizer of a certain objective function.

Theorem 1 (Theorem 5.7 of A.W. van der Vaart (2000)). Let $\hat{\theta}$ be defined as a maximizer of $Q_n(\theta)$. Suppose

1. $Q_n(\theta) \rightarrow_p Q(\theta)$ uniformly over $\theta \in \Theta$, i.e.,

$$\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| \rightarrow_p 0$$

2. θ_0 is a 'well-separated' maximum of $Q(\theta)$, i.e.,

$$\sup_{d(\theta, \theta_0) \geq \epsilon} Q(\theta) < Q(\theta_0)$$

for all $\epsilon > 0$.

Then, $\hat{\theta}_0 \rightarrow_p \theta_0$.

Proof. Note that $Q_n(\theta_0) \leq Q_n(\hat{\theta})$ by the definition of $\hat{\theta}$. Because $Q_n(\theta) \rightarrow_p Q(\theta)$ pointwise over Θ , we have

$$Q_n(\theta_0) \rightarrow_p Q(\theta_0)$$

This means that

$$Q(\theta_0) \leq Q_n(\hat{\theta})$$

in probability as $n \rightarrow \infty$. Thus, subtracting both sides by $Q(\hat{\theta})$, we have

$$\begin{aligned} Q(\theta_0) - Q(\hat{\theta}) &\leq Q_n(\hat{\theta}) - Q(\hat{\theta}) \\ &\leq \sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| \rightarrow_p 0 \end{aligned}$$

in probability as $n \rightarrow \infty$ by the uniform convergence. Thus, $Q(\theta_0) - Q(\hat{\theta}) \leq 0$ in probability as $n \rightarrow \infty$. However, by the well-separatedness of θ_0 , for all $\epsilon > 0$, we have

$$\sup_{d(\theta, \theta_0) \geq \epsilon} Q(\theta) < Q(\theta_0)$$

Suppose $d(\hat{\theta}, \theta_0) \geq \epsilon$. Then, we must have $Q(\hat{\theta}) < Q(\theta_0)$. However, the probability of the latter converges to zero as $n \rightarrow \infty$. This means that the probability of the event $\{d(\hat{\theta}, \theta_0) \geq \epsilon\}$, which is contained in the event $\{Q(\hat{\theta}) < Q(\theta_0)\}$, should also converge to zero. $\square$

The proof for the minimizer is essentially the same. Thus, to prove that GMM is consistent, we need to prove that (1) $Q_n(\theta) \rightarrow Q(\theta)$ uniformly, (2) θ_0 is a 'well-separated' maximum of $Q(\theta)$. Sufficient conditions for uniform convergence are

- (a) Θ is compact.
- (b) $\{w_i\}$ is i.i.d. or strictly stationary and ergodic.
- (c) $m(w_i, \theta)$ is continuous in $\theta \in \Theta$ almost surely.
- (d) $\mathbb{E}[\sup_{\theta \in \Theta} |m(w_i, \theta)|] < \infty$.
- (e) $W_n \rightarrow_p W$.

3 Asymptotic Normality

Notations

1. Population Jacobian matrix ($q \times p$):

$$M(\theta) = \mathbb{E} \left[\frac{\partial m(w_i, \theta)}{\partial \theta'} \right]$$

2. Sample Jacobian matrix ($q \times p$):

$$M_n(\theta) = \frac{1}{n} \sum_{i=1}^n \frac{\partial m(w_i, \theta)}{\partial \theta'}$$

In addition to (a) to (e) in the previous page, we further assume that

- (f) Θ is convex.
- (g) $m(w_i, \theta), m(\theta)$ are differentiable in $\theta \in \Theta$ almost surely.
- (h) $\sqrt{n}m_n(\theta_0) \rightarrow_d \mathbb{N}(0, V)$ where $V = \mathbb{E}[m(w_i, \theta_0)m(w_i, \theta_0)']$
- (i) $M_n(\theta) \rightarrow_p M(\theta)$ uniformly in Θ .

What does condition (i) give us? Suppose $\hat{\theta} \rightarrow_p \theta_0$ and $M_n(\theta) \rightarrow_p M(\theta)$ pointwise over all $\theta \in \Theta$. Does this guarantee $M_n(\hat{\theta}) \rightarrow_p M(\theta_0)$? Sadly, no... However, if we have uniform convergence, i.e.,

$$\sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)| \rightarrow_p 0$$

we can write

$$\begin{aligned} M_n(\hat{\theta}) - M(\theta_0) &= M_n(\hat{\theta}) - M(\hat{\theta}) + M(\hat{\theta}) - M(\theta_0) \\ &\leq |M_n(\hat{\theta}) - M(\hat{\theta})| + |M(\hat{\theta}) - M(\theta_0)| \\ &\leq \left(\sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)| \right) + |M(\hat{\theta}) - M(\theta_0)| \end{aligned}$$

The first term converges in probability to zero by uniform convergence. The second term converges to zero by continuous mapping theorem.

Theorem 2. The GMM estimator has the limit distribution

..

Proof. Using the mean value theorem, we have

$$m_n(\hat{\theta}) - m_n(\theta_0) = M_n(\tilde{\theta})(\hat{\theta} - \theta_0)$$

for some $\tilde{\theta}$ on the line segment connecting $\hat{\theta}$ and θ_0 . Pre-multiplying both sides by $M_n(\hat{\theta})'W_n$, we have

$$M_n(\hat{\theta})'W_n(m_n(\hat{\theta}) - m_n(\theta_0)) = M_n(\hat{\theta})'W_nM_n(\tilde{\theta})(\hat{\theta} - \theta_0)$$

Pre-multiplying both sides by $\sqrt{n}(M_n(\hat{\theta})'W_nM_n(\tilde{\theta}))^{-1}$ and rearranging, we have

$$\sqrt{n}(\hat{\theta} - \theta_0) = \sqrt{n}(M_n(\hat{\theta})'W_nM_n(\tilde{\theta}))^{-1}M_n(\hat{\theta})'W_n(m_n(\hat{\theta}) - m_n(\theta_0))$$

Since $\hat{\theta}$ is minimizer of $Q_n(\theta)$, it should satisfy the first order condition. In other words,

$$\frac{\partial Q_n(\hat{\theta})}{\partial \theta} = \frac{\partial(m_n(\hat{\theta})'W_nm_n(\hat{\theta}))}{\partial \theta} = 2M_n(\hat{\theta})'W_nm_n(\hat{\theta}) = 0$$

Thus, we have

$$\begin{aligned}\sqrt{n}(\hat{\theta} - \theta_0) &= \sqrt{n}(M_n(\hat{\theta})'W_nM_n(\tilde{\theta}))^{-1}M_n(\hat{\theta})'W_n(m_n(\hat{\theta}) - m_n(\theta_0)) \\ &= -(M_n(\hat{\theta})'W_nM_n(\tilde{\theta}))^{-1}M_n(\hat{\theta})'W_n\sqrt{n}m_n(\theta_0)\end{aligned}$$

Note that we've already proved consistency of $\hat{\theta}$. This means that the line segment connecting $\hat{\theta}$ and θ_0 gets smaller as $n \rightarrow \infty$ so that $\tilde{\theta}$ also converges in probability to θ_0 . Thus, we have

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d (M'WM)^{-1}M'WN(0, V)$$

where $M = M(\theta_0) = \mathbb{E} \left[\frac{\partial m(w_i, \theta_0)}{\partial \theta'} \right]$. In summary,

$$\begin{aligned}\sqrt{n}(\hat{\theta} - \theta_0) &\rightarrow_d (M'WM)^{-1}M'WN(0, V) \\ &= \mathbb{N}(0, \Sigma(W))\end{aligned}$$

$\sqrt{n}(\hat{\theta} - \theta_0)$ is asymptotically normally distributed with limit variance given by

$$\Sigma(W) := (M'WM)^{-1}M'WVWM(M'WM)^{-1}$$

□

3.1 Optimal Weight

The optimal weight is given by $W = V^{-1}$. To show this, we need to show that for any positive definite weight matrix W , we have

$$\Sigma(W) - \Sigma(V^{-1}) \geq 0$$

where $\geq$ denotes that $\Sigma(W) - \Sigma(V^{-1})$ is a positive semi definite matrix. This is equivalent to showing that

$$\Sigma(V^{-1})^{-1} - \Sigma(W)^{-1}$$

is positive semi definite. Notice that

$$\begin{aligned}\Sigma(V^{-1}) &= (M'V^{-1}M)^{-1}M'V^{-1}VV^{-1}M(M'V^{-1}M)^{-1} \\ &= (M'V^{-1}M)^{-1}\end{aligned}$$

Thus,

$$\begin{aligned}\Sigma(V^{-1})^{-1} - \Sigma(W)^{-1} &= M'V^{-1}M - (M'WM)(M'WVWM)^{-1}(M'WM) \\ &= M'V^{-1/2}(I_q - V^{1/2}WM(M'WVWM)^{-1}M'WV^{1/2})V^{-1/2}M\end{aligned}$$

Note that $I_q - V^{1/2}WM(M'WVWM)^{-1}M'WV^{1/2}$ is a projection matrix, whose eigenvalues only consist of only zeros or ones. Thus, it is a positive semi-definite matrix. Using the definition of positive semi-definiteness, we may show that $\Sigma(V^{-1})^{-1} - \Sigma(W)^{-1}$ is also positive semi-definite.