第 4 章付録 1：一般化モーメント法 (GMM) に関するノート

最終更新： 2025 年 11 月 3 日

上武康亮・遠山祐太・若森直樹・渡辺安虎
「実証ビジネス・エコノミクス」
第 4 章「プライシングの真髄は代替性にあり」
の付録

1　はじめに

本付録では、一般化モーメント法 (generalized method of moments、以下 GMM) に関する簡単な解説を行う。GMM は第 4 章で取り上げるランダム係数ロジットモデルの推定をはじめ、産業組織論の実証分析、および構造推定全般で広く用いられている。また、第 3 章における線形モデルの操作変数法や 2 段階最小 2 乗法も GMM の特殊ケースとして見ることができる。

本資料では、最初にモーメント法を解説した後、それの一般化である GMM について見ていく。なお、詳細な解説については以下の教科書等を参考にされたい。

- 末石直也 (2015)『計量経済学：ミクロデータ分析へのいざない』日本評論社。
- Hayashi, F (2000) *Econometrics*, Princeton University Prress.
- Newey, W. K. and McFadden, D. (1994) “Large Sample Estimation and Hypothesis Testing,” *Handbook of Econometrics*, Vol.4, North Holland: 2111–2245.

2　モーメント法

統計的推定における**モーメント法** (method of moment) とは、母集団におけるモーメントを標本 (サンプル) におけるモーメントで置き換えることから得られる条件を用いて未知パラメターを推定する方

法である。以下では具体例を通じて見ていこう。

2.1 例１：正規分布の平均と分散

確率変数 X を $X \sim N(\mu, \sigma^2)$ としよう。このとき、1 次と 2 次のモーメント $E[X], E[X^2]$ はそれぞれ以下で与えられる。

$$
\begin{aligned}
E[X] &= \mu, \\
E[X^2] &= \sigma^2 + \mu^2.
\end{aligned}
$$

これらのモーメントに関する条件を用いることでパラメター μ と σ^2 を推定することができる。具体的には、上二式に母集団モーメントを対応する標本のモーメント、すなわち標本平均で置き換えれば良い。今、母集団からのランダムサンプルを $\{x_1, \cdots, x_n\}$ とすると、モーメント推定量は以下で与えられる。

$$
\begin{aligned}
\hat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} x_i, \\
\hat{\sigma^2} &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2.
\end{aligned}
$$

2.2 例 2：線形回帰モデル

以下の線形回帰モデルを考えよう。

$$
y_i = x_i'\beta + u_i, \quad E[x_i u_i] = 0.
$$

ここで x_i と β は $(K \times 1)$ ベクトルである。また、条件 $E[x_i u_i] = 0$ が課されており、これは説明変数 x_i と誤差項 u_i が無相関であることを意味する。このモデルから以下の K 本のモーメント条件が得られる。

$$
E[x_i(y_i - x_i'\beta)] = 0.
$$

今、未知のパラメター β は K 個あり、モーメント条件が K 個存在する。適切な条件 (特に説明変数に関する多重共線性) のもとで、パラメター β について解くことができる。

推定に際しては、上と同様に母集団モーメントを標本で対応するものに置き換えてやればよい。

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

この式は最小 2 乗法における 1 階条件と等しい。

2.3　例 2：操作変数法

以下のモデルを考えよう。

$$y_i = x_i'\beta + e_i, \quad E[z_i e_i] = 0.$$

ここで、z_i は操作変数であり、$(L \times 1)$ ベクトルとしよう。操作変数と誤差項の独立性条件から、L 個のモーメント条件が得られる。

ここで、パラメター数 K と操作変数の個数 L に応じて、モーメント法で推定できるか否かが決まる。具体的には以下の 3 つのケースが考えられる。

■ケース 1：$L < K$　このケースでは、モーメント条件数がパラメター数よりも少ないため、モーメント条件を満たすパラメターが複数存在することとなる。これを**過小識別** (under-identification) と呼ぶ。

■ケース 2：$L = K$　モーメント条件数とパラメター数が同じであるため、一定の条件のもとで、モーメント条件を満たすパラメターを一意に特定することができる。これを**丁度識別** (just-identification) と呼ぶ。 具体的には、以下を満たすパラメターが操作変数推定量となる。

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

■ケース 3：$L > K$　この場合、満たすべき等式の数 (モーメント条件の数) がパラメター数よりも多いため、すべての等式を満たすようなパラメターを見つけることができない。このケースを**過剰識別** (over-identification) と呼ぶ。 このケースを考慮した推定方法が、以下で導入する GMM である。

3　線形モデルにおける GMM

3.1　GMM 推定量

以下の線形モデルを考えよう。

$$y_i = x_i'\beta + e_i, \quad E[z_i e_i] = 0.$$

モーメント条件は以下となる。

$$0 = E[z_i(y_i - x_i'\beta)] \equiv E[g_i(\beta)] \equiv \begin{pmatrix} E[g_{1i}(\beta)] \\ E[g_{2i}(\beta)] \\ \vdots \\ E[g_{Li}(\beta)] \end{pmatrix}.$$

このモーメント条件を標本のモーメント (標本平均) で置き換えたものは以下となる。

$$\bar{g}(b) \equiv \frac{1}{n}\sum_{i=1}^{n} g_i(b) = \frac{1}{n}(z_1, \ldots, z_n)(Y - Xb) = \frac{1}{n}Z'(Y - Xb) = 0.$$

ここで Y, X, Z は以下のように行列を用いて定義される。

$$\underbrace{Y}_{(n\times 1)} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \underbrace{X}_{(n\times K)} = \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{K1} \\ \vdots & \ddots & \vdots \\ x_{1n} & \cdots & x_{Kn} \end{pmatrix},$$
$$\underbrace{Z}_{(n\times L)} = \begin{pmatrix} z_1' \\ \vdots \\ z_n' \end{pmatrix} = \begin{pmatrix} z_{11} & \cdots & z_{L1} \\ \vdots & \ddots & \vdots \\ z_{1n} & \cdots & z_{Ln} \end{pmatrix}.$$

さて、すでに述べたように、$L > K$ の場合 (過剰識別) には、$\bar{g}(b) = 0$ を満たすような b は存在しない。そこで、モーメント条件がなるべくゼロに近づくようなパラメターを推定値とするというのが、GMM の考え方となる。

GMM 推定量を、以下の最小化問題を解くようなパラメターとして定義しよう。

$$\min_b Q_n(b) = \bar{g}(b)' \hat{W} \bar{g}(b).$$

ここで、$\hat{W}$ は荷重行列 (weighting matrix) と呼ばれ、 $(L \times L)$ 行列で、対称かつ正定値 (positive definite) である。この荷重行列は、L 個あるモーメント条件に重み付けを置く役割をもつ。この点については 3.3 項で説明する。

上記の線形モデルでは、目的関数は以下となる。

$$Q_n(b) = \left(\frac{1}{n}Z'(Y - Xb)\right)' \hat{W} \left(\frac{1}{n}Z'(Y - Xb)\right).$$

行列微分を用いることで 1 階条件は以下のように与えられる。

$$
\begin{aligned}
0 &= \left.\frac{\partial Q_n(b)}{\partial b}\right|_{b=\hat{\beta}} \\
&= \left(-\frac{1}{n}X'Z\right)\left(\hat{W}+\hat{W}'\right)\left(\frac{1}{n}Z'(Y-X\hat{\beta})\right) \\
&= -2\left(\frac{1}{n}X'Z\right)\hat{W}\left(\frac{1}{n}Z'(Y-X\hat{\beta})\right).
\end{aligned}
$$

この式を解くことで、GMM 推定量は以下の形で与えられる。

$$
\hat{\beta}_{GMM} = \left(X'Z\hat{W}Z'X\right)^{-1}X'Z\hat{W}Z'Y.
$$

3.1.1　丁度識別の場合 $(L=K)$

もし $L=K$ のとき、荷重行列 $\hat{W}$ はキャンセルされることがわかる。$Z'X$ が正方行列であり、かつ正則であるならば

$$
\begin{aligned}
\hat{\beta}_{GMM} &= \left(Z'X\right)^{-1}\hat{W}^{-1}\left(X'Z\right)^{-1}(X'Z)\hat{W}Z'Y \\
&= (Z'X)^{-1}Z'Y.
\end{aligned}
$$

これは、操作変数推定量と一致する。また、$Z=X$、すなわち内生性問題がない場合には、$\hat{\beta}_{GMM}$ は OLS 推定量 $\hat{\beta}=(X'X)^{-1}X'Y$ と一致する。

3.2　GMM 推定量の性質

一定の条件のもとで、GMM 推定量は以下の性質を持つことが示される (詳細は本資料の冒頭で挙げた教科書を参照)。

- 一致性: $\hat{\beta}_{GMM} \xrightarrow{p} \beta$.
- 漸近正規性

$$
\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} N\left(0, \left(Q'WQ\right)^{-1}Q'W\Omega WQ\left(Q'WQ\right)^{-1}\right),
$$

ここで、$Q=E[z_ix_i']$, $\Omega=E\left[z_iz_i'e_i^2\right]$, $W=\underset{\mathrm{n}\to\infty}{\mathrm{plim}}\,\hat{W}$ である。

3.3 荷重行列 $\hat{W}$ の選び方

上で導出した漸近分布は荷重行列 W に依存している。ここでは荷重行列 $\hat{W}$ のうまい選び方について考えよう。理論的な詳細は省くが、荷重行列を $g_i(\beta) = z_i e_i$ の分散共分散行列の逆行列にすることで、漸近分散が小さくなり、パラメターをより効率的に推定することができる。

$$W = \Omega^{-1} = E\left[z_i z_i' e_i^2\right]^{-1}.$$

この荷重行列は**最適荷重行列** (optimal weighting matrix) と呼ばれる。

しかしながら、行列 Ω は母集団のモーメントであるため推定する必要がある。ここでは、Ω の推定量として以下の $\hat{\Omega}$ を考えよう

$$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} z_i z_i' \hat{e}_i^2.$$

ここで、$\hat{e}_i$ は GMM 推定値から得られた残差である。$\hat{\Omega}$ は Ω に対する一致推定量であることが示される。この推定量を用いて、$\hat{W} = \hat{\Omega}^{-1}$ とする。

3.4 ２段階 GMM 推定量

以上のアイデアを実行するためには、残差 $\hat{e}_i$ を取得する必要がある。その実行方法として、以下のような 2 **段階 GMM 推定**と呼ばれる方法がある。

ステップ 1　荷重行列を単位行列、すなわち $W = I$ として GMM を行い、GMM 推定量 $\hat{\beta}_{GMM,1st}$ およびそれに基づく残差 $\hat{e}_i = y_i - x_i'\hat{\beta}_{GMM,1st}$ を計算する。その上で、$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} z_i z_i' \hat{e}_i$ を構築する。なお、$\hat{\Omega} \xrightarrow{p} \Omega$ である。

ステップ 2　荷重行列を $\hat{W} = \hat{\Omega}^{-1}$ とし、再び GMM 推定を行う。

２段階目で得られた推定量を**漸近的に効率的な GMM 推定量** (もしくは**最適 GMM 推定量**) $\hat{\beta}_{GMM,optimal}$ と呼ぶ。

一定の条件下で、$\hat{\beta}_{GMM,optimal}$ の漸近分布は以下のように得られる。

$$\sqrt{n}\left(\hat{\beta}_{GMM,optimal} - \beta\right) \xrightarrow{d} N\left(0, \left(Q'\Omega^{-1}Q\right)^{-1}\right).$$

ここで得られた漸近分散行列 $\left(Q'\Omega^{-1}Q\right)^{-1}$ は、通常の GMM 推定量で得られるのものよりも小さい。

具体的には、どのような対称かつ正定値行列 W についても、

$$\left(Q'WQ\right)^{-1}Q'W\Omega WQ\left(Q'WQ\right)^{-1}\geq\left(Q'\Omega^{-1}Q\right)^{-1},$$

が「行列の意味で」成立する [*1]。

3.5 特殊ケース：2 段階最小 2 乗法

ある条件のもとで、2 段階最小 2 乗法が漸近的に効率的な GMM の推定量と一致することを示そう。

誤差項について均一分散 (homoskedasticity) を仮定しよう。これは $E[e_i^2|z_i]=\sigma^2$ を意味し、このとき $\Omega=\sigma^2E[z_iz_i']$ となる。したがって、$\hat{W}=\left(\frac{1}{n}Z'Z\right)^{-1}$ とセットすることが、最適な荷重行列となり、このとき

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

となる。

これは、2 段階最小 2 乗法推定量と同じになる。というのも、2 段階最小 2 乗法推定量 $\hat{\beta}_{2SLS}$

$$\hat{\beta}_{2SLS}=\left(\hat{X}'\hat{X}\right)^{-1}\hat{X}'Y$$

として与えられる。ここで $\hat{X}$ は一段階目の最小二乗法から得られた説明変数 (の行列)X の予測値であり、

$$\hat{X}=Z\left(Z'Z\right)^{-1}Z'X$$

である。この式を $\hat{\beta}_{2SLS}$ の右辺に代入することで、上で導出した均一分散における漸近的に効率的な GMM 推定量と一致することがわかるであろう。

$$\hat{\beta}_{2SLS}=\left(X'Z\left(Z'Z\right)^{-1}Z'X\right)^{-1}X'Z\left(Z'Z\right)^{-1}Z'Y$$

[*1] A,B が正方行列とする。$A-B$ が正定値行列であるとき、行列の意味で $A\geq B$ が成立すると定義する。

4　一般ケースにおける GMM

以上のモデルはモーメント条件がパラメターに関して線形の形で書けることを仮定していたが、ここではモーメント条件がパラメターに関して非線形に依存するケースを含んだ一般形について考えよう。

モーメント条件を以下のように与える。

$$E[g(V_i;\theta_0)] \equiv E[g_i(\theta_0)] = 0.$$

ここで $\theta \in \mathbb{R}^K$, g は $(L \times 1)$ 要素、そして V_i はすべての変数を含んでいる。また $L \geq K$ とする。

GMM 推定量は上と同様の形で定義される。

$$\hat{\theta} = \arg\min_{\theta} Q_n(\theta) = \bar{g}(\theta)'\hat{W}\bar{g}(\theta).$$

なお、この場合は $\hat{\theta}$ について解析的に求まるとは限らず、数値計算によって数値解を探す必要がある。

この一般ケースにおいても同様に一致性と漸近正規性が得られ、漸近分布は以下のように与えられる。

$$\sqrt{n}(\hat{\theta}_{GMM} - \theta_0) \xrightarrow{d} N\left(0, \left(G'WG\right)^{-1} G'W\Omega WG\left(G'WG\right)^{-1}\right).$$

ここで、

$$G = E\left[\left.\frac{\partial g_i(\theta)}{\partial \theta'}\right|_{\theta=\theta_0}\right]$$

であり、$\Omega = E[g_i(\theta_0)g_i(\theta_0)']$ である。

なお、荷重行列の選択については、上の議論と同様に $\hat{W} = \Omega^{-1}$ となる。2 段階 GMM を用いることで、漸近的に効率的な GMM 推定量が得られる。