2010 年度 学部 エコノメトリックス
第 26 回講義メモ
(訂正版)

2011 年 1 月 19 日

§ 5.6 一般化モーメント法 (GMM)

一般化モーメント法 (GMM: Generalized Method of Moments) には、狭義と広義の二つの捉え方があり、その範囲に違いがある。狭義の GMM は、識別過剰の非線形操作変数法 (nonlinear IV with overidentification) を指すのに対し、広義の GMM は一般化したモーメント法 (genelazation of MM (OLS, GLS, NLS, MM just identification / over identification)) を指す。ここでは狭義の GMM について扱うが、非線形回帰モデルの推定はこの授業の範囲外であるので、線形回帰モデルの GMM 推定に限定して説明する。

1) 線形回帰モデルの GMM 推定

次のモデルを考えよう。

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$$
$$\text{where} \quad \mathrm{E}(\boldsymbol{\epsilon}) = \boldsymbol{0}, \mathrm{Var}(\boldsymbol{\epsilon}) = \mathrm{E}(\boldsymbol{\epsilon\epsilon}') = \boldsymbol{\Omega}$$

操作変数 $\boldsymbol{W}$ 行列は $n \times l$ とであが、$n > l \geq k$ であるものとする。$l > k$ のとき IV は $\boldsymbol{W}\underbrace{\boldsymbol{J}}_{l\times k}$ となる。

このとき直交条件は

$$\mathrm{E}(\boldsymbol{J}'\boldsymbol{W}'\boldsymbol{\epsilon}) = \boldsymbol{0}$$

で表されるが、標本モーメントでは

$$\frac{1}{n}\boldsymbol{J}'\boldsymbol{W}'\boldsymbol{\epsilon} = \boldsymbol{0}$$

となるから、これより

$$\boldsymbol{J}'\boldsymbol{W}'(\boldsymbol{y} - \boldsymbol{X\beta}) = \boldsymbol{0}$$

と書き換えることができる。ゆえに $\boldsymbol{\beta}$ の MME は

$$\begin{aligned}\boldsymbol{b}_{MM} &= (\boldsymbol{J}'\boldsymbol{W}'\boldsymbol{X})^{-1}\boldsymbol{J}'\boldsymbol{W}'\boldsymbol{y} \\ &= \boldsymbol{\beta} + (\boldsymbol{J}'\boldsymbol{W}'\boldsymbol{X})^{-1}\boldsymbol{J}'\boldsymbol{W}'\boldsymbol{\epsilon}\end{aligned}$$

として得られ、漸近分布を考えると、

$$\sqrt{n}(\boldsymbol{b}_{MM} - \boldsymbol{\beta}) = \left(\frac{1}{n}\boldsymbol{J}'\boldsymbol{W}'\boldsymbol{X}\right)^{-1}\frac{1}{\sqrt{n}}\boldsymbol{J}'\boldsymbol{W}'\boldsymbol{\epsilon}$$
$$\mathrm{Var}[\sqrt{n}(\boldsymbol{b}_{MM} - \boldsymbol{\beta})] = \left(\frac{1}{n}\boldsymbol{J}'\boldsymbol{W}'\boldsymbol{X}\right)^{-1}\frac{1}{n}\boldsymbol{J}'\boldsymbol{W}'\boldsymbol{\Omega}\boldsymbol{W}\boldsymbol{J}\left(\frac{1}{n}\boldsymbol{J}'\boldsymbol{W}'\boldsymbol{X}\right)^{-1'}$$

$\boldsymbol{J}$ として、$\boldsymbol{b}_{MM}$ の分散を最小にするものを採用すると、

$$\begin{aligned}\boldsymbol{J} &= \underset{\boldsymbol{J}}{\mathrm{argmin}}\{\mathrm{Var}[\sqrt{n}(\boldsymbol{b}_{MM} - \boldsymbol{\beta})]\} \\ &= (\boldsymbol{W}'\boldsymbol{\Omega}\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{X}\end{aligned}$$

となり、$\boldsymbol{\beta} = \boldsymbol{\beta}_0$（真値）、$\boldsymbol{\Omega} = \boldsymbol{\Omega}_0$（真値）用いて表現すると

$$\boldsymbol{b}_{GMM} = (\boldsymbol{X}'\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{\Omega}_0\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{\Omega}_0\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{y}$$

このとき

$$\mathrm{Var}[\sqrt{n}(\boldsymbol{b}_{GMM} - \boldsymbol{\beta}_0)] = \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{\Omega}_0\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{X}\right)^{-1}$$

となって、効率的な GMM 推定量 (efficient GMME) になる。このように efficient な GMM 推定量を得るためには、$\boldsymbol{\Omega}_0$ の情報が必要であることがわかる。

一般には $\boldsymbol{\Omega}_0$ のかわりに $\boldsymbol{\Omega}_0$ の一致推定量 $\hat{\boldsymbol{\Omega}}$ を代入して GMME を得る。

$\boldsymbol{\Omega}$の構造が既知 $\rightarrow$ パラメトリックに推定
$\boldsymbol{\Omega}$の構造が未知 $\rightarrow$ ノンパラメトリックに推定
HCCME, HAC, etc.

$\boldsymbol{\Omega}$ を推定するのではなく $\boldsymbol{W}'\boldsymbol{\Omega}\boldsymbol{W}$ を推定するところがポイント。

(1) HCCME (Heteroscedastic consistent covariance matrix estimator)

$$\frac{1}{n}\underbrace{\boldsymbol{W}'}_{l\times n}\underbrace{\boldsymbol{\Omega}}_{n\times n}\underbrace{\boldsymbol{W}}_{n\times l} = \frac{1}{n}\sum_{t=1}^{n} e_t^2 \underbrace{\boldsymbol{w}_t}_{l\times 1}\underbrace{\boldsymbol{w}_t'}_{1\times l}$$

ここで、$e_t = (\boldsymbol{e})_t = y_t - \tilde{\boldsymbol{x}}_t'\boldsymbol{b}_{GMM}$ である。

(2) HAC

$$\frac{1}{n}\boldsymbol{W}'\Omega\boldsymbol{W} = \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n} e_t e_s \boldsymbol{w}_t \boldsymbol{w}_s'$$

(2)' Hansen-White estimator

$$\frac{1}{n}\boldsymbol{W}'\boldsymbol{\Omega}\boldsymbol{W} = \hat{\Gamma}(0) + \sum_{j=1}^{p}(\hat{\Gamma}(j) + \hat{\Gamma}'(j))$$

これは HAC の変種で、p より離れたところの自己相関は 0 にしたものである。なお、$\Gamma(j)$ はベクトル自己共分散関数であり、その推定量は

$$\hat{\Gamma}(j) = \frac{1}{n}\sum_{t=j+1}^{n} e_t e_{t-j} \boldsymbol{w}_t \boldsymbol{w}_{t-j}'$$

として得ることができる。

2) GMM criterion function **とそれに基づく検定**

GMM criterion function とは一体何であろうか。efficient GMM の場合、定義は

$$Q(\boldsymbol{\beta}, \boldsymbol{y}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{W}(\boldsymbol{W}'\Omega_0\boldsymbol{W})^{-1}\boldsymbol{W}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \qquad (5.6.1)$$

で与えられる。ここで

$$\begin{aligned} \text{直交条件} \qquad & \mathrm{E}(\boldsymbol{w}_t\epsilon_t) = \boldsymbol{0} \\ & \Downarrow \\ & \mathrm{E}(\boldsymbol{W}'\boldsymbol{\epsilon}) = \mathrm{E}(\boldsymbol{W}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})) = \boldsymbol{0} \end{aligned}$$

となるが、識別過剰のとき

$$\boldsymbol{W}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) = \boldsymbol{0}$$

となる $\boldsymbol{\beta}$ は unique に存在するのだろうか。

答えを先回りすると存在しない。すなわち、

$$\boldsymbol{W}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \neq \boldsymbol{0}$$

である。したがって、直交条件との**近さ（距離）** を二次形式で表現する。その候補として、

(a) $(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{W}\boldsymbol{W}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$: l 次元空間における $\underbrace{\boldsymbol{W}'}_{l\times n}(\underbrace{\boldsymbol{y}}_{n\times 1} - \boldsymbol{X}\boldsymbol{\beta})$ のノルムの 2 乗

(b) $(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{W}\boldsymbol{A}^{-1}\boldsymbol{W}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$: $\boldsymbol{A}^{-1}$ を使った二次形式、すなわち、ウエイト $\boldsymbol{A}$ （どの軸に対してウエイトを大きくするか）で重み付けた“距離”

などが考えられる。

$\boldsymbol{W}'\boldsymbol{\epsilon}$ の原点からの近さ（距離）のウエイトを考える際、$\boldsymbol{W}'\boldsymbol{\epsilon}$ の分散共分散行列は一つの基準となる。

$$\mathrm{Var}(\boldsymbol{W}'\boldsymbol{\epsilon}) = \mathrm{E}(\boldsymbol{W}'\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\boldsymbol{W}) = \boldsymbol{W}'\boldsymbol{\Omega}\boldsymbol{W}$$

であるから、(5.6.1) 式の criterion function は、分散共分散行列で距離を基準化してることになる。

ところで、$\boldsymbol{\beta} = \boldsymbol{\beta}_0$ のとき

$$\mathrm{E}(\boldsymbol{W}'\boldsymbol{\epsilon}) = \boldsymbol{0}$$

であるから、$\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Omega}_0)$ のとき

$$Q(\boldsymbol{\beta}_0, \boldsymbol{y}) \sim \chi^2(l)$$

が成立する。そこでこの関係を用いれば、過剰識別の検定が構成できる。

$$Q(\boldsymbol{b}_{GMM}, \boldsymbol{y}) = \boldsymbol{e}'\boldsymbol{W}(\boldsymbol{W}'\hat{\boldsymbol{\Omega}}\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{e}$$

を考えると、

$$Q(\boldsymbol{b}_{GMM}, \boldsymbol{y}) \xrightarrow{d} \chi^2(l-k) \qquad \text{under } H_0$$

となる。