

条件付きモーメント制約モデルの幾何学

後 藤 文 廣

目次

- 1 はじめに
- 2 セミパラメトリック効率性限界
- 3 最適 GMM とワンステップ法
- 4 条件付き経験尤度法
- 5 可算無限近似法
- 6 幾何学的考察

1 はじめに

データ系列 $\{z_i\}_{i=1}^n$ を独立同一分布 (independent and identically distributed, i.i.d.) の p 次元確率変数ベクトルとして、以下のモデルを考える。

$$Eg(z_i; \theta_0) = 0 \quad (1.1)$$

ここで $g: \mathbb{R}^p \times \Theta \mapsto \mathbb{R}^r$ は既知の r 次元ベクトル値関数、 Θ はパラメータ空間、 θ は q 次元パラメータベクトル、 θ_0 はパラメータの真の値とする。(1.1) の制約をモーメント条件 (moment conditions) または直交条件 (orthogonality conditions)、またこのモデルをモーメント制約モデル (moment restriction model) と呼ぶ。

このモデルを推定する方法として Hansen (1982) によって提唱され、近年著しい発展を遂げたのが一般化モーメント法 (generalized method of moments, GMM) である。GMM 全般については計量経済学の標準的な教科書、例えば、Hamilton (1994), Hayashi (2000) などを参照されたい。

また同じくこのモデルを推定する方法として Owen (1988) によって提唱されたのが経験尤度法 (empirical likelihood, EL) である。Newey and Smith (2004) によって提唱された一般化経験尤度法 (generalized empirical likelihood, GEL) はそれを拡張したものである。これらの推定量は GMM 推定量と漸近的に同等になる。ただし小標本では GMM よりも優れた特性をもつことが知られている。EL 全般については Owen (2001) などを参照のこと。

一方、条件付きモーメント制約モデル (conditional moment restriction model) は次のように定義される。モーメント条件は、 $z_i = (x_i', y_i')'$ として、条件付き期待値

$$E[g(z_i; \theta_0) | x_i] = 0 \quad (1.2)$$

を用いて表される。なお前述したモーメント制約モデルを条件付きモーメント制約モデルと対比させる場合には、無条件モーメント制約モデル (unconditional moment restriction model) と呼ぶこともある。

Chamberlain (1987) は、条件付きモーメント制約モデルを推定する方法として、最適な (optimal) 操作変数行列を用いた GMM 推定量を提唱し、それがセミパラメトリック効率性の限界を達成することを示した。ただしそのためには操作変数行列の推定が必要になる。これを受けて Newey (1990) は実行可能な推定手順を示した。

Kitamura et al. (2004) は経験尤度の考えを条件付きモーメントモデルに拡張し、条件付き経験尤度法を提唱した。また後年になって、Severini and Wong (1992) の条件付きパラメトリックモデル (conditionally parametric model) の考え方を適用した一種の経験尤度法が、Sueishi (2016) により提案されている。

その他有力な方法として、Donald et al. (2003) による、標本の大きさ、別名サンプルサイズ (sample size) と共に増大する操作変数列を用いた推定法、また Ai and Chen (2003) や Newey and Powell (2003) によるシーブ (sieve) 法などがある。

この分野は理論面でも発展を続けている。例えば Sueishi (2016) はセミパラメトリック効率性の限界の新しい求め方を提示した。また Chen and Santos (2018)

は過剰識別に関する美しい定理を発見した。

条件付きモーメント制約モデルの簡単な例として、以下の線形回帰モデルを考える。

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

条件付きモーメント制約が

$$E[u_i | x_i] = E[y_i - x_i' \beta_0 | x_i] = 0 \quad (1.3)$$

であるとする。(1.3) ならば次の関係が成り立つ。

$$Ex_i u_i = Ex_i(y_i - x_i' \beta_0) = 0 \quad (1.4)$$

この場合、 x_i が操作変数行列であり、無条件モーメント制約 (1.4) に対応するパラメータ β の GMM 推定量は最小 2 乗 (OLS) 推定量である。この推定量の漸近分散共分散行列 V_1 は

$$V_1 = (Ex_i x_i')^{-1} E E[u_i^2 | x_i] x_i x_i' (Ex_i x_i')^{-1}$$

である。一方、(1.3) に対応する最適な GMM 推定量は

$$Ex_i(E[u_i^2 | x_i])^{-1} u_i = Ex_i(E[u_i^2 | x_i])^{-1} (y_i - x_i' \beta_0) = 0$$

を無条件モーメント制約として求められる。この場合 $x_i(E[u_i^2 | x_i])^{-1}$ が Chamberlain (1987) の言う最適な操作変数行列であり、これは実は一般化最小 2 乗 (GLS) 推定量に一致している。この推定量の漸近分散共分散行列 V_2 は

$$V_2 = (E(E[u_i^2 | x_i])^{-1} x_i x_i')^{-1}$$

と計算される。簡単な行列計算により $V_2 \leq V_1$ が示される。さらに V_2 はモデル (1.3) のパラメータ β の推定に関するセミパラメトリック効率性の限界となっている。つまり最適な GMM 推定量は (1.3) に対応する有効推定量である。

後藤 (2022) ではセミパラメトリック推測の解説の一環としてモーメント制約モデルについて論じ、条件付きモーメント制約モデルについても少し触れている。本論文は後藤 (2022) の続編として、条件付きモーメント制約モデルへのセ

ミパラメトリック理論の適用について論じることとしている。その際、モデルの背後に存在する幾何学的構造に着目し、直観的な理解に重点を置く。筆者の知る限りでは、条件付きモーメント制約モデルについて、幾何学的観点からアプローチした日本語のまとまった解説書はまだ無いようである。これが読者の計量経済学研究の一助となれば幸いである。

本稿の構成は以下の通りである。第 2 節では条件付きモーメント制約モデルのセミパラメトリック効率性限界を 3 通りの方法で求める。第 3 節では最適操作変数を用いた GMM とワンステップ法による有効推定量について論じる。第 4 節では条件付き経験尤度法による推定と仮説検定を紹介する。第 5 節は、標本の大きさと共に増大する基底関数列を用いた 2 つの代表的な推定法について論じる。最後に第 6 節では、モーメント関数の選択や過剰識別の問題など、モデルのもつ様々な興味深い特性について幾何学的な考察を行う。

2 セミパラメトリック効率性限界

セミパラメトリック効率性限界を求めるために、(1.2) のモデルをもう少し詳しく記述しよう。 $z = (x', y')'$ として、 $F(z)$ は未知の p 変数分布関数で、 $\mathbb{R}^p$ 上の Lebesgue 測度に関し密度関数 $f(z)$ をもつ。 (θ, F) をパラメータとして制約は

$$E_F[g(z; \theta)|x] = 0$$

また同等の表現として

$$\int g(z; \theta)f(z)dy = 0$$

である。 (θ_0, F_0) を真の値とし、真の値での期待値作用素 E_{F_0} を単に E で表す。さらに $\Omega(x) = E[g(z; \theta_0)g(z; \theta_0)'|x]$ および $G(x) = E[\partial g(z; \theta_0)/\partial \theta'|x]$ とおく。 $\Omega(x)$ は非特異とする。また観測データ $z_1, \dots, z_n$ は i.i.d. に従う。

条件付きモーメント制約モデルの効率性限界を 3 通りの方法で求めよう。

2.1 Qin and Lawless (1994) の方法

まず Qin and Lawless (1994) を発展させた方法を使う。パラメトリックサブモデル f_t とその関数 $\theta_t = \theta(f_t)$ を考える。次の制約がある。

$$\int g(z; \theta_t) f_t(z) dy = 0$$

両辺を t で微分して $t = 0$ で評価すると

$$G(x)\dot{\theta}_0 + E[gh|x] = 0 \quad (2.1)$$

ここで $h(z) = \partial \log f_t / \partial t|_{t=0}$ で、 h はスコア関数である。また g は $g(z; \theta_0)$ の省略形である。

接空間 $T(P_0)$ を、ある q 次元ベクトル c が存在して

$$E[gh|x] = G(x)c \quad (2.2)$$

を満たすような関数 $h(z) \in L_2^0(P_0)$ の集合と定義する。明らかに $T(P_0)$ はベクトル空間になる。また h に対応する 1 次元サブモデル $f_t = (1 + ht)f_0$ が存在するので $T(P_0)$ は確かに接空間となる。

ここで $k = (k_1, \dots, k_r)'$ かつ $k_i \in L_2^0(P_0)$ ($i = 1, \dots, r$) のとき $\{k\} = \{c'k : c \in \mathbb{R}^r\} \subset L_2^0(P_0)$ と定義する。また $c(x) \in C$ を x の r 次元ベクトル関数とする。このとき $\{k\}_x = \{c(x)'k : c(x) \in C\} \subset L_2^0(P_0)$ と定義する。ただし $E[k|x] = 0$ とする。

なお $L_2^0(P_0)$ や $T(P_0)$ などセミパラメトリック理論における基本的な概念は、後藤 (2022) を参照のこと。

補題 2.1 $\{g\}_x^\perp$ を $L_2^0(P_0)$ における $\{g\}_x$ の直交補空間とすると、次の関係が成立する。

$$T(P_0) = \{g\}_x^\perp \oplus \{G(x)'\Omega(x)^{-1}g\}$$

証明 まず明らかに $\{g\}_x^\perp \perp \{G(x)'\Omega(x)^{-1}g\}$ である。(2.2) の定義式より

$$T(P_0) \supset \{g\}_x^\perp \oplus \{G(x)'\Omega(x)^{-1}g\}$$

が成り立つ。次に任意の $h(z) \in T(P_0)$ について

$$\begin{aligned}
 h - \Pi(h|\{g\}_x^\perp) &= \Pi(h|\{g\}_x) \\
 &= E[hg'|x]\Omega(x)^{-1}g \\
 &= c'G(x)'\Omega(x)^{-1}g \\
 &\in \{G(x)'\Omega(x)^{-1}g\}
 \end{aligned}$$

を得る。ここで 3 番目の等号は (2.2) から出る。よって

$$T(P_0) \subset \{g\}_x^\perp \oplus \{G(x)'\Omega(x)^{-1}g\}$$

が成り立つ。 ■

任意の $r \times r$ 行列関数 $W(x)$ で $EG(x)'W(x)G(x)$ が非特異になるものを使って、(2.1) の両辺に左から $G(x)'W(x)$ を掛けて期待値をとれば

$$EG(x)'W(x)G(x)\dot{\theta}_0 + EG(x)'W(x)gh = 0$$

となり

$$\dot{\theta}_0 = -[EG(x)'W(x)G(x)]^{-1}EG(x)'W(x)gh$$

を得る。これは θ の影響関数 $\hat{\theta}$ が

$$\hat{\theta} = -[EG(x)'W(x)G(x)]^{-1}G(x)'W(x)g$$

であることを示している。

定理 2.2 θ の効率性限界 V_c は

$$V_c = \left[EG(x)'\Omega^{-1}(x)G(x) \right]^{-1}$$

である。 $F(z)$ の効率性限界 $W_c(z)$ は

$$\begin{aligned}
 W_c(z) &= E\phi^2 - E \left[E[\phi g'|x]\Omega^{-1}(x)E[g\phi|x] \right] \\
 &\quad + E \left[E[\phi g'|x]\Omega^{-1}(x)G(x) \right] V_c E \left[G'(x)\Omega^{-1}(x)E[g\phi|x] \right]
 \end{aligned}$$

である。ただし $\phi(z_i) = 1_{[z_i < z]} - F(z)$ とする。

証明 θ の有効影響関数 $\tilde{\theta}$ は

$$\tilde{\theta} = \Pi(\hat{\theta}|T_P(P_0))$$

$$\begin{aligned}
 &= -\Pi([EG(x)'W(x)G(x)]^{-1}G(x)'W(x)g|\{G(x)'\Omega(x)^{-1}g\}) \\
 &= -[EG(x)'\Omega(x)^{-1}G(x)]^{-1}G(x)'\Omega(x)^{-1}g
 \end{aligned}$$

効率性限界 V_c は

$$V_c = E\tilde{\theta}\tilde{\theta}' = \left[EG(x)'\Omega^{-1}(x)G(x)\right]^{-1}$$

により得られる。有効影響関数 $\tilde{\phi}$ は次のように計算される。

$$\begin{aligned}
 \tilde{\phi} &= \Pi(\phi|T(P_0)) \\
 &= \Pi(\phi|\{g\}_x^\perp) + \Pi(\phi|\{G(x)'\Omega(x)^{-1}g\}) \\
 &= \phi - E[\phi g'|x]\Omega^{-1}(x)g \\
 &\quad + E\left[E[\phi g'|x]\Omega^{-1}(x)G(x)\right]V_cG(x)'\Omega^{-1}(x)g
 \end{aligned}$$

そして効率性限界 $W_c(z)$ は $E\tilde{\phi}^2$ により得られる。 ■

2.2 Chamberlain (1987) の方法

効率性限界を求める第2の方法として、Chamberlain (1987) の多項分布近似を用いた方法と、それを発展させた Goto (2021a) の方法がある。

Goto (2021a) では無条件モーメント制約モデルの効率性限界を求める方法を提示した。設定は

$$\int g(z; \theta)f(z)dz = 0$$

のモーメント条件であるが、分布関数として $F(z; \theta, \eta)$ を考える。これが多項分布近似の大きな特長である。 F はパラメータ (θ, η) 以外は既知の関数とする。 η は有限または無限次元の局外パラメータである。 (θ_0, η_0) における θ と η のスコア関数をそれぞれ ℓ_θ および a とする。通常 $\ell_\theta = \partial \log f(z; \theta_0, \eta_0)/\partial \theta$ と計算される。また $f(z; \theta_0, \eta_t)$ を (θ_0, η_0) における η の1次元サブモデルとして、 $a = \partial \log f(z; \theta_0, \eta_t)/\partial t|_{t=0}$ と定義する。 η の接空間を A で表す。

Chamberlain (1987) によれば、多項分布の下では条件付きモーメント制約は、有限個の無条件モーメント制約で表現される。よって Chamberlain (1987) の多

項分布近似は条件付きモーメント制約モデルにも適用でき、結局

$$\int g(z; \theta) f(z; \theta, \eta) dy = 0 \quad (2.3)$$

を考えればよい。(2.3) の両辺を微分して

$$E[g\ell'_\theta|x] = -E\left[\frac{\partial g(z; \theta_0)}{\partial \theta'}\Big|_x\right] \quad (2.4)$$

および

$$E[ga|x] = 0 \quad (2.5)$$

を得る。

(2.5) は $L_2^0(P_0)$ において a が $c(x)'g(z|\theta_0)$ の形式のすべての関数と直交していることを意味する。ただし $c(x)$ は r 次元のベクトル値関数である。それらの関数の集合を $\{g\}_x$ で表す。 $\Omega(x) = E[gg'|x]$ および $G(x) = E[\partial g(z; \theta_0)/\partial \theta'|x]$ とおく。有効スコア関数 $\tilde{\ell}_\theta$ は

$$\tilde{\ell}_\theta = \Pi(\ell_\theta|A^\perp)$$

で得られる。(2.5) より $\{g\}_x = A^\perp$ なので

$$\tilde{\ell}_\theta = \Pi(\ell_\theta|\{g\}_x) = E[\ell_\theta g'|x]\{E[gg'|x]\}^{-1}g = -G'(x)\Omega^{-1}(x)g$$

となる。ここで 3 番目の等号は (2.4) により成り立つ。効率性限界 V_c は $[E\tilde{\ell}_\theta\tilde{\ell}'_\theta]^{-1}$ から得られ、

$$V_c = \left[EG(x)'\Omega^{-1}(x)G(x)\right]^{-1}$$

となる。これは定理 2.2 の前段と一致する。

接空間 $T(P_0)$ に関する補題 2.1 も同様に証明される。つまり

$$\begin{aligned} T(P_0) &= A + \{\ell_\theta\} \\ &= A \oplus \{\tilde{\ell}_\theta\} \\ &= \{g\}_x^\perp \oplus \{G(x)'\Omega(x)^{-1}g\} \end{aligned}$$

以上で示された。またこのことから定理 2.2 の後段も証明される。

2.3 Sueishi (2016) の方法

次に Sueishi (2016) の手法を使って最も不利なサブモデルを求めてみよう。 $\eta = \eta_{Y|X}\eta_X$ は $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ 上の分布関数で、その集合を H で表す。また η_X および $\eta_{Y|X}$ は、それぞれ $\mathcal{X}$ 上および $\mathcal{Y}$ 上の Lebesgue 測度に関し密度関数 $h(x)$ および条件付き密度関数 $h(y|x)$ をもつとする。さらに

$$\mathbf{Q}_\theta = \left\{ Q = Q_{Y|X}Q_X \in H : \int g(z; \theta) dQ_{Y|X}(y|x) = 0 \text{ almost surely } Q_X \right\}$$

と定義する。つまり $\mathbf{Q}_\theta$ は、 θ が与えられた下で条件付きモーメント制約を満たすような分布の集合を表している。 θ と η が与えられたとき、 $F_{\theta, \eta}(z) \equiv F(z; \theta, \eta)$ を次の条件を満たす分布として定義する。

$$- \int \log \frac{dF_{\theta, \eta}}{d\eta}(z) d\eta(z) = \min_{Q \in \mathbf{Q}_\theta} - \int \log \frac{dQ}{d\eta}(z) d\eta(z) \quad (2.6)$$

$F_{\theta, \eta}$ は、Kullback-Leibler ダイバージェンスに関する η の $\mathbf{Q}_\theta$ への射影として定義される。(2.6) の最小化問題を解いて、 $F_{\theta, \eta}$ を具体的に求めてみよう。まず (2.6) の右辺の目的関数は

$$- \int \int \log \frac{dQ_{Y|X}}{d\eta_{Y|X}}(y|x) d\eta_{Y|X}(y|x) d\eta_X(x) - \int \log \frac{dQ_X}{d\eta_X}(x) d\eta_X(x)$$

なので、これを最小にするような条件付き分布 $Q_{Y|X}$ と周辺分布 Q_X を求めればよい。第2項については、Jensen の不等式を使えば $Q_X = \eta_X$ が解となる。第1項については、各 x について

$$- \int \log \frac{dQ_{Y|X}}{d\eta_{Y|X}}(y|x) d\eta_{Y|X}(y|x)$$

を最小にするような $Q_{Y|X}(y|x)$ を求めればよい。ただし、 $Q_{Y|X}(y|x)$ は条件付きモーメント制約

$$\int g(z; \theta) dQ_{Y|X}(y|x) = 0$$

を満たさなければならない。また $Q_{Y|X}$ は $\mathcal{Y}$ 上の Lebesgue 測度に関し条件付き密度関数 $q(y|x)$ をもつとする。Lagrange 乗数 λ , t を用いてこれは次式の最大化問題に置き換えられる。

$$L = \int (\log q(y|x))h(y|x)dy - \lambda' \int g(z; \theta)q(y|x)dy \\ - t \left(\int q(y|x)dy - 1 \right)$$

これを解くと, $q(y|x)$ の任意の摂動 $\delta q(y|x)$ について

$$\int \frac{\delta q(y|x)}{q(y|x)} h(y|x)dy - \lambda' \int g(z; \theta) \delta q(y|x)dy - t \int \delta q(y|x)dy = 0$$

より,

$$\frac{h(y|x)}{q(y|x)} - \lambda' g(z; \theta) - t = 0 \quad (2.7)$$

が成立する。(2.7) の両辺に $q(y|x)$ を掛けて積分 ($\int \cdot dy$) して整理すれば, $t = 1$ を得る。よって

$$q(y|x) = \frac{h(y|x)}{1 + \lambda' g(z; \theta)}$$

かつ

$$\int g(z; \theta)q(y|x)dy = \int \frac{g(z; \theta)h(y|x)}{1 + \lambda' g(z; \theta)}dy = 0$$

となる。同時に

$$\int q(y|x)dy = \int h(y|x)dy - \lambda' \int \frac{g(z; \theta)h(y|x)}{1 + \lambda' g(z; \theta)}dy = 1$$

も成立する。 $\mathcal{Z}$ 上の Lebesgue 測度に関する $F_{\theta, \eta}$ の密度関数を $f_{\theta, \eta}$ とすれば, 最終的に

$$f(z; \theta, \eta) = \frac{h(z)}{1 + \lambda(x; \theta, \eta)' g(z; \theta)} \quad (2.8)$$

かつ

$$\int g(z; \theta)f(z; \theta, \eta)dy = \int \frac{g(z; \theta)h(z)}{1 + \lambda(x; \theta, \eta)' g(z; \theta)}dy = 0 \quad (2.9)$$

が得られる。ただし $h(z) = h(y|x)h(x)$ であり, これは $\eta(z)$ の密度関数になる。 λ は (θ, η) が与えられたとき (2.9) を満たすような調整パラメータと考えられる。 (θ, η) の真の値を (θ_0, η_0) とし, $\eta_0(z)$ の密度関数を $h_0(z)$ とする。 $\eta_0 = F_{\theta_0, \eta_0} = F_0$ とすれば, $h_0(z) = f(z; \theta_0, \eta_0)$ であり, $\lambda(x; \theta_0, \eta_0) = 0$ が成立する。

$f(z; \theta, \eta)$ が真のパラメータ (θ_0, F_0) を推定するための最も不利なサブモデルであることを示そう。 (θ_0, η_0) における θ と η のスコア関数をそれぞれ ℓ_θ および a

とする。通常 $\ell_\theta = \partial \log f(z; \theta_0, \eta_0) / \partial \theta$ と計算される。また $f(z; \theta_0, \eta_t)$ を (θ_0, η_0) における η の 1 次元サブモデルとして、 $a = \partial \log f(z; \theta_0, \eta_t) / \partial t|_{t=0}$ と定義する。 η の接空間を A で表す。モデルの接空間 $T(P_0)$ は

$$T(P_0) = \{\delta' \ell_\theta + a : \delta \in \mathbb{R}^q, a \in A\} = \{\ell_\theta\} + A$$

と表現される。

ℓ_θ を求めるために、まず $\partial \lambda(x; \theta_0, \eta_0) / \partial \theta'$ を計算する。

$$\int \frac{g(z; \theta) h(z)}{1 + \lambda(x; \theta, \eta)' g(z; \theta)} dy = 0$$

の両辺を (θ_0, η_0) において θ で偏微分して整理すれば

$$\frac{\partial \lambda(x; \theta_0, \eta_0)}{\partial \theta'} = \Omega(x)^{-1} G(x) \quad (2.10)$$

を得る。ただし $\Omega(x) = E[g(z; \theta_0) g(z; \theta_0)' | x]$ および $G(x) = E[\partial g(z; \theta_0) / \partial \theta' | x]$ である。

これと (2.8) より

$$\ell_\theta = \frac{\partial \log f(z; \theta_0, \eta_0)}{\partial \theta} = -G(x)' \Omega(x)^{-1} g \quad (2.11)$$

を得る。このサブモデルのスコア関数は、オリジナルモデルの有効スコア関数と一致することが分かる。

次に η の接空間 A が $\{g\}_x^\perp$ に等しいことを示そう。 $A \subset \{g\}_x^\perp$ であることは (2.9) から直ちに分かる。逆に任意の $i(z) \in \{g\}_x^\perp$ に対し $h_t(z) = (1 + ti(z))h_0(z)$ とおく。ただし $h_t(z)$ は $\eta_t(z)$ の密度関数である。このとき (2.9) を満たすための調整パラメータ $\lambda(x; \theta_0, \eta_t) = 0$ なので、 $f(z; \theta_0, \eta_t) = h_t(z)$ であり、

$$a = \left. \frac{\partial \log f(z; \theta_0, \eta_t)}{\partial t} \right|_{t=0} = i(z)$$

を得る。つまり任意の関数 $i(z) \in \{g\}_x^\perp$ に対し、それがスコア関数となるような道 η_t (path) が存在するので $A \supset \{g\}_x^\perp$ が示された。

これと (2.11) を合わせれば、このサブモデルの接空間 $T(P_0)$ は、補題 2.1 で提示されたオリジナルモデルの接空間と一致する。そしてこの $f(z; \theta, \eta)$ が真のパラメータ (θ_0, F_0) を推定するための最も不利なサブモデルであることが分かる。

3 最適 GMM とワンステップ法

3.1 最適 GMM による推定

次に θ の有効推定量の求め方について論じよう。有効スコア関数 $\tilde{\ell}_\theta$ は

$$\tilde{\ell}_\theta = -G'(x)\Omega^{-1}(x)g$$

なので、 $A(x) = -G'(x)\Omega^{-1}(x)$ が既知ならば、操作変数行列 $A(x)$ を使って $g(z; \theta)$ の代わりに $\tilde{\ell}_\theta(z; \theta) = A(x)g(z; \theta)$ をモーメント関数として GMM で推定すれば有効推定量が得られる。実際この GMM 推定量の漸近分散を計算すれば $[EG(x)' \Omega^{-1}(x)G(x)]^{-1}$ を得る。これを最適 (optimal) GMM 推定量と呼ぶ。これは Chamberlain (1987) により最初に示唆されたものである。

ただし 2 つ問題がある。1 点目は $A(x)$ は未知であり、推定しなければならないこと。2 点目は条件付きモーメント制約問題を無条件モーメント制約問題に緩めてしまうことにより、制約を満たす θ_0 が複数存在し、一致推定量にならないことである。後者については Dominguez and Lobato (2004) の指摘による。

3.2 最適 GMM による仮説検定

$A(x)$ の一致推定量 $\hat{A}(x)$ が得られ、

$$EA(x)g(z; \theta) = 0$$

が唯一の解 θ_0 をもつと仮定して、最適 GMM による仮説検定を考えてみよう。 $\theta = (\alpha', \gamma')'$ すなわち、パラメータ θ が q_1 次元ベクトル α と q_2 次元ベクトル γ に分解されるとする。次の帰無仮説を考える。

$$H_0 : \gamma = \gamma_0$$

対立仮説としては、次の局所対立仮説 (local alternatives) を考える。

$$H_1 : \gamma = \gamma_0 + \delta/\sqrt{n}$$

$g_o(z; \theta) = A(x)g(z; \theta)$, $\hat{g}_o(\theta) = n^{-1} \sum_{i=1}^n g_o(z_i; \theta)$ かつ $\Omega_o = E g_o(z; \theta_0) g_o(z; \theta_0)'$ とおく。

制約付き GMM 推定量 $\bar{\theta}$ は

$$\bar{\theta} = \arg \min_{\theta} \hat{g}_o(\theta)' \hat{\Omega}_o^{-1} \hat{g}_o(\theta) \text{ subject to } \gamma = \gamma_0$$

と定義される。ただし $\hat{\Omega}_o$ は Ω_o の初期推定量を表す。一方制約なし GMM 推定量 $\hat{\theta}$ は

$$\hat{\theta} = \arg \min_{\theta} \hat{g}_o(\theta)' \hat{\Omega}_o^{-1} \hat{g}_o(\theta)$$

と定義される。

検定統計量 DM を

$$DM = n \left[\hat{g}_o(\bar{\theta})' \hat{\Omega}_o^{-1} \hat{g}_o(\bar{\theta}) - \hat{g}_o(\hat{\theta})' \hat{\Omega}_o^{-1} \hat{g}_o(\hat{\theta}) \right]$$

とすれば、対立仮説の下で

$$DM \rightsquigarrow \chi^2(q_2, \kappa)$$

が成り立つ。ここで

$$\begin{aligned} \kappa = & \delta' [EK'(x)\Omega(x)^{-1}K(x) \\ & - EK'(x)\Omega(x)^{-1}H(x)(EH'(x)\Omega(x)^{-1}H(x))^{-1} \\ & \times EH'(x)\Omega(x)^{-1}K(x)]\delta \end{aligned} \quad (3.1)$$

かつ $H(x) = E[\partial g(z; \theta_0)/\partial \alpha | x]$, $K(x) = E[\partial g(z; \theta_0)/\partial \gamma | x]$ とする。証明は Goto (2021a) を参照されたい。なお $\chi^2(q_2, \kappa)$ は自由度 q_2 非心度 κ の非心カイ 2 乗分布を表す。 κ は検定の検出力 (power) を示している。

3.3 ワンステップ法

Newey (1990) は、均一分散つまり $\Omega(x) = \Omega$ の仮定の下で、 $A(x)$ をノンパラメトリック回帰を使って推定した。具体的には k -近傍 (k -nearest neighbor) 法とシリーズ (series) 法を使って推定し、比較した。 θ の推定には後述するワンステップ (one-step) 法を使用した。ただし $A(x)$ と θ の推定のいずれも初期推定量 $\bar{\theta}$ が

一様性をもつことが前提であるが, Dominguez and Lobato (2004) が指摘するように自明ではない。

ここでは Dominguez and Lobato (2004) の提案する初期推定量 $\hat{\theta}_{DL}$ を使ったワンステップ推定量を紹介する。

$$\hat{\theta}_{DL} = \arg \min_{\theta \in \Theta} \frac{1}{n^3} \sum_{j=1}^n \left(\sum_{i=1}^n g(z_i; \theta) 1\{x_i \leq x_j\} \right)^2$$

$\hat{\theta}_{DL}$ は漸近的有効性をもたないものの, $\sqrt{n}$ -一様性をもつ。

定理 2.2 より θ の有効影響関数 $\phi(z; \theta_0)$ は

$$\phi(z; \theta_0) = -I(\theta_0)^{-1} G(x)' \Omega(x)^{-1} g(z; \theta_0)$$

である。ここで $I(\theta_0) = EG(x)' \Omega^{-1}(x) G(x)$ とする。これを使ってワンステップ推定量 $\hat{\theta}$ は

$$\begin{aligned} \hat{\theta} &= \hat{\theta}_{DL} + \frac{1}{n} \sum_{i=1}^n \hat{\phi}(z_i; \hat{\theta}_{DL}) \\ &= \hat{\theta}_{DL} - \hat{I}(\theta_0)^{-1} \frac{1}{n} \sum_{i=1}^n \hat{G}(x_i)' \hat{\Omega}(x_i)^{-1} g(z_i; \hat{\theta}_{DL}) \end{aligned}$$

と構成される。ここで $\hat{I}(\theta_0)$, $\hat{G}(x)$ および $\hat{\Omega}(x)$ は, それぞれ $I(\theta_0)$, $G(x)$ および $\Omega(x)$ の一致推定量である。たとえば $\hat{G}(x)$ は $\partial g(z_i; \hat{\theta}_{DL}) / \partial \theta'$ を被説明変数, x_i を説明変数とするノンパラメトリック回帰で求める。これには Nadaraya-Watson 推定などが考えられる。

θ^* を $\hat{\theta}_{DL}$ と θ_0 の中間の値として

$$g(z_i; \hat{\theta}_{DL}) = \frac{\partial g(z_i | \theta^*)}{\partial \theta'} (\hat{\theta}_{DL} - \theta_0) + g(z_i; \theta_0)$$

を代入して整理すれば

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta_0) &= \left(I_q - \hat{I}(\theta_0)^{-1} \frac{1}{n} \sum_{i=1}^n \hat{G}(x_i)' \hat{\Omega}(x_i)^{-1} \frac{\partial g(z_i | \theta^*)}{\partial \theta'} \right) \sqrt{n}(\hat{\theta}_{DL} - \theta_0) \\ &\quad - \hat{I}(\theta_0)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{G}(x_i)' \hat{\Omega}(x_i)^{-1} g(z_i; \theta_0) \end{aligned}$$

となる。右辺の第 1 項は $\hat{\theta}_{DL}$ が $\sqrt{n}$ -一様性をもつので, $(1/n) \sum_{i=1}^n \cdot$ の一様収

束性を前提に, 0 に確率収束する。また第 2 項は確率的同程度連続性および不偏条件の下で

$$-I(\theta_0)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n G(x_i)' \Omega(x_i)^{-1} g(z_i; \theta_0) + o_p(1)$$

と変形され,

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightsquigarrow N(0, [EG(x)' \Omega^{-1}(x) G(x)]^{-1})$$

を得る。すなわちワンステップ推定量 $\hat{\theta}$ は有効推定量である。なお確率的同程度連続性などの概念は後藤 (2022) を参照されたい。

4 条件付き経験尤度法

Kitamura et al. (2004) は経験尤度の考え方を条件付きモーメントモデルに拡張し, $p_{ij} = P\{y = y_j | x = x_i\}$ なる条件付き多項確率を用いて, 以下の条件付き経験尤度最大化問題を定式化した。

$$\begin{aligned} & \max_{p_{ij}} \sum_{i=1}^n \sum_{j=1}^n w_{ij} \log p_{ij} \quad \text{subject to} \\ & p_{ij} \geq 0, \sum_{j=1}^n p_{ij} = 1, \sum_{j=1}^n g(z_j; \theta) p_{ij} = 0, i, j = 1, \dots, n \end{aligned} \quad (4.1)$$

ここで w_{ij} は

$$w_{ij} = \frac{K((x_i - x_j)/h)}{\sum_{j=1}^n K((x_i - x_j)/h)} \equiv \frac{K_{ij}}{\sum_{j=1}^n K_{ij}}$$

で定義されるウエイトであり, $K(\cdot)$ はカーネル関数, h はバンド幅である。

(4.1) の最大化問題を解くために Lagrange 乗数 $t_1, \dots, t_n$ および $\lambda_1, \dots, \lambda_n$ を使えば Lagrange 関数 L は

$$L = \sum_{i=1}^n \sum_{j=1}^n w_{ij} \log p_{ij} - \sum_{i=1}^n t_i \left(\sum_{j=1}^n p_{ij} - 1 \right) - \sum_{i=1}^n \lambda_i' \sum_{j=1}^n g(z_j; \theta) p_{ij}$$

となり, これを解くと

$$\hat{p}_{ij} = \frac{w_{ij}}{1 + \lambda_i' g(z_j; \theta)} \quad (4.2)$$

ここで各 $\theta \in \Theta$ に対し, λ_i は

$$\sum_{j=1}^n \frac{w_{ij}g(z_j; \theta)}{1 + \lambda_i'g(z_j; \theta)} = 0$$

を満たす。

(4.2) を使って, 平滑化された経験対数尤度 (smoothed empirical log-likelihood, SEL) を θ において定義する。

$$\text{SEL}(\theta) = \sum_{i=1}^n \sum_{j=1}^n w_{ij} \log \hat{p}_{ij} = \sum_{i=1}^n \sum_{j=1}^n w_{ij} \log \frac{w_{ij}}{1 + \lambda_i'g(z_j; \theta)}$$

ここで平滑化とはカーネル関数による平滑化の意味である。 θ_0 の条件付き経験尤度推定量 $\hat{\theta}$ は

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \text{SEL}(\theta) \quad (4.3)$$

と定義される。

次にこれと第 2 節で紹介した Sueishi (2016) の最も不利なサブモデルを使った θ_0 の推定法を比較してみよう。モデルの作り方から真の値 (θ_0, η_0) について $h_0(z) = f(z; \theta_0, \eta_0)$ かつ $\lambda(x; \theta_0, \eta_0) = 0$ が成立する。 $\eta = \eta_0$ とおいて (2.8) と (2.9) を再掲すると

$$f(z; \theta, \eta_0) = \frac{h_0(z)}{1 + \lambda(x; \theta, \eta_0)'g(z; \theta)} \quad (4.4)$$

および

$$\int g(z; \theta) f(z; \theta, \eta_0) dy = \int \frac{g(z; \theta) h_0(z)}{1 + \lambda(x; \theta, \eta_0)'g(z; \theta)} dy = 0 \quad (4.5)$$

となる。(4.5) の標本版は

$$\sum_{j=1}^n \frac{w_{ij}g(z_j; \theta)}{1 + \hat{\lambda}(x_i; \theta, \eta_0)'g(z_j; \theta)} = 0 \quad (4.6)$$

であり, (4.4) より最尤推定量 $\hat{\theta}_1$ は

$$\hat{\theta}_1 = \arg \max_{\theta \in \Theta} \sum_{i=1}^n \log \frac{1}{1 + \hat{\lambda}(x_i; \theta, \eta_0)'g(z_i; \theta)} \quad (4.7)$$

にて得られる。

(4.3) と (4.7) は平滑化だけの違いであり, $\hat{\theta}$ と $\hat{\theta}_1$ は漸近的には同値である。

4.1 漸近的性質

Kitamura et al. (2004) は $\hat{\theta}$ の一貫性, さらに漸近正規で有効推定量であることを示した。Sueishi (2016) は $\hat{\theta}_1$ の一貫性および有効性は, Severini and Wong (1992) の手法で証明可能であると述べている。ただ有効性の証明としては, いずれの手法も煩雑なので, ここでは一貫性を前提に簡明な証明を試みる。

(4.5) は任意のベクトル関数 $c(x)$ に対し

$$\int \frac{c(x)'g(z;\theta)h_0(z)}{1 + \lambda(x;\theta,\eta_0)'g(z;\theta)} dz = 0$$

と同値である。そしてその標本版は

$$\sum_{i=1}^n \frac{c(x_i)'g(z_i;\theta)}{1 + \hat{\lambda}(x_i;\theta,\eta_0)'g(z_i;\theta)} = 0 \quad (4.8)$$

となる。次に推定量 $(\hat{\theta}, \hat{\lambda}(x))$ を

$$(\hat{\theta}, \hat{\lambda}(x)) = \arg \max_{\theta, \lambda} \sum_{i=1}^n \log \frac{1}{1 + \lambda(x_i)'g(z_i;\theta)} \quad (4.9)$$

にて定義する。 $\lambda(x)$ の任意の摂動を $\delta\lambda(x)$ とすると (4.9) より

$$Q_1(\hat{\theta}, \hat{\lambda}) = \frac{1}{n} \sum_{i=1}^n \frac{\delta\lambda(x_i)'g(z_i;\hat{\theta})}{1 + \hat{\lambda}(x_i)'g(z_i;\hat{\theta})} = 0 \quad (4.10)$$

かつ

$$Q_2(\hat{\theta}, \hat{\lambda}) = \frac{1}{n} \sum_{i=1}^n \frac{(\partial g(z_i;\hat{\theta})/\partial \theta')' \hat{\lambda}(x_i)}{1 + \hat{\lambda}(x_i)'g(z_i;\hat{\theta})} = 0 \quad (4.11)$$

が得られる。ところが (4.8) を踏まえれば, $(\hat{\theta}, \hat{\lambda}(x)) = (\hat{\theta}_1, \hat{\lambda}(x; \hat{\theta}_1, \eta_0))$ であることが分かる。したがって Kitamura et al. (2004) や Sueishi (2016) の推定量の漸近正規性は (4.10) と (4.11) を調べればよい。(4.10) より

$$\begin{aligned} 0 &\cong Q_1(\theta_0, 0) + \frac{\partial Q_1(\theta_0, 0)}{\partial \theta'} (\hat{\theta} - \theta_0) + \frac{\partial Q_1(\theta_0, 0)}{\partial \lambda} (\hat{\lambda} - 0) \\ &= \frac{1}{n} \sum_{i=1}^n \delta\lambda_i' g_i + \frac{1}{n} \sum_{i=1}^n \delta\lambda_i' \frac{\partial g_i}{\partial \theta'} (\hat{\theta} - \theta_0) - \frac{1}{n} \sum_{i=1}^n \delta\lambda_i' g_i g_i' \hat{\lambda}_i \\ &\cong \frac{1}{n} \sum_{i=1}^n \delta\lambda_i' g_i + \frac{1}{n} \sum_{i=1}^n \delta\lambda_i' G(x_i) (\hat{\theta} - \theta_0) - \frac{1}{n} \sum_{i=1}^n \delta\lambda_i' \Omega(x_i) \hat{\lambda}_i \end{aligned} \quad (4.12)$$

また (4.11) より

$$\begin{aligned}
 0 &\cong Q_2(\theta_0, 0) + \frac{\partial Q_2(\theta_0, 0)}{\partial \theta'}(\hat{\theta} - \theta_0) + \frac{\partial Q_2(\theta_0, 0)}{\partial \lambda}(\hat{\lambda} - 0) \\
 &= \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial g_i}{\partial \theta'} \right)' \hat{\lambda}_i \\
 &\cong \frac{1}{n} \sum_{i=1}^n G(x_i)' \hat{\lambda}_i
 \end{aligned} \tag{4.13}$$

を得る。ただし、 $\lambda_i = \lambda(x_i)$ および $g_i = g(z_i; \theta_0)$ である。 $\delta\lambda(x)$ は任意のベクトル関数だったから、 q 個のベクトル関数 $\delta\lambda_1(x), \dots, \delta\lambda_q(x)$ を $(\delta\lambda_1(x), \dots, \delta\lambda_q(x)) = \Omega(x)^{-1}G(x)$ で定義して (4.12) に代入し (4.13) と合わせれば、

$$\sqrt{n}(\hat{\theta} - \theta_0) \cong - \left(\frac{1}{n} \sum_{i=1}^n G(x_i)' \Omega(x_i)^{-1} G(x_i) \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n G(x_i)' \Omega(x_i)^{-1} g(z_i; \theta_0)$$

が成立し、

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightsquigarrow N(0, V)$$

ここで

$$V = [EG(x)' \Omega(x)^{-1} G(x)]^{-1}$$

を得る。すなわち $\hat{\theta}$ は有効推定量である。

4.2 仮説検定

Kitamura et al. (2004) の SEL 推定量による仮説検定を紹介する。 θ を q 次元ベクトルとし、帰無仮説として q_1 次元ベクトル α が存在して $\theta = R(\alpha)$ であるとする。特に $\theta_0 = R(\alpha_0)$ である。ただし $q_1 < q$ かつ $q_1 + q_2 = q$ とする。また $\partial R(\alpha_0)/\partial \alpha$ の階数は q_1 であるとする。対立仮説は θ に制約なしとする。仮説検定のこのような設定は Amemiya (1985, Section 4.5.1) によるものである。

$$\hat{\alpha} = \arg \max_{\alpha} \text{SEL}(R(\alpha))$$

とする。 $\hat{\theta}^R = R(\hat{\alpha})$ とおく。尤度比統計量

$$\text{LR} = 2\{\text{SEL}(\hat{\theta}) - \text{SEL}(\hat{\theta}^R)\} \tag{4.14}$$

を考える。

$I(\theta_0) = EG(x)' \Omega(x)^{-1} G(x)$ とおく。

$$n^{1/2}(\hat{\theta} - \theta_0) = -I(\theta_0)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n G(x_i)' \Omega(x_i)^{-1} g(z_i; \theta_0) + o_p(1) \quad (4.15)$$

また $R_\alpha = \partial R(\alpha_0)/\partial \alpha'$ とおいて

$$n^{1/2}(\hat{\alpha} - \alpha_0) = -[R_\alpha' I(\theta_0) R_\alpha]^{-1} R_\alpha' \frac{1}{\sqrt{n}} \sum_{i=1}^n G(x_i)' \Omega(x_i)^{-1} g(z_i; \theta_0) + o_p(1)$$

が成り立つ。

$\hat{\theta}$ の周りでの Taylor 展開により

$$\text{SEL}(\hat{\theta}) - \text{SEL}(\theta_0) = -\frac{1}{2}(\hat{\theta} - \theta_0)' \nabla_{\theta\theta} \text{SEL}(\theta^*)(\hat{\theta} - \theta_0)$$

を得る。ここで $\nabla_{\theta} \text{SEL}(\hat{\theta}) = 0$ を使った。また θ^* は $\hat{\theta}$ と θ_0 の間の値をとる。同様に $\nabla_{\alpha} \text{SEL}(R(\hat{\alpha})) = 0$ を使って

$$\text{SEL}(R(\hat{\alpha})) - \text{SEL}(\theta_0) = -\frac{1}{2}(\hat{\alpha} - \alpha_0)' \nabla_{\alpha\alpha} \text{SEL}(R(\alpha^*))(\hat{\alpha} - \alpha_0)$$

を得る。ここで α^* は $\hat{\alpha}$ と α_0 の間の値をとる。

一様収束性を仮定して

$$\frac{1}{n} \nabla_{\theta\theta} \text{SEL}(\theta^*) = -I(\theta_0) + o_p(1) \quad (4.16)$$

および

$$\frac{1}{n} \nabla_{\alpha\alpha} \text{SEL}(R(\alpha^*)) = -R_\alpha' I(\theta_0) R_\alpha + o_p(1)$$

なので $\epsilon_n = (1/\sqrt{n}) \sum_{i=1}^n G(x_i)' \Omega(x_i)^{-1} g(z_i; \theta_0)$ とおいて

$$\text{LR} = [I(\theta_0)^{-1/2} \epsilon_n]' M [I(\theta_0)^{-1/2} \epsilon_n] + o_p(1)$$

が成立する。ここで

$$M = I_q - I(\theta_0)^{1/2} R_\alpha [R_\alpha' I(\theta_0) R_\alpha]^{-1} R_\alpha' I(\theta_0)^{1/2}$$

M は階数 q_2 のべき等行列かつ $I(\theta_0)^{-1/2} \epsilon_n \rightsquigarrow N(0, I_q)$ なので、連続写像定理により、帰無仮説の下で $\text{LR} \rightsquigarrow \chi^2(q_2)$ を得る。

これを第3節第2項の最適 GMM 検定の結果と比較するため、 θ が q_1 次元ベ

クトル α と q_2 次元ベクトル γ に分解されるとする。次の帰無仮説を考える。

$$H_0 : \gamma = \gamma_0$$

対立仮説としては、次の局所対立仮説を考える。

$$H_1 : \gamma = \gamma_0 + \delta/\sqrt{n}$$

このとき $R(\alpha) = (\alpha', \gamma_0')'$ と書いて

$$R_\alpha = \begin{pmatrix} I_{q_1} \\ 0 \end{pmatrix}$$

となるから

$$M = I_p - I(\theta_0)^{1/2} \begin{pmatrix} I(\theta_0)^{-1}_{11} & 0 \\ 0 & 0 \end{pmatrix} I(\theta_0)^{1/2}$$

を得る。ここで $I(\theta_0)_{11} = EH(x)' \Omega(x)^{-1} H(x)$ である。

対立仮説の下で

$$\begin{aligned} \epsilon_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^n G(x_i)' \Omega(x_i)^{-1} g(z_i; \alpha_0, \gamma_0 + \delta/\sqrt{n}) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n G(x_i)' \Omega(x_i)^{-1} g(z_i; \theta_0) + EG(x)' \Omega(x)^{-1} K(x) + o_p(1) \end{aligned}$$

なので、 $I(\theta_0)^{-1/2} \epsilon_n \rightsquigarrow N(I(\theta_0)^{-1/2} EG(x)' \Omega(x)^{-1} K(x) \delta, I_q)$ となり

$$\begin{aligned} \kappa &= \delta' EK'(x) \Omega(x)^{-1} G(x) I(\theta_0)^{-1/2} M I(\theta_0)^{-1/2} EG'(x) \Omega(x)^{-1} K(x) \delta \\ &= \delta' [EK'(x) \Omega(x)^{-1} K(x) \\ &\quad - EK'(x) \Omega(x)^{-1} H(x) (EH'(x) \Omega(x)^{-1} H(x))^{-1} \\ &\quad \times EH'(x) \Omega(x)^{-1} K(x)] \delta \end{aligned}$$

これは (3.1) と一致している。

5 可算無限近似法

本節は、標本の大きさと共に増大する基底関数列を用いた2つの代表的な推

定法について論じる。

5.1 増大操作変数列法

条件付きモーメント制約

$$E[g(z_i; \theta_0)|x_i] = 0$$

は、任意の関数 $c(x_i)$ に対し無条件モーメント制約

$$Ec(x_i)g(z_i; \theta_0) = 0 \quad (5.1)$$

が成立することと同値である。さらに $p^K(x_i)$ を K 次元の基底関数のベクトルとして無条件モーメント制約

$$Eg(z_i; \theta_0) \otimes p^K(x_i) = 0 \quad (5.2)$$

を定義すると、これは $K \rightarrow \infty$ のとき (5.1) と同値になる。なお $\otimes$ は Kronecker 積を表す。

Donald et al. (2003) は (5.2) を用いた推定方法を提案した。 K はサンプルサイズ n と共に増大する。 $\rho(z_i; \theta) = g(z_i; \theta) \otimes p^K(x_i)$ と定義すると $\rho(z_i; \theta)$ は rK 次元ベクトル関数となり、GMM 推定量 $\hat{\theta}$ は

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \hat{\rho}(\theta)' \hat{\Sigma}(\tilde{\theta})^{-1} \hat{\rho}(\theta) \quad (5.3)$$

と定義される。ここで

$$\hat{\rho}(\theta) = \frac{1}{n} \sum_{i=1}^n \rho(z_i; \theta)$$

および

$$\hat{\Sigma}(\theta) = \frac{1}{n} \sum_{i=1}^n \rho(z_i; \theta) \rho(z_i; \theta)'$$

であり、 $\tilde{\theta}$ は初期推定量である。適当な正則条件の下で、 $\hat{\theta}$ は一致推定量、さらには有効推定量になることが示される。後者について証明の概要を示そう。

$g_i = g(z_i; \theta_0)$ および $p_i = p^K(x_i)$ とする。(5.3) を解いて GMM 推定量 $\hat{\theta}$ の漸近分散の逆行列は

$$\begin{aligned}\frac{\partial \hat{\rho}(\theta_0)'}{\partial \theta} \hat{\Sigma}(\theta_0)^{-1} \frac{\partial \hat{\rho}(\theta_0)}{\partial \theta'} &= \frac{1}{n} \sum_{i=1}^n \frac{\partial g_i'}{\partial \theta} \otimes p_i' \left(\frac{1}{n} \sum_{i=1}^n g_i g_i' \otimes p_i p_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^n \frac{\partial g_i}{\partial \theta'} \otimes p_i \\ &\cong \frac{1}{n} \sum_{i=1}^n \frac{\partial g_i'}{\partial \theta} \otimes p_i' \left(\frac{1}{n} \sum_{i=1}^n \Omega_i \otimes p_i p_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^n \frac{\partial g_i}{\partial \theta'} \otimes p_i\end{aligned}$$

の極限として得られる。ここで

$$\Omega_i = \Omega(x_i) = E[g_i g_i' | x_i]$$

である。Donald et al. (2003) の Lemma A.3 より、この極限は

$$EG(x_i)' \Omega(x_i)^{-1} G(x_i)$$

に等しい。したがって $\hat{\theta}$ は有効推定量である。ただし彼らの Lemma A.3 は煩雑な代数計算の結果得られたものである。直観的な理解のため幾何学的に考えてみたい。第 2 節の結果から、(5.1) を制約にもつ条件付きモーメント制約モデルの有効スコア関数は $\Pi(\ell_\theta | \{g\}_x)$ である。一方、(5.2) を制約にもつ無条件モーメント制約モデルの有効スコア関数は $\Pi(\ell_\theta | \{g \otimes p^K\})$ である。GMM 推定量 $\hat{\theta}$ は有効推定量なので、漸近分散はこの有効スコア関数に対応する。ここで $p^K = p^K(x)$ である。次に

$$\{g \otimes p^K\} \subset \{g\}_x$$

であるが、 p^K は基底関数ベクトルなので、 $K \rightarrow \infty$ のとき両者は一致する。よって $\hat{\theta}$ は (5.1) の条件付きモーメント制約モデルの推定量としても有効推定量になる。

5.2 シーブ法

次のモデルを考える。

$$E[g(z_i; \theta_0) | x_i] = 0 \quad (5.4)$$

Ai and Chen (2003) および Newey and Powell (2003) の提案したシーブ最小距離 (sieve minimum distance, SMD) 法を使って (5.4) を推定してみよう。 $m(x_i; \theta) = E[g(z_i; \theta) | x_i]$ とすると、(5.4) より、真のパラメータ θ_0 は目的関数

$$Em(x_i; \theta)' \Omega(x_i)^{-1} m(x_i; \theta) \quad (5.5)$$

を最小化することがわかる。(5.5) の標本版を考えることによって、次のような最小距離推定量が得られる。

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n \hat{m}(x_i; \theta)' \hat{\Omega}(x_i)^{-1} \hat{m}(x_i; \theta) \quad (5.6)$$

ただし $\hat{\Omega}(x)$ は $\Omega(x)$ の一致推定量、 $\hat{m}(x; \theta)$ は以下のように定義される $m(x; \theta)$ のシーブ推定量である。

$$\hat{m}(x; \theta) = \sum_{j=1}^n g(z_j; \theta) p^K(x_j)' (P' P)^{-1} p^K(x) \quad (5.7)$$

ここで $P = (p^K(x_1), \dots, p^K(x_n))'$ である。

(5.7) の作り方から、 $n \rightarrow \infty$ のとき $\hat{m}(x; \theta)$ は

$$\hat{m}(x; \theta) \xrightarrow{P} \Pi(g(z; \theta) | \{p^K(x)\})$$

言い換えれば $g(z; \theta)$ の $p^K(x)$ への線形射影になる。さらに $p^K(x)$ が基底関数ベクトルだから、 $K \rightarrow \infty$ のとき、 $m(x; \theta) = E[g(z; \theta) | x]$ に近づく。

SMD 推定量 $\hat{\theta}$ が一致推定量であることを前提に、これが有効推定量であることを示そう。

$$M(x_i; \theta) = \hat{m}(x_i; \theta)' \hat{\Omega}(x_i)^{-1} \hat{m}(x_i; \theta)$$

とおく。(5.6) と平均値の定理から

$$0 = \frac{1}{\sqrt{n}} \sum_{i=1}^n M_{\theta}(x_i; \hat{\theta}) \quad (5.8)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^n M_{\theta}(x_i; \theta_0) + \frac{1}{\sqrt{n}} \sum_{i=1}^n M_{\theta\theta}(x_i; \theta^*) (\hat{\theta} - \theta_0) \quad (5.9)$$

ここで θ^* は $\hat{\theta}$ と θ_0 の間の値をとる。(5.9) の右辺第 1 項に (5.7) を代入して整理すれば、Ai and Chen (2003, Corollary C.3) より

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n M_{\theta}(x_i; \theta_0) &= \frac{2}{\sqrt{n}} \sum_{i=1}^n \hat{m}_{\theta}(x_i; \theta_0)' \hat{\Omega}(x_i)^{-1} \hat{m}(x_i; \theta_0) \\ &= \frac{2}{\sqrt{n}} \sum_{i=1}^n G(x_i)' \Omega(x_i)^{-1} g(z_i; \theta_0) + o_p(1) \end{aligned} \quad (5.10)$$

を得る。(5.9) の右辺第 2 項については

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n M_{\theta\theta}(x_i; \theta^*) &= \frac{2}{n} \sum_{i=1}^n \hat{m}_{\theta}(x_i; \theta_0)' \hat{\Omega}(x_i)^{-1} \hat{m}_{\theta}(x_i; \theta_0) + o_p(1) \\ &= \frac{2}{n} \sum_{i=1}^n G(x_i)' \Omega(x_i)^{-1} G(x_i) + o_p(1) \\ &\xrightarrow{P} 2EG(x)' \Omega(x)^{-1} G(x) = 2I(\theta_0) \end{aligned} \quad (5.11)$$

となる。ここで一様収束の仮定および $m(x; \theta_0) = 0$ であることを使った。

これらと (5.8) および (5.9) から

$$\sqrt{n}(\hat{\theta} - \theta_0) = -I(\theta_0)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n G(x_i)' \Omega(x_i)^{-1} g(z_i; \theta_0) + o_p(1) \quad (5.12)$$

が得られ

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightsquigarrow N(0, V)$$

ここで

$$V = [EG(x)' \Omega(x)^{-1} G(x)]^{-1}$$

を得る。すなわち SMD 推定量 $\hat{\theta}$ は有効推定量である。

Ai and Chen (2003) および Newey and Powell (2003) は実際には (5.4) よりも複雑なモデルについて考察している。

$$E[g(z_i; \alpha_0, h_0)|x_i] = 0$$

未知パラメータ $(\alpha_0, h_0) \in \mathcal{A} \times \mathcal{H}$ は有限次元のベクトル α_0 と無限次元の関数 $h_0(z_i)$ から成る。 α_0 を関心のあるパラメータ, h_0 を局外パラメータとする。 $\mathcal{H}$ は連続関数から成る関数空間である。 $q^{K_1}(z)$ を K_1 次元の基底関数ベクトルとすると, $h(z)$ は基底関数の線形結合 $q^{K_1}(z)' \beta \in \mathcal{H}_{K_1}$ により近似できるので,

$$g(z; \alpha, h) \cong g(z; \alpha, q^{K_1}(z)' \beta)$$

が成り立つ。 $\mathcal{H}_{K_1}$ を $\mathcal{H}$ のシーブ空間と呼ぶ。 $\theta = (\alpha, \beta)$ とすれば, θ は有限次元ベクトルなので (5.6) の最小化問題を解けばよい。その解を $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ とする。 K および K_1 は n と共に適当な速度で増加する。このとき適当な無限次

元ベクトル β_0 が存在して $q^\infty(z)'\beta_0 = h_0(z)$ が成り立つ。 $\theta_0 = (\alpha_0, \beta_0)$ である。 Newey and Powell (2003) はこの問題の SMD 推定量の一致性つまり $\hat{\alpha} \xrightarrow{P} \alpha_0$ および $q^{K_1}(z)'\hat{\beta} \xrightarrow{P} h_0(z)$ を示した。 さらに Ai and Chen (2003) は $\hat{\alpha}$ が有効推定量であることを示した。

Ai and Chen (2003) による有効性の証明は煩雑で長いが、セミパラメトリック理論と幾何学的手法を使って証明が可能である。実は基本的なアイデアは同じである。概略を示そう。第2節の結果から、条件付きモーメント制約モデル (5.4) におけるパラメータ θ の有効スコア関数は

$$\tilde{\ell}_\theta = -G(x)'\Omega(x)^{-1}g(z; \theta_0)$$

であった。

$$\tilde{\ell}_\theta(z; \theta) = -G(x)'\Omega(x)^{-1}g(z; \theta)$$

とおく。(5.10) より θ_0 の近傍で

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n M_\theta(x_i; \theta) = -\frac{2}{\sqrt{n}} \sum_{i=1}^n \tilde{\ell}_\theta(z_i; \theta) + o_p(1)$$

が成り立つ。これは (5.6) の SMD 推定量 $\hat{\theta}$ が、 θ の有効スコア方程式

$$\sum_{i=1}^n \tilde{\ell}_\theta(z_i; \theta) = 0$$

の解でもあることを示している。 $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ である。

$\tilde{\ell}_\theta = (\tilde{\ell}'_\alpha, \tilde{\ell}'_\beta)'$ と分解できるので、関心のあるパラメータ α の有効スコア関数は

$$\ell_\alpha^* = \tilde{\ell}_\alpha - \Pi(\tilde{\ell}_\alpha | \{\tilde{\ell}_\beta\})$$

と導出でき、 α の有効スコア方程式

$$\sum_{i=1}^n \ell_\alpha^*(z_i; \alpha, \hat{\beta}) = 0$$

の解は SMD 推定量 $\hat{\alpha}$ に一致する。これはプロファイル (profile) 尤度の考え方から明らかである。有効スコア方程式に関する van der Vaart (1998) の Theorem 25.54 を使って不偏条件 (no-bias condition) の下で $\hat{\alpha}$ の有効性が示される。不偏

条件については、局外パラメータ推定量 $\hat{\beta}$ の収束速度が $n^{-1/4}$ より大きいという条件下で得られるが、 K や K_1 の増加速度のコントロールなど適当な正則条件の下で、 $\hat{\theta}$ の収束速度を $n^{-1/4}$ より大きくすることができる。

6 幾何学的考察

最後に、条件付きモーメント制約モデルの、幾何学的に興味深い様々な性質を考察する。

6.1 モーメント関数の選択

モーメント制約モデルにおいては、どのモーメント関数を選択するかによって、推定量の効率性や検定の検出力に違いが出る。無条件モーメント制約モデルについては、Goto (2021a) において詳しく論じた。ここでは条件付きモーメント制約モデルのモーメント関数選択について考察したい。

2つのモーメント関数ベクトル $g(z; \theta)$ および $g^{(1)}(z; \theta)$ で、 $\{g^{(1)}\}_x \subset \{g\}_x$ なるものを考える。例えば $g^{(1)}(z; \theta) = C(x; \theta)g(z; \theta)$ で、 $C(x; \theta)$ は階数 r_1 の $r_1 \times r$ 行列 ($r_1 \leq r$) とする。最も簡単な例としては、 $g^{(1)}(z; \theta)$ は $g(z; \theta)$ からモーメント関数を1個除いたものでもよい。 Σ と Σ_1 を、それぞれ $g(z; \theta)$ および $g^{(1)}(z; \theta)$ に対応する、パラメータ θ の効率性限界とする。また $W(z)$ と $W_1(z)$ を、それぞれ $g(z; \theta)$ および $g^{(1)}(z; \theta)$ に対応する、分布関数 $F(z)$ の効率性限界とする。

命題 6.1 以下の関係が成立する。

$$(a) \Sigma \leq \Sigma_1$$

$$(b) W(z) \leq W_1(z)$$

証明 (a) について、第2節第2項によれば、 θ の有効スコア関数を $\tilde{\ell}_\theta$ とすれば効率性限界は $[E\tilde{\ell}_\theta\tilde{\ell}_\theta']^{-1}$ である。有効スコア関数はそれぞれ、 $\tilde{\ell}_\theta = \Pi(\ell_\theta|\{g\}_x)$ および $\tilde{\ell}_\theta^{(1)} = \Pi(\ell_\theta|\{g^{(1)}\}_x)$ であるが $\{g^{(1)}\}_x \subset \{g\}_x$ なので、任意の定数ベクトル δ について、

$$E(\delta' \tilde{\ell}_\theta^{(1)})^2 \leq E(\delta' \tilde{\ell}_\theta)^2$$

したがって $\Sigma \leq \Sigma_1$ が成立する。

(b) について、第 2 節第 2 項の議論から、接空間はそれぞれ $T(P_0) = \{g\}_x^\perp + \{\ell_\theta\}$ および $T^{(1)}(P_0) = \{g^{(1)}\}_x^\perp + \{\ell_\theta\}$ である。 $F(z)$ の影響関数を $\phi(z_i) = 1_{[z_i < z]} - F(z)$ とおくと、有効影響関数はそれぞれ $\tilde{\phi} = \Pi(\phi|T(P_0))$ および $\tilde{\phi}^{(1)} = \Pi(\phi|T^{(1)}(P_0))$ である。 $T(P_0) \subset T^{(1)}(P_0)$ なので

$$W(z) = E\tilde{\phi}^2 \leq E(\tilde{\phi}^{(1)})^2 = W_1(z)$$

が成り立つ。 ■

以上の結果は、モーメント条件が多いモデルの方が情報量も多いので、推定量の精度も高くなるだろうという直観に合っている。しかしモーメント条件が多くなると一致性の欠如等、別の弊害が生じることも知られている。サンプルサイズとモーメント条件の増加速度のバランスが重要になる。多数のモーメント条件をもつモデルに関する研究は、Koenker and Machado (1999) や Han and Phillips (2006)などを参照のこと。

次に有効スコア関数 $\tilde{\ell}_\theta$ の θ_0 を θ で置き換えたモーメント関数ベクトル $\tilde{\ell}_\theta(\theta)$ を考える。これは実は $-G(x)' \Omega(x)^{-1} g(z; \theta)$ と同値である。またこれは確かにモーメント条件

$$E[\tilde{\ell}_\theta(\theta_0)|x] = 0$$

を満たす。 Σ_A と $W_A(z)$ を、それぞれこのモデルに対応する θ および $F(z)$ の効率性限界とする。

さらに第 2 節第 2 項で扱ったサブモデル $f(z; \theta, \eta)$ について局外パラメータ η が既知のもの $f(z; \theta, \eta_0)$ を考える。この条件付き密度関数 $f(y|x; \theta, \eta_0)$ の対数をとって微分したもの

$$\ell_{\theta, y|x}(\theta) = \frac{\partial \log f(y|x; \theta, \eta_0)}{\partial \theta}$$

をモーメント関数ベクトルとする。これも確かにモーメント条件

$$E[\ell_{\theta,y|x}(\theta_0)|x] = 0$$

を満たす。 Σ_0 と $W_0(z)$ を、それぞれこのモデルに対応する θ および $F(z)$ の効率性限界とする。なお x の周辺密度関数の対数をとって微分したものを $\ell_{\theta,x}(\theta)$ とすると

$$\ell_{\theta}(\theta) = \ell_{\theta,y|x}(\theta) + \ell_{\theta,x}(\theta)$$

が成り立つ。

命題 6.2 以下の関係が成立する。

$$(a) \Sigma_0 \leq \Sigma = \Sigma_A$$

$$(b) W(z) \leq W_A(z)$$

証明 (a) について、まず $\tilde{\ell}_{\theta} = \Pi(\ell_{\theta}|\{g\}_x) = \Pi(\ell_{\theta}|\{\tilde{\ell}_{\theta}\}_x)$ なので $\Sigma = \Sigma_A$ が得られる。次に $\Pi(\ell_{\theta}|\{\ell_{\theta,y|x}\}_x) = \ell_{\theta,y|x}$ であるが⁵、 $E[g|x] = 0$ を踏まえて

$$\Pi(\ell_{\theta}|\{g\}_x) = \Pi(\ell_{\theta,y|x}|\{g\}_x)$$

なので、 $\Sigma_0 \leq \Sigma$ が成り立つ。

(b) について、 $\tilde{\ell}_{\theta}(\theta)$ をモーメント関数ベクトルにもつモデルの接空間は

$$T_A(P_0) = \{G(x)' \Omega(x)^{-1} g\}_x^{\perp} + \{\ell_{\theta}\}$$

である。また g をモーメント関数ベクトルにもつオリジナルモデルの接空間は

$$T(P_0) = \{g\}_x^{\perp} + \{\ell_{\theta}\}$$

である。 $\{G(x)' \Omega(x)^{-1} g\}_x \subset \{g\}_x$ なので $T(P_0) \subset T_A(P_0)$ より $W(z) \leq W_A(z)$ が得られる。 ■

以上の結果から、密度関数のモデル $f(z; \theta, \eta_0)$ が既知の場合には、これをモーメント条件に使った方が θ 推定の精度は高まるということであり、GMM より MLE の方が効率性が高いという直観にも通じる。なお $W_0(z)$ と $W(z)$ や $W_A(z)$ との大小は不明である。

次に仮説検定の検出力を比較する。最適 GMM による検定と他の手法による検定は同値になるので、ここでは代表して最適 GMM による検定で比較する。

$\theta = (\beta', \gamma')'$ すなわち, パラメータ θ が q_1 次元ベクトル β と q_2 次元ベクトル γ に分解されたとする。次の帰無仮説を考える。

$$H_0 : \gamma = \gamma_0$$

対立仮説としては, 次の局所対立仮説 (local alternatives) を考える。

$$H_1 : \gamma = \gamma_0 + \delta/\sqrt{n}$$

$g_A(z; \theta) = A(x)g(z; \theta)$ かつ $\Omega_A = E g_A(z; \theta_0)g_A(z; \theta_0)'$ とおく。ただし $A(x) = G(x)'\Omega(x)^{-1}$ である。

$g_A(z; \theta)$ をモーメント関数ベクトルとする無条件モーメント制約モデルに関し, 第 3 章第 2 項の (3.1) で与えられた DM 検定の検出力 κ は

$$\begin{aligned} \kappa &= \|\Pi(\delta'\ell_\gamma|\{g_A\}) - \Pi(\delta'\ell_\gamma|\{\Pi(\ell_\beta|\{g_A\})\})\|_P^2 \\ &= \|\Pi(\delta'\ell_\gamma|\{g_A\} \cap \{\ell_\beta\}^\perp)\|_P^2 \end{aligned} \quad (6.1)$$

と書ける。ここで $\phi \in L_2^0(P_0)$ のとき $\|\phi\|_P^2 = E\phi^2$ を表す。証明は Goto (2021a, Theorem 5) を参照のこと。なお (6.1) の κ は

$$\begin{aligned} \kappa &= \|\Pi(\delta'\ell_\gamma|\{g\}_x) - \Pi(\delta'\ell_\gamma|\{\Pi(\ell_\beta|\{g\}_x)\})\|_P^2 \\ &= \|\Pi(\delta'\ell_\gamma|\{g\}_x \cap \{\ell_\beta\}^\perp)\|_P^2 \end{aligned}$$

と同値である。

次に, g , $\tilde{\ell}_\theta(\theta)$ および $\ell_{\theta, y|x}(\theta)$ をモーメント関数ベクトルとする, 条件付きモーメント制約モデルの仮説検定の検出力を, それぞれ κ , κ_A および κ_0 とする。

命題 6.3 以下の関係が成立する。

$$\kappa = \kappa_A \leq \kappa_0$$

証明 モーメント関数ベクトル $\tilde{\ell}_\theta(\theta)$ に関し, g の g_A に対応するものは $\Pi(\ell_\theta|\{\tilde{\ell}_\theta\}_x) = \tilde{\ell}_\theta$ なので結局 $g_A = G(x)'\Omega(x)^{-1}g$ に等しく, よって $\kappa = \kappa_A$ を得る。次に $E[g_A|x] = 0$ を踏まえて (6.1) は次のように変形できる。

$$\kappa = \|\Pi(\delta'\ell_{\gamma, y|x}|\{g_A\}) - \Pi(\delta'\ell_{\gamma, y|x}|\{\Pi(\ell_{\beta, y|x}|\{g_A\})\})\|_P^2$$

$$= \|\Pi(\delta' \ell_{\gamma, y|x} | \{g_A\} \cap \{\ell_{\beta, y|x}\}^\perp)\|_P^2 \quad (6.2)$$

モーメント関数ベクトル $\ell_{\theta, y|x}(\theta)$ に関し、 g の g_A に対応するものは

$$\Pi(\ell_\theta | \{\ell_{\theta, y|x}\}_x) = \ell_{\theta, y|x}$$

より $\ell_{\theta, y|x}(\theta)$ 自身である。さらに $\Pi(\ell_{\gamma, y|x} | \{\ell_{\theta, y|x}\}) = \ell_{\gamma, y|x}$ などを踏まえ

$$\kappa_0 = \|\delta' \ell_{\gamma, y|x} - \Pi(\delta' \ell_{\gamma, y|x} | \{\ell_{\beta, y|x}\})\|_P^2 = \|\Pi(\delta' \ell_{\gamma, y|x} | \{\ell_{\beta, y|x}\}^\perp)\|_P^2$$

となり、これを (6.2) と比較して、 $\kappa \leq \kappa_0$ を得る。 ■

なお前述のモーメント関数ベクトル g と $g^{(1)}$ の比較に関し、検出力に関しての大小は不明である。

6.2 $\hat{\theta}$ と $\hat{\lambda}(x; \hat{\theta})$ の関係

第 4 節の Kitamura et al. (2004) による条件付き経験尤度法で現れた推定量 $\hat{\theta}$ および $\hat{\lambda}(x; \hat{\theta})$ について、もう少し深く考察してみよう。Kitamura et al. (2004, A.17) より

$$\sqrt{n}(\hat{\theta} - \theta_0) = -I(\theta_0)^{-1} \frac{1}{\sqrt{n}} \sum_{j=1}^n G(x_j)' \Omega(x_j)^{-1} g(z_j; \theta_0) + o_p(1) \quad (6.3)$$

また同 (2004, Lemma B.1) などから

$$\sqrt{n}\hat{\lambda}(x_i; \theta_0) = \Omega(x_i)^{-1} \sqrt{n} \sum_{j=1}^n w_{ij} g(z_j; \theta_0) + o_p(1) \quad (6.4)$$

また θ^* を $\hat{\theta}$ と θ_0 の中間値として

$$\hat{\lambda}(x_i; \hat{\theta}) = \hat{\lambda}(x_i; \theta_0) + \frac{\hat{\lambda}(x_i; \theta^*)}{\partial \theta'} (\hat{\theta} - \theta_0)$$

なので、第 2 節の (2.10) より

$$\sqrt{n}\hat{\lambda}(x_i; \hat{\theta}) = \sqrt{n}\hat{\lambda}(x_i; \theta_0) + \Omega(x_i)^{-1} G(x_i) \sqrt{n}(\hat{\theta} - \theta_0) + o_p(1) \quad (6.5)$$

を得る。

命題 6.4 $\sqrt{n}(\hat{\theta} - \theta_0)$ と $\sqrt{n}\hat{\lambda}(x_i; \hat{\theta})$ は漸近的に直交する。

証明 (6.3)–(6.5) から

$$\begin{aligned}\sqrt{n}\hat{\lambda}(x_i; \hat{\theta}) &= \Omega(x_i)^{-1} \frac{1}{\sqrt{n}} \sum_{j=1}^n (nw_{ij} - G(x_i)I(\theta_0)^{-1}G(x_j)' \Omega(x_j)^{-1})g(z_j; \theta_0) \\ &\quad + o_p(1)\end{aligned}\tag{6.6}$$

なのでこれと $\sqrt{n}(\hat{\theta} - \theta_0)$ の漸近共分散を求める。

$$nw_{ij} \cong \frac{K((x_i - x_j)/h)/h}{f(x_i; \theta_0)}$$

に注意すると, E_j を $z_j \sim f(z_j; \theta_0)$ に関する期待値として, $h \rightarrow 0$ のとき

$$\begin{aligned}E_j \frac{K((x_i - x_j)/h)/h}{f(x_i; \theta_0)} g(z_j; \theta_0) g(z_j; \theta_0)' \Omega(x_j)^{-1} G(x_j) \\ = \int_{x_j} \int_{y_j} \frac{K((x_i - x_j)/h)/h}{f(x_i; \theta_0)} g(z_j; \theta_0) g(z_j; \theta_0)' \Omega(x_j)^{-1} G(x_j) f(z_j; \theta_0) dy_j dx_j \\ = \int_{x_j} \frac{K((x_i - x_j)/h)/h}{f(x_i; \theta_0)} \Omega(x_j) \Omega(x_j)^{-1} G(x_j) f(x_j; \theta_0) dx_j \\ \rightarrow G(x_i)\end{aligned}$$

を得る。同様に

$$\begin{aligned}E_j G(x_i) I(\theta_0)^{-1} G(x_j)' \Omega(x_j)^{-1} g(z_j; \theta_0) g(z_j; \theta_0)' \Omega(x_j)^{-1} G(x_j) \\ = G(x_i) I(\theta_0)^{-1} E_j G(x_j)' \Omega(x_j)^{-1} G(x_j) \\ = G(x_i)\end{aligned}$$

が成立し, 漸近共分散は 0 となる。 ■

次に, 第 4 節 (4.4) を踏まえて, 分布関数 $F(z; \theta_0)$ の推定量 $\hat{F}(z)$ を

$$\hat{F}(z) = \frac{1}{n} \sum_{i=1}^n 1_{[z_i \leq z]} \frac{1}{1 + \hat{\lambda}(x_i; \hat{\theta})' g(z_i; \hat{\theta})}\tag{6.7}$$

と定義する。また $\phi(z_i) = 1_{[z_i \leq z]} - F(z; \theta_0)$ とする。(4.10) を踏まえれば

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{1 + \hat{\lambda}(x_i; \hat{\theta})' g(z_i; \hat{\theta})} = 1\tag{6.8}$$

になることが分かる。

命題 6.5 $\hat{F}(z)$ は $F(z; \theta_0)$ の有効推定量である。

証明 まず (6.7) で $1/n \sum_{i=1}^n \cdot$ の一様収束性を仮定して, $\hat{F}(z) \xrightarrow{P} F(z; \theta_0)$ を得

る。次に (6.8) を踏まえて

$$\begin{aligned}\sqrt{n}(\hat{F}(z) - F(z; \theta_0)) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi(z_i) \frac{1}{1 + \hat{\lambda}(x_i; \hat{\theta})' g(z_i; \hat{\theta})} \\ &= -\frac{1}{\sqrt{n}} \sum_{i=1}^n \phi(z_i) \hat{\lambda}(x_i; \hat{\theta})' g(z_i; \hat{\theta}) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi(z_i) + o_p(1)\end{aligned}$$

(6.6) より

$$\begin{aligned}\frac{1}{\sqrt{n}} \sum_{i=1}^n \phi(z_i) g(z_i; \hat{\theta})' \hat{\lambda}(x_i; \hat{\theta}) &= \frac{1}{\sqrt{n}} \sum_{j=1}^n E[\phi(z_i) g(z_i; \theta_0)' | x_j] \Omega(x_j)^{-1} g(z_j; \theta_0) \\ &\quad - E\phi(z_i) g(z_i; \theta_0)' \Omega(x_i)^{-1} G(x_i) I(\theta_0)^{-1} \\ &\quad \times \frac{1}{\sqrt{n}} \sum_{j=1}^n G(x_j)' \Omega(x_j)^{-1} g(z_j; \theta_0) + o_p(1)\end{aligned}$$

なので、直接的な計算により、 $\sqrt{n}(\hat{F}(z) - F(z; \theta_0))$ の漸近分散は、定理 2.2 で与えられた $F(z)$ の効率性限界に一致することが分かる。 ■

6.3 仮説検定と J 検定

条件付きモーメント制約モデルについて、ここまで様々な推定法を紹介してきた。最適 GMM 法、ワンステップ法、条件付き経験尤度法およびその類似法、増大操作変数列法、そしてシーブ法である。いずれの方法も有効推定量を与えることが分かった。また仮説検定については最適 GMM 法と条件付き経験尤度法について考察し、局所対立仮説に関し同等の検出力をもつことが示された。

ワンステップ法については構造上の問題から Wald 検定を使うことになるので、ここでは省略して、それ以外の手法について説明する。

シーブ法については、実は条件付き経験尤度法と同じ構造をもっている。帰無仮説として同じ制約設定 $\theta = R(\alpha)$ をシーブ法に適用してみよう。制約なし推定量は

$$\hat{\theta} = \arg \min_{\theta} \frac{1}{n} \sum_{i=1}^n M(x_i; \theta)$$

制約付き推定量は

$$\hat{\alpha} = \arg \min_{\alpha} \frac{1}{n} \sum_{i=1}^n M(x_i; R(\alpha))$$

である。検定統計量を

$$LR_1 = \sum_{i=1}^n M(x_i; R(\hat{\alpha})) - \sum_{i=1}^n M(x_i; \hat{\theta}) \quad (6.9)$$

とする。(4.16)を再掲して

$$\frac{1}{n} \nabla_{\theta\theta} \text{SEL}(\theta^*) = -I(\theta_0) + o_p(1)$$

(5.11) より

$$\frac{1}{n} \sum_{i=1}^n M_{\theta\theta}(x_i; \theta^*) = 2I(\theta_0) + o_p(1)$$

また (4.15) と (5.12) から両者の $\sqrt{n}(\hat{\theta} - \theta_0)$ は漸近的に等しい。以上のことから、(4.14) の LR と (6.9) の LR_1 は同値な検定であり、帰無仮説が正しいとき $\chi^2(q_2)$ に分布収束し、局所対立仮説の下でも両者同じ結果をもたらす。

次に増大操作変数列法については、基底関数の数 K を固定して考えれば、モーメント条件が rK 個の無条件モーメント制約モデルに帰着される。したがって第3節第2項の最適 GMM による仮説検定と同じ問題設定、つまり同じ帰無仮説と対立仮説の下で、最適 GMM の DM と全く同値な GMM 検定統計量 DM_1 を構成することができる。これも帰無仮説が正しいとき $\chi^2(q_2)$ に分布収束し、局所対立仮説の下でも $K \rightarrow \infty$ のとき両者同じ結果をもたらす。 K は n と共に増大するが、増加速度の適度なコントロールで望む結果が得られる。

J 検定すなわち過剰識別制約検定は、GMM 推定の副産物として発見され、Hansen (1982) により提示されたものである。J 統計量は、無条件モーメント制約モデルにおいては、効率的な GMM 推定や経験尤度推定を実施するときに自然に出現するものであるが、条件付きモーメント制約モデルでは必ずしもそうではない。例えば最適 GMM 法では、無条件モーメント制約モデルに変換するときに、必然的にモーメント制約数 r がパラメータ数 q に一致してしまうため、J 統計量は常に 0 になってしまうのである。

無条件モーメント制約モデルにおける仮説検定と J 検定の関係については Goto (2021a) および Goto (2021b) で詳しく論じている。帰無仮説としてモーメント制約数 r , q 次元パラメータ θ , ただし $q < r$, をもつモデルの GMM 推定量 $\hat{\theta}$ は

$$\hat{\theta} = \arg \min_{\theta} \hat{g}(\theta)' \hat{\Omega}^{-1} \hat{g}(\theta)$$

一方, 対立仮説として θ に $r - q$ 次元パラメータ σ を加えたモデルの GMM 推定量 $(\bar{\theta}', \bar{\sigma}')'$ は

$$(\bar{\theta}', \bar{\sigma}')' = \arg \min_{\theta, \sigma} \hat{g}(\theta, \sigma)' \hat{\Omega}^{-1} \hat{g}(\theta, \sigma)$$

検定統計量 DM は

$$DM = n \left[\hat{g}(\hat{\theta})' \hat{\Omega}^{-1} \hat{g}(\hat{\theta}) - \hat{g}(\bar{\theta}, \bar{\sigma})' \hat{\Omega}^{-1} \hat{g}(\bar{\theta}, \bar{\sigma}) \right]$$

であり, 帰無仮説の下で $DM \rightsquigarrow \chi^2(r - q)$ が成立する。しかしながら対立仮説のモデルでは制約方程式数=パラメータ数なので, $\hat{g}(\bar{\theta}, \bar{\sigma}) = 0$ が成立し, 結局帰無仮説の下で

$$n\hat{g}(\hat{\theta})' \hat{\Omega}^{-1} \hat{g}(\hat{\theta}) \rightsquigarrow \chi^2(r - q)$$

となる。これが J 検定である。仮説検定と J 検定の違いは, 仮説検定では対立仮説のモデルが特定されているのに対し, J 検定ではそれが特定されていないことである。J 検定では対立仮説のモデルが無数にある。Newey (1985) により, J 検定で検出できない対立仮説の存在が最初に指摘されて以来, J 検定の一致性について注意を呼びかける文献も見られる。例えば Hamilton (1994), Hayashi (2000) などを参照のこと。これに対し, Goto (2021a) は, 対立仮説のモデルが識別可能である限り J 検定で検出可能であることを示した。

最後に増大操作変数列法による J 検定を解説する。条件付きモーメント制約モデルは, 基底関数の数 K が固定されていると仮定すると, モーメント制約数 rK , パラメータ数 q の無条件モーメント制約モデルに帰着される。(5.3) を再掲すると GMM 推定量 $\hat{\theta}$ は

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \hat{\rho}(\theta)' \hat{\Sigma}(\bar{\theta})^{-1} \hat{\rho}(\theta)$$

と定義され、検定統計量 $\hat{T}$ は、帰無仮説の下で

$$\hat{T} = n\hat{\rho}(\hat{\theta})'\hat{\Sigma}(\hat{\theta})^{-1}\hat{\rho}(\hat{\theta}) \rightsquigarrow \chi^2(rK - q)$$

となる。 $K \rightarrow 0$ とすれば

$$\frac{\hat{T} - (rK - q)}{\sqrt{2(rK - q)}} \rightsquigarrow N(0, 1)$$

が成立する。つまりこれにより条件付きモーメント制約

$$E[g(z_i; \theta_0)|x_i] = 0$$

の J 検定が可能となる。詳細は Donald et al. (2003) を参照されたい。

6.4 過剰識別の問題

本論文の最後に、Chen and Santos (2018) が提示した「正則モデルにおける過剰識別の問題」について考察する。以下 Chen and Santos (2018) を CS と略す。CS はセミパラメトリックモデルにおける漸近的に線形な正則推定量 (asymptotically linear regular estimator) およびモデルの定式化の検定 (specification test) について幾何学的に興味深い命題を提示した。CS によれば、これらの挙動に決定的な影響を及ぼすのが、モデルが局所的に丁度識別 (locally just identified) か局所的に過剰識別 (locally overidentified) かの違いであり、計量経済学のいくつかの例も挙げて、この分析結果は十分実用に供するとしている。

モーメント制約モデルは本問題の典型的な例で、彼らも一般論を展開する前に、おそらくモーメント制約モデルを研究したのではないと思われる。そこでここではモーメント制約モデルを使って彼らの主張を考察してみたい。まず GMM 推定量や経験尤度推定量は漸的に線形な正則推定量である。さらに J 検定が CS の言う局所的に不偏な (locally unbiased) 定式化の検定に対応している。CS では Le Cam の第 3 補題に基づき $T(P_0)^\perp$ に属する関数を使った不偏検定を考察しているが、J 検定がこれを踏まえた実行可能な (feasible) 検定になっている。なお「漸的に線形な推定量」や「正則推定量」などの概念は Bickel et al. (1993) や van der Vaart (1998) などを参照のこと。

まず最初に無条件モーメント制約モデルを検討する。

$$Eg(z; \theta_0) = 0$$

モデルを $\mathcal{P}$, 真の分布を $P_0 \in \mathcal{P}$ とし, $T(P_0)$ を $\mathcal{P}$ の P_0 における接空間, $T(P_0)^\perp$ をその直交補空間とする。 $T(P_0) \oplus T(P_0)^\perp = L_2^0(P_0)$ が成立する。

定義 6.6 $T(P_0) = L_2^0(P_0)$ または $T(P_0)^\perp = \{0\}$ ならば, P_0 は $\mathcal{P}$ により局所的に丁度識別されるという。逆に $T(P_0) \neq L_2^0(P_0)$ または $T(P_0)^\perp \neq \{0\}$ ならば, P_0 は $\mathcal{P}$ により局所的に過剰識別されるという。

以下に彼らの定理に対応するいくつかの定理を提示する。一般論としてはかなり複雑な証明になるが, モーメント制約モデルにしほれば簡明な証明が得られる。

定理 6.7 P_0 が $\mathcal{P}$ により局所的に丁度識別されるとする。

(a) $\hat{\theta}$ および $\tilde{\theta}$ を θ_0 の任意の漸近的に線形な正則推定量とする。このとき $\sqrt{n}(\hat{\theta} - \tilde{\theta}) = o_p(1)$ が成り立つ。

(b) J 検定は検出力をもたない。

証明 (a) について, これは漸近的に線形な正則推定量についての Bickel et al. (1993) の Proposition 3.3.1 の直接の結果である。つまり ψ を任意の推定量 $\hat{\theta}$ に対応する影響関数, $\tilde{\psi}$ を有効影響関数とすると, $\psi - \tilde{\psi} \perp T(P_0)$ が成り立つ。ところが丁度識別性から $T(P_0)^\perp = \{0\}$ なので $\psi = \tilde{\psi}$ であり, 求める結果が得られた。幾何学的に説明すると, $T(P_0) = L_2^0(P_0)$ なので $\tilde{\psi} = \Pi(\psi|T(P_0)) = \psi$ となるからである。(b) について, J 検定の検出力 κ は, Goto (2021a, Theorem 5) より

$$\kappa = \|\Pi(\delta' \ell_\sigma | \{g\} \cap \{\ell_\theta\}^\perp)\|_P^2 = \|\Pi(\delta' \ell_\sigma | T(P_0)^\perp)\|_P^2 \quad (6.10)$$

ここで J 検定については, 第 6 節第 3 項の設定を使った。丁度識別性から $\kappa = 0$ を得る。 ■

定理 6.7 は CS の Theorem 3.1 に対応している。丁度識別性の例として最も簡単かつ分かり易い例を挙げよう。効率的な GMM 推定量は, ウェイト行列 $\hat{\Omega}^{-1}$ をもち, その漸近分散は $(G' \hat{\Omega}^{-1} G)^{-1}$ である。また一般的な GMM 推定量はウェイト

行列として正定値行列 W をもち、その漸近分散は $(G'WG)^{-1}G'W\Omega WG(G'WG)^{-1}$ である。 $q < r$ のとき簡単な行列計算により

$$(G'\Omega^{-1}G)^{-1} \leq (G'WG)^{-1}G'W\Omega WG(G'WG)^{-1}$$

が得られる。ところが $q = r$ の場合、 $(G'\Omega^{-1}G)^{-1} = G^{-1}\Omega(G')^{-1}$ 等と変形でき、結局、左辺=右辺が成立する。一方接空間については

$$T(P_0) = \{g\}^\perp \oplus \{G'\Omega^{-1}g\}$$

であるが $G'\Omega^{-1}$ が $r \times r$ の非特異行列なので $\{G'\Omega^{-1}g\} = \{g\}$ が成り立ち、結局 $T(P_0) = L_2^0(P_0)$ つまり丁度識別性が示された。また $q = r$ の場合、J 統計量は恒等的に 0 で、そもそも意味をなさない。

ウエイト行列 W を用いた GMM 推定量 $\hat{\theta}_W$ と効率的な GMM 推定量 $\hat{\theta}$ の関係をもう少し見てみよう。 $q < r$ とする。このとき $\{G'\Omega^{-1}g\} \neq \{g\}$ であり、 $T(P_0) \neq L_2^0(P_0)$ が言え、過剰識別となる。簡単な計算により

$$\sqrt{n}(\hat{\theta}_W - \theta_0) = -(G'WG)^{-1}G'W\sqrt{n}\hat{g}(\theta_0) + o_p(1)$$

および

$$\sqrt{n}(\hat{\theta} - \theta_0) = -(G'\Omega^{-1}G)^{-1}G'\Omega^{-1}\sqrt{n}\hat{g}(\theta_0) + o_p(1)$$

が得られる。これらの推定量が漸近的に線形で正則であることが分かる。また $\hat{\theta}_W$ の影響関数 ψ_W は

$$\psi_W = -(G'WG)^{-1}G'Wg$$

また $\hat{\theta}$ の影響関数 $\tilde{\psi}$ は

$$\tilde{\psi} = -(G'\Omega^{-1}G)^{-1}G'\Omega^{-1}g$$

であり、 $\Pi(\psi_W|T(P_0)) = \tilde{\psi}$ より、 $\tilde{\psi}$ は確かに有効影響関数である。また射影定理より $\psi_W - \tilde{\psi} \in T(P_0)^\perp$ が成り立つ。

定理 6.8 P_0 が $\mathcal{P}$ により局所的に過剰識別されるとする。

(a) θ_0 の任意の漸近的に線形な正則推定量 $\hat{\theta}$ に対し、ある漸近的に線形な正

則推定量 $\tilde{\theta}$ が存在して、 $\sqrt{n}(\hat{\theta} - \tilde{\theta}) \rightsquigarrow \Delta \neq 0$ が成り立つ。

(b) J 検定は検出力をもつ。

証明 (a) については、Chen and Santos (2018, Theorem 3.2) の証明を参照のこと。

(b) については (6.10) より $\delta' \ell_{\sigma} \in T(P_0)$ でなければ J 検定は検出力をもつ。Goto (2021a) の議論から、対立仮説のモデルが識別可能ならばこの条件は満たされる。 ■

定理 6.7 と定理 6.8 を合わせたものが CS の Theorem 3.2 に対応する。なお定理 6.8 (a) の証明であるが、CS では $T(P_0)^{\perp}$ に属する関数と $\hat{\theta}$ の影響関数を組み合わせて $\tilde{\theta}$ の影響関数を作成している。これには正則性の証明等、長い手続きが必要になる。しかしモーメント制約モデルに限れば次のように証明される。 $\hat{\theta}$ が有効推定量でなければ、ワンステップ法を使って有効推定量 $\tilde{\theta}$ を構築できる。逆に $\hat{\theta}$ が有効推定量ならば、 $\psi_W - \tilde{\psi} \neq 0$ となるウエイト W を使った GMM 推定量 $\tilde{\theta}$ を作ればよい。

次に段階的 (incremental) J 検定を考察する。これは Eichenbaum et al. (1988) により提案された検定を過剰識別問題のフレームワークで考えるものである。モーメント関数ベクトルを $g = (g_1, g_2)$ と分解する。ただし g_1 および g_2 の次元はそれぞれ r_1 および r_2 で $r_1 + r_2 = r$ である。 $g_1(z; \theta)$ の定式化は正しいものとして、 $g_2(z; \theta, \sigma_0)$ の定式化の正しさを検定するものである。 g と g_1 に対応する J 統計量をそれぞれ J および J_1 とすると、段階的 J 統計量は $IJ = J - J_1$ と表される。このとき局所対立仮説 $g_2(z; \theta, \sigma_0 + \delta/\sqrt{n})$ の下で

$$IJ \rightsquigarrow \chi^2(r_2, \kappa)$$

が成り立つ。ここで

$$\kappa = \delta' [K' \Omega^{-1} K - K' \Omega^{-1} H (H' \Omega^{-1} H)^{-1} H' \Omega^{-1} K] \delta$$

かつ $H = E \partial g(z; \theta_0, \sigma_0) / \partial \theta$, $K = E \partial g(z; \theta_0, \sigma_0) / \partial \sigma$ とする。証明は Goto (2021b) を参照されたい。通常の J 検定では $J \rightsquigarrow \chi^2(r - q, \kappa)$ である。モデル g_1 の識別条件から $q \leq r_1$ なので $r_2 \leq r - q$ が成り立ち、これは通常の J 検定よりも段階

的 J 検定の方が検出力が高いことを示している。なお帰無仮説の下では、それぞれ $\mathbf{IJ} \rightsquigarrow \chi^2(r_2)$ および $\mathbf{J} \rightsquigarrow \chi^2(r-q)$ である。

CS は以下のように考えた。 g をモーメント関数ベクトルとするモデルの P_0 における接空間を $T(P_0)$, g_1 をモーメント関数ベクトルとするモデルの P_0 における接空間を $T_1(P_0)$ とすると、次の関係が得られる。

$$T(P_0) \subset T_1(P_0) \subset L_2^0(P_0)$$

$T(P_0)^\perp$ と $T_1(P_0)^\perp$ の次元は有限で、それぞれ $r-q$ と r_1-q である。 $T(P_0)^\perp$ と $T_1(P_0)^\perp$ の正規直交基底をそれぞれ $\{p_k\}_{k=1}^{r-q}$ および $\{m_k\}_{k=1}^{r_1-q}$ とする。漸近表現定理 (asymptotic representation theorem) を使って、帰無仮説の下で段階的 J 検定は次のように表現される。

$$\begin{aligned} \mathbf{IJ} &= \mathbf{J} - \mathbf{J}_1 \\ &= \sum_{k=1}^{r-q} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n p_k(z_i) \right)^2 - \sum_{k=1}^{r_1-q} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n m_k(z_i) \right)^2 + o_p(1) \\ &= \sum_{k=1}^{r_2} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n f_k(z_i) \right)^2 + o_p(1) \end{aligned}$$

ここで、 $\{f_k\}_{k=1}^{r_2}$ を $T(P_0)^\perp \cap T_1(P_0)$ の正規直交基底として、3 番目の等号が成立する。よって帰無仮説の下で

$$\mathbf{IJ} \rightsquigarrow \chi^2(r_2)$$

が得られる。 $\mathbf{IJ}$ 検定が意味をもつのは $T_1(P_0)^\perp \neq \{0\}$ のときである。 $T(P_0)^\perp \neq \{0\}$ かつ $T_1(P_0)^\perp = \{0\}$ のときは J 検定と同等になってしまう。さらに $T(P_0)^\perp = \{0\}$ のときは両者意味をもたない。なお漸近表現定理については van der Vaart (1991) を参照されたい。

最後に条件付きモーメント制約モデルの過剰識別問題を考える。

$$E[g(z; \theta_0)|x] = 0$$

g を r 次元モーメント関数ベクトル、 θ を q 次元パラメータベクトルとする。モデルの識別性について調べてみよう。

$$T(P_0) = \{g\}_x^\perp \oplus \{G(x)' \Omega(x)^{-1} g\}$$

であった。

$\phi \in T(P_0)^\perp$ ならば x の r 次元関数ベクトル $c(x)$ が存在して $\phi = c(x)' g \perp \{G(x)' \Omega(x)^{-1} g\}$ である。これは

$$0 = EG(x)' \Omega(x)^{-1} g g' c(x) = EG(x)' c(x) \quad (6.11)$$

であることと同値である。 $T(P_0)^\perp \neq \{0\}$ であるためには, (6.11) を満たす $c(x) \neq 0$ が存在することが必要十分である。 $q < r$ のときは無条件モーメントモデルと同じで, 各 x について $G(x)' c(x) = 0$ を満たす $c(x) \neq 0$ が存在し当然 (6.11) は満たされる。しかし条件付きモーメント制約モデルは無条件モーメント制約モデルと違い, $q \geq r$ の場合でも (6.11) を満たす $c(x)$ が存在する。

この条件についてももう少し詳しく検討するため, 一般的な設定で考える。

$$E[g(z; \eta_0)|x] = 0$$

g を r 次元モーメント関数ベクトル, $\eta \in \mathcal{H}$ を無限次元パラメータとする。 $\mathcal{H}$ は Banach 空間である。 η に有限次元パラメータが混在していてもよい。モデルの識別性について調べてみよう。

$$G(x)[h] = E\left[\frac{\partial}{\partial t} g(z; \eta_0 + th)\right]_{t=0}|x]$$

と定義する。ここで $h \in \mathcal{H}$ とする。モデルの接空間は

$$T(P_0) = \{g\}_x^\perp \oplus \{G(x)[h]' \Omega(x)^{-1} g : h \in \mathcal{H}\}$$

である。

$L_2(x)$ を, x の関数だけで構成される, $L_2(P_0)$ の部分空間とする。関数ベクトル $f = (f_1, \dots, f_r)'$ ただし $f_i \in L_2(x)$ ($i = 1, \dots, r$) から成る空間を $L_2(x)^r$ で表す。さらに $\tilde{f} = (\tilde{f}_1, \dots, \tilde{f}_r)'$ ただし $\tilde{f}_i \in L_2(x)$ ($i = 1, \dots, r$) として内積 $\langle f, \tilde{f} \rangle$ を定義する。

$$\langle f, \tilde{f} \rangle = \sum_{i=1}^r E f_i(x) \tilde{f}_i(x)$$

これにより $L_2(x)^r$ は Hilbert 空間になる。 $G(x)[\cdot]$ は $\mathcal{H}$ から $L_2(x)^r$ への有界線形作用素である。その値域 $\mathcal{R}$ は

$$\mathcal{R} = \{f \in L_2(x)^r : f = G(x)[h] \text{ for some } h \in \mathcal{H}\}$$

と定義される。また $L_2(x)^r$ における $\mathcal{R}$ の直交補空間を $\mathcal{R}^\perp$ で表す。次の定理が、CS の Theorem 4.1 に対応した、条件付きモーメント制約モデルの過剰識別の判定に係る命題である。

定理 6.9 $T(P_0)^\perp$ は次を満たす。

$$T(P_0)^\perp = \left\{ \phi \in L_2^0(P_0) : \phi = c(x)'g \text{ for some } c \in \mathcal{R}^\perp \right\}$$

さらに、 $T(P_0)^\perp = \{0\}$ の必要十分条件は $\mathcal{R} = L_2(x)^r$ である。

証明 前述の有限次元パラメータ θ の議論で明らか。 ■

次の補題は CS の Lemma 4.1 に対応する。

補題 6.10 $\mathcal{H}^*$ を $\mathcal{H}$ の共役空間とし、 $G(x)^*[\cdot] : L_2(x)^r \rightarrow \mathcal{H}^*$ を $G(x)[\cdot]$ の共役作用素とする。 $\mathcal{R} = L_2(x)^r$ の必要十分条件は

$$\{f \in L_2(x)^r : G(x)^*[f] = 0\} = \{0\}$$

である。

証明 作用素と共役作用素の関係から明らか。 ■

いくつかの簡単な例で確認したい。

例 6.1 (非線形回帰モデル)

次の条件付きモーメント制約モデルを考える。

$$E[y - m(x; \theta_0) | x] = 0$$

y はスカラー、 θ は q 次元ベクトル、 x は p 次元ベクトルとする。

$$G(x)[h] = -m_\theta(x; \theta_0)h$$

となり、 $\mathcal{H} = R^q$ とすると、明らかに

$$\mathcal{R} = \{f \in L_2(x) : f(x) = -m_\theta(x; \theta_0)h \text{ for some } h \in R^q\} \neq L_2(x)$$

なので、モデルは過剰識別である。

例 6.2 (ノンパラメトリック回帰モデル)

次の条件付きモーメント制約モデルを考える。

$$E[y - m_0(x)|x] = 0$$

y はスカラー、 $m(x)$ は未知の関数、 x は p 次元ベクトルとする。

$$G(x)[h] = -h(x)$$

となり、 $\mathcal{H} = L_2(x)$ とすると、明らかに

$$\mathcal{R} = \{f \in L_2(x) : f(x) = -h(x) \text{ for some } h \in L_2(x)\} = L_2(x)$$

なので、モデルは丁度識別である。

例 6.3 (部分線形モデル)

$x = (w', v)'$ として、次の条件付きモーメント制約モデルを考える。

$$E[y - w'\beta_0 - m_0(v)|x] = 0$$

y はスカラー、 β は q 次元ベクトル、 $m(v)$ は未知の関数、 v はスカラーとする。

これを部分線形モデル (partially linear model) と呼ぶ。

$$G(x)[\delta, h] = -w'\delta - h(v) \quad (6.12)$$

となり、 $\mathcal{H} = R^q \otimes L_2(v)$ とすると、明らかに

$$\begin{aligned} \mathcal{R} &= \{f \in L_2(x) : f(x) = -w'\delta - h(v) \text{ for some } (\delta, h) \in R^q \otimes L_2(v)\} \\ &\neq L_2(x) \end{aligned}$$

なので、モデルは過剰識別である。なお (6.12) から有効スコア関数を構成すれば、モデルの識別のための必要十分条件は、行列 $E(w - E[w|v])(w - E[w|v])'$ が非特異なことであることが分かる。Severini and Tripathi (2001) を参照のこと。実際、有名な Robinson (1988) の推定量もここがポイントになっている。

例 6.4 (シングルインデックスモデル)

次の条件付きモーメント制約モデルを考える。

$$E[y - m_0(x'\beta_0)|x] = 0$$

y はスカラー, β は q 次元ベクトル, $m(w)$ はスカラー w の未知の関数とする。これをシングルインデックスモデル (single index model) と呼ぶ。モデルの識別性を満たすため, $\beta = (1, \beta'_{-1})'$ とする。ただし β_{-1} は $q-1$ 次元ベクトル。同様に $x = (x_1, x'_{-1})'$ とする。ただし x_{-1} は $q-1$ 次元ベクトルである。 $\nabla m_0(\cdot)$ を $m_0(\cdot)$ の微分として

$$G(x)[\delta, h] = -\nabla m_0(x'\beta_0)x'_{-1}\delta - h(x'\beta_0)$$

となり, $\mathcal{H} = R^{q-1} \otimes L_2(w)$ とすると, 明らかに

$$\begin{aligned} \mathcal{R} &= \{f \in L_2(x) : f(x) = -\nabla m_0(x'\beta_0)x'_{-1}\delta - h(x'\beta_0) \\ &\quad \text{for some } (\delta, h) \in R^{q-1} \otimes L_2(w)\} \\ &\neq L_2(x) \end{aligned}$$

なので, モデルは過剰識別である。

例 6.5 (操作変数モデル)

次の条件付きモーメント制約モデルを考える。

$$E[y - m_0(v)|x] = 0$$

y はスカラー, $m(v)$ は未知の関数, v は内生変数, x は操作変数とする。

$$G(x)[h] = -E[h(v)|x]$$

である。 $\mathcal{H} = L_2(v)$ とする。 $z = (x', y, v')'$ とし, $L_2(v)$ は v の関数だけから成る $L_2(P_0)$ の部分空間であり, Hilbert 空間である。よって $\mathcal{H}^* = \mathcal{H}$ となる。 $G(x)[h]$ の値域は

$$\mathcal{R} = \{f \in L_2(x) : f(x) = -E[h(v)|x] \text{ for some } h \in L_2(v)\}$$

また $G(x)[h]$ の共役作用素 $G(x)^*[f] : L_2(x) \rightarrow L_2(v)$ は, $-E[f(x)|v]$ である。これは任意の関数 $\phi(x)$ および $\psi(v)$ に関し,

$$E\phi(x)\psi(v) = E_x\phi(x)E[\psi(v)|x] = E_vE[\phi(x)|v]\psi(v)$$

の関係から容易に出る。ここで補題 6.10 を使うと、作用素 $E[f(x)|v]$ が単射 (injection) の場合 $\mathcal{R} = L_2(x)$ でありモデルは丁度識別、そうでない場合過剰識別になる。 $E[f(x)|v]$ が単射であることは、 (x, v) の分布が v に関して L_2 -完備 (L_2 -complete) であることと同値である。なお識別条件自体は、分布が x に関して完備なことである。指数型分布族 (exponential family) における L_2 -完備の例は、Newey and Powell (2003) で扱っている。

以上簡単な例を見てきたが、さらに複雑かつ具体的事例に基づくものに興味のある読者は、Chen and Santos (2018) を参照されたい。

参考文献

- Ai, C., and X. Chen (2003): “Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions,” *Econometrica*, 71, 1795-1843.
- Amemiya, T. (1985): *Advanced Econometrics*. Cambridge, Mass.: Harvard University Press.
- Bickel, P. J., C. A. J. Klaassen, Y. Ritov, and J. A. Wellner (1993): *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins University Press.
- Chamberlain, G. (1987): “Asymptotic Efficiency in Estimation with Conditional Moment Restrictions,” *Journal of Econometrics*, 34, 305-334.
- Chen, X., and A. Santos (2018): “Overidentification in Regular Models,” *Econometrica*, 86, 1771-1817.
- Dominguez, M. A., and I. N. Lobato (2004): “Consistent Estimation of Models Defined by Conditional Moment Restrictions,” *Econometrica*, 72, 1601-1615.
- Donald, S. G., G. W. Imbens, and W. K. Newey (2003): “Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions,” *Journal of Econometrics*, 117, 55-93.
- Eichenbaum, M. S., L. P. Hansen, and K. J. Singleton (1988): “A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice Under Uncertainty,” *The Quarterly Journal of Economics*, 103, 51-78.
- Goto, F. (2021a): “Large Sample Properties of GMM and GEL Estimators: A Geometric Approach,” *The Aoyama Journal of Economics*, Vol. 73, No. 1, 1-33.
- Goto, F. (2021b): “Hypothesis Testing with GMM and GEL Estimators,” *The Aoyama Journal of Economics*, Vol. 73, No. 2, 75-103.
- Hamilton, J. (1994): *Time Series Analysis*. Princeton: Princeton University Press.
- Han, C., and P. C. B. Phillips (2006): “GMM with Many Moment Conditions,” *Econometrica*, 74, 147-192.
- Hansen, L. P. (1982): “Large Sample Properties of Generalized Method of Moments Estimators,” *Econometrica*, 50, 1029-1054.
- Hayashi, F. (2000): *Econometrics*. Princeton: Princeton University Press.
- Kitamura, Y., G. Tripathi, and H. Ahn (2003): “Empirical Likelihood-Based Inference in Conditional Moment Restriction Models,” *Econometrica*, 72, 1667-1714.

- Koenker, R., and J. A. F. Machado (1999): “GMM Inference When the Number of Moment Conditions Is Large,” *Journal of Econometrics*, 93, 327-344.
- Newey, W. K. (1985): “Generalized Method of Moments Specification Testing,” *Journal of Econometrics*, 29, 229-256.
- Newey, W. K. (1990): “Efficient Instrumental Variables Estimation of Nonlinear Models,” *Econometrica*, 58, 809-837.
- Newey, W. K., and J. L. Powell (2003): “Instrumental Variable Estimation of Nonparametric Models,” *Econometrica*, 71, 1565-1578.
- Newey, W. K., and R. J. Smith (2004): “Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators,” *Econometrica*, 72, 219-255.
- Owen, A. B. (1988): “Empirical Likelihood Ratio Confidence Intervals for a Single Functional,” *Biometrika*, 75, 237-249.
- Owen, A. B. (2001): *Empirical Likelihood*. London: Chapman and Hall.
- Qin, J., and J. Lawless (1994): “Empirical Likelihood and General Estimating Equations,” *Annals of Statistics*, 22, 300-325.
- Robinson, P. M. (1988): “Root-N-Consistent Semiparametric Regression,” *Econometrica*, 56, 931-954.
- Severini, T. A., and G. Tripathi (2001): “A Simplified Approach to Computing Efficiency Bounds in Semiparametric Models,” *Journal of Econometrics*, 102, 23-66.
- Severini, T. A., and W. H. Wong (1992): “Profile Likelihood and Conditionally Parametric Models,” *Annals of Statistics*, 20, 1768-1802.
- Sueishi, N. (2016): “A Simple Derivation of the Efficiency Bound for Conditional Moment Restriction Models,” *Economics Letters*, 138, 57-59.
- van der Vaart, A. W. (1991): “An Asymptotic Representation Theorem,” *International Statistical Review*, 59, 97-121.
- van der Vaart, A. W. (1998): *Asymptotic Statistics*. Cambridge: Cambridge University Press.
- 後藤文廣 (2022): 「セミパラメトリック推測の理論と応用」『青山経済論集』第 74 巻第 3 号, pp. 3-70, 青山学院大学経済学会.