セミ・ノンパラメトリックモデルと内生性

末　石　直　也

Ⅰ　はじめに

この 20 年程の間に，セミ・ノンパラメトリックな統計的手法は目覚しい発展を遂げており，計量経済学においても重要性を増している。その理由のひとつとして，経済理論からは必ずしも変数間のパラメトリックな関係を導き出すことができない点が挙げられる。定式化を誤ると，推定量にバイアスが生じるため，極力仮定に依存しない，ロバストな方法が好まれている。また，その他の要因として，近年のデータ環境の整備やコンピュータの発達なども挙げられよう。セミ・ノンパラメトリックな手法は，その特性上多くのデータを必要とし，かつコンピュータ・インテンシブな手法であるため，かつては実行が困難であったが，今や実証分析においても頻繁に用いられている。

従来のセミ・ノンパラメトリックな推定問題は，専ら誘導型に関するものであった。しかし，経済学においてはしばしば，回帰変数の内生性，つまり説明変数と誤差項の相関の存在が問題となる。内生性は社会科学固有の問題であり，内生性が存在する場合の統計的推測に関する一連の研究は，計量経済学による統計学への大きな貢献のひとつである。特に，線形同時方程式モデルに関する研究は，計量経済学の黎明期より現在に至るまで中心的課題であった。1990 年代後半以降，線形モデルに限らず，セミ・ノンパラメトリックなモデルにおいても，内生性を考慮したモデルが考察されるようになり，現在，計量経済学において最も活発な研究が行われている分野のひとつである。

本論文は，内生性を考慮したセミ・ノンパラメトリックモデルの推定に関する最近の研究成果のサーベイ論文である。第Ⅱ章では，内生性が存在するときのノンパラメトリックモデルの識別性と推定量の一致性に関する一般的な問題について論じる。第Ⅲ章は未知関数を推定する代表的な手法のひとつである sieve 推定について論じる。第Ⅳ章ではノンパラメトリック操作変数法について考察し，第Ⅴ章ではセミパラメトリックモデルの推定方法を紹介する。

Ⅱ　識別性と一致性

以下では構造方程式の識別性と推定量の一致性について概要を論じ，パラメトリックモデルとノンパラメトリックモデルにおける相違点について考察する。Blundell and Powell［2003］に倣い，はじめに単純な線形モデル

$$Y_i = X_i'\beta + U_i \qquad (2.1)$$

を考える。データ $\{(Y_i, X_i)\}_{i=1}^n$ は iid で，Y_i はスカラー，X_i は k 次元のベクトル，β は k 次元の未知パラメータ，U_i は観測されない誤差項である。X_i は内生変数を含む，つまり，$E[X_iU_i] \neq 0$ であるとする。$X_i'\beta$ を Y_i の最良線形近似（線形射影）として解釈すれば，$E[X_iU_i]=0$ を満たす β は常に存在するので，(2.1) は単なる近似式ではなく，何らかの経済の構造を表現していると解釈すべきである。それゆえ，(2.1) は構造方程式と呼ばれる。

内生性が存在するとき，最小 2 乗推定量は一致性を持たない。パラメータ β を識別，推定す

るためには，操作変数の存在が仮定される。つまり

$$E[Z_iU_i]=0 \tag{2.2}$$

を満たす変数 Z_i が存在するものとする。ただし，$\dim(Z_i)\geq k$ である。すると，(2.1) と (2.2) より

$$0=P[U|Z]=P[Y|Z]-P[X|Z]\beta$$

あるいは，同じことであるが，

$$P[Y|Z]=P[X|Z]\beta \tag{2.3}$$

が満たされる。ただし，$P[Y|Z]$ と $P[X|Z]$ はそれぞれ，Y と X の Z への線形射影の係数であり，

$$P[Y|Z]=E[ZZ']^{-1}E[ZY]$$
$$P[X|Z]=E[ZZ']^{-1}E[ZX']$$

である。$P[Y|Z]$ 及び $P[X|Z]$ は常に識別可能であり，β の識別は (2.3) の解の一意性に帰着される。識別の必要十分条件は階数条件として知られている。仮に $P[X|Z]$ の逆行列が存在すれば，β は一意に定まり，間接最小2乗推定量，あるいは操作変数推定量

$$\hat{\beta}_{IV}=\hat{P}[X|Z]^{-1}\hat{P}[Y|Z]$$

によって推定できる。ただし，$\hat{P}[Y|Z]$ と $\hat{P}[X|Z]$ はそれぞれ $P[Y|Z]$ と $P[X|Z]$ の標本バージョンであり，

$$\hat{P}[Y|Z]=\left(\frac{1}{n}\sum_{i=1}^{n}Z_iZ_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}Z_iY_i$$
$$\hat{P}[X|Z]=\left(\frac{1}{n}\sum_{i=1}^{n}Z_iZ_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}Z_iX_i'$$

である。操作変数推定量 $\hat{\beta}_{IV}$ は $\hat{P}[Y|Z]$ 及び $\hat{P}[X|Z]$ について連続な関数なので，モデルが識別されていれば，$\hat{\beta}_{IV}$ の一致性は，$\hat{P}[Y|Z]$ と $\hat{P}[X|Z]$ の一致性により保障される。

次に，以下のようなノンパラメトリックモデルを考えよう。

$$Y_i=g(X_i)+U_i \tag{2.4}$$

ただし，g は未知の関数であり，特に関数形を仮定しない。また，$E[U_i|X_i]\neq 0$ である。操作変数法をノンパラメトリックモデルに適用するためには，(2.2) では条件が弱すぎるため，より強い仮定である

$$E[U_i|Z_i]=0$$

を満たすような操作変数 Z_i の存在が必要となる。(2.4) の両辺の条件付期待値を取ると，(2.3) に対応する条件

$$E[Y|Z]=E[g(X)|Z]=\int g(x)dF_{X|Z} \tag{2.5}$$

が成り立つ。ここで $F_{X|Z}$ は X の Z を条件とした条件付分布である。Y の誘導型である $E[Y|Z]$ と条件付分布 $F_{X|Z}$ は識別可能である。

今，K を積分作用素 $(Kg)(z)=\int g(x)dF_{X|Z=z}$，$r(z)=E[Y|Z=z]$ と定義すると，(2.5) より未知関数 g は積分方程式

$$(Kg)(z)=r(z) \tag{2.6}$$

の解として表され，識別性は積分方程式の解の存在と一意性に帰着される。線形モデルの場合には，識別性は (2.3) における行列の階数条件で表されるが，(2.5) の識別性の十分条件は

$$\mathcal{N}(K)\equiv\{g\in\mathcal{G}:Kg=0\}=\{0\} \tag{2.7}$$

である。ただし $\mathcal{G}$ は g が存在する関数空間で，$\mathcal{G}$ が2乗可積分な関数空間であれば，(2.7) は X の Z を条件とした分布が完備であることと同等である[1)]。

(2.6) の形式で与えられる積分方程式は，数学的には第1種フレドホルム型積分方程式と呼ばれるものである。g は逆問題の解として与えられる。一般に，逆問題は，(i)解の存在，(ii)解の一意性，(iii)解の連続性の3つの条件を満たすとき適切（well-posed）であると言われる。逆に，これらの条件が満たされないときには，非適切（ill-posed）であると言われる。第1種フレドホルム型積分方程式 (2.6) の問題点は，逆問題が非適切であること，特に解 g が r について不連続であることにある。

今，g が識別されていると仮定し，g を推定す

1）Newey and Powell［2003］を参照。

るために，(2.5) の $E[Y|Z]$ を一致推定量 $\hat{E}[Y|Z]$ で置き換えると[2]，

$$\hat{E}[Y|Z]=\int \hat{g}(x)dF_{X|Z} \tag{2.8}$$

を得る。(2.8) に解が存在するならば，解 $\hat{g}$ は

$$\hat{g}(z)=K^{-1}(\hat{r}(z))$$

で与えられる。ただし，$\hat{r}(z)=\hat{E}[Y|Z=z]$ である。しかし，逆作用素 K^{-1} は引数 $\hat{r}(z)$ について連続ではないため，$\hat{r}(z)$ の一致性からは必ずしも，$\hat{g}(z)$ の一致性が保障されない。これが内生性が存在するときの，線形モデルとノンパラメトリックモデルの大きな違いである。線形モデルでは，モデルが識別されれば，誘導型の一致性から推定量の一致性が保障されたが，ノンパラメトリックモデルでは誘導型の一致性だけからは，必ずしも $\hat{g}$ の一致性が保障されないのである。さらに，一般に積分作用素 K は未知なので，K も推定しなければならず，事態はさらに深刻である。

構造方程式の一致推定量を得るためには，逆問題の解が連続である必要がある。連続性を保障するための代表的な方法には，g の存在する関数空間にコンパクト性を課す方法と，チコノフの正則化と呼ばれる方法がある。これらについては第Ⅳ章で改めて論じることにする。

Ⅲ Sieve 推定

内生性の問題はひとまず脇に置いて，次のようなノンパラメトリック回帰モデルを考える。

$$Y_i=g(X_i)+\varepsilon_i,\ E[\varepsilon_i|X_i]=0 \tag{3.1}$$

ただし，$\{(Y_i,\ X_i)\}_{i=1}^n$ は iid で，Y_i はスカラーであるとする。(3.1) より $g(x)=E[Y|X=x]$ である。目的は未知関数 $g(x)$ を推定することである。

関数をノンパラメトリックに推定する方法には，大別してカーネル推定と sieve 推定の２つがある。本章では，基底関数を用いる推定方法である sieve 推定を紹介する。sieve 推定の背後にある基本的なアイデアは，充分に滑らかな関数は，基底関数の線形和によって任意の精度で近似が可能であるということである。関数を基底関数の線形和で置き換えることにより，関数を推定するという無限次元の問題は，線形関数の係数を推定するという有限次元の問題に帰着される。

推定方法を議論する前に，$g(x)$ が既知であるとして，基底関数の線形和によって近似することを考える。単純化のために，x は１変数とし，x のサポートは $[0,\ 1]$ であるとする。$g(x)$ が充分に滑らかな関数である場合，例えば以下のような有限次元の線形の sieve によって，$g(x)$ は任意の精度で近似が可能である（Chen［2006］の２節を参照)。

●べき関数

$$\mathrm{Pol}(J)=\left\{\sum_{k=0}^{J}\beta_k x^k,\ x\in[0,\ 1]:\beta_k\in\mathbb{R}\right\}$$

●スプライン

$$\mathrm{Spr}(r,\ J)=$$
$$\left\{\sum_{k=0}^{r-1}\beta_k x^k+\sum_{j=1}^{J}\gamma_j[\max\{x-t_j,\ 0\}]^{r-1},\ x\in[0,\ 1]:\beta_k,\ \gamma_j\in\mathbb{R}\right\}$$

スプラインの t_j はノットと呼ばれ，$0=t_0<t_1<\cdots<t_J<t_{J+1}=1$ を満たす。sieve による近似の精度は J 及び r を増加させることによって向上するが，様々な sieve について，収束の速度が知られている。べき関数を例にとると，$g(x)$ が p 階連続微分可能であれば，ある $\beta_0, \ldots, \beta_J$ が存在して

$$\sup_{x\in[0,\ 1]}\left|\sum_{k=0}^{J}\beta_k x^k-g(x)\right|=O(J^{-p})$$

が成り立つ。

次に，$g(x)$ は未知であるとして，sieve を用いた推定方法を考える。$p^K(x)=(p_1(x), \ldots, p_K(x))'$ を K 次元の基底関数のベクトルとする。例えば，べき関数を用いれ

2）条件付期待値のノンパラメトリックな推定方法については，第Ⅲ章で論じる。

ば，$p^K(x)=(1,x,x^2,\ldots,x^{K-1})'$ である。(3.1)の $g(x)$ を $p^K(x)$ の線形和で近似することで

$$Y_i=p^K(X_i)'\beta+\eta_i$$

と書ける。したがって，未知関数 $g(x)$ を推定する問題は，未知係数 β を推定する問題に帰着される。β は最小２乗法によって推定可能であり，

$$\hat{\beta}=(P'P)^{-1}P'Y$$

を得る。ただし，$P=(p^K(X_1),\ldots,p^K(X_n))'$，$Y=(Y_1,\ldots,Y_n)'$ である。$g(x)$ の推定量は

$$\hat{g}(x)=p^K(x)'\hat{\beta}$$

で与えられる。

Newey［1997］は一般的な条件の下で，sieve 推定量 $\hat{g}(x)$ の $g(x)$ への収束のレートについて論じている。有限標本における基底関数のベクトルの次元 K の選び方に関しては，Li［1987］などを参照されたい。一般に，K を大きく選ぶことによりバイアスは小さくなるが，推定するパラメータの数が多くなるため，$\hat{g}(x)$ の分散は大きくなる。したがって，例えば平均２乗誤差を最小化するような K の選び方が必要となる。

Ⅳ　ノンパラメトリック操作変数法

説明変数と誤差項に相関が存在する場合のノンパラメトリックモデルの推定方法について考察する。特に，有限次元の基底関数を用いる Newey, Powell, and Vella［1999］及び Newey and Powell［2003］による推定方法と，チコノフの正則化を用いる Hall and Horowitz［2005］の方法を紹介する。

基本的なモデルは

$$Y_i=g(X_i)+U_i$$

という形式で与えられる。前章との違いは，$E[U_i|X_i]\neq 0$ であることである。$g(x)$ は条件付期待値という単なる数学的な対象ではなく，ある経済の構造を反映した構造型であると解釈される。誤差項と説明変数の間に相関があるため，第Ⅲ章で紹介した方法を直接用いることはできない。この問題を解決するため，$E[U_i|Z_i]=0$ を満たすような操作変数 Z_i の存在が仮定される。

1　Newey, Powell, and Vella［1999］

次のような同時方程式モデルを考察する。

$$Y_i=g(X_i,\ Z_{1i})+U_i \tag{4.1}$$

$$X_i=\Pi(Z_i)+V_i \tag{4.2}$$

ただし，X_i は内生変数，$Z_i=(Z'_{1i},\ Z'_{2i})'$ は操作変数であり，$E[V_i|Z_i]=0$ を満たすものとする。したがって，(4.2)は誘導型である。(4.1)の両辺の X_i と Z_i を条件とする条件付期待値を取ると

$$E[Y_i|X_i,\ Z_i]=g(X_i,\ Z_{1i})+E[U_i|X_i,\ Z_i] \tag{4.3}$$

が成り立つ。ここで，(4.2)より X_i は Z_i 及び V_i の関数であるので，X_i と Z_i について条件付けるのは V_i と Z_i に関して条件付けるのと同等である。したがって，$E[U_i|X_i,\ Z_i]=E[U_i|V_i,\ Z_i]$ を得る。さらに，強い意味での外生性，つまり

$$E[U_i|V_i,\ Z_i]=E[U_i|V_i] \tag{4.4}$$

を仮定する。以上の仮定の下で(4.3)は

$$\begin{aligned} E[Y_i|X_i,\ Z_i]&=g(X_i,\ Z_{1i})+E[U_i|V_i]\\ &=g(X_i,\ Z_{1i})+\lambda(V_i)\equiv h(W_i) \end{aligned} \tag{4.5}$$

と書き換えられる。ただし，$\lambda(V_i)=E[U_i|V_i]$ で，$W_i=(X_i,\ Z_{1i},\ V_i)$ である。(4.5)より，誤差項と説明変数が相関を持たない通常の回帰モデル

$$Y_i=h(W_i)+\varepsilon_i,\ E[\varepsilon_i|W_i]=0$$

を得ることができる。V_i は観測できないが推定可能である。

Newey, Powell, and Vella［1999］は(4.5)に基づく２段階推定方法を提案した。第１段階で(4.2)を sieve 推定する。(4.2)は誘導型なので識別，推定が可能である。$r^L(z)=(r_{1L}(z),\ldots,r_{LL}(z))'$ を基底関数のベクトルとすると，$\Pi(z)$ の推定量

$\hat{\Pi}(z)=r^L(z)'\hat{\gamma},\ \hat{\gamma}=(R'R)^{-1}R'(X_1,\ \ldots,\ X_n)'$ (4.6)
を得る。ただし，$R=(r^L(Z_1), \ldots, r^L(Z_n))'$ である。(4.6) より，V_i の推定量 $\hat{V}_i=X_i-\hat{\Pi}(Z_i)$ が得られる。

次に，$\hat{V}_i$ を用いて $h(w)$ を推定する。$p^K(w)=(p_{1K}(w), \ldots, p_{KK}(w))'$ を $w=(x,\ z_1,\ v)$ の基底関数のベクトルとする。ただし，$p_{kK}(w)$ は $(x,\ z_1)$ あるいは v のみの関数であり，３変数 $(x,\ z_1,\ v)$ に同時に依存することはできないものとする。$\hat{W}_i=(X_i,\ Z_{1i},\ \hat{V}_i)$，$\hat{p}_i^K=p^K(\hat{W}_i)$ とする。Y_i を $\hat{p}_i^K$ に回帰することにより

$$\hat{h}(w)=p^K(w)'\hat{\beta},\ \hat{\beta}=(\hat{P}'\hat{P})^{-1}\hat{P}'(Y_1\ \ldots, Y_n)'$$

が求まる[3]。ただし，$\hat{P}=(\hat{p}_1^K, \ldots, \hat{p}_n^K)'$ である。

最後に $\hat{h}(w)$ を用いて，$g(x,\ z_1)$ と $\lambda(v)$ を推定する。$g(x,\ z_1)$ は $(x,\ z_1)$ にのみ依存するので，$\hat{h}(w)$ の $(x,\ z_1)$ に依存する項のみを集めることで推定できる。$\lambda(v)$ も同様である。今，$p_{1K}(w)=1$ とし，続く K_g 項，$p_{2K}(w), \ldots, p_{K_g+1K}(w)$ は $(x,\ z_1)$ のみの関数，残りの項は v のみの関数とする。$g(x,\ z_1)$ と $\lambda(v)$ の推定量は

$$\hat{g}(x,\ z_1)=\hat{c}_g+\sum_{j=2}^{K_g+1}\hat{\beta}_i p_{jK}(x,\ z_1)$$

$$\hat{\lambda}(v)=\hat{c}_\lambda+\sum_{j=K_g+2}^{K}\hat{\beta}_j p_{jK}(v)$$

となる。ただし，$\hat{c}_g+\hat{c}_\lambda=\hat{\beta}_1$ である。これらの推定量は定数項 $\hat{c}_g$ と $\hat{c}_\lambda$ を除いて一意に決定される。$\hat{h}$ の収束のレートが求められ，各点での漸近正規性が示されている。

２　Newey and Powell［2003］

Newey and Powell［2003］は Newey, Powell, and Vella［1999］で用いられた仮定 (4.4) を緩め，次のような一般的なモデルについて考察している。

$$Y_i=g(X_i,\ Z_{1i})+U_i,\ E[U_i|Z_i]=0,$$
$$Z_i=(Z_{1i}',\ Z_{2i}')' \qquad (4.7)$$

(4.7) の両辺の $Z=z$ を条件とする条件付期待値を取ると，積分方程式

$$r(z)\equiv E[Y|Z=z]=\int g(x,\ z_1)dF_{X|Z=z}\equiv(Kg)(z) \qquad (4.8)$$

が成り立つ。しかし，第Ⅱ章で解説した非適切な問題のため，一般的には $g(x,\ z_1)$ の一致推定量を得ることはできない。そこで，$g(x,\ z_1)$ がソボレフノルムに関してコンパクトな関数空間の要素であることを仮定する。関数空間のコンパクト性を仮定することで，K^{-1} が $r(z)$ について連続となるので，一致推定が可能となる。

今，$w=(x,\ z_1)$ とし，$\{p_1(w),\ p_2(w),\ \ldots\}$ を基底関数の列とする。さらに，$g(w)$ が

$$g(w)\cong\sum_{j=1}^{J}\gamma_j p_j(w)$$

によって近似されると仮定する。すると，(4.8) より

$$E[Y_i|Z_i]\cong\sum_{j=1}^{J}\gamma_j E[p_j(W_i)|Z_i] \qquad (4.9)$$

が成立する。

(4.9) から次のような２段階推定法が示唆される。まず，$p_j(W_i)$ を z の基底関数に回帰することによって，sieve 推定量 $\hat{E}[p_j(W_i)|Z_i]$ を得る。次に，Y_i を $\hat{E}[p_j(W_i)|Z_i]$ に回帰することによって，γ_j の推定量 $\hat{\gamma}_j$ が得られる。ただし，$g(w)$ の関数形について制約が必要であり，コンパクト性を保障するため，係数のサイズに上限を設けて γ を推定するよう提案されている。詳細は論文を参照されたい。２段階推定量の一致性は示されているが，収束のレートは求められておらず，各点での漸近正規性も示されていない。また有限標本において，w 及び z の基底関数のベクトルの次数をどのように選ぶべきかが議論されていない。

3）Newey, Powell, and Vella［1999］では trimming を行っているが，単純化のために省略した。

3　Hall and Horowitz［2005］

Hall and Horowitz［2005］はチコノフの正則化を用いた推定方法を提案している。論文中ではカーネルを用いた推定方法と sieve を用いた推定方法の２つを提案しているが，本章では sieve を用いた方法のみを紹介する。さらに単純化のため，内生変数 X_i と操作変数 Z_i はスカラーとし，サポートは $[0, 1]$ であるとする。

モデルは次のとおりである。

$$Y_i = g(X_i) + U_i, \quad E[U_i|Z_i] = 0 \tag{4.10}$$

今，f_X，f_Z，f_{XZ} をそれぞれ，X の周辺密度，Z の周辺密度，X と Z の同時密度とする。また，$[0, 1]$ 上で定義された２乗可積分な関数空間上の作用素 T を

$$(T\psi)(w) = \int t(x, w)\psi(x)dx$$

とする。ただし

$$t(x, w) = \int f_{XZ}(x, z) f_{XZ}(w, z)dz$$

である。(4.10) より

$$(Tg)(w) = E_Z[E[Y|Z] f_{XZ}(w, Z)] \tag{4.11}$$

が成り立つ。作用素 T は正則であると仮定すると，

$$g(x) = E_Z[E[Y|Z](T^{-1}f_{XZ})(x, Z)]$$

を得る。しかし，(4.11) は第１種フレドホルム型積分方程式なので，逆問題は非適切である。積分作用素 T は０を固有値の集積点に持ち，T^{-1} は有界ではないため連続ではない。この問題を解決するために，適当なリッジ・パラメータ a_n を用いて，T^{-1} を $(T+a_n)^{-1}$ で置き換えることが提案されている。

基本的なアイデアは，X 及び Z に適当な変換を施すことにより，X と Z の周辺分布を $[0, 1]$ 上の一様分布にすることにある。まず，X と Z が $[0, 1]$ 上の一様分布に従うと仮定しよう。$\{\phi_1, \phi_2 \ldots\}$ を $[0, 1]$ 上の２乗可積分な関数空間の正規直交系とする。f_{XZ} 及び g は次のようにフーリエ展開可能である。

$$f_{XZ}(x, z) = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} q_{jk}\phi_j(x)\phi_k(z)$$

$$g(x) = \sum_{j=1}^{\infty} \gamma_j \phi_j(x)$$

$Q=(q_{jk})$，$p_j = E[Y\phi_j(Z)]$，$p=(p_j)$，$\gamma=(\gamma_j)$ と定義する。(4.10) より $QQ'\gamma = Qp$，したがって，

$$\gamma = (QQ')^{-1}Qp$$

が成立する。したがって，Q 及び p を推定することにより，g のフーリエ係数である γ を推定することができる。

次に一様分布の仮定をはずす。$\hat{F}_Z$ と $\hat{F}_X$ を $Z_1, \ldots, Z_n$ と $X_1, \ldots, X_n$ の経験分布関数とする。$\hat{Z}_i = \hat{F}_Z(Z_i)$，$\hat{X}_i = \hat{F}_X(X_i)$ とすれば，q_{jk} ならびに p_j は

$$\hat{q}_{jk} = \frac{1}{n}\sum_{i=1}^{n} \phi_j(\hat{Z}_i)\phi_k(\hat{X}_i)$$

$$\hat{p}_j = \frac{1}{n}\sum_{i=1}^{n} Y_i \phi_j(\hat{Z}_i)$$

によって推定できる。$\hat{Q}$ を $\hat{q}_{jk}$ を (j, k) 要素に持つ $J \times J$ 行列とする。リッジ・パラメータ a_n を用いて，

$$\hat{\gamma} = (\hat{Q}\hat{Q}' + a_n I_J)^{-1}\hat{Q}\hat{p}$$

により γ を推定できる。ただし I_J は単位行列である。したがって，g の推定量は次のようになる。

$$\hat{g}(x) = \sum_{j=1}^{J} \hat{\gamma}_j \phi_j(x) \tag{4.12}$$

ある種の正則条件の下で，g の推定量が到達できる平均積分２乗誤差（mean integrated squared error）の収束レートのミニマックスリスクの下限が求められ，(4.12) はその下限に到達できることが示されている。Horowitz［2007］では，推定量が漸近正規性を持つための条件について論じている。

4　まとめ

ノンパラメトリック操作変数法の代表的な方法について解説した。紹介した方法の他にも，Darolles, Florens, and Renault［2002］や Gagliardini and Scaillet［2007］などがある。これらの方法は異なる正則条件の下で推定量の一致性と収束のレートについて議論しているため，比

較が困難である。Chen and Reiss［2007］が統一的なフレームワークの下で，平均積分２乗誤差のミニマックスリスクの下限について論じている。

Ⅴ　セミパラメトリックモデル

未知パラメータが有限次元のベクトルと無限次元の未知関数からなるセミパラメトリックモデルの推定方法を紹介する。ノンパラメトリックモデルにおいては未知の関数を推定することに興味があったが，セミパラメトリックモデルでは興味の中心は主として有限次元のパラメータにあり，未知関数はしばしば直接的には興味のない局外パラメータである。

本章前半では，Ai and Chen［2003］と Newey and Powell［2003］で提案された SMD 推定量について論じる。本章後半では，経験尤度法（empirical likelihood）を用いた推定量である Otsu［2007］の SCEL 推定量と Sueishi［2008］の SGEL 推定量を紹介する。

1　SMD 推定量

モデルは以下で与えられる。

$$E[\rho(X_i,\ \theta_0,\ h_0)|Z_i]=0 \tag{5.1}$$

$\{(Y_i,\ Z_i)\}_{i=1}^n$ は iid で，Z_{Xi} は Z_i の部分集合，$X_i=(Y_i,\ Z_{Xi})$，関数 ρ は d_ρ 次元の既知の（誤差）関数である。Y_i は内生変数で，Z_i は操作変数である。未知パラメータ $\alpha_0=(\theta_0,\ h_0)\in\Theta\times\mathcal{H}$ は，有限次元のベクトル θ_0 と無限次元の未知関数 h_0 からなる。$\mathcal{H}$ は連続関数からなる関数空間である。未知関数 h_0 は内生変数 Y_i に依存してもよい。

(5.1) は多くの既存のモデルを含んでいる。例えば，Robinson［1988］などで研究された部分線形モデルや，Ichimura［1993］などで考察されたインデックス・モデルなどがある。しかし，これらの例では未知関数は外生変数のみに依存し，(5.1) で与えられるモデルはより広いクラスのモデルである。

Ai and Chen［2003］と Newey and Powell［2003］は独立に，(5.1) の推定方法として sieve minimum distance（SMD）推定量を提案した。SMD 推定量のアイデアは以下のとおりである。$m(Z,\ \alpha)\equiv E[\rho(X,\ \alpha)|Z]$ とすると，(5.1) より，真のパラメータ $\alpha_0=(\theta_0,\ h_0)$ は目的関数

$$E[m(Z,\ \alpha)'\textstyle\sum(Z)^{-1}m(Z,\ \alpha)] \tag{5.2}$$

を最小化することがわかる。ただし，$\sum(Z)$ は適当なウエイト行列である。したがって，(5.2) の標本バージョンを考えることにより，次のような最小距離推定量を得ることができる。

$$\hat{\alpha}=\arg\min_{\alpha=(\theta,\ h)\in\Theta\times\mathcal{H}}\frac{1}{n}\sum_{i=1}^n m(Z_i,\ \alpha)'\hat{\textstyle\sum}(Z_i)^{-1}m(Z_i,\ \alpha) \tag{5.3}$$

しかし，(5.3) は２つの問題点を抱えている。第１に，条件付期待値 $m(z,\ \alpha)$ の関数形は未知であるために，(5.3) は実際には推定できない。第２の問題点は，最小化問題が関数空間上で定義されていることに起因する。もし関数空間 $\mathcal{H}$ が大きすぎる場合には，仮に $m(z,\ \alpha)$ の関数形が既知であっても，推定量は一致性を持たないか，あるいは一致性を持ったとしても，収束の速度が非常に遅いということが知られている。

第１の問題を回避するため，Ai and Chen［2003］と Newey and Powell［2003］は条件付期待値 $m(z,\ \alpha)$ を sieve を用いて推定することを提案している。$p^{k_n}(z)=(p_{01}(z),\ \ldots,\ p_{0k_n}(z))'$ を基底関数のベクトルとし，$m(z,\ \alpha)=(m_1(z,\ \alpha),\ \ldots,\ m_{d_\rho}(z,\ \alpha))'$ とする。$m(z,\ \alpha)$ の第 l 要素の sieve 推定量は，

$$\hat{m}_l(z,\ \alpha)=\sum_{j=1}^n\rho_l(X_j,\ \alpha)p^{k_n}(Z_j)(P'P)^{-1}p^{k_n}(z) \tag{5.4}$$

で与えられる。ただし，$P=(p^{k_n}(Z_1),\ \ldots,\ p^{k_n}(Z_n))'$ である。

第２の問題についても，未知関数を sieve により推定することで解決できる。$q^{k_{1n}}(x)$ を k_{1n}

次元の基底関数のベクトルとすると，$h(x)$ は基底関数の線形結合 $q^{k_{1n}}(x)'\beta$ により近似できるので，

$$\rho_l(x,\ \theta,\ h)\cong\rho_l(x,\ \theta,\ q^{k_{1n}}(x)'\beta) \tag{5.5}$$

が成り立つ。(5.4) の $\rho_l(x,\ \alpha)$ を (5.5) の右辺で置き換えることにより，関数空間 $\mathcal{H}$ 上の最小化問題は，適当な sieve 空間 $\mathcal{H}_n$ 上の最小化問題へと変換され，関数 h の推定問題は有限次元パラメータ β の推定問題となる。以上をまとめると，SMD 推定量は

$$\hat{\alpha}=\arg\min_{\alpha=(\theta,\ h)\in\Theta\times\mathcal{H}_n}\frac{1}{n}\sum_{i=1}^{n}\hat{m}(Z_i,\ \alpha)\hat{\Sigma}(Z_i)^{-1}\hat{m}(Z_i,\ \alpha) \tag{5.6}$$

と定義される。

SMD 推定量は GMM 推定量と解釈することもできる。(5.6) において，$\hat{\Sigma}(Z_i)=I$ とすると，SMD 推定量の最小化問題は

$$\min_{\alpha=(\theta,\ h)\in\Theta\times\mathcal{H}_n}\left(\sum_{i=1}^{n}\rho(X_i,\ \alpha)\otimes p^{k_n}(Z_i)\right)'(I\otimes(PP)^{-1})\left(\sum_{i=1}^{n}\rho(X_i,\ \alpha)\otimes p^{k_n}(Z_i)\right)$$

と書き換えられる。この目的関数は，無条件モーメント制約

$$E[\rho(X_i,\ \alpha_0)\otimes p^{k_n}(Z_i)]=0$$

について，ウエイト行列 $I\otimes(P'P)^{-1}$ を用いた GMM 推定量の目的関数そのものである。

Newey and Powell［2003］は適当な正則条件の下で，SMD 推定量の一致性を示した。Ai and Chen［2003］は，有限次元パラメータの推定量 $\hat{\theta}$ が漸近的に正規分布に従うことを証明した。一般に，有限次元パラメータの推定量の漸近正規性の条件として，ノンパラメトリックな推定量の収束の速度は $n^{-1/4}$ よりも早くなければならないことが知られている[4)]。しかし，未知関数が内生変数に依存する場合，L^2 ノルムなどの通常のノルムの下ではこの収束レートを達成することは困難である。そこで，新しいノルムを導入し，そのノルムの下で，ノンパラメトリック推定量の収束レートが $n^{-1/4}$ よりも早くなることを示している。さらに，(5.1) により定式化されるモデルのセミパラメトリック効率性（semiparametric efficiency）の下限を導出し，最適なウエイトを用いたとき，SMD 推定量の漸近分散が下限に到達できることを証明した。

2　経験尤度法

Otsu［2007］及び Sueishi［2008］では，経験尤度法を用いた (5.1) の推定方法が議論されている。彼らの手法を紹介する前に，条件付モーメント制約に未知関数が含まれない場合，すなわち，モーメント制約が

$$E[\rho(X_i,\ \theta_0)|Z_i]=0 \tag{5.7}$$

で与えられる場合の経験尤度推定について考察する。

経験尤度法では主として２つのアプローチが考えられる。ひとつは，Kitamura, Tripathi, and Ahn［2004］のカーネル法を用いる推定方法，もうひとつは，Donald, Imbens, and Newey［2003］の可算無限個の無条件モーメント制約を用いる推定方法である。

Kitamura, Tripathi, and Ahn［2004］の方法は次のとおりである。K と h をそれぞれ，カーネル関数とバンド幅とする。推定量は次のような制約付き最大化問題の解として与えられる。

$$\max_{\pi_{ji}0>,\ \theta\in\Theta}\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ji}\log\pi_{ji}\quad \text{s.t.}$$

$$\sum_{j=1}^{n}\pi_{ji}\rho(X_j,\ \theta)=0,\ \sum_{j=1}^{n}\pi_{ji}=1\quad for\quad i,\ j=1,\ldots,n \tag{5.8}$$

ただし，

$$w_{ji}=\frac{K\left(\frac{Z_j-Z_i}{h}\right)}{\sum_{j=1}^{n}K\left(\frac{Z_j-Z_i}{h}\right)}$$

である。π_{ji} は条件付確率 $P(X=X_j|Z=Z_i)$ と

4）例えば，Newey［1994］を参照。

解釈できる。w_{ji} は Z_i における Z_j へのウエイトである。したがって，$\sum_{j=1}^{n} w_{ji} \log \pi_{ji}$ は，Z_i の周辺の観測値の尤度をウエイト w_{ji} を用いて平滑化した，Z_i における局所尤度であると解釈することができる。

制約付き最大化問題(5.8)には双対問題が存在し，推定量は双対問題を解いて

$$\hat{\theta}=\arg\min_{\theta\in\Theta}\max_{\lambda_i\in\Lambda}\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ji}\log(1+\lambda_i'\rho(X_j,\ \theta)) \tag{5.9}$$

により得られる[5)]。λ_i は(5.8)における制約式のラグランジュ乗数と解釈できる。推定量(5.9)の漸近分散はモデル(5.7)のセミパラメトリック効率性の下限に到達できる。

Donald, Imbens, and Newey［2003］は無条件モーメント制約

$$E[\rho(X_i,\ \theta_0)\otimes p^{k_n}(Z_i)]=0$$

を用いた推定方法を提案している。$p^{k_n}(z)$ は k_n 次元の基底関数のベクトルである。推定量は以下の制約付き最大化問題を解くことで得られる。

$$\max_{\pi_i>0,\ \theta\in\Theta}\sum_{i=1}^{n}\log\pi_i \quad \text{s.t.} \quad \sum_{i=1}^{n}\pi_i g_i(\theta)=0,\ \sum_{i=1}^{n}\pi_i=1$$

ただし，$g_i(\theta)=\rho(X_i,\ \theta)\otimes p^{k_n}(Z_i)$ である。上記の制約付き最大化問題の双対問題は

$$\min_{\theta\in\Theta}\max_{\lambda\in\Lambda}\sum_{i=1}^{n}\log(1-\lambda' g_i(\theta)) \tag{5.10}$$

で与えられ，通常は(5.10)を解くことで推定量を求める。(5.10)はさらに一般化が可能である。関数 $s(\cdot)$ を0を含む開区間を定義域に持つ凹関数とする。さらに，$s_j(v)=d^j s(v)/dv^j$ とし，$s_1(0)=s_2(0)=-1$ となるように標準化する。(5.10)の対数関数を $s(v)$ で置き換えることで，一般化経験尤度（generalized empirical likelihood）推定量

$$\hat{\theta}=\arg\min_{\theta\in\Theta}\max_{\lambda\in\Lambda}\sum_{i=1}^{n}s(\lambda' g_i(\theta)) \tag{5.11}$$

が得られる。基底関数のベクトルの次元が観測個数とともに増加するとき，一般化経験尤度推定量(5.11)の漸近分散はセミパラメトリック効率性の下限に到達可能である。

最後に，条件付モーメント制約の中に未知関数が含まれるモデル(5.1)を考える。Otsu［2007］の方法は Kitamura, Tripathi, and Ahn［2004］の拡張であり，Sueishi［2008］の方法は Donald, Imbens, and Newey［2003］の拡張である。Ai and Chen［2003］などと同様に，未知関数 h は sieve を用いて推定される。

$\mathcal{H}_n$ を sieve 空間とする。Otsu［2007］の sieve conditional empirical likelihood（SCEL）推定量は

$$\min_{\alpha=(\theta,\ h)\in\Theta\times\mathcal{H}_n}\sum_{i=1}^{n}\max_{\lambda_i\in\Lambda}\sum_{j=1}^{n}w_{ji}\log(1+\lambda_i'\rho(X_j,\ \alpha)) \tag{5.12}$$

を解くことで得られる。一方，Sueishi［2008］の sieve generalized empirical likelihood（SGEL）推定量は，

$$\min_{\alpha=(\theta,\ h)\in\Theta\times\mathcal{H}_n}\max_{\lambda\in\Lambda}\sum_{i=1}^{n}s(\lambda' g_i(\alpha)) \tag{5.13}$$

の解として与えられる。ただし，$g_i(\alpha)=\rho(X_i,\ \alpha)\otimes p^{k_n}(Z_i)$ である。

それぞれの推定量について，一致性，有限次元パラメータの推定量の漸近正規性，漸近効率性が示されており，推定量の一次の漸近的性質は同等である。しかし，(5.12)と(5.13)を比較すればわかるように，SCEL 推定量は各観測値について最大化問題を解く必要があるのに対し，SGEL 推定量は1回の最大化問題を解けばすむので，計算量の観点からは後者が優れている。

3　今後の課題

本章で紹介した論文は，いずれも一次の漸近理論のみを考察しており，未知関数が内生変数に依存する場合，セミパラメトリックモデルの

5）Kitamura, Tripathi, and Ahn［2004］では trimming を行っているが，単純化のため無視する。

高次の漸近理論に関する研究は現在までのところ行われていない。Newey and Smith［2004］は未知関数が含まれない無条件モーメント制約モデルにおいて，経験尤度推定量は高次のバイアスの意味で，GMM 推定量よりも優れていることを示した。セミパラメトリックなモーメント制約の場合においても，同様の結果が期待される。Ai and Chen［2003］の推定量と比較して Sueishi［2008］の推定量の高次のバイアスが小さい可能性があるが，理論的には示されておらず，これは今後の研究課題である。

また，本章で紹介した論文では，実用上，どの程度の次数の sieve を選ぶべきかという問題が議論されていない。この点については，Sueishi［2009］がモーメント条件をもとにしたモデル選択の方法を提案しており，sieve を選択する問題にも適用可能である。

参考文献

Ai, C., and X. Chen［2003］"Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, pp. 1795–1843.

Blundell, R., and J. L. Powell［2003］"Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econometrics : Theory and Applications*, ed. by M. Dewatripont, L. P. Hansen, and S. J. Turnovsky, vol. 2, pp. 312–357, Cambridge University Press.

Chen, X.［2006］"Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6, Elsevier.

Chen, X., and M. Reiss［2007］"On Rate Optimality for Ill-Posed Inverse Problems in Econometrics," Working Paper, Yale University.

Darolles, S., J. -P. Florens, and E. Renault［2002］"Nonparametric Instrumental Regression," Working Paper, University of Toulouse.

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, pp. 55–93.

Gagliardini, P., and O. Scaillet［2007］"Tikhonov Regularization for Nonparametric Instrumental Variable Estimators," Working Paper, University of Lugano and Swiss Finance Institute.

Hall, P., and J. L. Horowitz［2005］"Nonparametric Methods for Inference in the Presence of Instrumental Variables," *Annals of Statistics*, 33, pp. 2904–2929.

Horowitz, J. L.［2007］"Asymptotic Normality of a Nonparametric Instrumental Variables Estimator," *International Economic Review*, 48, pp. 1329–1349.

Ichimura, H.［1993］"Semiparametric Least Squares（SLS）and Weighted SLS Estimation of Single Index Models," *Journal of Econometrics*, 58, pp. 71–120.

Kitamura, Y., G. Tripathi, and H. Ahn［2004］"Empirical Likelihood-Based Inference in Conditional Moment Restriction Models," *Econometrica*, 72, pp. 1667–1714.

Li, K. -C.［1987］"Asymptotic Optimality for C_p, C_L, Cross-Validation, and Generalized Cross-Validation : Discrete Index Set," *Annals of Statistics*, 15, pp. 958–975.

Newey, W. K.［1994］"The Asymptotic Variance of Semiparametric Estimation," *Econometrica*, 62, pp. 1349–1382.

――［1997］"Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, pp. 147–168.

Newey, W. K., and J. L. Powell［2003］"Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, pp. 1565–1578.

Newey, W. K., J. L. Powell, and F. Vella［1999］"Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67, pp. 565–603.

Newey, W. K., and R. J. Smith［2004］"Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, pp. 219–255.

Otsu, T.［2007］"Empirical Likelihood Estimation of Conditional Moment Restriction Models with Unknown Functions," forthcoming in *Econometric Theory*.

Robinson, P. M. [1988] "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, pp. 931-954.

Sueishi, N. [2008] "Generalized Empirical Likelihood Estimation via Conditional Moment Restrictions Containing Unknown Functions," Working Paper, University of Wisconsin.

—— [2009] "Information Criteria for Moment Restriction Models," Working Paper, University of Wisconsin.