離散・連続混合変数を含むセミパラメトリック動的パネルデータモデルの統計的推定

著者	杉原 左右一
雑誌名	商学論究
巻	60
号	1/2
ページ	397-409
発行年	2012-12-10
URL	http://hdl.handle.net/10236/10412

離散・連続混合変数を含むセミパラメトリック動的パネルデータモデルの統計的推定

杉　原　左 右 一

I　はじめに

パネルデータ分析（panel data analysis）は、クロスセクションデータ分析（cross section data analysis）と時系列データ分析（time series data analysis）を併合することにより、各分析方法が抱える問題点を相互補完すると共に、各分析方法の持つ特徴を活かしつつ、単独の方法のみでは十分に分析することが困難であった数多くの問題に強力な分析手段を提供出来る点に大きな特徴を見出すことが出来る。

パネルデータ分析については、これまでに広範な分野で統計理論的、実証的研究がなされてきたが、所謂個別効果（individual effects）を巡る、固定効果モデル（fixed effects model）と変量効果モデル（random effects model）の開発もその一例であろう。これに関しては、例えば、Anderson and Hsiao（1982）が、個別効果について、その要因を時間的に不変な独立変数を用いて線形近似したパラメトリック動的パネルデータモデル（parametric dynamic panel data model）を提示しているが、その後、同モデルは、この分野の基本モデルの一つとして実証分析で広く利用されている。

しかしながら、状況によって事情は大きく異なるとは言え、特に個別効果の特性に鑑みれば、線形近似もさることながら、非線形近似をも含めて、そもそも関数形が既知であるとする前提には大きな限界があると言わざるを得ない。その上更に、実証分析にあたっては、連続変数以外にもカテゴリカル

変数等の離散変数を取り扱う必要性が少なからず存在するのである。この様な観点から、離散・連続混合変数を含むセミパラメトリックモデルを構築することには重要な意味があると考えるのである。

以上を踏まえて、本稿では、従来のパラメトリックモデルに代わり、上記諸要素を組み入れたセミパラメトリックモデルを構築して、その統計的分析を行いたい。

以下、先ず第2節で、本稿で取り扱うセミパラメトリック動的パネルデータモデルを提示する。第3節で、このモデルにかかわる仮定について述べる。次に第4節で、モデルに含まれる未知母数、及び未知関数の推定方法と推定量の統計的性質について、既存のパラメトリック動的パネルデータ分析、及びノンパラメトリック回帰分析の分野で開発されている方法を利用することにより考察する。最後に第5節で、残された検討課題について述べる。なお、以下本稿では、推定量の漸近的性質が成立するために必要となる正則条件が満たされているものとして、導出過程の道筋を中心に述べることにする。

II　離散・連続混合変数を含むセミパラメトリック動的パネルデータモデル

i 及び t を、それぞれ個体（$i=1, 2, \cdots, N$）及び時点（$t=1, 2, \cdots, T$）を表わす添字記号とする。個体 i, 時点 t に依存する従属変数、h 次元独立変数ベクトル、及び誤差項をそれぞれ y_{it}, $x_{it}=(x_{i1t}, x_{i2t}, \cdots x_{iht})'$, 及び u_{it} と表わす。また、Z_i を個体 i のみに依存し、時点 t には依存しない k 次元独立変数ベクトルとする。特に、カテゴリカル変数の取り扱いを可能にすることを念頭に置いて、Z_i が p 次元の連続変数部分 $Z_i^c=(Z_{i1}^c, Z_{i2}^c, \cdots Z_{ip}^c)'$ と、q 次元の離散変数部分 $Z_i^d=(Z_{i1}^d, Z_{i2}^d, \cdots, Z_{iq}^d)'$ から構成されるものとし、$p+q=k$ として、$Z_i=(Z_i^{c\prime}, Z_i^{d\prime})'$ と表記することにする。特に Z_i^d については、$Z_{il}^d(l=1, 2, \cdots, q)$ が c_l 個（$c_l \geqq 2$）の値 $\{0, 1, 2, \cdots, c_{l-1}\}$ をとることが出来るものとして、$D^q=\prod_{l=1}^q\{0, 1, 2, \cdots, c_{l-1}\}$ とし、$Z_i=(Z_i^{c\prime}, Z_i^{d\prime})'\in R^p\times D^q$ とする。

そこで、ϕ をスカラーの未知母数、$\alpha=(\alpha_1, \alpha_2, \cdots, \alpha_h)'$ を h 次元未知母数ベクトル、$z=(z^{c\prime}, z^{d\prime})'\in R^p\times D^q$ について、$g(z)$: $R^p\times D^q\to R^1$ を未知関数として、次式で表わされるセミパラメトリック動的パネルデータモデルを考えよう。

$$(1)\quad \begin{aligned} y_{it}&=\phi y_{i,t-1}+x_{it}'\alpha+g(Z_i)+u_{it} \\ &=v_{it}'\theta_1+g(Z_i)+u_{it} \qquad i=1, 2, \cdots, N \quad t=1, 2, \cdots, T \end{aligned}$$

ここで、v_{it}, θ_1 はそれぞれ（$h+1$）次元ベクトル $v_{it}=(y_{i,t-1}, x_{it}')', \theta_1=(\phi, \alpha')'$ である。

また、誤差項 u_{it} については、次式(2)で表わされる一元配置誤差成分モデル（one-way error component model）を想定する。

$$(2)\quad u_{it}=\varepsilon_i+\nu_{it} \qquad i=1, 2, \cdots, N \quad t=1, 2, \cdots, T$$

上式で、ε_i は個体 i のみに依存する個別効果を示し、ν_{it} は ε_i 以外の個体 i , 時点 t に依存する誤差項を示している。

特に、$\beta=(\beta_1, \beta_2, \cdots, \beta_p)'$ を p 次元未知母数ベクトルとして、(1)式に於いて、未知関数 $g(Z_i)$ を連続変数 Z_i^c を用いて、

$$(3)\quad g(Z_i)=Z_i^{c\prime}\beta \qquad i=1, 2, \cdots, N$$

と表わしたモデルが、上記した Anderson and Hsiao（1982）によるパラメトリック動的パネルデータモデルに他ならない。

以後の便宜のために、ここで(1), (2)式をベクトル・行列表示しておこう。そのために、$y_i, y_{i,-1}, u_i, \nu_i, i_T$ をそれぞれ次式で表わされる T 次元ベクトルとし、X_i を（$h\times T$）次元行列、V_i を（($h+1$)$\times T$）次元行列とする。

$$(4)\quad \begin{aligned} &y_i=(y_{i1}, y_{i2}, y_{iT})' && i_T=(1, 1, \cdots, 1)' \\ &y_{i,-1}=(y_{i0}, y_{i1}, \cdots, y_{i,T-1})' && X_i=(x_{i1}, x_{i2}, \cdots, x_{iT}) \\ &u_i=(u_{i1}, u_{i2}, \cdots, u_{iT})' && V_i=\begin{pmatrix} y_{i,-1}' \\ X_i \end{pmatrix} \\ &\nu_i=(\nu_{i1}, \nu_{i2}, \cdots, \nu_{iT})' && \end{aligned}$$

$$i=1, 2, \cdots, N$$

そうすれば、(1), (2)式を個体 i 毎にまとめて次式(5), (6)で表わすことが出

来る。

$$(5)\quad y_i = y_{i,-1}\phi + X_i'\alpha + i_T g(Z_i) + u_i$$
$$= V_i'\theta_1 + i_T g(Z_i) + u_i \qquad i=1, 2, \cdots, N$$
$$(6)\quad u_i = i_T \varepsilon_i + \nu_i \qquad i=1, 2, \cdots, N$$

更に、$y, y_{-1}, g(Z), u, \nu$ を次式(7)で表わされる NT 次元ベクトル、ε を N 次元ベクトルとし、X を $(h\times NT)$ 次元行列、V を $((h+1)\times NT)$ 次元行列、Z を $(k\times N)$ 次元行列としよう。

$$y = (y_1', y_2', \cdots, y_N')' \qquad \varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_N)'$$
$$y_{-1} = (y_{1,-1}', y_{2,-1}', \cdots, y_{N,-1}')' \qquad X = (X_1, X_2, \cdots, X_N)$$
$$(7)\quad g(Z) = (i_T' g(Z_1), i_T' g(Z_2), \cdots, i_T' g(Z_N))' \qquad V = (V_1, V_2, \cdots, V_N)$$
$$u = (u_1', u_2', \cdots, u_N')' \qquad Z = (Z_1, Z_2, \cdots, Z_N)$$
$$\nu = (\nu_1', \nu_2', \cdots, \nu_N')'$$

そうすれば、(1), (2)式全体を次式(8), (9)で表わすことが出来る。

$$(8)\quad y = y_{-1}\phi + X'\alpha + g(Z) + u$$
$$= V'\theta_1 + g(Z) + u$$
$$(9)\quad u = (I_N \otimes i_T)\varepsilon + \nu$$

Ⅲ モデルの仮定

上記モデルに次の仮定１，２を設定する。

仮定１ $|\phi| < 1$

仮定２ $i, j = 1, 2, \cdots, N$ について、以下の(i)～(iv)を仮定する。

（ｉ） $E(\varepsilon_i | X_i, Z_i) = 0$

（ii） $E(\nu_j | X_i, Z_i, \varepsilon_i) = 0$

（iii） $E(\varepsilon_i \varepsilon_j | X_i, Z_i, X_j, Z_j) = \sigma_\varepsilon^2 \delta_{ij}$

（iv） $E(\nu_i \nu_j' | X_i, Z_i, \varepsilon_i, X_j, Z_j, \varepsilon_j) = \sigma_\nu^2 I_T \delta_{ij}$

仮定１は、動的パネルデータモデルの安定性を保証するものである。これを、例えばより一般の $s(\geqq 2)$ 時点前までのラグを持つ定常な場合に拡張す

ることは容易である。但し、$\phi=1$ とした単位根のある場合には、特に $T\to\infty$ のとき、様相が一変するために本稿では取り扱わない。

次に仮定 2 は、個体 $i, j(i, j=1, 2, \cdots, N)$ の ε_i, ν_j に関して、相関性や、分散不均一性、系列的相関性の存在を排除するものである。特に条件付き期待値の表現については、状況に応じてこれらを更に緩和、修正する必要があろうが、本稿では、議論の出発点としてこれらの諸仮定を設定することにする。なお、特に分散不均一性や系列的相関性に関しては、仮定 2 を緩めてモデルを拡張することは比較的容易である。

さて、以上の様に表現すれば、(1), (2)式（又は、(5), (6)式、ないし(8), (9)式）で表わされるモデルが、独立変数に離散・連続混合変数を含み、かつ、ラグ付き従属変数を含むセミパラメトリック部分的線形動的パネルデータモデル（semiparametric partially linear dynamic panel data model with discrete-continuous mixed independent variables and lagged dependent variables）と呼ぶべきモデルに相当していることが理解出来るだろう。

以下、本稿では、特に T を固定し、N が増大する場合（$N\to\infty$）を中心に考察することにしたい。

Ⅳ　未知母数及び未知関数の統計的推定

未知母数 $\theta_1=(\phi, \alpha')'$ 以外の、誤差項 u_{it} の分散部分を未知母数 $\theta_2=(\sigma_\varepsilon^2, \sigma_\nu^2)'$ と表わし、全ての未知母数を $\theta=(\theta_1', \theta_2')'$ と表わそう。そうすれば、モデルの全ての未知数は、未知母数 θ と未知関数 $g(z)$ となる。

これらの未知数を推定する方法として、例えば Wang, Carroll and Lin (2005) 等によるプロファイル尤度法（profile likelihood methods）を適用することが可能であるが、非線形反復推定法を用いなければならないために、必要となる計算は複雑なものとならざるを得ない[1)]。そこで、本稿では以下、

1)　固定効果パネルデータモデルの、未知母数、及び未知関数のセミパラメトリック推定量の漸近的性質、並びにそれにもとづく統計的仮説検定に関しては、例えば、Henderson, Carroll and Li (2008) がある。

これらの未知母数 θ 及び未知関数 $g(z)$ を、パラメトリック動的パネルデータ分析、及びノンパラメトリック回帰分析の分野で開発されている既存の方法を利用することにより、逐次段階的に推定することを考えたい。

まず、未知母数 $\theta_1=(\phi, \alpha')'$ に関しては、既にパラメトリック動的パネルデータ分析の分野で開発されている方法をそのまま用いて推定することが出来る。その際特に注意しなければならないのは、Anderson and Hsiao（1981, 1982）も指摘しているように様に、初期値 y_{i0} と個別効果 ε_i の相関関係である。この点を考慮した方法の一つとして、データに時間的な一階階差を施した上で GMM を適用する方法が挙げられる[2)]。すなわち、(5), (6)式に一階階差演算を施して ε_i を除去すれば次式(10)を得る。なお、その際同時に未知関数 $g(Z_i)$ も除去されることに注意したい。

$$
(10)\quad \begin{aligned} \Delta y_i &= \Delta y_{i,-1}\phi + \Delta X_i'\alpha + \Delta\nu_i \\ &= \Delta V_i'\theta_1 + \Delta\nu_i \qquad\qquad i=1, 2, \cdots, N \end{aligned}
$$

但し、上式で $\Delta y_i, \Delta y_{i,-1}, \Delta\nu_i$ はそれぞれ（$T-1$）次元ベクトルであり、$\Delta X_i, \Delta V_i$ はそれぞれ（$h\times(T-1)$）次元行列、（$(h+1)\times(T-1)$）次元行列である。

$$
(11)\quad \begin{aligned} \Delta y_i &= (y_{i2}-y_{i1}, y_{i3}-y_{i2}, \cdots, y_{iT}-y_{i,T-1})' \\ \Delta y_{i,-1} &= (y_{i1}-y_{i0}, y_{i2}-y_{i1}, \cdots, y_{i,T-1}-y_{i,T-2})' \qquad \Delta V_i=\begin{pmatrix}\Delta y_{i,-1}' \\ \Delta X_i\end{pmatrix} \\ \Delta\nu_i &= (\nu_{i2}-\nu_{i1}, \nu_{i3}-\nu_{i2}, \cdots, \nu_{iT}-\nu_{i,T-1})' \\ \Delta X_i &= (x_{i2}-x_{i1}, x_{i3}-x_{i2}, \cdots, x_{iT}-x_{i,T-1})' \qquad\qquad i=1, 2, \cdots, N \end{aligned}
$$

ここで、r_{it}、及び R_i をそれぞれ次式で定義される（$Th+(t-1)$）次元ベクトル、（$T(T-1)(h+1/2)\times(T-1)$）次元行列としよう。

$$
(12)\quad r_{it}=(y_{i0}, y_{i1}, \cdots, y_{i,t-2}, x_{i1}', x_{i2}', \cdots, x_{iT}')' \qquad i=1, 2, \cdots, N \quad t=2, 3, \cdots, T
$$

$$
(13)\quad R_i=Diag(r_{i2}, r_{i3}, \cdots, r_{iT}) \qquad\qquad i=1, 2, \cdots, N
$$

2） パネルデータモデルの種々の GMM に関する統計的性質ついては、例えば、Arellano（2003）が参考になる。

そうすれば、

⒁　$E(R_i \Delta\nu_i)=0 \qquad i=1, 2, \cdots, N$

が成立するから、$R_i(i=1, 2, \cdots, N)$ を操作変数行列として用いることにより、θ_1 の GMM 推定量 $\hat{\theta}_1$ を次式により求めることが出来る。

⒂
$$\hat{\theta}_1=\left\{\left(\sum_{i=1}^{N}(\Delta V_i)R_i'\right)\left(\sum_{i=1}^{N}R_iD_TD_T'R_i'\right)^{-1}\left(\sum_{i=1}^{N}R_i(\Delta V_i)'\right)\right\}^{-1}$$
$$\left(\sum_{i=1}^{N}(\Delta V_i)R_i'\right)\left(\sum_{i=1}^{N}R_iD_TD_T'R_i'\right)^{-1}\left(\sum_{i=1}^{N}R_i\Delta y_i\right)$$

但し、上式で D_T は次式⒃で与えられる（$(T-1)\times T$）次元階差行列である。

⒃
$$D_T=\begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}$$

よく知られている様に、$\hat{\theta}_1$ は θ_1 の $\sqrt{N}$ 一致推定量であり、その漸近的分散共分散行列は、

⒄
$$Avar(\hat{\theta}_1)=\sigma_\nu^2\left\{\left(\sum_{i=1}^{N}(\Delta V_i)R_i'\right)\left(\sum_{i=1}^{N}R_iD_TD_T'R_i'\right)^{-1}\left(\sum_{i=1}^{N}R_i(\Delta V_i)'\right)\right\}^{-1}$$

で与えられる。なお、誤差項 ν_{it} に系列的相関がある場合には、推定量 $\hat{\theta}_1$ は一致性を失うことに注意しなければならない。

次に、未知関数 $g(z)$: $R^p\times D^q\rightarrow R^1$ の推定について考えよう。これに関しては直接にノンパラメトリック推定することも可能であるが、本稿では、特に N が大きい場合に有効な方法として、通常のノンパラメトリック回帰法を用いて $g(z)$ を推定することを考えたい。

そのために、まず⑴式の各成分毎の時間平均を求めれば次式⒅を得る。

⒅
$$\begin{aligned}\bar{y}_i&=\phi\bar{y}_{i,-1}+\bar{x}_i'\alpha+g(Z_i)+\bar{u}_i\\&=\bar{v}_i'\theta_1+g(Z_i)+\bar{u}_i \qquad i=1, 2, \cdots, N\end{aligned}$$

但し、$\bar{y}_i, \bar{y}_{i,-1}, \bar{u}_i, \bar{\nu}_i$ はそれぞれ次式⒆で表わされるスカラーであり、$\bar{x}_i, \bar{v}_i$ はそれぞれ h 次元ベクトル、$(h+1)$ 次元ベクトルである。

$$(19)\quad \begin{aligned} \bar{y}_i &= \frac{1}{T}\sum_{t=1}^{T} y_{it} & \bar{\nu}_i &= \frac{1}{T}\sum_{t=1}^{T} \nu_{it} \\ \bar{y}_{i,-1} &= \frac{1}{T}\sum_{t=1}^{T} y_{i,t-1} & \bar{x}_i &= \frac{1}{T}\sum_{t=1}^{T} x_{it} \\ \bar{u}_i &= \frac{1}{T}\sum_{t=1}^{T} u_{it} & \bar{v}_i &= (\bar{y}_{i,-1}, \bar{x}_i')' \\ &= \varepsilon_i + \bar{\nu}_i & & \qquad i=1, 2, \cdots, N \end{aligned}$$

次に、(1)式より m_{it} を

$$(20)\quad \begin{aligned} m_{it} &\equiv y_{it} - v_{it}'\theta_1 \\ &= g(Z_i) + u_{it} \qquad i=1, 2, \cdots, N \quad t=1, 2, \cdots, T \end{aligned}$$

と表わし、m_{it} の時間平均 $\overline{m}_i$（スカラー）を求めれば、

$$(21)\quad \begin{aligned} \overline{m}_i &= \frac{1}{T}\sum_{t=1}^{T} m_{it} \\ &= \bar{y}_i - \bar{v}_i'\theta_1 \\ &= g(Z_i) + \bar{u}_i \qquad i=1, 2, \cdots, N \end{aligned}$$

を得る。

ここで、特に上式(21)より、

$$(22)\quad E(\overline{m}_i|Z_i) = g(Z_i) \qquad i=1, 2, \cdots, N$$

が成立することに注意しよう。

また、$\overline{m}_i (i=1, 2, \cdots, N)$ は、$\hat{\theta}_1$ を用いて次式により推定することが出来る。

$$(23)\quad \widehat{\overline{m}}_i = \bar{y}_i - \bar{v}_i'\hat{\theta}_1 \qquad i=1, 2, \cdots, N$$

そうすれば、$g(z)=E(\overline{m}_i|Z_i=z)$ を、離散・連続混合変数を含むノンパラメトリック回帰分析の分野で開発されたカーネル推定法を直接用いて以下の様に推定することが出来る[3)]。

3) 離散・連続混合変数を含むノンパラメトリック回帰モデルの統計的分析については、Aitchison and Aitken (1976), Silverman (1986), Simonoff (1996), Li and Racine (2003, 2007), Racine and Li (2004), Li, Quyang and Racine (2009) 等が参考になる。

すなわち、まず連続変数 Z_i^c については、平滑化定数を $h_l(>0)$, $h=(h_1, h_2, \cdots, h_p)'$ とし、通常の連続形1次元2次カーネル関数を $w(\cdot)$ として、次式(24)で表わされる連続形積カーネル関数 $W(Z_i^c, z^c, h)$ を考える。

(24)　$$W(Z_i^c, z^c, h)=\prod_{l=1}^{p}\frac{1}{h_l}w\left(\frac{Z_{il}^c-z_l^c}{h_l}\right) \qquad i=1, 2, \cdots, N$$

次に、離散変数 Z_i^d についても、所謂“疎なデータ”(sparse data)の問題に対処するために平滑化することを考えて、平滑化定数を λ_l, $0\leqq\lambda_l\leqq1$, $\lambda=(\lambda_1, \lambda_2, \cdots, \lambda_q)'$ として、非序列形（unordered）離散変数に有効な方法として、次式で与えられる離散形積カーネル関数 $L(Z_i^d, z^d, \lambda)$ を考える[4)]。

(25)　$$L(Z_i^d, z^d, \lambda)=\prod_{l=1}^{q}\lambda_l^{1(Z_{il}^d\neq z_l^d)} \qquad i=1, 2, \cdots, N$$

但し、1(E) はEが真であれば1、偽であれば0となる指示関数である。

以上をもとにして、平滑化定数ベクトルを $\gamma=(h', \lambda')'$ として、離散・連続混合変数に対する一般化積カーネル関数 $K(Z_i, z, \gamma)$ を、

(26)　$$K(Z_i, z, \gamma)=W(Z_i^c, z^c, h)L(Z_i^d, z^d, \lambda) \qquad i=1, 2, \cdots, N$$

と定義しよう。そうすれば、$g(z)$ を次式の Nadaraya-Watson 形カーネル推定量（局所定数カーネル推定量）$\hat{g}(z)$ により推定することが考えられる。

(27)　$$\hat{g}(z)=\sum_{i=1}^{N}\hat{\hat{m}}_i\frac{K(Z_i, z, \gamma)}{\sum_{i=1}^{N}K(Z_i, z, \gamma)}$$

すなわち、$\hat{g}(z)$ は $\hat{\hat{m}}_i(i=1, 2, \cdots, N)$ の加重平均に他ならない。

この様にして求めた $g(z)$ の推定量 $\hat{g}(z)$ について、Li, Quyang and Racine (2009) より、$\hat{g}(z)-g(z)=O_p\left((\sum_{\ell=1}^{p}h_\ell^2+\sum_{\ell=1}^{q}\lambda_\ell)+(N\prod_{\ell=1}^{p}h_\ell)^{-\frac{1}{2}}\right)$ であり、$N\to\infty$ のとき、(i) $(\sum_{\ell=1}^{p}h_\ell^2+\sum_{\ell=1}^{q}\lambda_\ell)\to0$, (ii) $(N\prod_{\ell=1}^{p}h_\ell)\to\infty$ であれば、$\hat{g}(z)$ は $g(z)$ の一致推定量となり、更に加えて、(iii) $(N\prod_{\ell=1}^{p}h_\ell)(\sum_{\ell=1}^{p}h_\ell^2+\sum_{\ell=1}^{q}\lambda_\ell)^3\to0$ であれば、$N\to\infty$ のとき、次式が成立することがわかる[5)]。

4）なお、序列形（ordered）離散変数の場合には、(25)式で、$\lambda_l^{1(Z_{il}^d\neq z_l^d)}$ に代わって、$\lambda_l^{|Z_{il}^d-z_l^d|}$ とすればよい。

5）詳細については、Li, Quyang and Racine (2009) を参照されたい。

$$(28)\quad \left(N\prod_{l=1}^{p}h_l\right)^{\frac{1}{2}}\left(\hat{g}(z)-g(z)-\sum_{l=1}^{p}B_{1l}(z)h_l^2-\sum_{l=1}^{q}B_{2l}(z)\lambda_l\right)$$

$$\xrightarrow{d}N\left(0,\ \kappa^p\frac{\sigma^2}{f(z)}\right)$$

但し、上式で $B_{1l}(z)$, $B_{2l}(z)$ はそれぞれ、次式で与えられる。

$$(29)\quad B_{1l}(z)=\frac{\mu_2(w)}{2}\left(\frac{\partial^2 g(z)}{\partial z_l^2}+2\frac{\partial g(z)}{\partial z_l}\frac{\partial f(z)}{\partial z_l}f(z)^{-1}\right)\quad l=1,\ 2,\ \cdots,\ p$$

$$(30)\quad B_{2l}(z)=\frac{1}{c_l-1}\sum_{x^d\in D^q}1_l(z^d,\ x^d)(g(z^c,\ x^d)-g(z))f(z^c,\ x^d)f(z)^{-1}$$

$$l=1,\ 2,\ \cdots,\ q$$

また、$1_l(z^d, x^d)$ は、

$$(31)\quad 1_l(z^d,\ x^d)=1(z_l^d\neq x_l^d)\prod_{r\neq l}^{q}1(z_r^d=x_r^d)\qquad l=1,\ 2,\ \cdots,\ q$$

であり、$f(z)$ は $Z_i=(Z_i^{c\prime}, Z_i^{d\prime})'$ の同時確率密度関数を示し、κ, $\mu_2(w)$, σ^2 はそれぞれ次式で与えられる。

$$(32)\quad \kappa=\int_{-\infty}^{\infty}w(v)^2dv$$

$$(33)\quad \mu_2(w)=\int_{-\infty}^{\infty}v^2w(v)dv$$

$$(34)\quad \sigma^2=\sigma_\varepsilon^2+\frac{1}{T}\sigma_\nu^2$$

なお、最適な平滑化定数ベクトル γ については、例えばこれを最小２乗クロスバリデーション法（least squares cross-validation methods）を用いて選択することが出来る。

最後に、誤差項 u_{it} の分散部分の未知母数 $\theta_2=(\sigma_\varepsilon^2, \sigma_\nu^2)'$ については、上で求めた推定量 $\hat{\theta}_1$、及び $\hat{g}(z)$ を用いて、例えばこれを次の様にして推定すればよい。

(35)　$$\hat{\sigma}_{\nu}^{2}=\frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}\{y_{it}-\bar{y}_{i}-\hat{\phi}(y_{i,t-1}-\bar{y}_{i,-1})-(x_{it}-\bar{x}_{i})'\hat{\alpha}\}^{2}$$

(36)　$$\hat{\sigma}_{\varepsilon}^{2}=\frac{1}{N}\sum_{i=1}^{N}\{\bar{y}_{i}-\hat{\phi}\bar{y}_{i,-1}-\bar{x}_{i}'\hat{\alpha}-\hat{g}(z_{i})\}^{2}-\frac{1}{T}\hat{\sigma}_{\nu}^{2}$$

この様にして求めた θ_2 の推定量を $\hat{\theta}_2=(\hat{\sigma}_{\varepsilon}^2, \hat{\sigma}_{\nu}^2)'$ と表わそう。

そうすれば、以上の方法により全ての未知数である未知母数 $\theta=(\theta_1', \theta_2')'$ と未知関数 $g(z)$ について、それらの推定量 $\hat{\theta}=(\hat{\theta}_1', \hat{\theta}_2')'$ と $\hat{g}(z)$ を求めることが出来た。

上記方法は、既存の方法を利用した比較的簡便な推定方法であることが理解出来るだろう。

V　おわりに

本稿では、従来のパラメトリック動的パネルデータモデルを、特に離散・連続混合変数を含むセミパラメトリック動的パネルデータモデルに拡張した上で、モデルに含まれる未知母数、並びに未知関数を、パラメトリック動的パネルデータ分析、及びノンパラメトリック回帰分析の分野で開発された既存の方法を利用することにより逐次段階的に推定し、その統計的性質について考察した。

本稿を閉じるにあたって、幾つかの残された検討課題について述べておくことにしたい。

まず、本稿では未知関数 $g(z)$ のノンパラメトリック推定に際して、時間平均を施したデータをもとにして、これをカーネル推定する方法を考えた。この推定方法は簡便な方法ではあるものの、時間平均を施す結果、データ数が TN から N に激減するという問題を抱えている。また、本稿では時間幅 T を固定し、個体数 N が増大する場合（$N\to\infty$）を念頭に置いて考察したのであるが、上述した問題と合わせて、今後特に T, N が共に小さい場合の推定量の小標本特性について検討しなければならない。また、これとは別に、T, N が共に増大する場合（$T, N\to\infty$）の推定量の漸近的性質についても合

わせて考察しなければならないであろう。また、カーネル推定法に代わる方法として、未知関数を、冪級数関数やB-スプライン関数等の基底関数（base functions）を用いて近似する方法が考えられる。今後同方法との比較検討を行わなければならない。

更に、本稿の方法と、Wang, Carroll and Lin（2005）等によるプロファイル尤度法との比較検討も重要である。また、統計的推定問題のみならず統計的仮説検定問題についても検討しなければならない。

本稿ではひとまず仮定1, 2を設定して議論したのであるが、これらの仮定の妥当性の統計的仮説検定や、これらの仮定を緩めた場合の検討が必要であろう。更に、より一般的なモデルについて検討することも重要である。例えば、$\varepsilon_i, \nu_j(i, j=1, 2, \cdots, N)$ について、個体間に相関構造を導入し、これをノンパラメトリック分析することは今後の検討課題の一つである。また、加算モデル（additive model）、単一指標モデル（single index model）や、分位点回帰モデル（quantile regression model）等との融合についても検討しなければならない。

ノンパラメトリック、並びにセミパラメトリック分析の研究は、今後の統計科学の発展にとって極めて重要な役割を果たすものであると考えている。しかしながら、この分野の研究は未だ発展途上にあり、未解決な問題が山積している。この様な現状を踏まえて、今後とも引き続きこの分野の統計科学的研究に従事する所存である。

（筆者は関西学院大学商学部教授）

＜参考文献＞

Aitchison, J. and C. G. G. Aitken.（1976）. "Multivariate Binary Discrimination by the Kernel Method," Biometrika, 63, 413-420.

Anderson, T. W. and C. Hsiao.（1981）. "Estimation of Dynamic Models with Error Components," Journal of the American Statistical Association, 76, 598-606.

Anderson, T. W. and C. Hsiao.（1982）. "Formulation and Estimation of Dynamic Models Using Panel Data," Journal of Econometrics, 18, 47-82.

Arellano, M.（2003）. Panel Data Econometrics, Oxford University Press, Oxford.

Henderson, D., R. J. Carroll and Q. Li. (2008). "Nonparametric Estimation and Testing of Fixed Effects Panel Data Models," Journal of Econometrics, 144, 257-275.

Li, C., D. Quyang and J. S. Racine. (2009). "Nonparametric Regression with Weakly Dependent Data: The Discrete and Continuous Regressor Case," Journal of Nonparametric Statistics, 21, 697-711.

Li, Q. and J. Racine. (2003). "Nonparametric Estimation of Distributions with Categorical and Continuous Data," Journal of Multivariate Analysis, 86, 266-292.

Li, Q. and J. S. Racine. (2007). Nonparametric Econometrics, Theory and Practice, Princeton University Press, Princeton.

Racine, J. and Q. Li. (2004). "Nonparametric Estimation of Regression Functions with Both Categorical and Continuous Data," Journal of Econometrics, 119, 99-130.

Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis, Chapman and Hall, New York.

Simonoff, J.S. (1996). Smoothing Methods in Statistics, Springer Verlag, New York.

Wang, N., R. J. Carroll and X. Lin. (2005). "Efficient Semiparametric Marginal Estimation for Longitudinal / Clustered Data," Journal of the American Statistical Association, 100, 147-157.