第 2 講　ミクロ計量経済学の基礎理論

2.1　比較静学と因果律

経済分析が最終的に目指しているのは経済変数間の因果関係を明らかにすることにある。経済学では、他の条件を一定にした上で（centeris paribus)、因果関係を確定する。さもなければ他の変数の変化が結果に影響を与えることになる。では実際に他の条件をコントロールするにはどうすればばいいのだろうか。

もしある経済変数 y の平均的反応に関心があるとすれば、他の条件を一定にする分析とは経済変数とコントロール変数 $\mathbf{c}$ を条件とした y の期待値、$E(y \mid w, \mathbf{c})$ を求めればいい。ここで、コントロール変数は w の y に対する影響をみる場合に固定されているものである。第 1 講の例でいえば、家計の属性などが $\mathbf{c}$ に相当する。ここで $\mathbf{c}$ をコントロールするのは、$\mathbf{c}$ が w の水準に関連しているからである。

もし w が連続変数であれば w の偏微効果（partial effect）は $\partial E(y \mid w, \mathbf{c})／\partial w$ として求めることができる。w が離散変数であれば、$E(y \mid w, \mathbf{c})$ は $\mathbf{c}$ を固定して w の異なった値で評価できる。

ここでコントロール変数 $\mathbf{c}$ がすべて変数としてそろっていれば問題はないが、多くの場合、$\mathbf{c}$ は完全に調査されているわけではないし、観察不可能なものもある。また長期的には $\mathbf{c}$ 自体も変化する。

例えば、教育の賃金に関する効果を推計する場合、次のようなモデルを考える。$E(wage \mid educ, \exp er, abil)$ ここで $educ$ は教育年数、$exper$ は労働経験年数、$abil$ は能力とする。この場合、コントロール変数は $\mathbf{c} = (exper, abil)$ であり、$exper$ は観察可能だが、$abil$ は不可能である。実は、ミクロ計量経済学を用いることによって、逆にこの観察不可能な要素を抽出することができるのである。これがミクロ計量経済学を用いることの利点の一つである (この方法については後で詳述する)。

また $\mathbf{c}$ に関しては完全にデータが取れたとしても、y と w が事後的な均衡値だけしかわからない場合、w の変化によって y がどう変化したかを推計することは難しい。これは１年に１回、あるいは５年に１回程度で行われる統計調査について当てはまることであるが、その年の平均値 (均衡値) が記録されているものを複数年集めたところで、必ずしも動学的な推計ができることは意味しないのである。

2.2　データ

実証分析において適切なデータを用いることは必須であるが、経済モデルは母集団全体に当てはまるものであること (population model)、母集団から無作為に抽出され独立同一分析 (independent, identically distributed = i.i.d) に従う (標本サンプル) が得られることが前提になる。

経済モデルが対象としている母集団を明確に把握してモデル化することが大切である。また逆に標本 (サンプル) の持っている属性が母集団から無作為抽出されたものであるということは常に意識する必要がある。例えば標本 (サンプル) に 24 − 34 歳の女性の消費行動に関するデータが入っているとすると、この統計の母集団は社会全体の中で 24 − 34 歳の女性に限定されていることを意識しなければならない。例えば 24 歳以下の女性、35 歳以上の女性、すべての年齢層の男性の行動はここには含まれていないということである。

無作為抽出（random sampling）はある一時点のクロスセクション・データの場合、説明がつきやすい。説明変数は母集団全体の変数から無作為に抽出された結果であると解釈できる。説明変数が定数 (ダミーも含む) の場合、独立同一分布に従うことにはならないことにも注意が必要である (失業者の賃金収入はゼロである場合など)。無作為抽出の仮定はプールリング・クロスセクション・データ（pooled cross section）には当てはまらない。なぜならば異なった時点の母集団から無作為に抽出されているとしても、それは全時点の母集団から無作為に抽出されていることにはならないからである。これは独立非同一分布（independent, not identically distributed = i.n.i.d）に従っていることになる。

クロスセクション・データが独立でない場合にも注意する必要がある。地域的な相関がある場合、標本間には必ず相関が入ってくる。例えば失業率は、ある地域内では何らかの相関があるはずであるが、これは OLS や TSLS で推計する場合には工夫が必要となる。

クロスセクション・データが無作為抽出に従わないケースとしてサンプル・セレクション・バイアス (sample selection bias) の問題も考慮されている。現代の統計調査では被調査者の自発的協力に基づいていたり、調査主体の特定化のプロセスで必ずサンプル・セレクション・バイアスの問題がつきまとう。

クロスセクション・データに含まれている同一標本 (例えば個人、企業、政府など) に対して複数年調査を繰り返したデータを**パネル・データ (panel data)** という。このデータでは無作為抽出の仮定は初年度に関しては満たされたとしても、時間の経過とともに満たされなくなる。しかしパネル・データの場合、無作為抽出の仮定はクロスセクション方向には成り立っていると考えることによって、この問題を処理している。[1]

2.3 実 例

賃金関数[2]

賃金（$wage^0$）の自然対数を次のようなクロスセクション・モデルで表すとしよう。

$$\log(wage_i^0) = \beta_0 + \beta_1 educ_i + \beta_2 \exp er_i + \beta_3 married_i + u_i \tag{1}$$

ここで $educ$ は教育年数、$exper$ は労働経験年数、$married$ は結婚しているかどうかの二項選択、u は誤差項。

このモデルが適切かどうかは誤差項 u が他の説明変数と相関しているかどうか、u の分散は一定かどうか、u が i と j の間で誤差が相関しているかどうか等によって判断される。

生産関数

製造業について 3 年間のパネル・データがあるとする。生産関数を次のように定義しよう。

まずコブ・ダグラス型の生産関数を考える。

$$Y = AL^{\beta_1} K^{\beta_2} \tag{2}$$

対数をとる。

$$\log Y = \log A + \beta_1 \log L + \beta_2 \log K \tag{3}$$

[1] 一般に、パネル・データはクロスセクション方向にはたくさんのサンプルが含まれているが、それと比べると時系列方向のサンプルは少ない。このことは、クロスセクション方向での無作為性が満たされていることがはるかに重要であることを意味している。

[2] ここでの実例は Wooldridge (2002),pp.8-9 を用いている。

ここで $TFP = logA$ は次のような関数であるとする。

$$A = e^{(\delta_t + \beta_3 spillover + quality_i)}$$

これを（3）式に代入すると、（4）式を得る。

$$\log(Y) = \delta_t + \beta_1 \log(L_{it}) + \beta_2 \log(K_{it}) + \beta_3 spillover_{it} + quality_i + u_{it} \qquad t = 1, 2, 3 \tag{4}$$

ここで、$spillover$ は同一地域に存在する外国企業の集中度、$quality$ は視察不可能な生産性に影響を与える要因で、時間を通して固定されている。δ_t は生産性ショックを捉える時間ダミーである。

このようなモデルをパネル推計することが望ましいかどうかは統計的にテストによって決定される。すなわち、固定効果モデルが選択されれば $quality_i$ を推計することは意味があるが、ランダム効果モデルが選択された場合には $quality_i$ は意味をなさない。固定効果を個別企業に想定するのか、産業別につけるのかも実証上の問題である。

2.4 条件付期待値[3]

被説明変数 Y は確率変数、説明変数 $\mathbf{x} \equiv (x_1, x_2, \ldots x_k)$ は $1 \times k$ 行列の確率変数ベクトルで表されるとする。このとき、次のような関数 $\mu(\mathbf{x})$ を考える。

$$E(Y \mid x_1, x_2, \ldots x_k) = \mu(x_1, x_2, \ldots x_k) \Leftrightarrow E(Y \mid \mathbf{x}) = \mu(\mathbf{x}) \tag{5}$$

$\mu(\mathbf{x})$ は Y の期待値が $\mathbf{x}$ の変化によってどのように変化するかを決める関数である。例えば Y が賃金で $\mathbf{x}$ には教育、労働経験、IQ などの個人属性が含まれているとすると、$E(wage \mid educ, \exp er, IQ)$ は $educ, \exp er, IQ$ を所与としたときの平均賃金を表している。

$E(Y \mid \mathbf{x})$ は期待値であり、$\mathbf{x}$ の積分や集計を通して Y の条件付記確率分布を決めることによって求めることができる。計量経済学的にはパラメータを推計して $E(Y \mid \mathbf{x})$ を決める方法を用いることが多いが、これは関数 $\mu(\mathbf{x})$ の特定化の一種ではあるが、他の計量経済学で用いる期待値 $E(Y \mid \mathbf{x})$ のように数学的関数 $y = f(\mathbf{x})$ として厳密に一対一対応しているわけではない。$\mu(\mathbf{x})$ は漸近的に微分可能で、x_j は連続変数であると仮定すると、偏微分係数 $\partial\mu(\mathbf{x}) / \partial x_j$ は $E(Y \mid \mathbf{x})$ において x_j が変化したとき ($x_1 \ldots x_{j-1}, x_{j+1} \ldots x_k$ は固定）の限界効果を近似的に表していると考えることができる。

$$\Delta E(Y \mid \mathbf{x}) \approx \frac{\partial \mu(\mathbf{x})}{\partial x_j} \Delta x_j \tag{6}$$

$E(Y \mid \mathbf{x})$ の x_j に関する偏微分は x_j の $E(Y \mid \mathbf{x})$ に関する偏微効果 (partial effect) と呼ばれている。

もし x_j が離散変数であれば、部分効果は偏微分ではなく、x_j の異なった値を代入した（ここで $x_{i \neq j}$ は固定) を比較することによって求める。

偏微効果の中で弾力性（elasticity）に関心がある場合、x_j に関する $E(Y \mid \mathbf{x})$ の（偏微）弾力性は、$x_1 \ldots x_{j-1}, x_{j+1} \ldots x_k$ を固定すると、

$$\frac{\partial E(Y \mid \mathbf{x})}{\partial x_j} \cdot \frac{x_j}{E(Y \mid \mathbf{x})} = \frac{\partial \mu(\mathbf{x})}{\partial x_j} \cdot \frac{x_j}{\mu(\mathbf{x})} \tag{7}$$

[3] 本節は Wooldridge (2002, Chapter2) に基づいている。

もし $E(Y \mid \mathbf{x}) > 0$ で $x_j > 0$ であれば、（７）は次の式と同値である。

$$\frac{\partial \log\left[E(Y \mid \mathbf{x})\right]}{\partial \log(x_j)} \tag{8}$$

これは x_j の１%の変化が $E(Y \mid \mathbf{x})$ を何%変化させるかを表している。
x_j の１単位の変化に対する $E(Y \mid \mathbf{x})$ の何%かの変化は次のように近似できる。

$$100 \cdot \frac{\partial E(Y \mid \mathbf{x})}{\partial x_j} \cdot \frac{1}{E(Y \mid \mathbf{x})} = 1.00 \frac{\partial \log\left[E(Y \mid \mathbf{x})\right]}{\partial x_j} \tag{9}$$

ここで $E(Y \mid \mathbf{x}) > 0$ 　これは $E(Y \mid \mathbf{x}) >$ の x_j に対する準弾力性（semielasticity）と呼ばれる。

誤差項の構造

確率変数 Y は次のように分解できる。

$$Y = E(Y \mid \mathbf{x}) + u \tag{10}$$

$$E(u \mid \mathbf{x}) = 0 \tag{11}$$

(10) より、Y は条件付期待値 $E(Y \mid \mathbf{x})$ と条件付期待値ゼロの誤差項 u の和で表せることを意味している。

$E(u \mid \mathbf{x}) = 0$ は次の含意を持っている。(1)$E(u \mid \mathbf{x}) = 0$,　u は無条件で期待値ゼロであることを意味している。(2)u は $x_1, x_2, \ldots x_k$ のどの値とも相関していないし、$x_1, x_2, \ldots x_k$ のどの関数 (例えば $x_1^2, x_2^2, x_1 x_2, \exp(x_1)$ など）とも相関していない。

観察不可能な変数

$\mathbf{w}$ は確率ベクトル、Y は確率変数とする。$\mathbf{x}$ は確率ベクトルでの関数であるとする。$\mathbf{x} = f(\mathbf{w})$ 最も一般的な**期待収束法則**（law of iterated expectations = LIE）は次のように表せる。

$$E(Y \mid \mathbf{x}) = E\left[E(Y \mid \mathbf{w}) \mid \mathbf{x}\right] \tag{12}$$

つまり、$\mu_1(\mathbf{w}) \equiv E(Y \mid \mathbf{w})$, $\mu_2(\mathbf{x}) \equiv E(Y \mid \mathbf{x})$ とすれば、$\mu_2(\mathbf{x})$ は $\mathbf{x}$ を所与として $\mu_2(\mathbf{w})$ の期待値を計算することで求めることができる。$\mu_2(\mathbf{x}) = E\left[(\mu_1(\mathbf{w}) \mid \mathbf{x}\right]$
(12) より簡単な結果に次のようなものがある。

$$E(Y \mid \mathbf{x}) = E\left[E(Y \mid \mathbf{x}) \mid \mathbf{w}\right] \tag{13}$$

$\mathbf{x}$ は $\mathbf{w}$ の関数であり、$\mathbf{w}$ がわかれば $\mathbf{x}$ もわかる。$\mu_2(\mathbf{x}) = E(Y| \mathbf{x})$ は $\mathbf{x}$ の関数である。$\mathbf{w}$ を所与とする $\mu_2(\mathbf{x})$ の期待値は $\mu_2(\mathbf{x})$ に他ならない。

(12) と (13) の結果から得られる含意は、情報セットが小さい方に規定されるということである。つまり $\mathbf{x}$ は $\mathbf{w}$ より情報量が少ないとすれば、$\mathbf{w}$ を知っていれば、$\mathbf{x}$ は必ず知っているということを意味するが、逆は成り立たないということである。

具体例を考えてみよう。

$$E(Y \mid x_1, x_2, z) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 z \tag{14}$$

ここで z は観察不可能だとする。線形の条件付期待は次のように表せる。

$$\begin{aligned} E(Y \mid x_1, x_2) &= E(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 z \mid x_1, x_2) \\ &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 E(z \mid x_1, x_2) \end{aligned} \tag{15}$$

ここで $E(z \mid x_1, x_2)$ について線形近似できるとする。

$$E(z \mid x_1, x_2) = \delta_0 + \delta_1 x_1 + \delta_2 x_2 \tag{16}$$

(16) を (15) に代入して、整理すると

$$\begin{aligned} E(Y \mid x_1, x_2) &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3(\delta_0 + \delta_1 x_1 + \delta_2 x_2) \\ &= (\beta_0 + \beta_3\delta_0) + (\beta_1 + \beta_3\delta_1)x_1 + (\beta_2 + \beta_3\delta_2)x_2 \end{aligned} \tag{17}$$

z の情報がわからない場合は、x_1 と x_2 で z の値を近似して、できるだけその効果を考慮しようという考え方を用いる。

z とその他の観察可能な変数（例えば x_1）が交差項 $x_1 z$ として入っている場合、

$$E(Y \mid x_1, x_2, z) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 z + \beta_4 x_1 z \tag{18}$$

期待収束法則（LIE) によって、

$$E(Y \mid x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 E(z \mid x_1, x_2) + \beta_4 x_1 E(z \mid x_1, x_2) \tag{19}$$

$E(z \mid x_1, x_2)$ は (16) によって線形近似できるとすると、(19) は、

$$\begin{aligned} E(Y \mid x_1, x_2) &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3(\delta_0 + \delta_1 x_1 + \delta_2 x_2) + \beta_4 x_1(\delta_0 + \delta_1 x_1 + \delta_2 x_2) \\ &= (\beta_0 + \beta_3\delta_0) + (\beta_1 + \beta_3\delta_1 + \beta_4\delta_0)x_1 + (\beta_2 + \beta_3\delta_2)x_2 + \beta_4\delta_1 x_1^2 + \beta_4\delta_2 x_1 x_2 \end{aligned} \tag{20}$$

ここでも z の情報がわからない場合でも、既知の x_1 と x_2 を用いての効果を近似的に推計できる。

平均偏微効果

被説明変数が観察不可能な変数 q（unobserved heterogeneity）に依存している場合、すなわち、$E(Y \mid \mathbf{x}, q) = \mu_1(\mathbf{x}, q)$ とすると、連続変数に関するの偏微分効果は次のように表せる。

$$\theta_j(\mathbf{x}, q) \equiv \partial E(Y \mid \mathbf{x}, q) / \partial x_j = \partial \mu_1(\mathbf{x}, q) / \partial x_j \tag{21}$$

$\theta_j(\mathbf{x}, q)$ は q に依存しているが、異なった q に対して $\theta_j(\mathbf{x}, q)$ を推計することは不可能である。たとえそれができたとしても、$\theta_j(\mathbf{x}, q)$ の平均値として適切な q の値が何なのかはわからない。一般的には q の変化に応じて変わる偏微分効果の平均をとる（average partial effect）のが適切だろう。これは次のように定義できる。

$$\delta_j(\mathbf{x}^0) \equiv E_q\left[\theta_j(\mathbf{x}, q)\right] \tag{22}$$

ここで $E_q[\cdot]$ は q に関する期待値オペレータである。

q と $\mathbf{x}$ は観察できる変数ベクトル $\mathbf{w}$ を条件として独立であるとする。

$$D(q \mid \mathbf{x}, \mathbf{w}) = D(q \mid \mathbf{w}) \tag{23}$$

ここで $D(\cdot \mid \cdot)$ は条件付分布である。(23) は $\mathbf{w}$ は q の代理変数ベクトルであることを意味する。また $\mathbf{w}$ は $E(Y \mid \cdot)$ の構造方程式の中では無視できるとする。

$$E(Y \mid \mathbf{x}, q, \mathbf{w}) = E(Y \mid \mathbf{x}, q) \tag{24}$$

(23)(24) より、(21) は次のように書き換えられる。

$$\delta_j(\mathbf{x}^0) = E_w\left[\partial E(Y \mid \mathbf{x}^0, \mathbf{w}) \big/ \partial x_j\right] \tag{25}$$

$E_w[\cdot]$ は $\mathbf{w}$ の分布に関する期待値であるとする。

ここで明らかにしておきたいのは、観察不可能な変数 q は (25) 式から完全に消え、条件付期待値 $E(Y \mid \mathbf{x}, \mathbf{w})$ は $(Y, \mathbf{x}, \mathbf{w})$ が独立確率変数である限り容易に推計できる。

所与の $\mathbf{x}^0$ に対して、確率変数に関する平均偏微効果は $\partial\hat{\mu}_2(\mathbf{x}^0, \mathbf{w})\partial x_j$ と表せる。ここで $\mu_2(\mathbf{x}, \mathbf{w}) \equiv E(Y \mid \mathbf{x}, \mathbf{w})$。すなわち、次式を得る。

$$E_w\left[\partial\mu_2(\mathbf{x}^0, \mathbf{w}) \big/ \partial x_j\right] = E\left\{E\left[\theta_j(\mathbf{x}^0, q) \mid \mathbf{w}\right]\right\} = \delta_j(\mathbf{x}^0) \tag{26}$$

2.5 線形モデル

2.5.1 単一方程式の OLS 推計[4]

母集団が次のような線形モデルに従っているとしよう。

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u \tag{27}$$

ここで $Y, x_1, x_2, \ldots x_k$ は観察できる確率変数であり、u は誤差項、$\beta_0, \beta_1, \beta_2, \ldots \beta_k$ は推計されるパラメータである。

誤差項 u には様々な要素が含まれる。それには欠落変数 (omitted variables) や測定誤差 (measurement error) も含まれる。

パラメータ β_j の一致推定を得るための大前提は、誤差項の平均がゼロであり、説明変数と無相関であることである。

$$E(u) = 0, \quad Cov(x_j, u) = 0 \quad j = 1, 2, \ldots, k \tag{28}$$

説明変数 x_j は誤差項 u と無相関であれば**外生変数**であるという。相関している場合には、内生変数と考えるべきである。

問題 1　欠落変数（omitted variables）

例えば、$E(Y \mid \mathbf{x}, q)$ が本来の条件付期待値であるときに、q が観察不可能化あるいはデータとして集められていないために、推計できない。q と x_j が相関している場合には、x_j は内生だということになる。

説明変数と観察不可能な欠落変数の間の相関には自己選択 (self-selection) の問題があるといわれている。例えば、賃金を被説明変数として、説明変数を教育年数、観察不可能な変数を能力だとすれば、教育年数は能力の関数である可能性が高い。

問題 2　観測誤差（measurement error）

偏微効果を知りたい変数が不完全にしか測定されない場合には、観測誤差が誤差項と相関を持つ可能性がある。例えば、関心のある変数が実効限界税率であるのに、平均税率しかわからない場合や、個別企業の資本の実際の減価償却率を用いたいのに、わかるのは産業全体の平均減価償却率である場合などが当てはまる。

[4] 本節の議論は Wooldridge (2002, Chapter 4) に基づいている。

問題 3　同時決定（simultaneity）

説明変数 x_j が被説明変数 Y によって決定されているという側面があれば、誤差項と被説明変数が相関している可能性が高い。例えば Y を都市殺人率として、x_j を警察人員だとすれば、x_j は Y によって決定されてる可能性はある。

これらの問題にどのように対処すればいいのだろうか。

解法 1　欠落変数に関する対応

次のようなモデルを考えよう。

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \gamma q + u \tag{29}$$

$$E(v \mid x_1, x_2, \ldots x_k, q) = 0 \tag{30}$$

ここで q が欠落変数、v は構造誤差（structural error）。

解法 1-1　q を誤差項に入れると (29) は次のように表せる。

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u \tag{31}$$

$$u \equiv \gamma q + v \tag{32}$$

この場合の問題は q が説明変数 x_i と相関している場合には、誤差項 u と説明変数が相関を持ち、x_i の外生性が問題になる。またその結果、推定される β_i も一致推定ではなくなる。

欠落変数を無視した場合の OLS の確率極限（plims）を求めてみよう。

q が説明変数 $x_1, x_2, \ldots x_k$ で説明される線形モデルを考える。

$$q = \delta_0 + \delta_1 x_1 + \ldots + \delta_k x_k + r \tag{33}$$

ここで $E(r) = 0,\ cov(x_j, r) = 0,\ j = 1, 2, \ldots, k$

(33) を (31)(32) に代入すると、

$$Y = (\beta_0 + \gamma\delta_0) + (\beta_1 + \gamma\delta_1)x_1 + (\beta_2 + \gamma\delta_2)x_2 + \ldots + (\beta_k + \gamma\delta_k)x_k + v + \gamma r \tag{34}$$

ここで、誤差項 $u + \gamma r$ は平均ゼロで説明変数とは相関していない。一般に係数の確率極限は $plim\hat{\beta}_j = \beta_j + \gamma\delta_j$ と表せる。q と相関しているのは変数 x_k のみであるとすると、δ_j $(j \neq k)$ はゼロである。したがって

$$\begin{aligned} &plim\ \hat{\beta}_j = \beta_j \ \ j \neq k \\ &plim\ \hat{\beta}_k = \beta_k + \gamma\left[cov(x_k, q) / var(x_k)\right] \end{aligned} \tag{35}$$

ここで $\delta_k = cov(x_k, q)$。この式によってバイアスの方向 (sign）やその大きさがわかる。例えば $\gamma > 0$ であれば、x_k と q は生の相関があり、バイアスは正である。x_k と q との共分散に比べて、x_k の分散が十分大きい場合には、バイアスは比較的小さいことになる。一般に x_k の符号条件を決めることは難しい。

解法 1-2　q の代理変数 (proxy variable) を用いる。

代理変数の条件として次の二つがある。
（1）代理変数は構造モデルの中で重複 (redundant あるいは ignorable) していること。
z を q の代理変数として、z が構造モデルの中で重複しているとして、

$$E(Y \mid \mathbf{x}, q, z) = E(Y \mid \mathbf{x}, q) \tag{36}$$

が成り立つことを意味している。すなわち Y を説明するのに z は関係がないということである。
（2）欠落変数 q と説明変数 x_j の相関は z の影響をコントロールすると (partial out) ゼロになること。これは次のように表せる。

$$L(q \mid 1, x_1, \ldots, x_k, z) = L(q \mid 1, z) \tag{37}$$

q と r を誤差項よりなる線形モデルで表す。

$$q = \theta_0 + \theta_1 z + r \tag{38}$$

ここで $E(r) = 0,\ cov(z, r) = 0$
z が q の適切な代理変数であるとすれば $\theta_1 \neq 0$ であり、(37) は次の関係と同値となる。

$$cov(x_j, r) = 0 \quad j = 1, 2, \ldots, k \tag{39}$$

z が q に強い相関を持っていることで、z がモデルに含まれると、x_j と q の相関が弱められるということである。
(29) に (38) を代入すると、構造モデルは次のように表せる。

$$Y = (\beta_0 + \gamma\theta_0) + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \gamma\theta_1 z + (\gamma r + v) \tag{40}$$

誤差項 $\mathbf{u} \equiv \gamma r + v$ はすべての x_j と無相関である。この場合、OLS 推計は一致推定となり、偏微効果を測ることができる。

解法 2　観測誤差に関する対応

解法 2-1　被説明変数に誤差が含まれる場合

$$Y^* = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + v \tag{41}$$

真の被説明変数を Y とすると

$$Y^* = y - e_0 \tag{42}$$

(42) を (41) に代入して、

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + v + e_0 \tag{43}$$

$x_1, x_2, \ldots x_k$ は観察可能であり、Y も観察できれば、単純に OLS を推計すればいいことになる。e_0 は x_j とは独立であり、無相関だとすれば、OLS 推定は一致推定である。
もし e_0 と v が無相関であれば、$var(v + e_0) = \sigma_v^2 + \sigma_0^2 > \sigma_v^2$ となり、観測誤差を含む推定のほうが分数は大きくなり、パラメータの推定誤差も拡大する。しかし分数の拡大によって OLS 誤差が無効になることは無い。

解法 2-2　説明変数に誤差が含まれる場合

$$Y^* = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k^* + v \tag{44}$$

$x_1, x_2, \ldots x_{k-1}$ は観察可能だが、x_k^* は観測不可能であるとする。v は平均ゼロで x_k^* を含むすべての説明変数と無相関であるとする。観測誤差が次のように表せるとする。

$$e_k = x_k - x_k^*, \quad E(e_k = 0) \tag{45}$$

v は x_k^* とも x_k、e_k とも無相関であると仮定する。
次の二つの仮定が重要となる。

1)

$$Cov(x_k, e_k) = 0 \tag{46}$$

(45) より e_k は観測不可能な x_k^* と相関していることになる。
ここで $x_k^* = x_k - x_k$ とし、(44) に代入すると、

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + (v - \beta_k e_k) \tag{47}$$

v と e_k は平均ゼロですべての x_j と無相関であり、$v - \beta_k e_k$ も平均ゼロですべての x_j と無相関である。この場合 x_k^* の代わりに x_k を入れて OLS 推定してもすべての β_j に対して一致推定が得られる。

2)

$$Cov(x_k^*, e_k) = 0 \tag{48}$$

この条件は古典的観測誤差問題（classical errors-in-variables=CEV）の仮定であり、観測誤差が観測不可能な変数 x_k^* と無相関であることを意味している。観測される変数にも観測誤差が含まれているとすると、$x_k = x_k^* + e_k$ と表せ、x_k^* と e_k は無相関であるということである。
(48) が成り立つと、x_k と e_k は相関していなければばならない。

$$Cov(x_k, e_k) = E(x_k e_k) + E(e_k^2) = \sigma_{ek}^2 \tag{49}$$

この場合、x_k と e_k の共分散は観測誤差の分散と等しくなる。ここで x_k と e_k が相関しているということは (47) の OLS 推定に問題をもたらす。v と x_k は無相関で、x_k と $(v - \beta_k e_k)$ の共分散は $Cov(x_k, v - \beta_k e_k) = -\beta_k Cov(x_k, e_k) = -\beta_k \sigma_{ek}^2$ となる。OLS 推定はすべてのの推定で不一致となる。

References

[1] Wooldridge, J.M. (2002) *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: The MIT Press.