ノンパラメトリック操作変数推定における変数選択規準

末石直也

京都大学経済学研究科

IBIS2010 企画セッション
「計量経済と機械学習」

内生性とは？

- 線形モデル

$$y_i = x_i'\beta + e_i \tag{1}$$

を考える。ただし、$x_i \in \mathbb{R}^p$ である。

- 誤差項と説明変数に相関があるとき、つまり

$$E[x_i e_i] \neq 0 \tag{2}$$

であるとき、内生性が存在するという。

- (1) を単なる近似（線形射影）ではなく、経済学的に意味のあるモデルとして考えるとき、内生性が生じる。

内生性が生じる例 (omitted variable)

- 賃金が次のように決定されるとする。

$$\log(w_i) = \beta_0 + \beta_1 edu_i + \beta_2 abil_i + \eta_i \tag{3}$$

- 観測できない *ability* を無視して

$$\log(w_i) = \delta_0 + \delta_1 edu_i + e_i \tag{4}$$

を LS で推定すると、線形射影の係数に収束する。

$$\hat{\delta}_1 \xrightarrow{p} \beta_1 + \beta_2 \frac{\mathrm{Cov}[edu_i, abil_i]}{\mathrm{Var}[edu_i]} \tag{5}$$

- 高学歴のために高所得を得ているのか、高い潜在能力のために高所得を得ているのか、判別できない。

操作変数法 (2 段階最小 2 乗法)

- 説明変数 x_i とは相関があるが、誤差項 e_i とは相関がない変数 (操作変数) z_i を用いる。

$$
\begin{aligned}
y_i &= x_i'\beta + e_i \quad (6)\\
x_i &= \Gamma' z_i + u_i \quad (7)
\end{aligned}
$$

ただし、$z_i \in \mathbb{R}^l$ で、$l \geq p$ とする。

- (7) を (6) に代入すれば

$$
y_i = z_i'\Gamma\beta + e_i + u_i'\beta \equiv z_i'\lambda + v_i \quad (8)
$$

を得る。$E[z_i v_i] = 0$ なので、λ は識別可能。

識別と推定

- パラメータ β は

$$\lambda = \Gamma\beta \tag{9}$$

の解として識別される。

- $\hat{x}_i = \hat{\Gamma}' z_i$ とすると、β は y_i を $\hat{x}_i$ に回帰することで推定できる。

$$\hat{\beta} = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'y \tag{10}$$

操作変数の例

- Angrist and Krueger (1991) は賃金方程式を推定するために、Quarter of birth を操作変数に使った。
- アメリカでは中退できる時期が生まれた時期によって異なるので、教育年数と誕生日には相関がある。しかし、誕生日と個人の能力に相関があるとは考えにくい。

ノンパラメトリックモデル

- 次のようなノンパラメトリックモデルを考える。

$$y_i = g(x_i) + e_i \tag{11}$$

- $E[e_i|x_i] \neq 0$、つまり $g(x_i) \neq E[y_i|x_i]$ なので、通常のノンパラメトリック回帰の手法を用いることができない。

ノンパラメトリック操作変数モデル

- 操作変数 z_i の存在を仮定する。

$$\begin{aligned} y_i &= g(x_i) + e_i \qquad (12) \\ E[e_i|z_i] &= 0 \qquad (13) \end{aligned}$$

- 関数 $g(\cdot)$ は積分方程式

$$E[y_i|z_i = z] = \int g(x) dF_{x_i|z_i=z} \qquad (14)$$

の解として特徴づけられる。

Newey and Powell (2003)

- 基底関数 $\{\psi_1(x), \psi_2(x), \cdots\}$ を用いて、$g(x)$ を近似する。

$$y_i \approx \sum_{j=1}^{p} \beta_j \psi_j(x) + e_i \tag{15}$$

- 基底関数 $\{q_1(z), q_2(z), \ldots\}$ を用いて、$E[\psi_j(x)|z]$ を近似する。

$$\psi_j(x_i) \approx \sum_{k=1}^{l} \gamma_{jk} q_k(z_i) + u_i \tag{16}$$

- 実用上、どの程度の p と l を選べばよいだろうか？

ノンパラメトリック回帰モデル

- 真のモデルは無限次元。

$$y_i = \mu(x_i) + e_i \tag{17}$$

$$\mu(x_i) = \sum_{j=1}^{\infty} \beta_j \psi_j(x_i) \tag{18}$$

$$E[e_i|x_i] = 0 \tag{19}$$

$$E[e_i^2|x_i] = \sigma^2 \tag{20}$$

有限次元モデルによる近似

- 有限次元のモデル h で近似する。

$$y_i = \psi^h(x_i)'\beta_h + e_i \tag{21}$$

ただし、$\psi^h(x_i) = (\psi_1(x_i), \ldots, \psi_{p_h}(x_i))'$ である。

- LS 推定量は

$$\hat{\mu}(h) = \Psi_h\hat{\beta}_h = \Psi_h(\Psi_h'\Psi_h)^{-1}\Psi_h y \equiv M(h)y \tag{22}$$

ただし、$\Psi_h = (\psi^h(x_1), \ldots, \psi^h(x_n))'$。

損失関数

- 説明変数の数 p_h を増やせばバイアスは小さくなるが、分散は大きくなる。
- 最もポピュラーな損失関数は平均 2 乗誤差

$$L_n(h) = \frac{1}{n} \|\mu - \hat{\mu}(h)\|^2 \tag{23}$$

である。ただし、$\mu = (\mu(x_1), \ldots, \mu(x_n))'$。

Mallows の C_p

- Mallows の規準 (Mallows (1973))

$$\begin{aligned} C_n(h) &= \frac{1}{n}\|y - \hat{\mu}(h)\|^2 + \frac{\sigma^2}{2n}\mathrm{tr}(M(h)) \\ &= \frac{1}{n}\|y - \hat{\mu}(h)\|^2 + \frac{\sigma^2}{2n}p_h \end{aligned} \tag{24}$$

- $C_n(h)$ はリスクの不偏推定量になっている。

$$E[C_n(h)|X] = E[L_n(h)|X] + const. \tag{25}$$

Li の最適性

- $\mathcal{H}_n$ を候補となっているモデルの集合、$\hat{h} = \arg\min_{h \in \mathcal{H}_n} C_n(h)$ とする。Li (1987) は

$$\frac{L_n(\hat{h})}{\inf_{h \in \mathcal{H}_n} L_n(h)} \xrightarrow{p} 1 \tag{26}$$

を示した。

- 漸近的に最適なモデルを選ぶことができる。
- 真のモデルが有限次元であるときには成り立たない。

ノンパラメトリック操作変数モデル

- 操作変数モデルに話を戻す。

$$y_i = g(x_i) + e_i, \tag{27}$$
$$E[e_i|z_i] = 0 \tag{28}$$
$$E[e_i^2|z_i] = \sigma^2 \tag{29}$$

- 無限次元モデル

$$y = \sum_{j=1}^{\infty} \beta_j \psi_j(x) + e_i \tag{30}$$
$$\psi_j(x) = \sum_{k=1}^{\infty} \gamma_{jk} q_j(z_i) + u_{ji} \tag{31}$$

を推定する。

有限次元モデルによる近似

- 有限次元モデル h は

$$
\begin{aligned}
y_i &= \psi_h(x_i)'\beta_h + e_i && (32) \\
\psi^h(x_i) &= \Gamma_h' q^h(z_i) + u_i && (33)
\end{aligned}
$$

ただし、$q^h(z_i) = (q_1(z_i), \ldots, q_{l_h}(z_i))'$ である。

- $Q_h = (q^h(x_1), \ldots, q^h(x_n))'$、$P(h) = Q_h(Q_h'Q_h)^{-1}Q_h'$ とする。2 段階最小 2 乗推定量は

$$
\hat{g}(h) = \Psi_h(\Psi_h' P(h)\Psi_h)^{-1}\Psi_h' P(h) y \tag{34}
$$

である。

損失関数

- 別の損失関数を考える。

$$L_n(h) = \frac{1}{n}\,\|P(h)(g - \hat{g}(h))\,\|^2 \tag{35}$$

ただし、$g = (g(x_1), \ldots, g(x_n))'$ である。

- 関数 $g(\cdot)$ は

$$g(\cdot) = \arg\min_{\phi(\cdot)} E\left[\left|E[y_i|z_i] - \int \phi(x)dF_{x_i|z_i}\right|^2\right] \tag{36}$$

として特徴づけられる。

変数選択規準

- 次の選択規準を提案する。

$$C_n(h) = \frac{1}{n}\|P(h)(y - \hat{g}(h))\|^2 - \frac{\sigma^2}{n}(l_h - 2p_h) \tag{37}$$

- $C_n(h)$ は漸近的に $E[L_n(h)|Z]$ の不偏推定量になっている。

最適性

- $\mathcal{H}_n$ を候補となっている説明変数と操作変数の組み合わせ、$\hat{h} = \arg\min_{h\in\mathcal{H}_n} C_n(h)$ とする。

$$\frac{L_n(\hat{h})}{\inf_{h\in\mathcal{H}_n} L_n(h)} \xrightarrow{p} 1 \tag{38}$$

が成り立つ。

- 漸近的に最適な説明変数と操作変数の組み合わせを選ぶことができる。

ill-posed inverse problem

- 変数選択だけでは、良い振る舞いをする推定量が得られないことがある。
- 関数 $g(\cdot)$ が第 1 種フレドホルム型積分方程式の解になっていることが原因。

$$r(z) \equiv E[y_i|z_i = z] = \int g(x) dF_{x_i|z_i=z} \equiv (Kg)(z) \tag{39}$$

- $g(z) = K^{-1}(r(z))$ だが、K^{-1} は一般に連続ではない。

Blundell, Chen and Kristensen (2007)

- BCK (2007) は次のような推定量を提案した。

$$\hat{\beta}_{h,\lambda} = \arg\min_{\beta_h} \frac{1}{n} \|P(h)(y - \Psi_h \beta_h)\|^2 + \lambda \beta_h' \Omega_h \beta_h \tag{40}$$

ただし、$\Omega_h = \Omega_{0h} + \Omega_{2h}$ で

$$\begin{aligned} \Omega_{0h} &= \frac{1}{n} \sum_{i=1}^{n} \psi^h(x_i) \psi^h(x_i)' \\ \Omega_{2h} &= \int \left\{ \frac{d^2 \psi^h(x)}{dx^2} \right\} \left\{ \frac{d^2 \psi^h(x)}{dx^2} \right\}' dx \end{aligned}$$

選択問題

- 最小化問題には解析的な解が存在する。

$$\hat{g}_\lambda(h) = \Psi_h(\Psi_h' P(h)\Psi_h + n\lambda\Omega_h)^{-1}\Psi_h P(h)y \tag{41}$$

- 損失関数

$$L_n(h,\lambda) = \frac{1}{n}\|P(h)(g - \hat{g}_\lambda(h))\|^2 \tag{42}$$

を最小にするモデル h と λ の組み合わせを考える。

選択規準と最適性

- 選択規準は

$$C_n(h,\lambda) = \frac{1}{n}\|P(h)(y - \hat{g}_\lambda(h))\|^2 - \frac{\sigma^2}{n}\mathrm{tr}(P(h) - 2W_\lambda(h)) \qquad (43)$$

ただし、$P(h)\hat{g}_\lambda(h) = W_\lambda(h)y$。

- $(\hat{h},\hat{\lambda}) = \arg\min_{h\in\mathcal{H}_n,\lambda\in\Lambda_n(h)} C_n(h,\lambda)$ とする。

$$\frac{L_n(\hat{h},\hat{\lambda})}{\inf_{h\in\mathcal{H}_n,\lambda\in\Lambda_n(h)} L_n(h,\lambda)} \xrightarrow{p} 1 \qquad (44)$$

が成り立つ。

参考文献

- Angrist J. D., and A. B. Krueger (1991): “Does Compulsory School Attendance Affect Schooling and Earnings?,” *Quarterly Journal of Economics*, 106, 979-1014.
- Blundell, R., X. Chen, and D. Kristensen (2007): “Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves,” *Econometrica*, 75, 1613-1669.
- Li, K.-C. (1987): “Asymptotic Optimality for C_p, C_L, Cross-Validation and Generalized Cross-Validation: Discrete Index Set,” *Annals of Statistics*, 15, 958-975.
- Mallows, C. L. (1973): “Some Comments on C_p,” *Technometrics*, 15, 661-675.
- Newey, W. K., and J. L. Powell (2003): “Instrumental Variable Estimation of Nonparametric Models,” *Econometrica*, 1565-1578.