

8. 実証分析の考え方（前半）

今までは、行動理論のモデル（主に図解）によって
外生変数の変化が内生変数に与える影響を分析してきた。

紙と鉛筆だけの世界
現実（データ）と直接的関係なし

経済学の応用分野においては、理論と現実との関係を考えることが重要。

数個の（恣意性を疑われがちな）例で話しを終わらせるのではなく、
科学的な手法で行いたい。（統計学的手法）

「理論と実証の橋渡し」（この教科書の帯）

“理論に基づいた実証”

“実証を志向する理論”

8.1 実証モデルとは

焦点を当てる変数

観察されない部分

$$L_i = \alpha + \beta w_i + \gamma P_i + \delta I_i + \kappa X_i + \varepsilon_i$$

i さんの
労働供給時間

その他の（観察される）要素

と仮定。（誘導形）

理論モデル

$$L_i = L(w_i, P_i, I_i)$$

制約条件付き

最適化問題の解
（構造推定アプローチ）

静学的な関係を
考えている。

個票データ	L_1	w_1	P_1	I_1	X_1
(n 人の個人)	L_2	w_2	P_2	I_2	X_2

L_n	w_n	P_n	I_n	X_n
-------	-------	-------	-------	-------

δ は、他の条件 (w_i 、 P_i 、 X_i 、 ε_i) が一定のとき、
 I_i の 1 単位の増加が L_i に与える因果的効果

(その他についても同様)

* 実証分析の目標は、 β 、 γ 、 δ といった、注目する外生変数の係数を
 データから統計的に推測することによって、それぞれの外生変数の
 因果的効果を推し測ることである。

8.2 実験データと識別問題・識別戦略

8.2.1 実験データ

非現実的だが、「実験」が出来るとする。

全労働者

データ作成
 (サンプル)

無作為抽出
 (ランダム・サンプリング)

再び、無作為に二つのサブサンプルに分割
 (グループ)

グループ 1

グループ 2

ε_{i1}

ε_{i2}

分布は大きくは異ならないと考えられる。

実験の開始

グループ 1

グループ 2

外生変数を

$$I_{i1}^T$$

$$I_{i1}^C$$

実験者が変化させる

$$I_{i2}^T = I_{i1}^T + \Delta I$$

$$I_{i2}^C$$

$$i = 1, \dots, n_T$$

$$i = 1, \dots, n_C$$

treatment

control

group

group

(実験群)

(対照群)

$$L_{i1}^T$$

$$L_{i2}^T = L_{i1}^T + \Delta L_i^T$$

i さんの変化分

構築されるパネルデータ

$$L_{11}^T, w_{11}^T, P_{11}^T, I_{11}^T, X_{11}^T$$

$$L_{12}^T, w_{12}^T, P_{12}^T, I_{12}^T, X_{12}^T$$

$$L_{21}^T, w_{21}^T, P_{21}^T, I_{21}^T, X_{21}^T$$

$$L_{22}^T, w_{22}^T, P_{22}^T, I_{22}^T, X_{22}^T$$

$$L_{n_T 1}^T, w_{n_T 1}^T, P_{n_T 1}^T, I_{n_T 1}^T, X_{n_T 1}^T$$

$$L_{n_T 2}^T, w_{n_T 2}^T, P_{n_T 2}^T, I_{n_T 2}^T, X_{n_T 2}^T$$

$$L_{11}^C, w_{11}^C, P_{11}^C, I_{11}^C, X_{11}^C$$

$$L_{12}^C, w_{12}^C, P_{12}^C, I_{12}^C, X_{12}^C$$

8.2.2 差の差推定法

ここで、出てくる「時間」は、推定上から生ずるもので、対象としている行動モデル（静学的労働供給モデル）には含まれていないもの。

T

C

$t = 1$

$t = 2$

$$L_{it} = \alpha + \beta w_{it} + \gamma P_{it} + \delta I_{it} + \kappa X_{it} + \varepsilon_{it}$$

For i in T

$$I_{i1} \neq I_{i2}$$

For i in C

$$I_{i1} = I_{i2}$$

観察されない 観察されない

選好 出来事

$$\varepsilon_{it} = (\theta_i + \phi_t) + \mu_{it}$$

新たな誤差項（他の変数とは相関しない）

t を通じて 全ての i に

変化しない 共通

これらは全て、観察者にとって見えない変数

各 i について、 $t = 2$ から $t = 1$ の差を考えると

$$\begin{aligned} L_{i2} - L_{i1} &= \beta(w_{i2} - w_{i1}) + \gamma(P_{i2} - P_{i1}) + \delta(I_{i2} - I_{i1}) + \kappa(X_{i2} - X_{i1}) \\ &\quad + (\theta_i - \theta_i) + (\phi_2 - \phi_1) + (\mu_{i2} - \mu_{i1}) \end{aligned}$$

Tに属する個人を考えると

$$L_{i2} - L_{i1} = \delta(I_{i2} - I_{i1}) + (\phi_2 - \phi_1) + (\mu_{i2} - \mu_{i1}) \quad \text{仮定より、これらには}$$

相関関係はない。

もし観察不可能な出来事がなければ
($\phi_2 = \phi_1$)

$$L_{i2} - L_{i1} = \delta(I_{i2} - I_{i1}) + (\mu_{i2} - \mu_{i1})$$

これ全体を誤差項と看做して
通常の実二乗法で δ の推定値を
求める。

もし ($\phi_2 - \phi_1$) の値がいくらか分からない (観察不可能) ときは
誤差項の部分を除いて
[$\delta(I_{i2} - I_{i1}) + (\phi_2 - \phi_1)$] は、 $\delta(I_{i2} - I_{i1})$ の部分による貢献と
($\phi_2 - \phi_1$) の部分による貢献とが
区別が出来ない (δ も未知であることに注意)。
(δ が識別されない)

そこで、($\phi_2 - \phi_1$) の影響を除くため、control group を使う。

$$L_{i2}^T - L_{i1}^T = \delta(I_{i2}^T - I_{i1}^T) + (\phi_2 - \phi_1) + (\mu_{i2}^T - \mu_{i1}^T)$$

$$L_{j2}^C - L_{j1}^C = (\phi_2 - \phi_1) + (\mu_{j2}^C - \mu_{j1}^C)$$

“差の差” $[(L_{i2}^T - L_{i1}^T) - (L_{j2}^C - L_{j1}^C)] = \delta(I_{i2}^T - I_{i1}^T) + [(\mu_{i2}^T - \mu_{i1}^T) + (\mu_{j2}^C - \mu_{j1}^C)]$

をとる。

ここの部分は、($I_{i2}^T - I_{i1}^T$) と

相関を持たない誤差項

8.3 観察データと識別問題

実験データを使って、やりたかったこと。

注目したい外生変数の内生変数に対する因果的効果を統計学の道具で推し量る。

見えない (データで観察されていない) 部分が、内生変数に与える影響を除去

したい。

$$\varepsilon_{it} = \theta_i + \phi_t + \mu_{it}$$

ここの部分は見えないが、
系統的に他の変数に影響を
与えていないので（相関が
ない）、（多くの個人の）「平
均をとる」ことによって、
除去できる

8.3.1 観察データの特徴

実験は、金銭的

倫理的に 無理な場合が多い。

実験データで仮定できた仮定が

満たされているとはいいにくくなる（後述：8.4）

8.3.2 脱落変数バイアス（1.イントロダクションの1.4「実証分析」も見直すこと）

観察データ上においては、通常、説明変数と労働者の観察されない属性
との間に相関関係があると考えられる。

非労働所得と労働に対する選好との間に正の相関が疑われる。

労働者 i と労働者 j との間の差分を考える。

この二人の非労働所得以外の観察される変数の値が同一だとすると、

$$L_i - L_j = \delta(I_i - I_j) + (\theta_i - \theta_j) + (v_i - v_j)$$

よって、

$$\delta = \frac{(L_i - L_j) + (\theta_i - \theta_j) + (v_i - v_j)}{I_i - I_j}$$

ここで、 $(v_i - v_j)$ と $(I_i - I_j)$ は無相関なので、平均的に無視することが

できるが、 $(\theta_i - \theta_j)$ と $(I_i - I_j)$ の正の相関より、他の条件を全て

一定とすれば、 δ は、 $(L_i - L_j)$ と $(I_i - I_j)$ の比率として与えられる

べきにも関わらず、例えば

$$\frac{-20}{2} = -10$$

となるべきところが、

$$\frac{-20+10}{2} = -5$$

となって、絶対値で過少に評価されてしまう。

推定値が真の値と比べて、過大である場合は、上方バイアス

過少である場合は、下方バイアス

があると言う。

上記の例は、上方バイアス。

8.3.3 サンプルセクションバイアス

例 各労働者は、 w 、 P 、 I を見て、労働供給 L を決めている。

ここで、 w は、働いたら得られたであろう賃金 ($L = 0$ を選んだとしても)

ところが、データは、 $L = 0$ を選んだ労働者が選んだであろう賃金 w を記入していないことが多い。

そのような労働者を取り除いて、 $L > 0$ の労働者だけで、サンプルを構成してしまうと、 θ_i は、たとえ I_i と相関関係がもともとなかったとしても、構成されたサンプルにおいては、正の相関関係が発生してしまう。

8.3.4 測定誤差バイアス

推定値にバイアスをもたらす第 3 の原因。