労働経済学

Lecture 6 実証研究における因果的効果の識別

張　俊超

11th May 2017

実証モデル

- **単回帰**

$$L_i = \alpha + \beta w_i + \varepsilon_i$$

誤差項以外、労働供給時間は賃金だけに依存する。(明らかに間違っている)

- **重回帰**

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

誤差項以外、労働供給時間は賃金、消費者物価、非労働所得、他の観察可能の変数に依存する。X_i は複数の変数を含めるベクトル。

条件付き期待値

回帰モデルは条件付き期待値で表現できる

$$L_i = E(L_i|w_i, P_i, I_i, X_i) + \varepsilon_i = \alpha + \beta w_i + \gamma P_i + \delta I_i + \kappa X_i + \varepsilon_i$$

$E(L_i|w_i, P_i, I_i, X_i)$ **は賃金、消費者物価、非労働所得、その他の変数が一定の場合、労働供給時間** L_i **の平均値。図でも説明できる (板書だけ)。**

実証モデル

一般的に、実証分析では、一つの内生変数を着目して、その変数が従属変数に与える効果を推定する。仮に、賃金が労働供給時間に与える効果を見るために、

$$L_i = \alpha + \beta w_i + \tau T_i + \varepsilon_i \quad (1)$$

- w_i 賃金は内生変数となる。（理論モデルでの内生変数と違う）
- T_i はコントロール変数のベクトル、データ上観察可能な、P_i, I_i, X_i などを含む。説明を簡単化するために、これからは (1) 式のような表記を扱う。（教科書と違う）

識別問題と識別戦略

- **識別問題**

  データから、推定したい未知のパラメータ（β、τ など）を一意的に定めることができるどうかという理論的問題です。

- **識別戦略**

  未知のパラメータ（β、τ など）を一意的に定めるための統計的方法。

データの種類

- 実験データ

  自然科学分野によく使われる。無作為に処置群、対照群を抽出し、その二つのグループの間の平均の差を比較する。実験は完璧な場合、処置変数以外のすべての変数は一定のまま、識別戦略は不要。(教科書では、実験データでの識別戦略も説明したが、実験が完璧でない時の統計的方法だと考えてよい)

- 観察データ

  非実験データ。観察データは労働経済学において、賃金、非労働所得、消費者物価などの観察できるものを記録したデータ。特徴１、記録したデータは、労働者が自分が選択した結果。特徴２、一つの変数が変わる時に、一般的に、その他の変数も同時に変わる。

観察データにおける識別問題

単回帰を考えて、賃金が労働供給時間に与える効果を見る

$$L_i = \alpha + \beta w_i + \varepsilon_i$$

- 最小二乗法（OLS）で推定した $\hat{\beta}$ は BLUE 推定量であるために、五つの仮定を満たさないといけない。
- その中、$Cov(w_i, \varepsilon_i) = 0$ という仮定が極めて重要。$Cov(w_i, \varepsilon_i) = 0$ が満たされない場合、OLS は不偏推定量ではなく、一致推定量でもない。
- $\hat{\beta} = \frac{\sum(w_i - \bar{w})(L_i - \bar{L})}{\sum(w_i - \bar{w})^2} = \beta + \frac{\sum(w_i - \bar{w})\varepsilon_i}{\sum(w_i - \bar{w})^2}$
- $Cov(w_i, \varepsilon_i) = 0$ が満たさない場合、$E(\hat{\beta}) = \beta + E(\frac{\sum(w_i - \bar{w})\varepsilon_i}{\sum(w_i - \bar{w})^2})$ のため、第二項は 0 にならず、不偏ではない。
- 一致性について、$plim\hat{\beta} = \beta + \frac{Cov(w_i, \varepsilon_i)}{Var(w_i)}$ のため、第二項は 0 にならず、一致ではない。

内生性

説明変数と誤差項との間の相関は内生性と呼ぶ。一般的に、以下の三種類の内生性がある。

- 脱落変数バイアス
- サンプルセレクッションバイアス
- 測定誤差バイアス

脱落変数バイアス

$$L_i = \alpha_0 + \beta_0 w_i + \varepsilon_i$$

$$L_i = \alpha_1 + \beta_1 w_i + \tau_1 T_i + u_i$$

上の式は間違って、下の式は正しいとする。下の式から見れば、労働供給時間は、賃金と T_i に含むいろんな変数で正しく解釈できる。

上の式は、入れるべき変数 T_i を脱落し、w_i と ε_i の間に相関がある。

しかし、現実には、すべての変数を T_i ベクトルに入れるのが難しい。観測できない変数もある。（やる気、野心など）

バイアスの方向

$Cov(w_i, \varepsilon_i) = 0$ が満たされない以下の推定式を考える

$$L_i = \alpha_0 + \beta_0 w_i + \varepsilon_i$$

バイアスの方向は $Cov(L_i, \varepsilon_i)$ と $Cov(w_i, \varepsilon_i)$ から予測できる。
上方バイアス：$\beta_0 < \hat{\beta}_0$
下方バイアス：$\beta_0 > \hat{\beta}_0$
表で説明。（板書だけ）