ミクロ計量経済早わかり (2)

横浜国立大学国際社会科学研究院教授
大森 義明

今回は、脱落変数問題の標準的解決手法を解説します。

固定効果モデル

説明変数と相関する、(誤差項に含まれる) 観察不可能な属性を「観察可能にする」のが、固定効果モデルの識別戦略です。まず、(1) 式[1]の誤差項 ε_i を観察不可能な個人の属性 θ_i とそれ以外の部分 v_i の和として表わせると仮定します。

$$\ln w_i = \alpha + \beta educ_i + \gamma \exp_i + \delta ten_i + \kappa X_i + \theta_i + v_i \quad (2)$$

ここで、個人 i が個人 j である ($j = i$) 場合には1、そうでない場合には0の値を取る、各個人を識別する (n−1) 個の (ダミー) 変数 D_i^i ($j = 2, \cdots, i, \cdots, n$ であり、n はサンプルサイズ) を考えてください。個人 i については、(n−1) 個のダミー変数の内、D_i^i のみが1の値を取り、他は全て0の値を取ります。これらの変数は観察可能であり、説明変数として用いることができ、他の説明変数との相関 (例えば、他の個人よりも高い能力を持つ個人 i が他の個人よりも学歴が高いなど) も問題にならないことがポイントです。

これらのダミー変数を用いると、(2) 式を次のように表わすことができます。

$$\ln w_i = \theta_1 + \beta educ_i + \gamma \exp_i + \delta ten_i + \kappa X_i$$
$$+ \theta_2 D_i^2 + \theta_3 D_i^3 \cdots + \theta_i D_i^i \cdots + \theta_n D_i^n + v_i \quad (3)$$

ここで、D_i^i の係数が θ_i になっていることに注意してください。$\beta, \gamma, \delta, \kappa$ と同様に、$\theta_1, \theta_2, \cdots, \theta_i, \cdots \theta_n$ も定数であると仮定すると、(3) 式は標準的な回帰モデルになるので、この仮定を設けます。これが固定効果の「固定」の由来です。

しかし、(3) 式の係数を識別できるためには、同一個人を複数時点で観察したパネルデータが必要です。各個人の観察値が2つ以上なければ、$\theta_1, \theta_2, \ldots, \theta_i, \ldots \theta_n$ の識別は不可能だからです。次式は (3) 式のパネルデータ版です。

$$\ln w_{it} = \theta_1 + \beta educ_{it} + \gamma \exp_{it} + \delta ten_{it} + \kappa X_{it}$$
$$+ \theta_2 D_i^2 + \theta_3 D_i^3 \cdots + \theta_i D_i^i \cdots + \theta_n D_i^n + v_{it} \quad (4)$$

各変数の下添え字が個人を表わす i から個人と時点の組み合わせを表わす it になっていることに注意してください。

あるいは、(2) 式に似た形で (4) 式を表わすと次式になります。

$$\ln w_{it} = \theta_1 + \beta educ_{it} + \gamma \exp_{it} + \delta ten_{it} + \kappa X_{it}$$
$$+ \theta_i + v_{it} \quad (5)$$

これを**固定効果モデル**と呼びます。

①誤差項 v の条件付き期待値 (説明変数の値を一定に保ったときの誤差項の期待値) が説明変数の値に依存しない、②多重共線性がない (説明変数が互いに完全相関しない) などのOLSの標準的な識別条件が満たされれば、モデル (4) の係数は識別できます。誤差項の仮定については、特に、個人の $educ$ が誤差項 v と相関してはならない (学歴の変化は無作為でなければならない、例えば、賃金低下を経験した後、キャリアアップのために大学院進学するような行動パターンがあってはならない) ことに注意してください。また、多重共線性に関しては、特に、同一個人の $educ$ が2時点間で異なる (復学歴がある) 必要があることに注意してください。異ならなければ、$educ$ は個人を表わすダミー変数と完全相関してしまい、係数の識別ができません。

固定効果モデルは、計量経済分析用の汎用ソフトで推定できます。以下はstataの例です。まず、*iis* ステートメントを用い、i にあたる変数名を、また、tis ステートメントを用い、t にあたる変数名を宣言します。 例えば、IDとYEARがそれらの変数である場合には、次のようにします。

```
iis ID
tis YEAR
```

固定効果モデルを推定するには、次のようにします。

xtreg 被説明変数名　説明変数のリスト, fe

末尾の “, fe” は固定効果 (fixed effect) を用いることを意味します。(5) 式の場合は、次のようにします。

xtreg lnwage educ exp ten other, **fe**

1　11ページ参照。

差分の差分

パネルデータを扱う際には、時間の経過に伴って生じる、観察不可能な環境変化にも注意が必要です。例えば、賃金に影響する、賃上げ圧力、景気感などを説明変数のみで十分に捉えることは困難なので、これらは誤差項に含まれると考えるべきです。これらが説明変数と相関する場合には、脱落変数バイアスが生じてしまいます。バイアス回避には、時点に特殊な観察不可能な環境属性も固定効果ϕ_tとしてモデルに加えます。

$$\ln w_{it} = \theta_1 + \beta educ_{it} + \gamma \exp_{it} + \delta ten_{it} + \kappa X_{it} + \theta_i + \phi_t + v_{it} \qquad (6)$$

これを差分の差分(Difference in Difference、DID)モデルと呼びます。単純な固定効果モデルを推定すべき特殊な事情がない限り、差分の差分モデルを推定すべきです。

固定効果モデルの説明変数のリストに時間を表わすダミー変数を加えれば、差分の差分推定ができます。stataで行うには、次のようにします。

xtreg 被説明変数名　説明変数のリスト 時間を表わすダミー変数のリスト, fe

例えば、2000年から2005年のパネルデータを用い、2000年を基準年とし、(6)式を推定するには、年ダミー変数YR2001、YR2002、…、YR2005を用い、次のようにします。

xtreg lnwage educ exp ten other YR2001 YR2002 YR2003 YR2004 YR2005, **fe**

操作変数法

固定効果モデルや差分の差分モデルは、各個人について複数の観察値を含む、パネルデータなどがなければ利用できません。これに対し、操作変数法は、各個人について一つの観察値しか含まないクロスセクショナルデータでも利用できます。そこで、再度、(1)式を考えます。

$$\ln w_i = \alpha + \beta educ_i + \gamma \exp_i + \delta ten_i + \kappa X_i + \varepsilon_i \qquad (1)$$

操作変数法の識別戦略は「自然実験」です。例えば、Angrist and Krueger (1991) は次のような有名な自然実験を考えました。米国の多くの州では、満6歳になる暦年に子どもが小学校に秋入学します。第4四半期生まれは6歳未満で入学し、第1四半期生まれは6.5歳程度で入学します。 多くの州では16歳の誕生日で義務教育が終わります。入学と義務教育に関するこれらの規定により、個人は義務教育終了時点で異なる教育年数(あるいは、同じ教育年数の異なる月数)を達成することになります。一部の個人は、義務教育期間が終わり次第、学校を中退するので、第1四半期生まれは他の四半期生まれと比べ平均最終教育年数が低くなります。更に、誕生日は無作為に決まると考えられるので、誕生日は(個人の能力のような、個人の観察不可能な属性を含む)誤差項とは相関しないと考えられます。まるで実験者が誕生日を個人に無作為に割り当て(従って、個人の能力など、個人の観察不可能な属性を一定に保ったままで)、個人の教育年数にばらつきを生じさせ、将来、賃金がどうなるかを確かめる実験をしているかのようです。操作変数(instrumental variable、IV)とは、この例の誕生四半期のように、誤差項との相関が疑われる説明変数(この例では、教育年数)とは相関するが、誤差項とは相関しない変数のことです。

自然実験に基づく因果効果の識別は極めて直感的です。この例では、第1四半期生まれの賃金が他の四半期生まれの賃金より低いならば、それは教育年数が賃金を増加させる因果効果を持つからにほぼ間違いありません。第1四半期生まれと他の四半期生まれの間で賃金と教育年数がそれぞれどれだけ異なるかを調べれば、その比から教育年数の賃金への因果効果を識別することができます。

操作変数法推定をstataで行うには、次のようにします。

ivreg 被説明変数名　(誤差項との相関が疑われる説明変数＝操作変数) 他の説明変数のリスト

例えば、第1四半期生まれを示すダミー変数Q1をeducの操作変数として用いる場合には、次のようにします。

ivreg lnwage (educ = Q1) exp ten other

あるいは、educの因果効果だけに関心があるのであれば、次のようにしても構いません。

ivreg lnwage (educ = Q1)

なぜならば、exp、ten、otherを故意に説明変数のリストから脱落させ、誤差項に含ませてしまったとしても、Q1はそれらとは相関を持たず、従って、educの因果効果をバイアスなく推定できるからです。

操作変数法は、従来から知られていましたが、近年、

ミクロ計量経済分析の中心的役割を果たすようになりました。先述したように、操作変数は（誤差項との相関が疑われる）説明変数と相関するだけでなく、誤差項と無相関である必要があります。従来は、後者の条件を厳格に満たさない変数であっても、操作変数として用いてしまう悪しき慣習が散見されました。例えば、教育年数の操作変数として父親の教育年数を用いるなどです。しかし、今日では、父親の教育年数は操作変数として相応しくないと多くの実証研究者は考えています。なぜならば、父親の教育年数は、父親の能力と相関し、父親の能力は子どもの能力と相関するからです。

近年、ミクロの実証分析では自然実験が注目されています。これは、操作変数と誤差項の無相関の条件を厳格に満たす究極的な操作変数は自然実験にしか見出しようがないからです。このため、ミクロ実証分析では、自然実験とみなせる現象を探すため、制度、法律、文化、慣習、自然災害などの研究が重要性を増しています。

参考文献

Angrist, Joshua D. and Alan B. Krueger (1991). "Does Compulsory Schooling Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics* 106, pp.976-1014.

Joshua D. Angrist and Jorn-Steffen Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, Princeton, New Jersey.

大森 義明（おおもり よしあき）