

## 多重回帰と操作変数法

松下 幸敏

(東京工業大学准教授)

本稿では、多重回帰と操作変数 (instrumental variables, IV) 法について、特にそれぞれの手法のアイデアを中心に解説したい。

例として教育年数が賃金に与える影響を計測することを考えよう。そのために母集団から無作為に標本を抽出し、教育年数ごとに平均賃金を比較したとする。この比較によって、教育が賃金に与える影響を正しく測ることは可能であろうか？

答えは、以下の理由により No である。その理由を理解するための鍵となる事実は、「相関関係」と「因果関係」とは異なるということである。例えば、上の方法で教育年数の長い人ほど賃金が高くなる傾向があるという結果を得たとする。ところが、賃金はその人の能力や環境などの要因によっても影響を受ける。また、これらの要因は教育年数にも影響を与えると考えられる。そうであるならば、教育年数と賃金の「相関関係」はそれらの要因、すなわち能力や環境の違いによってもたらされた可能性がある。従って、この結果からでは教育年数を伸ばすことは賃金に対して正の影響を与えると結論づけることはできない。すなわち、上記の比較方法では教育年数と賃金の「相関関係」は測れたとしても、教育年数が変化したことによる賃金への「因果関係」を測っていることにはならないのである。

そこで、この賃金と教育年数の両方に影響を与える変数 (「交絡因子 (confounding factor)」あるいは「共変量 (covariate)」と呼ぶ) をコントロールした (一定とした) 上で比較することを考える。そのために有力な方法の一つが、多重回帰である。多重回帰は、上記の交絡因子を線形モデルに追加的に組み入れることにより、自動的に交絡因子をコントロールした上で教育年数と賃金の関係を測る方法である<sup>1)</sup>。例えば、賃金を教育年数、労働経験年数、年齢の線形和で表現したモデルを、多重回帰モデルと呼ぶ。ここで、賃金を「従属変数 (dependent variable)」あるいは「被説明変数 (explained variable)」, 教育年数、労働経験年数、年齢を「独立変数 (independent variables)」あるいは「説明変数 (explanatory variables)」と呼ぶ。それぞ

れの説明変数の係数は「最小二乗法 (Ordinary Least Squares, OLS)」という手法によって推定される。最小二乗法は、誤差の二乗和を最小にするように係数の値を決める推定法であり、いくつかの仮定の下で不偏性、一致性、効率性などの望ましい性質を持つことが知られている。そして、推定された教育年数の係数の値は、「経験年数、年齢を一定とした上で教育年数を一年伸ばすことが賃金に与える (平均的な) 効果」として解釈される。

しかしここで重要なことは、多重回帰によっても「因果効果」を測ることができない場合が存在するということである。その主な場合は以下の3つである。

第一に、交絡因子を全て観測することができない場合である。この場合、多重回帰によって観測できる交絡因子をコントロールしたとしても、観測されない交絡因子の影響を取り除くことができない。従って、(観測されない交絡因子の影響が無視できないほど大きい場合) 多重回帰によっては因果関係を正しく測ることはできない。

第二は、説明変数に観測誤差がある場合である。説明変数に観測誤差が含まれる場合、上述の最小二乗法はその観測誤差の影響を考慮していないため、推定値が偏り (バイアス) を持つてしまうことが知られている。

そして第三は、興味のある説明変数 (教育年数) と被説明変数 (賃金) が「同時」に決まっている場合である。それは以下のような意味である。教育年数は外生的に与えられるものではなく、各個人の選択によって決まる。例えば、能力や野心の高い人は教育を長く受ける傾向があると考えられる。しかし、能力や野心の高さは同時にそれ自身が賃金を押し上げる効果を持つと考えられる。その意味で、教育年数と賃金は「同時に」(あるいは「内生的に」) 決まっている。このような場合、最小二乗法によって推定された係数値はやはり偏りを持つてしまう。以上三つの場合に最小二乗法による推定が偏りを持つてしまうという問題は、「説明変数の内生性の問題」と呼ばれる。

このような場合にも偏りなく「因果関係」を推定する方法として提案されたのが操作変数法である。操作

変数法は、モデルに含まれている変数以外に「説明変数と相関があり」かつ「誤差との相関がゼロである」ような変数（「操作変数 (instrumental variable)」と呼ぶ）を用いる推定法である。例えば、冒頭の例において、親の教育年数はその人の教育年数と相関していると考えられるだろう。そして、もし親の教育年数と観測されない全ての交絡要因（能力、野心など）との相関がゼロであるならば、親の教育年数は操作変数として用いることができる。

操作変数法として最も良く知られている推定法は、二段階最小二乗法 (Two Stage Least Squares, 2SLS) と呼ばれる推定法である。これは文字通り「最小二乗法」を二回繰り返すことによって推定値を得る方法であり、以下のように計算される。第一段階で、説明変数を操作変数を含む全ての外生変数に回帰し、「説明変数の予測値」を作る。そして、第二段階で被説明変数を第一段階で求めた「説明変数の予測値」に回帰する。この方法は、次のように解釈することができる。第一段階で、操作変数を用いて教育年数から「教育年数と賃金に影響を与える要因 (交絡因子の影響)」を取り除く。その上で、第二段階でその交絡因子の影響を取り除いた説明変数を用いて回帰モデルを推定する方法であると解釈することができる。

あるいは別の解釈として、二段階最小二乗法は全ての内生変数と外生変数の関係を表した方程式体系を考慮に入れた推定方法として解釈することができる。それゆえ、説明変数の内生性を無視した最小二乗法とは異なり、説明変数が「内生的」に決まっている場合にも偏りのない推定値を得ることができる。

最後に操作変数法を用いる際の注意点を簡単に述べておきたい。それは、二段階最小二乗法はいくつかの問題点も持っているということである。上述のように、操作変数とは「説明変数と相関があり」かつ「誤差との相関がゼロである」ような変数のことである。しかし、現実にはこの二つの条件を同時に満たす変数を見つけることは容易ではない。実際、「誤差との相関がゼロである」変数は説明変数との相関も「弱い」こと

が多い。そのような場合、二段階最小二乗法による推定値は偏りを持つことが知られており、「弱操作変数問題 (weak instrument problem)」と呼ばれている<sup>2)</sup>。一方、上記の方程式体系に基づいて最尤法のアイデアによって導かれる制限情報最尤法 (Limited Information Maximum Likelihood, LIML) という推定法がある。この推定法は実は二段階最小二乗法よりも前に Anderson and Rubin (1949) によって提案された方法であるが、実証分析で用いられることは比較的少なかった。しかし、制限情報最尤法は二段階最小二乗法に比べて弱操作変数の場合にも偏りが小さいことが知られており、特に弱操作変数問題が危惧される場合には、二段階最小二乗法と合わせて計算されるべき推定方法である。

また近年では、より効率的な推定方法として、一般化積率法 (Generalized Method of Moments, GMM) や経験尤度法 (Empirical Likelihood Method) といった手法も提案されるなど、操作変数法の改良についての研究は継続中である。

最後に、2000年頃から弱操作変数の場合にも信頼できる統計的推測法 (検定及び信頼区間の構成方法) の開発がさかんに行われ、いくつかの有力な手法が提案されている。弱操作変数問題が懸念される場合の統計的推測に利用されたい。<sup>3)</sup>

- 1) 交絡因子をコントロールした比較を行うことは実際には容易ではない。一つの方法として、交絡因子の値が似ている個体を同じグループとして、グループごとに比較する方法も考えられるが、そのグループ分けの基準には、多くの場合恣意性が残ってしまう。
- 2) 弱操作変数の場合、二段階最小二乗法に基づく統計的推測も不正確になる。
- 3) 詳しくは Stock-Wright-Yogo (2002, *Journal of Business and Economic Statistics*) 等の論文を参照されたい。

まつした・ゆきとし 東京工業大学情報理工学研究所准教授。最近の主な著作に "On the Asymptotic Optimality of the LIML Estimator with Possibly Many Instruments", (with Anderson, T. W. and N. Kunitomo), *Journal of Econometrics* 2010, 157, pp.191-204。計量経済学・統計学専攻。