応用科目・経済開発論　(講義資料)

農業経営規模と土地生産性の「逆関係」命題 (2)

——除外変数問題の観点から——

佐　藤　隆　広

1　はじめに

別稿（佐藤（2002））において，不確実性と生産要素市場の不完全性が農業経営規模と土地生産性との間に「逆関係」をもたらす理論的可能性があることを，ハウスホールドモデル（農業世帯モデル：agricultural household model）を利用して解説した。本稿では，この「逆関係」命題を前回とは違う角度から再検討してみたい。すなわち，別稿が開発のミクロ経済学アプローチの理論を解説しているとすれば，今回はその応用を紹介するものである[1)]。

2　除外変数問題

「逆関係」命題は，通常，次式のような推計式にもとづいて議論されている。

$$\ln\left(\frac{Q_i}{A_i}\right)=\alpha+\beta\ln A_i+u_i$$

ここで，Qは農業生産，Aは農業経営規模（耕地面積），uは確率誤差項を意味する。したがって，農業経済データを利用して推計されたパラメータβの符号はマイナスが想定されている。

別稿で取りまとめたインドの農業調査データ（佐藤（2002：第1表））を利用して，実際に上述のモデルを最小自乗法（OLS）で推計してみよう。その推定結果はつぎの通りである[2)]。推定係数のつぎの括弧内はt値（ただしその絶対値）を意味し，州ダミー変数の基準はマドラス州としていることに注意されたい。

$$\begin{aligned}\ln\left(\frac{Q}{A}\right)=&5.34(40.47)-0.18(4.64)\ln A\\&-0.60(4.09)\text{MP州ダミー}\\&-0.79(6.25)\text{ボンベイ州ダミー}\\&+0.34(2.71)\text{西ベンガル州ダミー}\\&+0.53(3.24)\text{UP州ダミー}\\&-0.056(0.34)\text{パンジャブ州ダミー}\end{aligned}$$

$$Adj.R^2=0.82,\quad F(6,51)=44.12,\quad NOB=58$$

たしかに，推計結果によれば，農業経営規模の符号はマイナスで統計的に有意である。農業経営規模が10%上昇すると，土地生産性が1.8%低下することがわかる[3)]。

しかしながら，この結果を額面通りに受け入れることはできない。ここでの問題は，説明変数Aと誤差項uに相関がある場合である。たしかに，農業世帯はAが固定されていて動か

1）本稿を作成するにあたって大いに参考にした文献としては，「逆関係」命題の応用についてはDeaton（1997：95-97），操作変数法の解説についてはWooldridge（2002：chapter 5）である。また，邦語文献としては，藤田（1992）が最も重要であろう。

2）推計にあたっては，統計パッケージソフトStata 8を用いた。

3）パンジャブ州ダミーの符号が有意ではないもののマイナスとなっている。これは，ここで利用したデータが「緑の革命」直前のものだからである。当然，「緑の革命」の後では，パンジャブ州は，著しく土地生産性を向上させ，インドのなかで最も高い水準の生産性を実現している。

すことができないものとみなしているかもしれないが，そのことからただちに A と u に相関が存在しないということにはならない。具体的には，農業経営規模や土地生産性の決定にあたって，推計式では考慮されていない「土地の質」(land quality) がかかわっている可能性があるということである。たとえば，農地が砂地なら役畜があまり利用されないであろうし，圃地なら肥沃で生産性が高いかもしれない，ということである。

「土地の質」のように推計式から「除外された変数」(omitted variable) が存在する場合には，説明変数と誤差項の間に相関関係が生じてしまい，OLS で推計したパラメータは不偏性と一致性を満たさない。これが，除外変数問題である。

こうした除外変数の存在によって誤差項と相関を持ってしまう説明変数のことを内生変数という。除外変数問題は，同時方程式 (simultaneous equations) や観測誤差 (measurement error) の存在によって引き起こされる内生性問題 (endogenous problem) の一種である[4]。

そこで，除外変数問題を定式化することにしよう。いま，正しいモデルが次式で与えられるとする。

$$y_i=\alpha+\beta x_i+\gamma l_i+u_i$$

ここで，y と x のデータは存在するが，l はそうでないとする。ここでの事例では，y が土地生産性，x が農業経営規模，l が土地の質となる。l を無視して OLS で回帰分析を行うことにしよう。推計式は次式になる。

$$y_i=\alpha+\beta x_i+e_i$$

ここで，$e_i\equiv\gamma l_i+u_i$ である。この時，推計された β は，

$$\hat{\beta}=\frac{Cov(x,y)}{Var(x)}$$

$$=\frac{Cov(x,\alpha+\beta x+\gamma l+u)}{Var(x)}$$

$$=\frac{E[(x-E[x])(\alpha+\beta x+\gamma l+u-E[\alpha+\beta x+\gamma l+u])]}{Var(x)}$$

$$=\frac{E[(x-E[x])(\beta(x-E[x])+\gamma(l-E[l]))}{Var(x)}$$

$$=\beta\frac{Var[x]}{Var[x]}+\gamma\frac{Cov(x,l)}{Var(x)}$$

$$=\beta+\gamma\frac{Cov(x,l)}{Var(x)}$$

となる。いま，$\beta=0$ であるが，$\gamma>0$, かつ，$Cov(x,l)<0$ であるなら，$\hat{\beta}$ はマイナスになり，規模と生産性に逆関係が発生してしまう[5]。すなわち，本来なら規模と生産性の間には逆関係が存在しないにもかかわらず，除外変数を無視して推計したために，あたかも逆関係が存在しているかのような誤った結果が得られてしまうのである。

3 操作変数法

それでは，こうした除外変数問題をどのように解決したらいいのだろうか。最も簡単な答えとしては，除外変数のデータそのものを入手す

4) 経済理論でいう内生変数とは経済モデルのなかで決定される変数を意味するが，計量経済学でいう内生変数とは誤差項と相関を持つ説明変数のことを意味していることに注意されたい。詳しくは，Wooldridge (2002：50-51) を参照されたい。

5) $Cov(x,l)<0$ はいわば，経営規模が大きくなればなるほど土地の質が悪くなっていくという関係を意味するが，このことはそれほど不自然な仮定ではないだろう。たとえば，その理由の1つとしては，つぎのように考えることができる。貧しい小農が，借入金返済のために土地を売却せざるを得ないという状況に追い詰められている。この場合，土地の売却先が地主だとして，小農は土地の質が相対的に劣悪な土地から売却していくだろう。こうしたプロセスが継続すると，地主に土地が集積していく反面，その質は平均的に低落していくはずである。こうした論点についての批判的検討としては，藤田(1992) のとくに第5章を参照されたい。

ることである。本稿の事例でいえば，「土地の質」に関するデータである。たとえば，バングラデシュ農業における「逆関係」命題を緻密に分析している藤田（1992）は，土地の質の変数として，洪水の季節変動をもとにした区分である高位地・中位地・低位地を用いている。土地の質を説明変数に追加しても，依然として，「逆関係」命題が成立していることが明らかにされている。Bhalla and Ray (1988) や Bhalla (1988) は，土壌特性の詳細がわかる新しい統計資料を用いて，インド農業における「逆関係」命題を再検証している[6]。その結果，土地生産性と経営面積の間の逆関係が多くのケースで消滅し，「逆関係」命題が成立している度合いが従来言われてきたような100％ではなく30％程度に過ぎないことが明らかにされた。

次に，除外変数そのもののデータが入手不可能な場合はどうすべきであろうか。この場合の解決方法としては，第1にパネルデータ（panel data）の利用，第2に操作変数（instrumental variable: IV）法の利用が考えられる。まず，パネルデータを考えてみることにしよう。パネルデータとは，同一の個体に対して複数時点のデータが利用可能であるようなデータのことをいう。ある時点をとればクロスセクションのデータとなり，ある個体をとればタイムシリーズのデータとなる。このようなパネルデータは，クロスセクションとタイムシリーズの両方の性質を持ち合わせている。いま，2時点のデータが利用可能であり，時間を通じて土地の質が不変であるならば，

$$y_{it}=\alpha+\beta x_{it}+\gamma l_i+u_{it}$$

$$y_{it+1}=\alpha+\beta x_{it+1}+\gamma l_i+u_{it+1}$$

となるだろう。うえの2式の $t+1$ 期の式から t 期を差し引けば，除外変数をキャンセルアウトすることができる。すなわち，

$$y_{it+1}-y_{it}=\beta(x_{it+1}-x_{it})+(u_{it+1}-u_{it})$$

となる。この式に OLS を実行すれば，β の不偏でかつ一致推定量を求めることができる[7]。

次に操作変数法を考えてみることにしよう。操作変数法とは，モデルの誤差項とは相関しないが，問題となっている内生変数とは相関している変数（これを操作変数（IV）という）を利用することで，除外変数問題を解決するものである。

操作変数法を定式化してみることにしよう。

$$y_i=\beta_0+\beta_1x_{1i}+\beta_2x_{2i}+\cdots+\beta_kx_{ki}+\epsilon_i \quad (1)$$

ここで，$E(\epsilon_i)=0$, $Cov(x_{ji},\epsilon_i)=0$, $\forall j=1,2,...,k-1$ を仮定する。このことは，x_k と ϵ の間に相関があることを意味する ($Cov(x_{ki},\epsilon_i)\neq0$)。すなわち，$x_k$ が内生変数である。すでに述べたように，この問題を解決するためには操作変数 z_1 を必要とするが，z_1 は以下の2つの条件を満たさなければならない。すなわち，

$Cov(z_{1i},\epsilon_i)=0$,

$x_{ki}=\gamma_0+\gamma_1x_{1i}+\gamma_2x_{2i}+\cdots+\gamma_{k-1}x_{k-1i}+\theta_1z_{1i}+\eta_i$ において，$\theta_1\neq0$,

である。式(1)の x_k に第2の条件を代入して整理すれば，次式を得る。

$$y_i=\alpha_0+\alpha_1x_{1i}+\alpha_2x_{2i}+\cdots+\alpha_{k-1}x_{k-1i}+\lambda_1z_{1i}+\nu_i$$

ここで，$\nu_i=\epsilon_i+\beta_k\eta_i$, $\alpha_0=\beta_0+\beta_k\gamma_0$, $\alpha_j=\beta_j+\beta_k\gamma_j$, $\lambda_1=\beta_k\theta_1$ である。仮定から，ν_i は右辺のすべての説明変数と無相関であるため，α_j と λ_1 の一致推定量を求めることができる。

いま，n 個の観測値が存在するとし，操作変数法による推定量を行列を用いて求めてみることにしよう。

$$y=X\beta+\epsilon \quad (2)$$

ここで，被説明変数 y，説明変数 x_j，誤差項 ϵ は $(n\times1)$ の列ベクトル，β は $((k+1)\times1)$ の列

6) 新資料は，National Council for Applied Economic Research, *Fertilizer Demand Survey* である。

7) ただし，α や β が時間を通じて一定であると仮定している。また，農業経営規模 (x) が時間を通じて変化することが前提となっていることにも注意されたい。

ベクトル，X は $(n\times(k+1))$ の行列であり，第1列の要素はすべて1となっている。x_j は行列 X の j 番目の列ベクトルであり，最後の列は内生変数 x_k である。ここで，$Z=[1, x_1, x_2, ..., x_{k-1}, z_1]$ とする。さらに，$Cov(x_{ji}, \epsilon_i)=0\ \forall j=1,2,...,k-1$ と $Cov(z_{ji}, \epsilon_i)=0$ を仮定する。これは，直交条件（orthogonality condition）である $E(Z'\epsilon)=0$ を意味する[8]。

さて，式(2)の両辺に前から Z' を掛ければ，$Z'y=Z'X\beta+Z'\epsilon$ を得る。期待値をとり，直交条件を利用すればつぎのようになる。

$$E[Z'y]=E[Z'X]\beta+E[Z'\epsilon]=E[Z'X]\beta$$

ここから求められる β が操作変数法の推定量 b_{IV} である。すなわち，

$$b_{IV}=E[Z'X]^{-1}E[Z'y]$$

である[9]。さらに，

$$\begin{aligned}b_{IV}&=E[Z'X]^{-1}E[Z'y]\\&=E[Z'X]^{-1}E[Z'(X\beta+\epsilon)]\\&=\beta E[Z'X]^{-1}E[Z'X]+E[Z'X]^{-1}E[Z'\epsilon]\\&=\beta\end{aligned}$$

となる。最後の項は直交条件からゼロになり，操作変数法の推定量が不偏性を持つことがわかる。

いま，操作変数法による推定量が不偏性を持つことがわかったが，次に一致性を考えてみることにしよう。$(x_i, y_i, z_{1i}: i=1,2,...,N)$ が母集団からのランダムサンプルであるなら，操作変数法による推定量は次式で与えられる。

$$b_{IV}=\left(\frac{1}{N}\sum_{i=1}^{N}z_i'x_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}z_i'y_i\right)$$

ここで，x_i と z_i は $(1\times(k+1))$ の行ベクトルで，y_i がスカラーであることに注意しておこう。これを行列表記すれば，$b_{IV}=(Z'X)^{-1}Z'y$ となり，Z と X は $(n\times(k+1))$ の行列，y は $(n\times1)$ の列ベクトルである。さらに，この式を変形すれば次式を得る。

$$\begin{aligned}b_{IV}&=(Z'X)^{-1}Z'y\\&=(Z'X)^{-1}Z'(X\beta+\epsilon)\\&=\beta(Z'X)^{-1}(Z'X)+(Z'X)^{-1}Z'\epsilon\\&=\beta+\left(\frac{1}{N}\sum_{i=1}^{N}z_i'x_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}z_i'\epsilon_i\right)\end{aligned}$$

いま，サンプル数を無限大に持っていくと，$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}z_i'\epsilon_i=E[z_i'\epsilon_i]=0$ となるため，右辺第2項目はゼロになる[10]。したがって，操作変数による推定量 b_{IV} は一致性を持つのである。

以上，操作変数推定量が不偏性と一致性という優れた特性を持つことがわかった。しかしながら，実際には，「良い操作変数」(good instrument) を見つけることは，必ずしも容易ではない。操作変数の第2の条件である $\theta_1\neq0$ については，帰無仮説を検証することができるが，誤差項と操作変数に相関がないという条件は元の誤差項そのものがわからないので直接にテストすることができないからである。

実際，内生変数である農業経営規模とは相関があり，誤差項と無相関である操作変数として，何が適切であろうか。あるいは，これはつぎのように言い換えてもよい。すなわち，内生変数とは相関があり，除外変数そのものとは相関を持っていない変数は，どのようなものが考えら

8) まず，i 番目の観測値を取り上げることで，共分散ゼロと直交条件の関係を確認しておきたい。

$$\begin{aligned}Cov(x_{ji}, \epsilon_i)&=E[(x_{ji}-E[x_{ji}])(\epsilon_i-E[\epsilon_i])]\\&=E[x_{ji}\epsilon_i-x_{ji}E[\epsilon_i]-E[x_{ji}]\epsilon_i\\&\quad+E[x_{ji}]E[\epsilon_i]]\end{aligned}$$

$E[\epsilon_i]=0$ であるから，$Cov(x_{ji}, \epsilon_i)=E[x_{ji}\epsilon_i]=0$ となる。ここで，$E(Z'\epsilon)=E[(1, x_1, x_2, ..., x_{k-1}, z_1)'\epsilon]=0$ であるから，その行列は $((k+1)\times1)$ となり，j 行目の要素を取り出せば $E[x_{j1}\epsilon_1+x_{j2}\epsilon_2+\cdots+x_{ji}\epsilon_i+\cdots+x_{jn}\epsilon_n]=E[x_{j1}\epsilon_1]+E[x_{j2}\epsilon_2]+\cdots+E[x_{ji}\epsilon_i]+\cdots+E[x_{jn}\epsilon_n]=0$ となっている。したがって，共分散ゼロならば直交条件が成立していることがわかるだろう。

9) ここで $E[Z'X]$ が逆行列を持つことを仮定している。

10) $\frac{1}{N}\sum_{i=1}^{N}z_i'\epsilon_i$ は $z_i'\epsilon_i$ の標本平均とみなすことができる。ここでは，観測数を多くしたときに，標本平均が母集団平均に限りになく近づくという大数の法則を応用しているのである。

れるであろうか。一見簡単そうに見えるパズルではあるが，良い操作変数を思い浮かべることは意外に困難である。

例を挙げよう。Benjamin (1995) は，インドネシアのジャワにおける「逆関係」命題を検証する際に，「土地の質」という除外変数問題に対処するために，人口密度を操作変数に採用している。ここでの文脈に即して，人口密度が「良い操作変数」なのかどうかを考えてみよう。人口密度が高ければ，農地の細分化が起こっていると考えられるので，経営規模と人口密度には相関がありそうである。したがって，操作変数の第2の条件はクリアーできそうだ。それでは，第1の条件はどうであろうか。これは，先に述べた通り，直接検証できない。元の推計式における誤差項そのものが，分からないからである。ここで，第1の条件を「除外変数である土地の質と人口密度に相関がない」として考えるならばどうだろうか[11]。もちろん異論はあるだろうが，ひとまず，そのようにみなして分析を進める価値は十分あるように思われる。こうして，Benjamin (1995) は，操作変数法を用いることで，ジャワにおいて土地生産性と経営規模の間にみられる逆関係が消滅することを明らかにした。

しかしながら，除外変数である土地の質と人口密度には相関があるとも考えられる。なぜなら，歴史的に土地の質が良好な地域では，人口を多く養うことができるだろうし，人口も増加しているはずだからである。そうすると，除外変数と人口密度には相関がでてきてしまい，人口密度が操作変数の第1の条件を満たさない可能性もある[12]。

また，操作変数の候補として，人口密度の他にも，データの利用可能性の問題がもちろんあるが「地価」や「農地の分散的保有」なども考えられるのかもしれない。いずれのケースにおいても，操作変数が満たすべき第2の条件を実証的に検証することは可能であるが，第1の条件は必ずしもそうではないことに注意すべきである[13]。それでは，「地価」や「農地の分散的保有」は良い操作変数であろうか。おそらく，そうではないだろう。その理由について，読者自ら検討されたい。

4　おわりに

本稿では，農業経営規模と土地生産性の「逆関係」命題を手がかりに，ごく限られた範囲ではあるが，開発のミクロ経済学における実証分析手法を紹介した。データを利用して実証分析を行う場合に，ここで検討したような除外変数（土地の質）の存在のため，誤差項と説明変数との間に相関が発生する場合がある。このことを無視して OLS を実行すると，得られたパラメータは不偏性も一致性も満たさない。この除外変数問題の解決法として，(1) 除外されている変数データを入手する，(2) パネルデータを利用する，(3) 操作変数法を利用する，の3つを解説してきた。

とくに，操作変数法を利用するためには，「誤差項と相関をもたず，内生変数とは相関を持つ」あるいは「内生変数とは相関を持つが，除外変数とは相関を持たない」という操作変数を発見しなければならない。

操作変数法を利用した最近の研究動向を紹介している Angrist and Krueger (2001) が「われ

11）実証分析を行う研究者が操作変数を試行錯誤しながら探索する際には，「誤差項と相関を持たない変数」というよりもむしろ「除外変数と相関を持たない変数」という基準を用いることが多い。

12）こうした土地の質と人口密度の関係については，藤田（2002：134-135）を参照されたい。

13）操作変数が内生変数とそれほど相関が高くない場合，それを「弱い操作変数」(weak instrument) といい，除外変数との相関が高い場合，それを「悪い操作変数」(bad instrument) という。

われの見解では，良い操作変数というものは，分析対象になっている説明変数を決定付けている経済メカニズムや経済制度に関する詳細な知識からしばしばやって来るものである」(Angrist and Krueger (2001：73))，「ここでの研究上の挑戦は，新しい理論や推定方法を必要とするという意味において第一義的にはテクニカルなものではない。むしろ，研究の進歩は，詳細な制度に関する知識や特定の状況下で機能している力学についての注意深い検証からやって来るのである」(Angrist and Krueger (2001：83)) と述べているように，操作変数法を利用するためには経済制度に対する深い理解を必要とする。

開発経済学に引き付けて解釈すれば，このことは，研究者個人が持つ開発途上国における現場感覚を磨き，経済学だけでなく，それ以外の途上国地域研究の成果をも同時に取り入れる必要があることを示唆している。とくに，近年，操作変数法の応用として，「自然実験」(natural experiment) や「無作為化された実験」(randomized experiment) などがよく利用されているが，これは，こうした文脈で理解すべきであろう[14]。なぜなら，操作変数として利用できそうな「自然実験」や「無作為化された実験」は，必ずしも経済の領域に限定されるものではなく，経済以外の分野にまで目配りしなければ発見するのが困難であると思われるからである。

実際，現在の開発経済学は，歴史学・政治学・法律学・地理学・文化人類学・社会学・医学などの途上国地域研究の関連分野からの成果を貪欲に消化し，その分析を精緻化するだけではなく分析領域をも同時に広げているのである。

14) 近年のこうした操作変数法の新しい利用方法については，たとえば，Angrist and Krueger (2001) や Rosenzweig and Wolpin (2000) などの優れたサーベイ論文を参照されたい。

参考文献

佐藤隆広 (2002) 「農業経営規模と土地生産性の『逆関係』命題：ハウスホールドモデルの観点から」『経済学雑誌』第103巻別冊 (前期)。

藤田幸一 (1992) 『バングラデシュ農業発展論序説』農業総合研究所。

Angrist, J. D. and A. B. Krueger (2001) "Instrumental Variables and the Search for Identification," *Journal of Economic Perspectives*, 15 (4), 69-85.

Benjamin, D. (1995) "Can Unobserved Land Quality Explain the Inverse Productivity Relationship?" *Journal of Development Economics*, 46 (1), 51-84.

Bhalla, S. S. (1988) "Does Land Quality Matter," *Journal of Development Economics*, 29, 45-62.

Bhalla, S. S. and P. Roy (1988) "Mis-Specification in Farm Productivity Analysis," *Oxford Economic Paper*, 40, 55-73.

Deaton, A. (1997) *The Analysis of Household Surveys*, The Johns Hopkins University Press.

Rosenzweig, M. R. and K. I. Wolpin (2000) "Natural 'Natural Experiments' in Economics," *Journal of Economic Literature*, 38 (4), 827-874.

Wooldridge, J. (2002) *Econometric Analysis of Cross Section and Panel Data*, The MIT Press.