

ヘックマン

「サンプルセレクションによるバイアスは特定化の誤謬によるバイアスと解釈できる」¹⁾

【労働経済】

市村 英彦

1 背景

個票による実証分析が大きく進展した1950年代以降、回帰モデルの従属変数がある特定の値を正の確率でとる場合を十分考慮した分析の重要性が様々な問題で認識されてきた。このような問題は集計されたデータを分析しているなら生じない。集計されたデータを使って需要や供給分析をしているなら、買われない財や提供されない財はないし、投資行動を分析しているなら投資が行われない年や産業はないだろう。また、集計されたデータを使って、例えば通勤手段の選択を分析するなら、どの通勤手段も一定程度は用いられている。しかし、個票を用いて、例えば女性の労働供給の分析を行うと、働いていない女性の割合は高いし、投資行動の分析を行うと、ある年に投資を行っていない企業もある。また、通勤手段の選択を分析するなら、特定の個人であれば、選択肢のうち、通勤手段は一つの場合がほとんどだろう。

このような場合を含むデータの分析の枠組みはTobin (1958), Gronau (1974), Heckman (1974), McFadden (1973) などにより、それぞれ、家計の投資行動、女性の賃金と労働参加あるいは労働供給、通勤手段の選択の文脈で開発され、理論的にはAmemiya (1973, 1974) により一般的な分析手法が開発された²⁾。

本稿ではこのような一連のモデルの概要をまとめ、その流れの中におけるHeckman (1979) 論文の意義とその後の展開を概観する。

2 問題の定式化と意義

通常の線形回帰モデルは従属変数を y_i^* 、説明変数ベクトルを x_i 、未知の定数を a 、係数ベクトルを β 、誤差項を u_i で表すと (添字 i は特定のデータを表し、 i の値は1から n の整数をとる。ここで n

はサンプル数を表す整数)

$$y_i^* = a + x_i' \beta + u_i$$

と書ける。ここで x_i を所与とする誤差項の条件付き期待値は $0 (E(u_i | x_i) = 0)$ とする。

従属変数を理想的な投資量で、右辺がそれを説明するモデルだと考えると、実際に投資できる量が正の場合だけなら観察される投資量 y_i が理想的な投資量 y_i^* に等しいのは y_i^* が正の場合だけである。従って、観察される投資量のモデル (これが所謂 Tobit モデルである) は

$$y_i = \begin{cases} a + x_i' \beta + u_i & \text{if } a + x_i' \beta + u_i > 0 \\ 0 & \text{if } a + x_i' \beta + u_i \leq 0 \end{cases}$$

となる。Tobin (1958) は u_i が平均0、分散 σ^2 の正規分布に従うことと、 u_i と x_i が独立であることを仮定したモデルを提示し、さらにこのモデルパラメーターの最尤推定法を提案した。またこれらを用いて家計の耐久財投資量が0の場合を含む個票に関する実証分析を行った³⁾。

Amemiya は Tobit モデルを含む一般的な同時方程式モデルにおけるパラメーターの最尤推定量が一致性と漸近正規性をもつ十分条件を示し、さらに最尤推定量を計算する際の初期値となりうる一貫性をもつ推定量を開発している。Tobit モデルの場合は、標準正規分布の密度関数を ϕ 、分布関数を Φ と書いたとき

$$\begin{aligned} & E(y_i | x_i, x_i' \beta + u_i > 0) \\ &= a + x_i' \beta + \sigma \frac{\phi((a + x_i' \beta) / \sigma)}{1 - \Phi((a + x_i' \beta) / \sigma)} \\ & E(y_i^2 | x_i, x_i' \beta + u_i > 0) \\ &= (a + x_i' \beta)^2 + \sigma (a + x_i' \beta) \times \frac{\phi((a + x_i' \beta) / \sigma)}{1 - \Phi((a + x_i' \beta) / \sigma)} + \sigma^2 \\ &= (a + x_i' \beta) \times E(y_i | x_i, x_i' \beta + u_i > 0) + \sigma^2 \end{aligned}$$

となるので、

$$y_i^2 = y_i(a + x_i'\beta) + \sigma^2 + v_i$$

というモデルの a , β と σ^2 を、Amemiya は 1 と \hat{y}_i と $x_i\hat{y}_i$ (ここで \hat{y}_i は y_i を x_i に回帰して得られる x_i による predictor) を操作変数として推定することを提案した。Amemiya の、条件付き期待値を元に推定問題を考える手法はその後、Heckman (1976) と Lee (1976) によりサンプルセレクションモデルへも応用された⁴⁾。

Gronau (1974) による潜在的賃金モデルでは、Tobit モデルのように潜在的賃金が 0 という固定点より大の場合に賃金が観察されるのではなく、潜在的賃金が留保賃金より大きい場合に賃金が観察されるとする。従って、潜在的賃金と留保賃金の差を $z_i'\theta + v_i$ で表すと (ここで z_i は定数項を含むとする)、

$$y_i = \begin{cases} a + x_i'\beta + u_i & \text{if } z_i'\theta + v_i > 0 \\ 0 & \text{if } z_i'\theta + v_i \leq 0 \end{cases}$$

となる。ここで $y_i = 0$ の場合は i 番目の人が働いていないことを表す。

このモデルで、 x_i と z_i を所与とした時、 (u_i, v_i) が平均が 0、

$$\begin{pmatrix} \sigma_u^2 & \sigma_u\sigma_v\rho \\ \sigma_u\sigma_v\rho & \sigma_v^2 \end{pmatrix}$$

を分散共分散行列とする二変数の正規分布に従うとき、

$$\begin{aligned} & E(y_i | x_i, z_i, z_i'\theta + v_i > 0) \\ &= a + x_i'\beta + E(u_i | x_i, z_i, z_i'\theta + v_i > 0) \\ &= a + x_i'\beta + E[E(u_i | v_i, x_i, z_i, z_i'\theta + v_i > 0) | x_i, z_i, z_i'\theta + v_i > 0] \\ &= a + x_i'\beta + E\left(\frac{\sigma_u\sigma_v\rho}{\sigma_v^2}v_i | x_i, z_i, z_i'\theta + v_i > 0\right) \end{aligned}$$

最後の等号は、同時分布が正規分布の場合には条件付き期待値は所与とした正規分布をもつ変数の線形関数となり、係数は回帰係数となるという結果を用いている。この結果と Amemiya の結果を用いると

$$E(y_i | x_i, z_i, z_i'\theta + v_i > 0) = a + x_i'\beta + \rho\sigma_u \frac{\phi(z_i'\theta/\sigma_v)}{1 - \Phi(z_i'\theta/\sigma_v)}$$

Heckman (1976) と Lee (1976) は、この結果に基づいて β の二段階推定法を開発した。Gronau のモデルの尤度は大域的に凸な目的関数ではなく、尤度の最大値で定義される最尤推定量は確実には求められないので、二段階推定法は非常に広く利用されることとなった⁵⁾。

このモデルは Heckman (1979) 以降サンプルセレクションモデルと呼ばれるが、以上概観してきたように、この論文が書かれた段階で既に理論的な意味でも実証的に必要な手法の開発という観点からもこのモデルに関するものは出揃っていた。

Heckman (1976, 1979) 論文の最大の貢献は、Gronau (1974), Heckman (1974) の潜在的賃金関数の研究に留まらず、労働組合が賃金に与える影響、移住が所得に与える影響、ジョブトレーニングが賃金に与える影響、パネルデータで常に観察されるデータに限定して実証分析を行うことに関する問題など、非常に幅広い問題に対する統一的な対処方法としてこのモデルが使えることを指摘し、それがサンプルセレクションという一般的な問題であることを提示したことにあると思う。例えば、潜在的賃金関数の研究に関しては、ランダムに選ばれた労働者の賃金が観察されているのではなく、留保賃金を超える賃金をオファーされて働いている人の賃金が観察されているのである。それまではサンプルセレクション問題は無視されて線形回帰モデルにより実証分析されるか、無視されない場合も、追加的な変数を線形回帰式に加えて対応されるか、あるいは操作変数を用いて実証分析されてきた。しかし、この論文の後は様々な分野にわたる数多くの実証問題が、サンプルセレクションという視点から再吟味される必要性があることを多くの研究者が認識したのである。

Goldberger (1981) が指摘しているように、サンプルセレクションという問題が存在すること自体は勿論古くから知られていた。しかし、この論文はそれに対して、有効と考えられる非常に簡明

な手法を提供した。70年代後半から80年代にかけて世界中で文字どおり千を超える実証論文がサンプルセレクションモデルを用いて書かれた。

3 LaLondeによる観察データを用いたプログラム評価手法の批判

こういった実証分析の多くはプログラム評価と一般的に呼ばれる計量経済学分野に属し、例えば労働組合やジョブトレーニングが賃金に与える効果を実証分析する。社会実験が難しい中、Heckman (1979)以降、経済学における実証分析では現実には観察されるデータとサンプルセレクションモデルに代表される計量経済学的手法を用いたプログラム評価が80年代では主流となった。

その中、LaLonde (1986)はサンプルセレクションモデルを含む多くの計量経済学的手法を用いて得られた実証結果が、社会実験によって得られた結果と一致しないことを示して、このような手法の一般的妥当性に疑問を呈した。

LaLondeのこの結果は計量経済学的手法の脆弱性と社会実験データの重要性を多くの研究者に印象づけた。さらに Angrist 等が、所謂「自然実験」的なデータを見いだすことでプログラム評価に関しては複雑な計量経済学的手法が回避できることを示すと、プログラム評価の主流は計量経済学のものから、社会実験あるいは自然実験的なものへと移行していった⁶⁾。

しかし、どうして計量経済学的手法が社会実験の結果と乖離する結果しか出さなかったのか、また、「自然実験」的なデータがない場合にはプログラム評価をどう行えば良いのか、という二つの疑問は残った。

4 セミパラメトリック分析とマッチング分析: サポートの重要性

サンプルセレクションモデルは上で説明したように、二変数正規分布を前提として定式化されている。そのような関数型の仮定が計量経済学的手法の問題なのではないかと思われた。1980年代半ば頃からこのようなパラメトリックな仮定をより柔軟な仮定で置き換えてなお β を推定する様々な手法が開発された。例えば Ichimura-Lee (1991)

は正規性の仮定がなくても未知の関数 φ を使うと

$$E(y_i | z_i, z_i'\theta + v_i > 0) = a + x_i'\beta + \varphi(z_i'\theta)$$

と書けることを利用して、 z_i に x_i に含まれていない連続型の変数が含まれている時、 β と θ を同時に推定する手法を開発した。但し、 φ が未知な関数なので、 a は ϕ と区別できず、 θ のうち一つの係数は基準化する必要がある。また Ahn-Powell (1993)はセレクションを決定するモデルをノンパラメトリックにしても、従属変数が観察される確率を $P(z)$ と書いて

$$E(y_i | x_i, z_i, z_i'\theta + v_i > 0) = a + x_i'\beta + \varphi(P(z_i))$$

となることを示し、このモデルで β の推定法を示した。

これらの研究の焦点は、特定の誤差項の分布を仮定することなく β をパラメトリックな場合と同じ収束スピードで推定することができるか、という点にあった。この問題は肯定的に解決されたのだが、プログラム評価の観点からは a が焦点となる。この先をみてみよう。

プログラム評価の際に用いられるモデルではサンプルセレクションのモデルで0が観察されるケースも何らかの従属変数が観察される:

$$\begin{aligned} y_{1i} &= a_1 + x_i'\beta_1 + u_{1i} \text{ if } z_i'\theta + v_i > 0 \\ y_{0i} &= a_0 + x_i'\beta_0 + u_{0i} \text{ if } z_i'\theta + v_i \leq 0 \end{aligned}$$

以降、 $z_i'\theta + v_i > 0$ なら $d_i=1$ で表し、 $z_i'\theta + v_i \leq 0$ なら $d_i=0$ で表す。また、プログラムに参加した場合、即ち $d_i=1$ の時得られる結果が y_{1i} で、プログラムに参加しなかった場合、即ち $d_i=0$ の時得られる結果が y_{0i} だと解釈する。プログラム評価によく用いられる指標として平均的なプログラム効果があるが、それは

$$E(y_{1i} - y_{0i} | x_i) = a_1 - a_0 + x_i'(\beta_1 - \beta_0)$$

となる。即ち、定数項を推定することが重要となるが、それまでの推定方法で定数項は推定されない。この点について Andrews-Schafgans (1998)により、 x_i を所与として、 $P(z_i)=1$ 、即ちプログラムへの参加確率が1となるような z_i を用いて、 a_1 が推定でき、 x_i を所与として、 $P(z_i)=0$ 、即ち

プログラムへの参加確率が0となるような z_i を用いて、 a_0 が推定できることが示された。

一方 Heckman-Ichimura-Smith-Todd (1996) はジョブトレーニングの所得への影響に関する社会実験データを用いて、計量経済学的手法を用いて得られたプログラム効果と、社会実験によって得られたプログラム効果とが一致しないという LaLonde の発見にはどのような要素が影響しているかを吟味した。その結果、所与とする変数が的確に選ばれているなら、サンプルセレクション問題の影響は大きくないことが明らかにされた。圧倒的に大きな影響があるのはトレーニングを受けたグループと受けていないグループの比較をする際に所与とされる変数の分布の違い、特にサポートの違いであった。これらの要素は観察可能なので、調整可能である。Heckman-Ichimura-Todd (1997, 1998) は従来から統計の分野で用いられていたマッチング手法をサポートの違いを明示的に考慮する手法へと改良し、LaLonde の批判以降、懸案となっていた二つの疑問に一定の解答を与えた。

5 バウンド分析

Heckman-Ichimura-Todd の提案した手法は、観察されない要素によるサンプルセレクション問題がないことを前提としている。Andrews-Schafgans の手法はこのような要素を許容するが、Heckman-Ichimura-Todd の手法が線形回帰モデルを前提としないノンパラメトリックな手法であるのに対して、Andrews-Schafgans の手法は観察されない要素が加法的に影響している従来の線形回帰モデルに基づくサンプルセレクションモデルを前提としている。Costa Dias-Ichimura-Van den Berg (2013) はこの結果には線形回帰モデルを前提とする必要はないことを示しているが、 $P(z_i)$ が極端な値を取る必要があることに変わりはない。

そのような極端な値がないときにはどのような分析が可能なのかという点について Manski (1990) はバウンド分析を提唱しているが、それだけでは興味のあるパラメータの取り得る範囲が多くの場合広すぎる。Blundell-Gosling-Ichimura-

Meghir (2007) は Gronau の吟味した状況について、理論的に有効と考えられる様々な制約がどのようにバウンドを縮めるかを吟味し、バウンド分析がこの状況で有効であることを示した。

6 結びに代えて

Heckman (1976, 1979) は先人達の理論的な研究成果と様々な分野における実証分析を吸収し、そのエッセンスをサンプルセレクション問題というかたちに昇華させてみせた。さらにそれに対する簡便な対処法を提供することでその後の研究に大きな影響を与えた。LaLonde (1986) の研究とその後の自然実験「ブーム」により、一時的に衰えた計量経済学的手法も、LaLonde の結果はサポート問題が理由だということが明らかになり、改良されたマッチング法、バウンド分析など観察データを使う実証分析も復権してきている。これらの研究ではある「プログラム」にどれほどの効果があったのか、という点に焦点があてられているが、当然のことながら、経済学にはどのようなメカニズムでそのような効果もたらされたのか、それはどのような場合に大きいのか、また、それはどのような理論と整合的かといった点にも焦点はあてられるべきである。これらの点については構造推定を用いた分析が1980年代半ばから Miller, Rust, Pakes, Wolpin などにより始められ、現在も着実に進められている。

J. J. Heckman "Sample Selection Bias as a Specification Error," *Econometrica*, 47 (1), (1979), 153-161.

- 1) 元のタイトルは「J. J. Heckman (1979) による *Econometrica* 論文 "Sample Selection Bias as a Specification Error" の意義とその後の展開」であるが、本誌企画のため編集者の意向により、上記のタイトルとした。
- 2) McFadden (1973) は多項 Logit モデルを開発した。二項の Probit モデルや Logit モデルについては Berkson や Finney による研究や書籍が1940年代からある。
- 3) Tobin は家計の2カ年にわたる耐久消費財支出が所得に占める割合を従属変数としているが、それが0である家計が約25%ある。隔世の感がある。
- 4) Gronau (1974) も Mills's ratio ($\phi(x_i\beta/\sigma)/[1-\Phi(x_i\beta/\sigma)]$) を含む定式化を行っている。
- 5) このモデルでも Amemiya の方法で α , β , ρ それに σ^2 の一致推定量を作ることができる。
- 6) 経済学の分野に限っても Angrist 以前から Rosenzweig や Wolpin 等が自然実験的アプローチを用いている。

参考文献

- Ahn, H. and J. L. Powell (1993) "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," *Journal of Econometrics*, 58 (1-2), pp.3-29.
- Amemiya, T. (1973) "Regression Analysis When the Dependent Variable is Truncated Normal," *Econometrica*, 41 (6), pp.997-1016.
- (1974) "Multivariate Regression and Simultaneous Equation Models When the Dependent Variables are Truncated Normal," *Econometrica*, 42 (6), pp.999-1012.
- Andrews, D. W. K. and M. M. A. Schafgans (1998) "Semiparametric Estimation of the Intercept of a Sample Selection Model," *Review of Economic Studies*, 65 (3), pp.497-517.
- Blundell, R., A. Gosling, H. Ichimura and C. Meghir (2007) "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds," *Econometrica*, 75 (2), pp.323-363.
- Costa Dias, M., H. Ichimura and G.J. Van den Berg (2013) "Treatment Evaluation with Selective Participation and Ineligibles," *Journal of the American Statistical Association*, 108, pp.441-455.
- Goldberger, A. (1981) "Linear Regression after Selection," *Journal of Econometrics*, 15 (3), pp.357-366.
- Gronau, R. (1974) "Wage Comparisons: A Selectivity Bias," *Journal of Political Economy*, 82 (6), pp.1119-1143.
- Heckman, J. J. (1974) "Shadow Prices, Market Wages, and Labor Supply," *Econometrica*, 42 (4), pp.679-694.
- (1976) "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement*, 5 (4), pp.475-492.
- Heckman, J. J., H. Ichimura, J. Smith and P. E. Todd (1996) "Sources of Selection Bias in Evaluating Social Programs: An Interpretation of Conventional Measures and Evidence on the Effectiveness of Matching as a Program Evaluation Method," *Proceedings of the National Academy of Sciences*, 93 (23), pp.13416-13420.
- Heckman, J. J., H. Ichimura and P. E. Todd (1997) "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *Review of Economic Studies*, 64 (4), pp.605-654.
- (1998) "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, 65 (2), pp.261-294.
- Ichimura, H. and L. Lee (1991) "Semiparametric Estimation of Multiple Index Models: Single Equation Estimation," in International Symposia in Economic Theory and Econometrics, W. A. Barnett, J. L. Powell, and G. Tauchen (eds.), Cambridge University Press, pp.3-49.
- LaLonde, R. (1986) "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *American Economic Review*, 76 (4), pp.604-620.
- Lee, L. (1976) "Estimation of Limited Dependent Variable Models by Two Stage Methods," University of Rochester, PhD Thesis.
- McFadden, D. L. (1973) "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka (ed.), *Frontiers in Econometrics*, pp.105-142, Academic Press: New York.
- Manski, C. F. (1990) "Nonparametric Bounds on Treatment Effects," *American Economic Review Papers and Proceedings*, 80 (2), pp.319-323.
- Tobin, J. (1958) "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, 26 (1), pp.24-36.

(いちむら・ひでひこ 東京大学大学院経済学研究科教授)

労働関係法規集 2016年版

B6判変型 854頁 2016年3月25日刊行 ISBN978-4-538-14028-5 定価：1,389円+税

労働関係の法令を幅広く収録

社会生活に必携の労働関係法規を持ち運べるコンパクトサイズに収めました。

基本的な法令のほか、必要な告示や指針等も収録し、労働法の学習だけでなく実務にも役立つよう編集しています。

企業の人事担当者、労働組合の方はもちろん、広く一般の皆様にもご活用いただけます。

2016年版の主な改正法令等

「労働者派遣事業の適正な運営の確保及び派遣労働者の保護等に関する法律」「労働者の職務に応じた待遇の確保等のための施策の推進に関する法律」「女性の職業生活における活躍の推進に関する法律」「青少年の雇用の促進等に関する法律」など



お問合せ先) 独立行政法人 労働政策研究・研修機構(JILPT)

〒177-8502 東京都練馬区上石神井 4-8-23 成果普及課

Tel : 03-5903-6263 Fax : 03-5903-6115 E-mail : book@jil.go.jp

ホームページ : <http://www.jil.go.jp/>