

科学研究費助成事業 研究成果報告書

令和 6 年 6 月 3 日現在

機関番号：12601
研究種目：基盤研究(C)（一般）
研究期間：2020～2023
課題番号：20K01586
研究課題名（和文）Identification and Estimation of Regression Models with a Misclassified and Endogenous Binary Regressor
研究課題名（英文）Identification and Estimation of Regression Models with a Misclassified and Endogenous Binary Regressor
研究代表者
下津 克己（Shimotsu, Katsumi）
東京大学・大学院経済学研究科（経済学部）・教授
研究者番号：50547510
交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：本研究は、内生的な二値説明変数が誤分類されている計量経済モデルにおける識別問題を研究した。経済学の実証研究において、二値説明変数が誤分類されて記録されることは、自己申告による職業訓練への参加などでしばしば見られる。既存研究は、このモデルを、外生性と二値説明変数の測定誤差からの独立性という二つの除外制約を満たす操作変数を用いて識別している。しかしながら、実証研究においては、これらの除外制約を満たす操作変数を見つけることは容易ではない。本研究は、内生的二値説明変数が誤分類されている計量経済モデルを、一つの二値操作変数と一つの二値共変量を用いてノンパラメトリックに識別できることを証明した。

研究成果の学術的意義や社会的意義

本研究によって得られた成果は、理論的にも実証的な応用にも有用である。既存研究においては、内生的な二値説明変数が誤分類されている計量経済モデルを識別するためには、回帰式と誤分類誤差の両方から除外される操作変数を見つけることが必要であった。本研究の結果を適用すれば、このモデルの識別のためには、操作変数は回帰式から除外されていれば十分であり、そして回帰式に含まれる多くの共変量の中から条件をみだす共変量を1つ選べば良い。このため、実証研究におけるこのモデルの識別が著しく容易となる。

研究成果の概要（英文）：This study investigates the identification problem in econometric models in which endogenous binary explanatory variables are misclassified. In empirical studies in econometrics, the misclassification and recording of binary explanatory variables is often observed, for example, in self-reported participation in job training. Existing studies have identified this model with an instrumental variable that satisfies two exclusion restrictions: exogeneity and independence from measurement error of the binary explanatory variable. In empirical studies, however, it is not easy to find instrumental variables that satisfy these exclusion restrictions. This study proves that econometric models in which endogenous binary explanatory variables are misclassified can be identified nonparametrically using a single binary instrumental variable and a single binary covariate.

研究分野：計量経済学

キーワード：識別 内生的二値説明変数 操作変数 共変量

1. 研究開始当初の背景

多くの計量経済学の実証研究において、内生的な二値回帰変数が誤分類されている場合が観察される。代表的な例としては、自己申告による学歴や自己申告による職業訓練への参加などが挙げられる。例えば、Black et al. (2023)は、米国の1990年の国勢調査で専門職学位を持っていると申告した人のうち、実際には66.4%しか専門職学位を持っていないことを発見している。

誤分類された内生的な二値回帰変数が存在するノンパラメトリック回帰モデルは以下で表される。Yを結果変数(例えば賃金)、Xを外生コントロール変数、Uを観測不可能な誤差項、T*を観測不可能な二値回帰変数(例えば真の学歴)、Zを操作変数とすると、以下の式がモデルを表す。

$$Y = a(X) + b(X)T^* + U, E[U | X, Z] = 0.$$

ここで、T*はUと相関する意味で内生的である。TをT*の誤分類を伴う観測値(例えば自己申告の学歴)とする。回帰変数T*は二値であるため、その測定誤差T*-Tは必然的にT*と相関する。このT*-TとT*の相関は、このモデルの識別と操作変数法の適用を困難としてきた。

既存研究の多くは、誤分類された二値回帰変数T*が外生的である、すなわちUと相関しないモデルを研究してきた。しかし、誤分類された二値回帰変数T*が内生的であるモデルの識別を分析した論文は少なく、T*が内生的である場合の上記のモデルの識別可能性は未解決問題であった。

近年、DiTraglia and Garcia-Jimeno (2019)とNguimkeu et al. (2019)が、T*が内生的である場合の上記のモデルの識別条件を導出した。彼らの識別結果は、操作変数Zが、モデルの誤差項Uと独立であり、さらに二値回帰変数T*の誤分類からも独立であること必要とする。

しかし、実証研究では、Uと誤分類の両方から独立である操作変数Zを見つけることは簡単ではない。特に、内生的二値回帰変数T*が自己申告される変数の場合、虚偽の供述をするインセンティブが操作変数Zに影響される可能性があるため、操作変数ZとT*の誤分類が相関する可能性がある。

例えば、学歴が所得に与える効果の分析を考える。ここで、ZはCard (1993)のように大学と自宅との距離、Yは賃金、T*は真の学歴、Tは自己申告された学歴とする。自己申告による学歴は、回答者が虚偽の申告をした場合に誤分類される。しかし、大学と自宅の所在地が同じ都市である場合は、自分の学歴を確認されやすいと考えるかもしれないので、自分の学歴について虚偽の申告をする確率は大学と自宅との距離に影響される。これは、ZがT*の誤分類から独立であるという仮定に反する。

このように、内生的な二値回帰変数が誤分類されている計量経済モデルの識別は、計量経済学理論・実証研究の両方において目下の重要な課題となっている。

2. 研究の目的

本研究の目的は、内生的な二値回帰変数が誤分類されている計量経済モデルにおいて、操作変数ZがT*の誤分類と独立であることを必要としない識別条件を導出することである。

3. 研究の方法

本研究は、結果変数を説明する回帰式の共変量の中の1つであるVが、T*の誤分類と独立である場合に、内生的な二値回帰変数が誤分類されている計量経済モデルが識別可能であることを示す。すなわち、以下のモデルを分析する。

$$Y = a(X, V) + b(X, V)T^* + U, E[U | X, V, Z] = 0.$$

VはT*の誤分類と独立であることが必要であるが、Zとは違い、Vは回帰式の共変量の中から任意に選ぶことができる。操作変数Zは回帰式の共変量となることはできないため、実証研究においてT*の誤分類と独立であるVを見つけることは、T*の誤分類とUの両方から独立である操作変数Zを見つけることよりもはるかに簡単である。

本研究の鍵となるのは、上記のモデルを、観測される変数であるYとTの同時分布が潜在変数T*によって決定される有限混合モデルとして定式化することである。この定式化により、YとTに含まれる情報を有効に活用することが可能となる。Vの(T*, Z)に条件づけた分布がTから独立であると仮定すると、上記のモデルはYとTの(X, Z, V)に条件づけた同時分布の有限混合モデルとして表される。そして、YとTの(X, Z, V)に条件づけた同時分布に、De Lathauwer (2006), Hu (2008), Kasahara and Shimotsu (2009)で用いられた固有値分解を適用することにより、Yの(T*, X, V)に条件づけた分布とTの(T*, X, V)に条件づけた分布を導出することが可能となる。最終的に、Yの(T*, X, V)に条件づけた分布を用いて、上記のモデルにおけるa(X, V)とb(X, V)が識別できる。

4 . 研究成果

本研究は、以下の2つの主要な研究成果をあげた。(1)内生的な二値回帰変数が誤分類されている計量経済モデルにおいて、操作変数と二値回帰変数の誤分類の独立性を必要としない識別条件の導出。(2)(1)で識別されるモデルのノンパラメトリック推定。(3)打ち切りを伴う Weibull duration model における観測されない異質性の検定手法の構築。

(1) 内生的な二値回帰変数が誤分類されている計量経済モデルの識別

本研究は、内生的な二値回帰変数が誤分類されている計量経済モデルを、結果変数と観測される二値回帰変数の同時分布が真の二値回帰変数によって決定される有限混合モデルとして定式化することにより、操作変数と二値回帰変数の誤分類の独立性を必要としない識別条件を導出した。本研究の識別条件は、二値回帰変数の誤分類から独立である共変量 V を用いることを特徴とする。共変量 V は回帰式の共変量の中から任意に選ぶことができるため、既存研究よりも弱い条件の下で識別が可能となる。

(2)(1)で識別されるモデルのノンパラメトリック推定

本研究の識別手法は、観測可能な変数の分布関数により構成される行列に固有値分解を適用することによってモデルの要素を識別する。そのため、識別されるモデルの要素は固有値分解の解として表記される。したがって、観測可能な変数の分布関数をノンパラメトリック推定することにより、モデルの要素もノンパラメトリックに推定することが可能である。この観察に基づき、本研究では、上記のモデルのノンパラメトリック推定量を構築し、その漸近分布を導出した。

(3) 打ち切りを伴う Weibull duration model における観測されない異質性の検定手法の構築

関連する研究として、打ち切りを伴う Weibull duration model における観測されない異質性の検定手法を構築した。打ち切りを伴う Weibull duration model において、観察されない異質性の検定は重要な問題である。しかしながら、このモデルにおける観察されない異質性の検定問題は、その非正規性のために分析が困難であることが知られていた。さらに、打ち切りが共変量に依存する場合には、観察されない異質性の検定手法は存在しなかった。本研究では、尤度に基づいた修正 EM 統計量を用いて、打ち切りを伴う Weibull duration model における観察されない異質性を検定することを提案した。本研究が提案する検定は、打ち切りが共変量に依存する場合にも簡単に適用可能である。また、コンピューター・シミュレーションにより、本研究が提案する検定は、Cho and White (2010)が提案した尤度比検定などの既存の検定よりも優れた性能を持つことが確認された。

参考文献

- Black, D., Sanders, S., and Taylor, L. (2003), "Measurement of Higher Education in the Census and Current Population Survey," *Journal of American Statistical Association*, 98, 545-554.
- Card, D. (1993), "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," Tech. Rep. Working Paper No. 4483, NBER.
- Cho, J. S. and White, H. (2010), "Testing for Unobserved Heterogeneity in Exponential and Weibull Duration Models," *Journal of Econometrics*, 15, 458-480.
- De Lathauwer, L. (2006), "A Link between the Canonical Decomposition in Multilinear Algebra and Simultaneous Matrix Diagonalization," *SIAM Journal on Matrix Analysis and Applications*, 28, 642-666.
- DiTraglia, F. J. and Garcia-Jimeno, C. (2019), "Identifying the Effect of a Mis-Classified, Binary, Endogenous Regressor," *Journal of Econometrics*, 209, 376-390.
- Hu, Y. (2008), "Identification and Estimation of Nonlinear Models with Misclassification Error Using Instrumental Variables: A General Solution," *Journal of Econometrics*, 144, 27-61.
- Kasahara, H. and Shimotsu, K. (2009), "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices," *Econometrica*, 77, 135-175.
- Nguimkeu, P., Denteh, A., and Tchernis, R. (2019), "On the Estimation of Treatment Effects with Endogenous Misreporting," *Journal of Econometrics*, 208, 487-506.

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件/うち国際共著 1件/うちオープンアクセス 1件）

1. 著者名 Hiroyuki Kasahara and Katsumi Shimotsu	4. 巻 38
2. 論文標題 Identification of Regression Models with a Misclassified and Endogenous Binary Regressor	5. 発行年 2022年
3. 雑誌名 Econometric Theory	6. 最初と最後の頁 1117 - 1139
掲載論文のDOI（デジタルオブジェクト識別子） 10.1017/S0266466621000451	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

〔学会発表〕 計5件（うち招待講演 1件/うち国際学会 3件）

1. 発表者名 Katsumi Shimotsu
2. 発表標題 Testing for unobserved heterogeneity in censored duration models: EM approach
3. 学会等名 6th International Conference on Econometrics and Statistics (EcoSta 2023)（国際学会）
4. 発表年 2023年

1. 発表者名 Katsumi Shimotsu
2. 発表標題 Testing for unobserved heterogeneity in censored duration models: EM approach
3. 学会等名 2023 Asian Meeting of the Econometric Society in East and Southeast Asia（国際学会）
4. 発表年 2023年

1. 発表者名 Katsumi Shimotsu
2. 発表標題 Testing the Order of Multivariate Normal Mixture Models
3. 学会等名 TRANSDISCIPLINARY ECONOMETRICS & DATA SCIENCE SEMINAR（招待講演）（国際学会）
4. 発表年 2022年

1. 発表者名 下津克己
2. 発表標題 Identification of Regression Models with a Misclassified and Endogenous Binary Regressor
3. 学会等名 日本経済学会2021年度秋季大会
4. 発表年 2021年

1. 発表者名 下津克己
2. 発表標題 Testing the Order of Multivariate Normal Mixture Models
3. 学会等名 2021年度関西計量経済学研究会
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関