

令和 5 年 4 月 25 日現在

機関番号：34316

研究種目：基盤研究(C)（一般）

研究期間：2019～2022

課題番号：19K01595

研究課題名（和文）二つのデータセットの接合を伴う回帰モデルの推定及び推論

研究課題名（英文）Estimation and Inference of Regression Models in Which Two Data Sets Are Combined

研究代表者

蛭川 雅之（Hirukawa, Masayuki）

龍谷大学・経済学部・教授

研究者番号：10597628

交付決定額（研究期間全体）：（直接経費） 3,400,000円

研究成果の概要（和文）：マイクロ経済データを利用して線形回帰モデルを推定する際、手元にあるデータセットがモデル内の変数を完備していない状況がしばしば起こる。この場合、必要な変数を記録する別のデータセットを組み合わせてモデルを推定する必要がある。本研究では、研究代表者がこれまで調査・研究を行ってきた2標本推定問題の多方面にわたる深化・発展を目指し、特に、(i)最近傍法によらない2標本回帰推定、(ii)データ接合の際の次元の呪いの回避策を中心に取り組んだ。同時に、標準的な統計パッケージ上でコマンドを開発し、実務家の利便も図った。

研究成果の学術的意義や社会的意義

2標本の接合を前提とする線形回帰モデル推定に関してより効率性の高い推定法を提案し、その大標本・小標本特性を探求した。また、2標本推定は「生成された説明変数」を用いた推定法の特殊ケースである。高次元データから説明変数を生成する際の次元の呪いを回避する策の一つとして、スパースモデリングの発想を取り入れて説明変数を生成する手法を開発した。加えて、これら推定手法を標準的な統計ソフト上で実装するためのコマンドも開発・公開した。さらに、本研究の副産物として、ある非対称カーネル関数を用いたノンパラメトリック推定量の一致収束及びその収束速度を証明・導出し、非対称カーネル関数の利用可能性を拡大した。

研究成果の概要（英文）：Suppose that we are interested in estimating linear regression models using microeconomic datasets. Often a single dataset does not contain all variables required for the estimation. In this case, we are tempted to estimate the regression models by combining the primary dataset at hand with an auxiliary one, if any, that includes missing variables. This research project has aimed at extending and developing two-sample regression estimation methods studied by the principal investigator, with emphases on (i) exploring an estimation method that does not rely on the nearest neighbor method and (ii) circumventing the curse of dimensionality in data combination. At the same time, commands for these new methods executed in a standard statistical package have been developed and made public, for the benefit of practitioners.

研究分野：社会科学・経済学・経済統計

キーワード：接合データの計量経済学 2標本回帰推定 セミパラメトリック推定 非対称カーネル関数 最近傍法 生成された説明変数 スパースモデリング サンプルセレクションモデル

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

## 1. 研究開始当初の背景

匿名マイクロ経済データ（個票データ）を利用して線形回帰モデルを推定する際、手元にあるデータセットがモデル内の変数を完備していない状況がしばしば起こる。この場合、必要な変数を記録する別のデータセットの情報を利用してモデルを推定する必要が生じる。一例を挙げると、教育効果を測定する際の重要な説明変数の一つが個人の能力を表すものである。なお、能力は直接観測されないため、知能・職能テストの得点をその代理変数として用いるのが通例である。しかし、賃金関数の推定に使用される賃金・所得のサーベイではテストの得点を調査しない。そこで、テストの得点を記録した別のデータセットを組み合わせることで回帰分析を行うことになる。

本研究の出発点となる Hirukawa and Prokhorov (2018, “HP18”) では、主たるデータセットが説明変数の一部を欠いているため、別のデータセットを最近傍法 (nearest neighbor method) により接合して得られるデータセットを構築して回帰推定する状況を想定した。HP18 はこのようなデータセットを用いた最小 2 乗推定に一致がないことを示し、ノンパラメトリック法を応用したバイアス修正推定法を提案した。

HP18 はこの種のデータ接合を伴う回帰推定に関する初の理論的研究であり、その成果を発展・拡張させる余地は大きい。特に、HP18 を完成させる途上で浮き彫りになった以下の課題に対する解決策を見出したいと考えた。

(1) バイアス修正推定量の標準誤差が大きい。【⇒最近傍法にこだわらず、どうすれば 2 標本推定の精度を高めることができるか。】

(2) 「次元の呪い (curse of dimensionality)」の影響により、バイアス修正推定量のパラメトリック収束を確保するにはデータ接合に用いる共通変数を最大 3 種類までと限定せざるを得ない。【⇒より多くの共通変数を用いてデータ接合を行いたい場合どうするか。】

## 2. 研究の目的

(1) 第一の目的は、マイクロ経済分析に利用可能な新たな 2 標本推定法を提案し、その統計的特性を明らかにすることであった。2 標本推定は包括的なサーベイ (Ridder and Moffitt, 2007) が存在するほど重要な計量経済学の一分野である。しかし、最近傍法によるデータ接合と回帰推定を組み合わせた場合の統計的特性に関する分析は HP18 まで行われていなかった。本研究では、主たるデータセットに存在しない説明変数を別のデータセットから補填する枠組みの中で、2 標本推定理論の深化・発展を企図した。

(2) もう一つの目的は、標準的な統計パッケージ上で各 2 標本推定法に関するコマンドを開発し、実務家への利便を図ることであった。2 標本推定法を計量経済学・統計学の専門誌で公開するだけでは、インパクトは理論家の間に留まる。実務家が本研究の成果を利用しやすい環境を整えるために、統計パッケージ上でのコマンド開発は必須である。本研究では、経済学者間で最も広く利用されている統計パッケージ Stata 上でのコマンド開発も目指した。

## 3. 研究の方法

(1) 研究対象：匿名マイクロ経済データを接合して推定する経済モデルを線形回帰モデルに特化した。2 標本推定問題に固有の特徴として、モデル識別・漸近理論の統一的取り扱いが困難であり、モデル毎の個別対応が必要となることが知られている。そのため、2 標本推定に関する研究の大部分は見通しの立てやすい線形回帰モデルを対象としており、本研究も線形回帰モデルの分析に注力した。また、推定手法導出の過程でノンパラメトリック推定に関する新たな理論を構築する必要も生じた。これらも 2 標本回帰推定に関連するものとして研究対象に含めた。

(2) 研究手順：新たに提案する推定量の大標本特性（例：一致性、収束速度、極限分布）を導出し、さらに、モンテ・カルロ実験を通してそれらの小標本特性を確認した。加えて、経済データへの応用例も紹介した。なお、データ分析に使用する匿名マイクロ経済データは米国のもので、米国在住の共同研究者経由で入手したものである。

(3) 研究組織：本研究は、研究代表者及び海外の研究者数名による国際共同研究として遂行された。本研究は HP18 の延長線上に位置づけられ、計量経済学に関する幅広い知識（例：回帰分析、ノンパラメトリック法、計測誤差の統計的処理等）を要求される課題である。なお、研究代表者及び海外の共同研究者は、これらの話題に関する論文を計量経済学・統計学の主要英文学術誌に多数発表している。

(4) 成果公開：本研究では、理論家・実務家両者への成果の浸透を意識した。具体的には、2 標本推定理論に関連する論文を主要な計量経済学・統計学の英文査読誌へ投稿する一方、Stata コ

マンドの解説論文についてのみ英文査読誌 *Stata Journal* での掲載を目指した。

#### 4. 研究成果

(1) HP18 で開発した最近傍法による 2 標本推定量を *Stata* で実装するためのコマンド開発を行った。当該コマンド“*msreg*”の解説論文は英文査読誌 *Stata Journal* に掲載された。

(2) HP18 で提案したバイアス修正推定量の漸近分散にはバイアス項推定値の分散が加わっており、その結果、標準誤差が大きくなりがちである。標準誤差をより小さく推定する方法の一つとして、主たるデータセットにない説明変数の条件付き期待値を別のデータセットから一致推定し、これを当該説明変数の代理変数として利用して回帰モデルを最小 2 乗推定するという手法を開発した。なお、条件付き期待値の関数形は特定しないため、ノンパラメトリック法、具体的には、カーネル回帰推定を利用した。この最近傍法によらない 2 標本推定法でも、HP18 のバイアス修正推定量と同様、パラメトリック収束を確保するにはデータ接合に用いる共通変数を最大 3 種類までと限定せざるを得ないことが判明した。なお、2 標本推定法は従来の操作変数法に代わる手法の一つとも考えられる。実際、説明変数の一部が主たるデータセットにない場合、省略変数バイアスを避けるために操作変数を利用して推定を行うことが通例である。数多くのモンテ・カルロ実験を通して、今回提案する推定量は HP18 のバイアス修正推定量及び操作変数推定量より効率的であることを確認した。また、この推定量の経済データへの応用例も紹介した。研究結果をまとめた論文は計量経済学の英文査読誌 *Econometric Reviews* に掲載された。

(3) 成果(2)の論文では、非対称ベータ・カーネル関数を使用したノンパラメトリック回帰推定量を代入する 2 標本推定量の方が、一般に用いられる対称カーネル関数を使用したものに比べて優れた小標本特性を持つことを示唆している。尤も、ベータ・カーネル関数を用いたノンパラメトリック推定量の一樣収束及びその収束速度を証明・導出した文献は皆無であった。そこで、これらノンパラメトリック推定量に関する一樣収束の結果をまとめた論文を別途執筆した。本論文は統計学の英文査読誌 *Scandinavian Journal of Statistics* に掲載された。

(4) 本研究で取り扱う 2 標本推定は、主たるデータセットに存在しない説明変数の代理変数を別のサンプルから生成・代入する手法をとる。この手法は「生成された説明変数 (*generated regressor*)」の特殊ケースと見做すことができる。また、HP18 及び(2)で提案した手法では、回帰係数推定量のパラメトリック収束を確保するためにデータ接合に用いる共通変数を最大 3 種類までと限定していた。高次元データから代理変数を生成するには、このような次元の呪いの問題を解決する必要もある。これらの点を踏まえ、生成された説明変数とスパースモデリングによる多段階推定を組み合わせた手法、具体的には、スパースモデリングを用いてサンプルセレクションモデルを 2 段階推定する手法の開発に着手した。この手法は優れた小標本特性を持つことを確認しており、推定手法全体の理論的裏付け及び経済データへの応用例を加えて論文にまとめた。本論文は実証経済学の英文査読誌 *Empirical Economics* に掲載されることが決定しており、オンライン版が既に刊行されている。同時にこの推定手法を *Stata* 上で実装するためのコマンド“*ds-heck*”も公開した。

(5) 成果(1)-(4)に加え、更に 2 本論文を刊行した。それらは、非対称カーネル関数を用いたノンパラメトリック回帰推定量にある種のバイアス修正を施すと収束速度が改善することを証明する論文、及びカーネル法を中心とする長期共分散行列推定に関するサーベイ論文である。いずれも計量経済学・統計学の英文査読誌 *Econometrics and Statistics* からの招待を受けて執筆した。両論文ともノンパラメトリック法が応用され、内容的に本研究の課題と関連が深い点を付記しておく。

(6) 上記の他、研究代表者は令和 4 年 6 月に所属研究機関において開催された国際学会 5th *International Conference on Econometrics and Statistics (EcoSta 2022)* の大会実行委員長・現地責任者を務めた。本学会は計量経済学・統計学・データサイエンスの分野で世界屈指の規模を誇るものである。コロナ禍に伴い当初計画とは異なる対面・オンラインを併用したハイブリッド形式での開催となったが、その運営に携わることにより、通常の学会発表とは違った貢献ができ、また、今後の研究の進展にも大きな刺激となった。

#### <引用文献>

- ① Hirukawa, M., and A. Prokhorov (2018): “Consistent Estimation of Linear Regression Models Using Matched Data,” *Journal of Econometrics*, 203, 344-358.
- ② Ridder, G., and R. Moffitt (2007): “The Econometrics of Data Combination,” in J. J. Heckman and E. E. Leamer (eds.), *Handbook of Econometrics*, Vol. 6, Part B. Elsevier, 5469-5547.

5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 6件／うち国際共著 5件／うちオープンアクセス 1件）

1. 著者名 Masayuki Hirukawa, Irina Murtazashvili, Artem Prokhorov	4. 巻 42
2. 論文標題 Yet Another Look at the Omitted Variable Bias	5. 発行年 2023年
3. 雑誌名 Econometric Reviews	6. 最初と最後の頁 1-27
掲載論文のDOI（デジタルオブジェクト識別子） 10.1080/07474938.2022.2157965	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する
1. 著者名 Masayuki Hirukawa, Irina Murtazashvili, Artem Prokhorov	4. 巻 49
2. 論文標題 Uniform Convergence Rates for Nonparametric Estimators Smoothed by the Beta Kernel	5. 発行年 2022年
3. 雑誌名 Scandinavian Journal of Statistics	6. 最初と最後の頁 1353-1382
掲載論文のDOI（デジタルオブジェクト識別子） 10.1111/sjos.12573	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する
1. 著者名 Masayuki Hirukawa, Di Liu, Irina Murtazashvili, Artem Prokhorov	4. 巻 -
2. 論文標題 DS-HECK: Double-Lasso Estimation of Heckman Selection Model	5. 発行年 2023年
3. 雑誌名 Empirical Economics	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s00181-023-02406-w	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
1. 著者名 Masayuki Hirukawa	4. 巻 -
2. 論文標題 Robust Covariance Matrix Estimation in Time Series: A Review	5. 発行年 2022年
3. 雑誌名 Econometrics and Statistics	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.ecosta.2021.12.001	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Masayuki Hirukawa, Di Liu, Artem Prokhorov	4. 巻 21
2. 論文標題 msreg: A Stata Command for Consistent Estimation of Linear Regression Models Using Matched Data	5. 発行年 2021年
3. 雑誌名 Stata Journal	6. 最初と最後の頁 123-140
掲載論文のDOI (デジタルオブジェクト識別子) 10.1177/1536867X211000008	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 Benedikt Funke, Masayuki Hirukawa	4. 巻 20
2. 論文標題 Bias Correction for Local Linear Regression Estimation Using Asymmetric Kernels via the Skewing Method	5. 発行年 2021年
3. 雑誌名 Econometrics and Statistics	6. 最初と最後の頁 109-130
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.ecosta.2020.01.004	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

〔学会発表〕 計9件 (うち招待講演 7件 / うち国際学会 6件)

1. 発表者名 蛭川雅之
2. 発表標題 Nonparametric Threshold Detection for Cost Distributions
3. 学会等名 第30回関西計量経済学研究会
4. 発表年 2023年

1. 発表者名 蛭川雅之
2. 発表標題 Nonparametric Threshold Detection for Cost Distributions
3. 学会等名 研究会「第23回ノンパラメトリック統計解析とベイズ統計」(招待講演)
4. 発表年 2023年

1. 発表者名 Masayuki Hirukawa
2. 発表標題 Robust Covariance Matrix Estimation in Time Series: A Review
3. 学会等名 41st International Symposium on Forecasting (ISF 2021) (招待講演) (国際学会)
4. 発表年 2021年

1. 発表者名 Masayuki Hirukawa
2. 発表標題 Yet Another Look at the Omitted Variable Bias
3. 学会等名 1st International Conference on Econometrics and Business Analytics (iCEBA 2021) (招待講演) (国際学会)
4. 発表年 2021年

1. 発表者名 Masayuki Hirukawa
2. 発表標題 DS-HECK: Double-Lasso Estimation of Heckman Selection Model
3. 学会等名 14th International Conference on Computational and Financial Econometrics (CFE 2020) (招待講演) (国際学会)
4. 発表年 2020年

1. 発表者名 Masayuki Hirukawa
2. 発表標題 Yet Another Look at the Omitted Variable Bias
3. 学会等名 15th International Symposium on Econometric Theory and Applications (SETA 2019) (国際学会)
4. 発表年 2019年

1. 発表者名 Masayuki Hirukawa
2. 発表標題 Bias Correction for Local Linear Regression Estimation Using Asymmetric Kernels via the Skewing Method
3. 学会等名 3rd International Conference on Econometrics and Statistics (EcoSta 2019) (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 蛭川雅之
2. 発表標題 Yet Another Look at the Omitted Variable Bias
3. 学会等名 データサイエンス・サマーキャンプ福島 (招待講演)
4. 発表年 2019年

1. 発表者名 Masayuki Hirukawa
2. 発表標題 GMM Estimation Using Time-Series Data with Missing Observations
3. 学会等名 13th International Conference on Computational and Financial Econometrics (CFE 2019) (招待講演) (国際学会)
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

Masayuki Hirukawa - Home  
<https://www.econ.ryukoku.ac.jp/~hirukawa/>

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			
オーストラリア	University of Sydney			
米国	Drexel University	Stata Corp.		
ドイツ	TH Koln			