

令和 6 年 6 月 12 日現在

機関番号：32689

研究種目：基盤研究(C)（一般）

研究期間：2020～2023

課題番号：20K01598

研究課題名（和文）形状制約を用いたノンパラメトリック、セミパラメトリック推定の理論と応用

研究課題名（英文）Theory and its application of nonparametric and semiparametric estimation and inference under shape constraint

研究代表者

荒井 洋一（Arai, Yoichi）

早稲田大学・社会科学総合学院・准教授

研究者番号：50376571

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）：本研究では、不均一分散が存在する回帰モデルにおいて、単調性という形状制約を用いた一般化最小二乗法を提案した。第一段階の推定において isotonic regression を用いた一般化最小二乗推定量が現実的な仮定のもとで漸近的にセミパラメトリック効率的であることを示した。提案された一般化最小二乗推定量は平滑化パラメータを必要とせず計算量もノンパラメトリックな推定量と比較して非常に少ない。シミュレーションに基づく結果ではノンパラメトリックな一般最小化二乗法と比較しても優れていること、広く用いられている頑健な分散推定量を用いた方法よりもかなり優れていることが示された。

研究成果の学術的意義や社会的意義

本研究ではいくつかの重大な欠点のため最近用いられることの少なくなった一般化最小二乗推定量に注目した。欠点を形状制約を用いることにより解決し、現実的で非常に優れた推定量を提案した。現在、広く用いられている頑健な分散推定量を用いたアプローチよりかなり効率的な推定を行うことができ、実証研究において広く用いられることが期待される。本研究で提案された一般化最小二乗推定量の第一段階目では推定された変数を被説明変数として用いている。この一般化は新しい結果であり、一般化最小二乗法だけでなくその他の統計的な問題に広く用いることができる。

研究成果の概要（英文）：We proposed a generalized least squared (GLS) estimator with monotone restriction under the heteroskedasticity. We employ the isotonic regression in the first stage and use the isotonic regression estimator as a weight for the second stage. We show that the GLS estimator is semiparametrically efficient. The proposed GLS estimator does not require any smoothing parameters and it is computationally attractive. Our simulation shows that the GLS estimator is superior to nonparametric GLS estimators and robust approaches.

研究分野：計量経済学

キーワード：セミパラメトリック ノンパラメトリック 推定 検定 形状制約 単調性

### 1. 研究開始当初の背景

経済学においては経済変数間の関係を分析するにあたり回帰モデルを推定することが一般的である。回帰モデルを推定するにあたっては事前に関数形を定める必要があるが、経済理論では経済変数間の関係について具体的な関数形が定まらない場合が多い。このような場合には近似的に線形回帰モデルのようなパラメトリックなモデルを推定する機会が多いが、実際のデータから変数間の関係が想定とはかけ離れたものであることが示唆されることも多い。

このような状況で実証分析を行うときには経済変数間の関係について強い仮定を置かないノンパラメトリックモデルが有力な分析手段となる。ノンパラメトリックモデルを用いた手法には、関数形に関する仮定が緩いという大きな長所がある一方で、その長所ゆえの短所も存在する。短所としては、実用上問題となりうる計算量の多さと手順の複雑さが存在する。関数形の柔軟性のために計算量が負担になる場合も多く、計算が収束するのに長い時間がかかる場合がある。複雑な関数形に対応するためにサンプル数が大きい必要がある場合も多い。また、ノンパラメトリックモデルを用いた分析にはバンド幅 (bandwidth) のような平滑化パラメーター (smoothing parameter) が必要になるが最適な平滑化パラメーターの選択は容易ではない。しかし推定量の性質は平滑化パラメーターの選択に大きく依存することが広く知られている。

ノンパラメトリックモデルの推定量の短所を克服するための方法の一つが形状制約の導入である。形状制約とは、モデルの関数形についての単調性、凸性、凹性や滑らかさのような制約である。経済理論では具体的な関数形に関する情報を与えることは少ないが、質的な情報を与えることは多い。たとえば需要関数は価格の減少関数であり、生産関数は投入材に関して非減少かつ凹であることなどが経済理論と整合的であると考えられる。よってこれらのような形状制約を導入することによりパラメトリックモデルやノンパラメトリックモデルの短所を克服しようとする試みがされている。

### 2. 研究の目的

本研究においては不均一分散が存在する線形回帰モデルに注目する。不均一分散は実証分析で多く用いられるクロスセクションデータにおいて存在する機会が非常に多い。線形回帰モデルの推定においては最小二乗法が用いられることが多いが、不均一分散が存在する場合には最小二乗推定量は依然として一致性を持つ一方で非効率となり分散が大きくなってしまふ。最小二乗法を用いる場合には現在では頑健な分散推定量を報告する機会が多い。不均一分散が存在する場合、以前は一般化最小二乗法が用いられることが多かった。一般化最小二乗法では、誤差項の分散が説明変数に依存すると仮定する。そのような場合の一般化最小二乗法には、大きく分けて2つの方法が存在する。それらは、分散と説明変数の関係に関してパラメトリックなモデルを仮定する場合とノンパラメトリックなモデルを仮定する場合である。パラメトリックなモデルを仮定する場合は推定の手順は単純で利用しやすい。ただし、経済理論が分散と説明変数の関係を明示的に示唆することは極めて稀であるため定式化の誤りへの懸念が付きまとう。一方でパラメトリックな関係を仮定せずに推定を行うノンパラメトリックなモデルにおいては計算量が多くなり、手順が複雑になるという欠点がある。また平滑化パラメーターの選択という難しい問題も存在する。

そこで本論文においては分散と説明変数の間に単調性という形状制約を想定し、形状制約を用いたセミパラメトリックな推定量を提案する。単調性は変数間の関係として自然に成り立つと考えられる機会が多く、形状制約を用いたセミパラメトリックな推定量はノンパラメトリックな推定量と違い平滑化パラメータを用いずに推定できる場合がある。また、計算量が少なく手順を簡素化することが可能となるなど実用上大きな利点がある。また、不均一分散が存在する場合に現在多く用いられる頑健な分散推定量を用いた方法よりも効率的な推定が可能となる。

### 3. 研究の方法

本研究では、不均一分散に関する単調性制約のもとで一般化最小二乗法を提案する。例えば、クロスセクションデータに基づく消費関数の推定においては、誤差項の分散と説明変数の一つである所得の間に単調な関係が存在すると仮定するのは非常に自然である。

まず、誤差項の分散と説明変数の関係を推定するにあたり統計学においては古くから用いられている isotonic regression という統計手法を用いる。そのためには、モデルをまず最小二乗法で推定し残差を求める。そして残差の二乗を被説明変数とした isotonic regression により条件付き分散関数を推定する。次に推定された分散を重みとして用いる一般化最小二乗法 (GLS) によりモデルを推定する。次に大標本における GLS 推定量の統計的な性質を導出する。本研究の学術的独自性と新規性は、第一段階の isotonic regression において被説明変数が残差のように誤差をもって推定されたものであること (generated variable)、第二段階においては isotonic regression の推定量を重みとして用いた一般化最小二乗法を提案し統計的な性質を吟味していることにある。

一般的に isotonic regression に基づく推定量は求める手順こそ非常に単純であるが、非連続の関数となる性質を持っているため、漸近的な性質が非常に複雑なものとなる。本研究においては isotonic regression を単純に用いるのではなく、推定された被説明変数を用いて推定を行い、さらにその推定されたものを重みとして用いることになるため漸近的な性質がさらに複雑となる。そのために経験過程に

関する最新の理論を用いて一般化最小二乗推定量の漸近的な性質を導出する。また提案された推定量の有限標本における性質をシミュレーションを用いて検証する。

#### 4. 研究成果

本研究では、不均一分散が存在する回帰モデルにおいて、単調性という形状制約を用いた一般化最小二乗法を提案した。まずは説明変数が一つの単回帰モデルにおいて推定方法や統計的性質を吟味した。第一段階の条件付き分散関数の推定において isotonic regression を用いた一般化最小二乗推定量が、現実的な仮定のもとで漸近的にセミパラメトリックな意味で効率的であることを示した。つまり一般的なクラスの中で提案された推定量が最小の分散をもつことを示した。その際に、三乗根漸近理論に基づく条件付き分散関数推定量の漸近的な性質も導出した。提案された一般化最小二乗推定量は平滑化パラメータを必要とせず計算量もノンパラメトリックな推定量と比較して非常に少ない。モンテカルロシミュレーションに基づく結果ではノンパラメトリックな一般最小化二乗法と比較して優れていること、広く用いられている頑健な分散推定量を用いた統計的推測よりもかなり優れていることが示された。特に 95%信頼区間の平均的な長さが頑健な分散推定量を用いた方法と比べ経験カバレッジを犠牲にすることなく、かなり短くなっていることは注目に値する。

次にここまでの結果を説明変数が内生的な場合への拡張を行なった。実験のように説明変数を操作できない状況の多い社会科学の研究においては説明変数が内生的となる場合も多い。これは説明変数が個人の選択によって決定される場合が多いためでもある。このような状況では操作変数を用いて操作変数法や二段階最小二乗法(TSLS)にもとづいて線形モデルを推定するケースが多い。そこで本研究では不均一分散が存在し、説明変数が内生的な状況における TSLS 推定量を提案した。手順としてはまず通常の TSLS を用いてモデルを推定し TSLS 残差を求める。その残差の二乗を被説明変数として isotonic regression により単調性制約に基づく条件付き分散関数を推定する。そして推定された分散を重みとして、重みつき TSLS 推定量を求めるという手順になる。説明変数が外生的である場合と同様に提案された推定量がセミパラメトリックの意味で漸近的に効率的であることを示した。

次に単調性という形状制約に対して懸念が存在する場合どのように対処すべきであるか議論を行なった。単調性という形状制約の下で提案された一般化最小二乗推定量や TSLS 推定量は単調性が満たされない場合においても一貫性を持つ。しかしそれらの漸近的理論に基づく分散推定量は正しいものではない。そこで単調性が正しくない場合においても漸近的に妥当となる頑健な分散推定量を提案した。この頑健な分散推定量は、単調性が正しい場合には先に提案された分散推定量と漸近的に同等となっていることも示した。

ここまでの研究はすべて条件付き分散推定量が一つの説明変数に依存する場合において行われた。しかし現実的には条件付き分散関数が複数の説明変数に依存する可能性も十分にある。そこでそのような場合に対処するために条件付き分散関数にシングルインデックスモデルを想定して議論を行なった。シングルインデックスモデルは、条件付き分散関数については以前と同様に一つの変数を引数としてとるが、その一つの変数の引数は複数の説明変数の線型結合となっている。よって結果的に条件付き分散関数は複数の説明変数に基づくこととなっている。シングルインデックスモデルはパラメトリックなモデルよりより広い関数形に対応することができるという利点がある。またノンパラメトリックなモデルと異なり次元の呪いに苦しまないのも、説明変数の数が多い場合にも対処が可能となる利点も存在する。

条件付き分散関数にシングルインデックスモデルを想定した場合の推定のアルゴリズムを提案した。まずはさまざまな線型結合が与えられた場合の条件付き分散関数を各々推定し、それらの中で最小の二乗誤差を与える線型結合を選択するという手順となる。ここまでで開発された一般化最小二乗法の理論や TSLS 推定量の議論を発展させて、条件付き分散関数にシングルインデックスを想定した場合の推定量の漸近的な性質を導出した。特にその推定量がセミパラメトリックの意味で効率的であることを示した。また三乗根漸近理論を用いてシングルインデックスモデルの条件付き関数推定量の漸近的な分布も導出した。また複数のシングルインデックス構造が加法的に現れるようなより一般的なモデルへの拡張も議論した。またシングルインデックス構造を想定した条件付き分散関数が単調性を満たさない場合に妥当となるような頑健な分散推定量も導出した。そして有限標本における推定量の性質を吟味するためにモンテカルロシミュレーションを行なった。そこでは頑健な分散推定量に基づく通常の最小二乗法を用いた方法やノンパラメトリックな方法に基づく方法と比較して優れた性質を持つことが確認できた。またシングルインデックス構造が成り立っていない場合においても他の推定量と同等の性質を持つことも確認ができた。

これまでの理論的な結果を現実のデータへ適用し実証分析を行なった。経済学において一人当たりの GDP と高齢者人口比率の間には負の関係があると伝統的に考えられていたが最近の研究では負の関係が検証されないという結果を報告しているものもある。それらの結果は通常の最小二乗法や TSLS を用いて頑健な分散推定量に基づいている。そこで本研究で提案された新しい推定量を用いてより正確な推定を行ない再検証した。世界 169 カ国のデータを用いた結果では概ね既存の研究を追随する結果となったが、OECD35 カ国のデータを用いて検証した場合には負の関係が存在することを発見した。これは新たな結果でありさらなる議論が必要であることを示している。

本研究で提案された一般化最小二乗推定量の第一段階目では推定された変数を被説明変数として用いている。この一般化は統計学の分野において新しい結果であり、一般化最小二乗法だけでなくその他の統計的な問題に広く用いることができる。また生成された変数を用いて isotonic regression を行うというアイデアも統計学においても新しく、本研究において示された結果はより広い応用へ適用が可能であり、その結果自体も統計学的に重要である。

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 Arai Yoichi, Hsu Yu-Chin, Kitagawa Toru, Mourifi? Ismael, Wan Yuanyuan	4. 巻 13
2. 論文標題 Testing identifying assumptions in fuzzy regression discontinuity designs	5. 発行年 2022年
3. 雑誌名 Quantitative Economics	6. 最初と最後の頁 1~28
掲載論文のDOI（デジタルオブジェクト識別子） 10.3982/QE1367	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3件（うち招待講演 0件/うち国際学会 3件）

1. 発表者名 Yoichi Arai
2. 発表標題 GLS under Monotone Heteroskedasticity
3. 学会等名 Econometric Society European Meeting (Barcelona) (国際学会)
4. 発表年 2023年

1. 発表者名 Yoichi Arai
2. 発表標題 GLS under Monotone Heteroskedasticity
3. 学会等名 Spring Econometrics Forum (国際学会)
4. 発表年 2024年

1. 発表者名 Yoichi Arai
2. 発表標題 GLS under Monotone Heteroskedasticity
3. 学会等名 Econometrics Seminar, University of Arizona (国際学会)
4. 発表年 2024年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

GLS under Monotone Heteroskedasticity  
<https://arxiv.org/abs/2210.13843>

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			
英国	London School of Economics			
ドイツ	University of Mannheim			