

令和 3 年 5 月 31 日現在

機関番号：34504

研究種目：基盤研究(C)（一般）

研究期間：2018～2020

課題番号：18K01553

研究課題名（和文）モード回帰に対する計量手法の開発とその応用

研究課題名（英文）Econometric Analysis of Modal Regression Models and Their Applications

研究代表者

古澄 英男（KOZUMI, Hideo）

関西学院大学・経済学部・教授

研究者番号：10261273

交付決定額（研究期間全体）：（直接経費） 2,900,000円

研究成果の概要（和文）：通常回帰分析に代わる計量手法として、モード回帰とよばれる方法がある。モード回帰は、データ分布のモードの構造を分析するための方法であり、近年注目を浴びている。本研究では、このモード回帰を取り上げ、新たなモード回帰モデルの開発を行うとともに、提案する計量モデルを効率的に推定するため、シミュレーションを利用した推定方法の開発を行った。また、数値実験や実データによる実証分析を行うことによって、本研究で提案する方法の有用性について比較・検討した。

研究成果の学術的意義や社会的意義

本研究によって有用かつ実用的な計量モデルや推定方法が開発されたことにより、経済学におけるモード回帰モデルの利用が進むことが期待される。また、モード回帰分析の適用範囲は経済学だけでなく、生物学、医学、工学、天文学など多岐にわたるため、本研究の学術的波及効果は大きいと考えられる。

研究成果の概要（英文）：Modal regression has gained increasing attention in recent years since it offers an important alternative to classical mean regression. Unlike mean regression, modal regression aims to explore the relationship between a response variable and its covariates. In this research project, we have developed new modal regression models and efficient simulation-based methods for analyzing economic data. The proposed methods were examined in comparison with some other existing methods, and their effectiveness were clarified via both simulated and real data sets.

研究分野：経済統計学

キーワード：ベイズ統計学 マルコフ連鎖モンテカルロ法 モード回帰モデル

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。

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

(1) 経済学の実証分析では、分布が歪んでいたり、あるいは外れ値を含んでいたりするデータを分析しなければならないことがある。よく知られているように、分布が歪んでいる場合、平均値は必ずしもデータの中心的傾向を表しておらず、データの特徴を適切に捉えることができない。また別の問題として、平均値は外れ値の影響を受けやすく、得られる分析結果は不安定なものとなる。そのため、データの平均構造に注目する通常の回帰モデルでは、こうした経済データを適切に分析することができない。そこで最近になり、通常の回帰分析に代わる計量手法としてモード回帰に関心が向けられるようになってきた。

(2) モード回帰は、データ分布のモードの構造を分析するための計量手法であり、古くから研究が行われてきた(モードとは、データや確率分布において度数や確率が最大となる値のことである)。しかし、モデルの推定では、絶対値を伴う目的関数を最大化しなければならないため、複数の局所解が存在し大局解を得るのが困難であるなど、数値計算上の問題点が指摘されていた。そのため、モード回帰モデルの実データへの適用は、データ数や説明変数の数が少ない場合などに限られており、既存の方法に代わる新たな推定方法の開発が必要とされていた。

(3) 最近の経済学では、打ち切りデータ、質的データ、時系列データ、パネルデータなど様々な種類のデータが実証分析で用いられている。しかし、既存の接近方法ではこうした多様な経済データをモード回帰によって分析することは困難であり、そのためこれまでに十分な研究成果が得られているとは言えなかった。こうした状況を解決するため、推定方法の開発とあわせてモード回帰モデルの拡張や新たなモデルの開発も求められていた。

(4) 「平均値」、「中央値」、「モード」は、データの中心的傾向を表す重要な代表値であり、通常、回帰分析が「平均値」、分位点回帰が「中央値」、モード回帰が「モード」に対応する。研究を開始した当初、データの分位点を分析する分位点回帰については、理論的にも実証的にも着実に研究が進んだことから、その有用性が示され様々な分野で利用されるようになっていた。経済学分野において、モード回帰も通常、回帰分析や分位点回帰のように重要な計量手法となり得るのであるか? この問いに対する答えを探るため、モード回帰モデルに対する計量分析の枠組みを新たに構築することが重要な課題の一つとなっていた。

2. 研究の目的

本研究では、以下に示す研究課題に対して研究を実施する。

(1) モード回帰モデルの開発

本研究では、歪んだ分布を持つデータを念頭に置き、既存のモード回帰モデルの見直しや拡張を行うことによって、新たなモード回帰モデルの開発を行う。また、計量分析でよく用いられるデータである打ち切りデータとパネルデータを取り上げ、それら进行分析するために必要なモード回帰モデルの開発を行う。こうしたモデルの開発と並行して、提案するモード回帰モデルの理論的性質についても明らかにする。

(2) モード回帰モデルに対する統計的分析方法の開発

「1. 研究開始当初の背景」で述べたように、既存の推定方法では数値計算上の問題がある。また、本研究で提案するモード回帰モデルでは、尤度関数が複雑で解析的に解けないことも予想される。そこで新たな接近方法として、シミュレーションによる推定方法(主としてマルコフ連鎖モンテカルロ法)の開発を目指す。研究代表者は、これまでシミュレーションによる推定方法の開発について研究を行っており、いくつかの研究成果を得ている。これらの研究成果を一層発展させることにより、既存の方法よりも効率かつ実行が容易な推定方法を提案する。

(3) モード回帰モデルを用いた現実の経済問題への応用

本研究で提案するモード回帰モデルならびにその統計的分析方法の有用性を示すため、実際のデータを用いた計量分析を行う。主たる応用分野としては、生産効率性と賃金分布の分析を考えることにする。また、既存の計量モデル(主として通常、回帰モデルと分位点回帰モデル)との比較を行うことにより、本研究で提案するモード回帰モデルやそれに対する推定方法の有用性を検証する。

これらの研究課題に取り組むことによって、モード回帰モデルに対する新たな計量分析の枠組みを構築することが本研究の目的である。

3. 研究の方法

(1) 本研究の遂行に必要な文献およびデータの収集を行う。また、必要に応じて収集したデータを整理・加工する。

(2) モード回帰に対する新たな計量モデルならびにシミュレーションによる推定方法の開発を行うため、既存研究においてどのようなモデルや推定方法が提案されてきたかを精査し、その問題点を明らかにする。また、関連する実証研究についても調べ、どのような研究成果が得られているのかを整理する。

(3) モード回帰モデルの開発を行う。具体的には以下のとおりである。

歪んだ分布を持つデータを想定し、新たなモード回帰モデルの開発を行う。そのために、これまでのモード回帰モデルの定式化を見直し、誤差項に対してツーピース正規分布 (two-piece normal distribution) を仮定したモード回帰モデルの構築を行う。このツーピース正規分布は、正と負の領域で異なるパラメータを持つ正規分布である。

で構築したモード回帰モデルを基礎とし、ディリクレ過程事前分布を利用した混合モデルを考えることによって、誤差項に対して確率分布を仮定しないセミパラメトリックなモード回帰モデルへと拡張を行う。また、モードの関数形を特定化しないモード回帰モデルの構築も検討する。

打ち切りのある被説明変数に対し、潜在変数を導入することによって推定が複雑にならないモード回帰モデルの開発を行う。また、変量効果さらには説明変数にラグ付き変数を導入することにより、パネルデータに対するモード回帰モデルの開発を行う。

(4) 本研究で提案する計量モデルに対して、シミュレーションによる推定方法を開発する。具体的には、マルコフ連鎖モンテカルロ法の適用を考える。また、より効率的な推定を行うため、モデルの再定式化や崩壊型ギブスサンプリング法の適用可能性についても検討を行う。

(5) 本研究で提案する計量モデルやその推定方法を数値実験や実データへの応用を通して検証し、問題が見つければ適宜改善を行う。

4. 研究成果

本研究の主な研究成果は、以下のとおりである。

(1) モード回帰モデルをベイズ法によって推定する場合、これまでの研究では、誤差項に対して一様分布、あるいは三角分布が仮定されてきた。一様分布を仮定したモード回帰モデルでは、誤差項の平均とモードが一致するため、歪んだ分布を持つデータを分析するには必ずしも適切であるとは言えない。一方、誤差項に三角分布を仮定した場合には、誤差項の分布が右に歪んでいるのか、あるいは左に歪んでいるのかを事前に指定しなければならず、実際のデータに適用するには困難が伴うことがある。またモデルの推定に関しても、効率性に関して問題があることが知られている。

既存研究におけるこうした問題を解決するため、本研究では、一様分布や三角分布を用いた既存モデルの定式化を見直すことから始めた。すなわち、「3. 研究の方法」で述べたように、誤差項の確率分布としてツーピース正規分布を用いることによって、新たなモード回帰モデルの構築を行った。ツーピース正規分布を用いた理由として、モードの位置を自由に設定できること、確率分布の形状がモードの左右で異なること、の2点を挙げることができ、いずれもモード回帰モデルを構築する上で重要かつ必要な性質である。また、一様分布や三角分布に比べて、確率分布の形状が柔軟であること、通常の正規分布を特別な場合として含むこともツーピース正規分布を採用した理由となっている。新たなモード回帰モデルの開発と併せて、ツーピース正規分布の平均値や分散などの理論的性質についても導出した。

(2) 本研究で提案するツーピース正規分布を仮定したモード回帰モデルに対し、シミュレーション法による推定方法、より具体的にはマルコフ連鎖モンテカルロ法を用いた推定方法の開発を行った。特に回帰係数に対しては、2つのアルゴリズムによるサンプリング方法を提案している。1つ目のアルゴリズムでは、ツーピース正規分布が一様分布とガンマ分布による混合表現が可能であることを利用して、回帰係数をギブス・サンプリングによって逐次的にサンプリングしている。一方、2つ目のアルゴリズムは、すべての回帰係数を一度にサンプリングする方法であり、本研究では最適化を伴う Acceptance-Rejection Metropolis-Hastings (ARMH) アルゴリズムによってサンプリングすることを提案している (以下では、前者を第1アルゴリズム、後者を第2アルゴリズムとよぶことにする)。いずれの方法も、実装が簡便なアルゴリズムであり、本研究で提案する推定方法は、実用上非常に有用であると考えられる。

本研究で提案する推定方法の有効性を調べるために多くの数値実験を行い、既存の方法よりも効率的であることを確認することができた。また、回帰係数に対する2つのアルゴリズムにつ

いても比較を行い、第1アルゴリズムの計算速度は速いが、サンプリングに関しては極めて非効率である、第2アルゴリズムの計算速度は第1アルゴリズムに劣るものの、サンプリングに関しては第1アルゴリズムよりも効率的であることを明らかにした。また、計算速度と効率性の両者を加味した場合、説明変数の数が少ないとき(例えば5以下のとき)には第1アルゴリズムを、説明変数の数が多いときには第2アルゴリズムを利用することが望ましいという結論を得た。

(3) (1)の研究成果を基礎として、誤差項に特定の確率分布を仮定しないセミパラメトリックなモデルへの拡張を行った。具体的には、ディリクレ過程事前分布を用いた混合モデルの構築を行った。ここで提案するモード回帰モデルの特徴として、尺度パラメータだけでなく形状パラメータについても階層構造を持っていることが挙げられる。このことにより、既存のモード回帰モデルや(1)で提案するモデルよりも柔軟なモデルとなっている。さらに、(2)で示したシミュレーションによる推定方法を改良することにより、計算負荷をあまり増やすことなくモデルが推定できることも示した。本研究で提案するセミパラメトリックモデルと推定方法の有用性を調べるため、ここでも数値実験や実データを用いた分析を行い、これまでのモデルよりも予測精度が良いこと、またサンプリング方法について実用上問題ないことを確認した。

(4) これまでのモード回帰モデルに関する研究の多くは、モードの構造に対して線形性を仮定してきた。しかし、データによっては説明変数に対して非線形であることも考えられる。そこで、モードを説明変数の関数として捉え、これに対して特定の関数形を仮定しないモード回帰モデルの開発を行った(誤差項の分布については、(1)と同じくツーピース正規分布を仮定している)。また、提案する計量モデルをより効率的に推定するためのマルコフ連鎖モンテカルロ法も開発した。

本研究では、未知のモード関数をスプライン基底関数によって表現したモード回帰モデルを構築した。その際、罰則付きスプラインの方法を適用することによって、過剰適合の問題を回避している。またモデルの推定に関しては、(2)で示した推定方法をそのまま適用できることを明らかにした。さらに数値実験を通して、本研究で提案する計量モデル・推定方法がうまく機能することを明らかにした。

(5) これまでに述べた研究は、クロスセクションデータに関するものであった。近年の計量分析ではパネルデータもよく用いられていることから、本研究ではパネルデータに対するモード回帰モデルについても研究を行った。ここでは、既存の個体効果をとまなうパネルデータモデルを参考にし、ラグ付き被説明変数を導入することで動学モデルへと拡張を行い、さらに打ち切りデータに対するモード回帰モデルの開発も併せて行った。

本研究の特徴は、動的パネルデータモデルにおける初期値問題に対応するため、被説明変数の初期値と分析期間中の説明変数に依存する形で個体効果を導入したこと、打ち切りデータを扱うため、潜在変数を用いてモデルを定式化し直している点である。ここで提案する計量モデルは、潜在変数を多く含むため最尤法などで推定することは不可能である。しかしながら、(2)の推定方法を修正することにより、モデルが複雑であるにも関わらず、比較的簡単なマルコフ連鎖モンテカルロ法によって推定することが可能となっている。

本研究のパネルデータに対するモード回帰モデルの有用性を示すために、生産効率性に関するパネルデータを用いて実証研究を行った。その結果、静的モデルよりも動的モデルの方がモデルの説明力が改善されることが明らかとなった。

(6) 回帰分析の方法の一つに、expectile回帰とよばれる方法がある。これは、通常回帰分析と分位点回帰を合わせたような方法であり、現在関心を集めている方法の一つである。本研究で提案するツーピース正規分布にもとづくモード回帰モデルが、このexpectile回帰を特別な場合として含んでいることが明らかとなった。すなわち、モード回帰モデルでは形状パラメータを推定する必要があるが、このパラメータを固定するとexpectile回帰となることがわかった。この関係性の発見は、本研究の副次的成果であると考えられる。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計0件

〔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
---------	---------