

令和元年6月1日現在

機関番号：34504

研究種目：若手研究(B)

研究期間：2015～2018

課題番号：15K17041

研究課題名(和文)複数選択及び重複選択を考慮した離散選択モデルへのベイズアプローチ

研究課題名(英文) Bayesian approach to the discrete choice model with multiple selection of alternatives

研究代表者

宮脇 幸治 (MIYAWAKI, Koji)

関西学院大学・経済学部・准教授

研究者番号：40550249

交付決定額(研究期間全体)：(直接経費) 2,400,000円

研究成果の概要(和文)：複数選択が可能な状況における離散選択モデルに関する研究を行なった。経済理論に基づく素朴な統計的モデルの推測は、計算量が膨大になるため事実上不可能である。そのため、本研究では経済理論に基づく部分を出来るだけ残しつつ、計算可能となるような統計的モデルの構築と推定手法の開発を行なった。得られたモデルのパラメータの推測に関して識別問題があることも分かったため、いくつかの推定手法を構築し比較を行なった。その結果、より計算量が少ない推定手法が良いのではないかと結論に至った。得られた統計的モデル及び推定手法を実データを用いて分析を行なった。

研究成果の学術的意義や社会的意義

個人の行動を分析する場合、離散選択問題はしばしば観察される。また複数選択が可能である状況もある。そのような行動を分析するにあたって、経済理論に基づいた統計的な分析を行うことは限定的な状況でしかなされてこなかった。本研究では、分析枠組みをより一般的な状況に拡張したとともに、推定可能な統計的モデルの構築及びその推定手法の開発を行うことができたと考えている。また推定上の課題がいくつかあるが、それらに対してもより現実的な解を提案できたと考えている。応用例としては、自動車の選択行動の分析が考えられ、本研究の手法を用いれば、例えばガソリン税の効果を分析することが可能であると考えている。

研究成果の概要(英文)：This project focused on the discrete choice model where the multiple selection of alternatives is allowed. The inference based on the simple statistical model directly derived from the economic theory is virtually infeasible because of its computational burden. To address this issue, the project generalizes the simple one preserving the relationship with the economic theory, and develops the estimation method for this generalized statistical model. Because the model cannot identify part of model parameters, several solutions are considered. Comparison of them suggests the estimation method with relatively lower computational burden is favorable. The statistical model and its estimation method are applied to the real dataset.

研究分野：計量経済学

キーワード：経済統計

1. 研究開始当初の背景

離散選択モデルとは、個人の選択行動を統計的に分析するためのツールであり、経済理論で仮定される効用最大化とも関連がある。例えば喫煙行動の分析では、喫煙するかしないかの選択行動があり、どのような要因が喫煙行動に影響を与えるかを分析する。また銀行の融資行動、女性の労働市場への参加行動なども同じような枠組みで分析されている。多くの場合、いくつかの選択肢の中から一つを選ぶということが仮定されているが、同じ選択肢を複数選ぶような状況もしばしば観察される。

アメリカの自動車保有はそのような状況の一つの例である。データから無視できない割合の家計が複数台の自動車を保有していることが知られている。また旅行プランの選択も同じように考えることができる。選択肢を旅行先とした場合、どこに何日滞在するかということは複数選択の問題と見ることができる。これらの例から分かる通り、複数選択の問題は現実には観察されるものである。

従来の研究では、追加的な仮定をおいて推定を行う分析、もしくは選択できる集合を制限し推定を行う分析などが行われてきた。これらの特定化を自動車の保有問題において簡単に述べると、短い期間を考えれば従来の一つの選択肢を選ぶ状況であると仮定して分析を行うこと、もしくは自動車を事前に主要ないくつかのカテゴリに分けその上での選択を考えた分析などが行われてきた。

しかしこれらの特定化は、必要な結論が得られないもしくは得られた結論にバイアスがあるかもしれないといった問題がある。複数選択の問題を考える意味の一つとしては、選択肢間の代替・補完の有無を知ることが挙げられる。前者のような特定化ではそのような点を議論することができない。また自動車を事前にカテゴリ分けしてしまうことは、カテゴリ分けの誤りによるバイアスを引き起こすことが懸念される。特に自動車の保有問題で得られた結論はガソリン税などの政策上の議論に用いられることが多いため、そのような問題は異なる結論に至る可能性を孕んでいる。そのため、明示的に複数選択を考慮した分析を行う必要があると考えられる。

2. 研究の目的

上で述べたような問題意識の下、可能な限り一般的な状況を考え、代替・補完の推定を行うための統計的モデル及びその推定手法を開発することが本研究の目的となる。特に実データを用いた分析が可能である必要があるため、推定が可能であるような範囲での一般的な統計的モデルとする。

3. 研究の方法

一般的な状況での研究はほとんどなされていないため、まず簡単なケースにおける統計的モデルから始める。その後計算上の問題、統計的推測上の諸問題を克服しつつ、より現実的な状況下での統計的モデルを構築及びその推定手法の開発を行う。関連する推定・分析手法の研究も合わせて行う。

4. 研究成果

まず最も一般的な状況での統計的モデルの研究を行なった。この統計的モデルは経済理論より導出されているため、そのパラメータの経済学的な解釈が可能である。その中でも代替・補完に関するパラメータは特に着目すべき部分である。研究の結果、その計算量が膨大であり実用的ではないことが分かった。選択肢の数を大幅に制限すれば可能であるが、それは例えば自動車の保有問題などでは現実的な状況ではない。

そのため新しいパラメータを統計的モデルに導入し、計算量の大幅な削減を行なった。新しいパラメータは選択肢に関する潜在的なカテゴリを考えるものである。また極限として最も一般的な状況での統計的モデルを含むため、元のモデルの特徴（経済理論との関連）を引き継いでいると考えられる。結果、元のモデルの特徴を残しつつ、実用的な範囲に計算量を収めることが可能となった。

新しいパラメータを導入した結果、統計的モデルに識別性の問題が発生した。この場合、識別性の問題とはデータのみからはパラメータを一意に決めることができない問題を指す。そのため識別性の問題に関する研究を行なった。その結果、3つほどのアプローチを考えることができた。それぞれ、(1) 識別制約を課して推定を行うアプローチ、(2) 事後的に識別制約を課すアプローチ、(3) 識別されているパラメータ（不変量）に着目するアプローチである。

次に複数選択の問題において、それぞれのアプローチを試した。まず識別制約を課すアプローチであるが、完全に識別することが難しく一部の識別制約のみ可能であることが分かった。そして実データを用いた結果、計算時間が長く、あまりうまく機能していないと判断された。より具体的には、得られたパラメータの解釈が難しいことが分かった。また事後的に識別制約を課すアプローチも試したが、どのような制約が合理的かということが問題となり、制約によって得られる結論が変わることは望ましくないため、こちらもうまく機能していないと判断された。

最後のアプローチは、識別されないパラメータを識別されたもの（以下、不変量）に変換するアプローチである。モデルの識別性に関する先行研究で明らかにされていることは、識別されないパラメータの推定には計算時間がかかるが、識別されているもののそれは比較的短い計算時間で推定可能であるということである。本研究課題のモデルでも同じ現象が確認され、不変量の推定は比較的短い計算時間でできることが分かった。また、代替・補完のパラメータは識別されないパラメータの一つであるが、それも不変量を考えることで容易に統計的推測を行えることが分かった。

また推定のためのアルゴリズムの改良を行なった。統計的モデルは経済理論と関連があるため、最大化条件が課せられている。推定されるパラメータはこの条件を満たしながら推定される必要があるが、そのためにはソートアルゴリズムの利用が最も簡単な実装であった。その結果推定手法のアルゴリズムは単純となる一方で、計算時間の無視できない部分がこのソートアルゴリズムの計算で占められていることが見受けられたため、改良を行なった。結果ソートアルゴリズムを用いない効率的なアルゴリズムを開発することができた。これは一般的なソートアルゴリズムの改良ではなく、考えている問題特有の構造を生かした改良となっている。実データを用いた場合であるが、ソートアルゴリズムを用いた場合と用いなかった場合とで同じ結果が得られること及び約 40%程度の計算時間の短縮が確認された。

さらに実データを用いた実証分析を開始した。先に述べた通り、データはアメリカの自動車保有状況に関するものである。データは雑誌 American Economic Review の査読論文で用いられたものであり、インターネットを通じて公開されている。データは様々な家計の属性及び自動車保有状況が記録されている。すべてのデータを用いた推定は計算時間がかかるため、同論文での分析にならぬように、家計の属性によってグループ分けし、それぞれのグループ内での推定を行うこととした。そして、以下の二点について主に研究を行なった。それぞれ、(1) 事前分布の影響、(2) 変数選択の影響である。

まず事前分布の影響であるが、本研究ではベイズのアプローチを用いているため、その影響を確認する必要がある。いくつかの特定化を試したが、選択肢の数が多いためにも影響して、計算効率のためにはやや強めの事前分布を仮定する必要があることが分かった。ただし選択肢の数がそれほど多くない状況では、弱い事前分布でも構わないと考えられる。また変数選択についても、いくつかの特定化を試した。結果、より弱い仮定に基づく特定化が、着目すべき代替・補完のパラメータに関する議論をより容易にするという観点から優れていることが分かった。この特定化に付随して、パラメータ数を減らすことも可能となった。つまり同じような情報を推測することを担っているパラメータが複数ある場合、一方を落とすことによりモデルの説明力を落とさず、より正確な推定が可能となると考えられる。ただし弱い仮定に基づく本特定化の予測力は低いと考えられるため、いくつか変数を追加した、より正確なモデルの特定化を行うことが今後の課題である。

これらの結果を踏まえた上で実証分析の結果を取りまとめている最中であり、今後査読誌への投稿を予定している。

最後に、これらの問題に関して、離散選択モデルではない一般的な回帰モデルでの研究も合わせて行なった。より具体的には、事前分布の選択及び変数選択に関する研究を行なった。事前分布の選択については、モデルの一致性の観点から研究を行なった。また変数選択に関しては、決定理論の枠組みを利用した方法論の可能性を探った。得られた結果を用いて、実データ（オゾン、糖尿病、賃金）への応用を行い、合理的な結論が得られることを確認した。この研究に関しては、研究結果をまとめディスカッションペーパーとして公開し、査読誌への投稿を行なっている。また、今後の課題ではあるが、この研究結果を本研究課題に結びつけることも考えている。

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〕
出願状況(計 0 件)

名称：
発明者：
権利者：
種類：
番号：
出願年：
国内外の別：

取得状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
取得年：
国内外の別：

〔その他〕

ホームページ等

“Economic variable selection” (<https://arxiv.org/abs/1903.02136>)

6 . 研究組織

(1)研究分担者

研究分担者氏名：

ローマ字氏名：

所属研究機関名：

部局名：

職名：

研究者番号（8桁）：

(2)研究協力者

研究協力者氏名：

ローマ字氏名：

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、研究者個人に帰属されます。