

平成 31 年 4 月 24 日現在

機関番号：14501

研究種目：基盤研究(C)（一般）

研究期間：2015～2018

課題番号：15K03396

研究課題名（和文）罰則付き経験尤度推定量による高次元データ解析

研究課題名（英文）Penalized Empirical Likelihood Estimation for High-Dimensional Data Analysis

研究代表者

末石 直也（Sueishi, Naoya）

神戸大学・経済学研究科・教授

研究者番号：40596251

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）：候補となる変数の数が多いとき、罰則付き推定は有用な変数選択の方法である。罰則付き推定量を用いる際に応用上問題となるのは、正則化パラメータの選び方である。ところが、モーメント制約モデルの罰則付き推定量に関しては、正則化パラメータの選択方法に関する研究はこれまでほとんど行われてこなかった。そこで本研究では、罰則付き経験尤度推定量の正則化パラメータを選択するための、経験尤度情報量規準と呼ばれる新しい情報量規準を提案した。赤池情報量規準のアイデアを基に、カルバックライブラー情報量の漸近的に不偏な推定量として、情報量規準を導出した。

研究成果の学術的意義や社会的意義

近年、ビッグデータと呼ばれる大規模データの分析が、産業界においても脚光を浴びている。計量経済学でも徐々に高次元データの利用が進みつつあるものの、応用研究はそれほど多くないのが実情である。そのひとつの理由として、従来の正則化法の研究の多くが相関関係の発見を目的としているのに対し、経済学者は因果関係の発見に興味があるという点が考えられる。本研究は、回帰モデルのみならず、より広いクラスのモデルに対して正則化法の利用可能性を広げるものであり、因果効果の分析など経済学的見地から重要な問題を考察するうえで有用な手段となりうるものである。

研究成果の概要（英文）：Penalized estimation is a useful technique for variable selection when the number of potential variables is large. A crucial issue in penalized estimation is the selection of the regularization parameter. However, there has been little study on the selection method of the regularization parameter for the penalized estimation of moment restriction models. This study proposes a new information criterion, which we call the empirical likelihood information criterion, to select the regularization parameter of the penalized empirical likelihood estimator. On the basis of the idea of the Akaike information criterion, our information criterion is derived as an asymptotically unbiased estimator for the Kullback-Leibler information criterion.

研究分野：計量経済学

キーワード：罰則付き経験尤度推定量 情報量規準 変数選択 高次元データ

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

データ環境の向上により、高次元データと呼ばれる大量の変数を含むデータが、経済学の実証研究においても用いられるようになりつつある。高次元データには多くの情報が含まれる一方で、特定の分析には不要な変数も多く含まれており、これらの変数を適切に除去しなければ、正確な分析を行うことはできない。ところが、どの変数が有用であるかは、経済理論からは必ずしも特定することはできず、観測されたデータに基づいて、使用する変数を選択する必要が生じる。

統計モデルに含まれる変数をどのように選ぶかという変数選択の問題は、統計学では古くから考察されてきた研究課題である。古典的アプローチは、赤池情報量規準 (AIC; Akaike information criterion) などの情報量規準に基づいて、変数選択を行うものである。このような方法は現在でも広く用いられているが、変数の組み合わせの数は、変数の数とともに指数的に増大する。情報量規準を用いて変数選択を行うには、候補となっているすべて変数の組み合わせに対してモデルを推定する必要がある。そのため、高次元データでは計算量が膨大になり、実行不可能となってしまう。

Tibshirani (1996) は、線形回帰モデルの推定方法として、Lasso と呼ばれる L1 罰則項を用いた正則化法を提案した。Lasso には、いくつかの係数パラメータを 0 と推定するスパース性と呼ばれる性質がある。対応するパラメータが 0 と推定された変数はモデルに含まれないので、推定と変数選択を同時に行うことが可能である。そのため、説明変数の数が多いときには、計算量の観点から非常に有用な方法である。Tibshirani (1996) 以降、スパース正則化に関する研究は急速に進み、現在でも統計学で最も盛んに研究が行われている分野のひとつである。

計量経済学でも、正則化法の研究は行われている。特に計量経済学では、モーメント制約によってモデルが定式化されることが多いため、モーメント制約モデルの変数選択の方法として、罰則付き GMM 推定量や罰則付き経験尤度 (PEL; penalized empirical likelihood) 推定量などが提案されている。先行研究では、罰則の度合いを決める正則化パラメータがサンプルサイズの増加とともに一定のレートで 0 に収束するという条件のもとで、変数選択の一致性や、推定量の漸近正規性といった性質が示されている。ところが、実際に推定を行う際には、正則化パラメータの値はデータの分析者が決定しなければならない。推定量の有限標本特性は、正則化パラメータの選び方に大きく依存するため、適切な選択方法が求められている。しかしながら、罰則付き GMM 推定量や PEL 推定量では、正則化パラメータの実践的な選択方法はほとんど考察されていないのが現状である。

2. 研究の目的

先行研究によって示された PEL 推定量の漸近的性質は理論的に重要であるが、漸近理論は実際上どのように正則化パラメータを選べばよいかということに関しては指針を与えない。正則化パラメータの値を分析者が自ら決めなければならないということは、実証研究者が推定量を用いるうえでの大きな足かせとなる。そこで本研究では、PEL 推定量の正則化パラメータを選択するための情報量規準を提案する。さらに可能であれば、提案された情報量規準を用いた PEL 推定量が、ある種の基準において望ましい性質を満たすことを示す。

3. 研究の方法

罰則付き最尤推定量に関して、Konishi and Kitagawa (以下 KK) (1996) が、正則化パラメータを選択するための一般化情報量規準 (GIC; generalized information criterion) を提案している。本研究では、GIC のアイデアを拡張することにより、カルバックライブラー情報量の漸近的に不偏な推定量として、PEL 推定量のための情報量規準を導出する。ただし、本研究で考察する対象は、KK (1996) とは以下の 2 点で異なるため、独自の解決策が必要となる。第一に、KK (1996) ではパラメトリックモデルを考察しているのに対して、本研究で考察するモーメント制約モデルはセミパラメトリックモデルである。第二に、KK (1996) では原点で微分可能な罰則項を考察しており、L1 罰則項などの推定量にスパース性をもたらす罰則項は除外されている。これに対し、本研究の目的は変数選択であるので、原点で微分不可能な罰則項が興味の対象となる。これらの問題を解決することによって、新たな情報量規準を導出する。

4. 研究成果

主たる研究成果は、以下の 3 つである。

(1) PEL 推定量の漸近的性質の解明

PEL 推定量の漸近特性については、Leng and Tang (2012) と Chang et al. (2015) によって明らかにされたと考えられていた。しかし、論文を再検証するうちに、両論文において設定されている仮定に問題があることが明らかになった。そこでまず、比較的妥当と考えられる仮定の下で、SCAD 罰則付きの経験尤度推定量の漸近的性質を再検証した。

設定としては、Leng and Tang (2012) と同様に、モーメント制約の数 (r) とパラメータの数 (p) がサンプルサイズ (n) の増加とともに緩やかに増加しうるケースを考えた。得られた結果としては、先行研究よりも仮定を緩められただけでなく、先行研究で示されていた PEL 推定量の収束レートよりも早いレートである $\text{root-}n/p$ 一致性を満たすことが示された。また、真のデータ生

成過程がスパースであるときには、正則化パラメータに関する一定の仮定の下で、変数選択の一致性や、オラクル性と呼ばれる性質を満たすことが示された

本研究で明らかになったことは、PEL 推定量は罰則なしの通常の経験尤度推定量と同じ収束のレートを実現することができるということである。これは当然予想される結果ではあるのだが、Leng and Tang (2012)では証明されることなく仮定されていた。一方、Chang et al. (2015)は PEL 推定量の一致性は示しているが、 $\text{root-n}/r$ 一致性までしか示されていないうえに、論理的に満たされない仮定を置いていた。本研究で示された $\text{root-n}/p$ 一致性は、特に p が固定されている場合においては root-n 一致性をもたらし、Donald, Imbens and Newey (2003) などの先行研究と整合的な結果になっている。これらの成果は、経験尤度推定量の理論研究の分野に新たな知見をもたらすものである。

本研究成果をまとめた論文は、Econometrics 誌に掲載されている(雑誌論文 を参照)。

(2) 正則化パラメータ選択のための経験尤度情報量規準の提案

実証研究で正則化法を用いる際には、正則化パラメータの選び方が重要となる。しかし、得られた PEL 推定量の漸近理論からは正則化パラメータの値を決定することはできないため、実証研究で PEL 推定量を用いるのは困難である。そこで、KK (1996) で提案された GIC を拡張して、PEL 推定量の正則化パラメータを選択するための経験尤度情報量規準 (ELIC; empirical likelihood information criterion)を提案した。

「研究の方法」において述べたとおり、ELIC を導出するために2つの問題に対処する必要がある。第一の問題は、Chen (2007) などの結果を用いて、モーメント制約モデルを分布の集合として表現することで、パラメトリックモデルと同様にモーメント制約モデルにおいてもカルバックライブラー情報量を定義することができる。一方、第二の問題は、先行研究においては考察されていない問題である。この点については、ゼロと推定されるパラメータの推定量と、非ゼロと推定されるパラメータの推定量で、収束のスピードが異なることを利用することで解決される。ゼロと推定されるパラメータの推定量の分布は退化するため、ゼロパラメータが既知である場合とほぼ同様に扱うことが可能となる。非ゼロと推定されるパラメータの推定量は M 推定量の形式で表されるため、結果として、GIC と類似した形式をもつ情報量規準が導出された。

本研究は、Akaike(1973) 以来続く、カルバックライブラー情報量の不偏推定量として導かれる AIC タイプの情報量規準の一連の研究に、新たな知見をもたらすものである。従来の研究ではモデルはパラメトリックモデルに限定されていたが、本研究は分布の仮定を必要としないモーメント制約モデルにおいても情報量規準を導出できることを示しており、AIC タイプの情報量規準の適用範囲を大きく広げるものである。

また、本研究で提案された ELIC は、モーメント制約モデルの罰則付き推定量の正則化パラメータの選択方法としてははじめて、理論的な裏付けをもつものである。これまでも、モーメント制約モデルの AIC もしくは BIC タイプの情報量規準は提案されていたが、パラメトリックモデルの AIC や BIC の形式を模しただけで、理論的な根拠がないものであった。本研究で示された ELIC の導出方法は、Hong, Preston and Shum (2003) などの既存の結果に対しても理論的な裏付けを与えるものになっている。

本研究成果をまとめた論文は、Economics Letters 誌に掲載されている(雑誌論文 を参照)。

(3) 一般化線形モデルの罰則付き推定量と正則化パラメータの選択方法

において提案した ELIC について、ある種の損失関数の下での漸近的な最適性を示すことを試みたが、技術的な問題に直面し、期間内に成果を挙げることはできなかった。しかしながら、この課題に取り組む過程で、派生的な研究成果を得ることができた。

本研究では、一般化線形モデルの2段階の罰則付き推定量を考察し、その正則化パラメータの選択方法を示した。推定方法としては、第一段階で尤度関数に $L1$ 罰則項をつけて推定を行い、変数選択のみを行う。次に、第一段階で選択されたモデルを用いて、尤度関数に $L2$ 罰則項をつけて再び推定を行う。このように推定を2段階にわけると、特にノイズに比べてシグナルが強い場合には、 $L1$ 罰則付きの推定量のバイアスを減らすことができる。この推定を行うためには、 $L1$ 罰則と $L2$ 罰則の2つの罰則に対応する正則化パラメータを決定する必要がある。これに対しては、Mallows の C_p タイプの基準 (GC_p ; generalized C_p) を新たに導出した。さらに、 GC_p はカルバックライブラー損失の意味で、漸近的に最適な正則化パラメータを選択することが可能であることを証明した。

一般化線形モデルの罰則付き推定に関しては、Zhang et al. (2010) と Flynn et al. (2013) が AIC の漸近的な最適性を証明している。しかし、両論文の証明を検討するうちに、彼らの課しているモデルに関する仮定は非常に強く、実質的に変数選択を行う必要がないことを仮定していることが明らかになった。本研究は、妥当な仮定の下で正則化パラメータの選択方法の漸近的な最適性を示した最初の研究である。

本研究はシミュレーションなどに未完成的な点があるが、完成次第投稿予定である。

5 . 主な発表論文等

〔雑誌論文〕(計 2 件)

Ando, T. and N. Sueishi (2019): “On the Convergence Rate of the SCAD-Penalized Empirical Likelihood Estimator,” *Econometrics*, Vol.7, 15. 査読有

DOI: 10.3390/econometrics7010015

Ando, T. and N. Sueishi (2019): “Regularization Parameter Selection for Penalized Empirical Likelihood Estimator,” *Economics Letters*, Vol. 178, pp.1-4. 査読有

DOI: 10.1016/j.econlet.2019.02.011

〔学会発表〕(計 4 件)

末石直也, “Efficient Regularization Parameter Selection for Maximum Likelihood Post-Selection Estimation in Generalized Linear Models,” Summer Workshop on Economic Theory, 2018 年

末石直也, “Efficient Regularization Parameter Selection for Maximum Likelihood Post-Selection Estimation in Generalized Linear Models,” Workshop on Advances in Econometrics 2018, 2018 年

末石直也, “Penalized Empirical Likelihood Estimation and Model Selection for High-Dimensional Misspecified Moment Restriction Models,” Summer Workshop on Economic Theory, 2016 年

末石直也, “Penalized Empirical Likelihood Estimation and Model Selection for High-Dimensional Misspecified Moment Restriction Models,” The 2nd Annual International Conference on Applied Econometrics in Hawaii, 2016 年

〔その他〕

ホームページ等

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。