

平成 21 年 6 月 1 日現在

研究種目：若手研究(B)

研究期間：2006～2008

課題番号：18730146

研究課題名（和文） バグGINGを用いた推定の理論的・実証的分析

研究課題名（英文） Theoretical and empirical analysis of estimation using the bagging

研究代表者

難波 明生 (NAMBA AKIO)

神戸大学・大学院経済学研究科・准教授

研究者番号：60324901

研究成果の概要：経済モデルに含まれる変数をより高い精度で推定することができる方法として近年注目されているバグGINGと、バグGINGと深い関わりのある手法である、ブートストラップ法および縮小推定量を分析対象とし、これらの手法を用いることにより、既存の結果よりも精度の高い結果が得られる可能性があることを理論的分析・コンピュータを用いた数値計算やシミュレーション分析により示した。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2006年度	1,300,000	0	1,300,000
2007年度	800,000	0	800,000
2008年度	1,000,000	300,000	1,300,000
年度			
年度			
総計	3,100,000	300,000	3,400,000

研究分野：社会科学

科研費の分科・細目：経済学・経済統計学

キーワード：バグGING・ブートストラップ・リサンプリング・縮小推定量

1. 研究開始当初の背景

計量経済学においては、理論分析であるか実証分析であるかにかかわらず、線形回帰モデルが多く用いられている。線形回帰モデルにおいては、最小自乗推定量がすべての線形不偏推定量の中で最小の分散を持つことが良く知られている。

しかし、推定量のクラスを不偏でなく、かつ非線形のクラスまで拡張したとき、Stein (1956)および James and Stein (1961)は、いわゆる Stein 型推定量が最小自乗推定量を平均自乗誤差において優越することを証明した。平均自乗誤差において優越すると言うことは、その推定量を用いればより良い予測

が得られる可能性が高いと言うことを意味している。このことから、Stein (1956) および James and Stein (1961)の研究以降、多くの推定量が提案され、その性質についての研究が行われてきた。これらの推定量は最小自乗推定量を原点方向に縮小して得られるため、縮小推定量と呼ばれている。これらの研究結果が意味するように、縮小推定量を用いればよりよい予測を得られる可能性は高い。しかし、実証研究においては、縮小推定量はほとんど用いられてこなかった。その理由の一つは、縮小推定量の分布が非常に複雑であるため、実証研究で重要な検定、推定に用いることができないためである。

このような複雑な分布を持つ統計量の分布を近似する際に有効であるのが、Efron (1979) によって提案されたブートストラップ法である。Brownstone (1990), Kazimi and Brownstone (1999) 等によって示されているように、ブートストラップ法はいくつかの縮小推定量にも応用可能である。

ブートストラップ法はさまざまな問題に応用可能な方法ではあるが、Andrews (2000) 等いくつかの研究で示されているように、必ずしもすべての問題に応用できるわけではない。特に、推定量が予備検定を含むものである場合(このような推定量は予備検定推定量と呼ばれる)、ブートストラップ法による近似は無効である可能性が高い。縮小推定量の中には予備検定推定量であるものも数多く存在し、このような推定量にはブートストラップ法を行ってもよい近似は得られないと考えられる。しかし、Breiman (1996) により、近似が無効の場合でも、ブートストラップ法を行うことにより予測の精度が通常の縮小推定量よりもさらに改善される可能性があることが示された。この方法はブートストラップ・アグリゲーション(Bootstrap Aggregation)の頭文字を取ってバグギング(Bagging)と呼ばれている。

前述のように、縮小推定量についての理論的な研究は多く行われているが、実際のデータにはこれまでほとんど応用されてこなかった。しかし、近年になって、バグギングの有効性が分かって来るとともに、実際のデータを用いた予測にも応用されるようになり始めた。縮小推定量およびブートストラップ法、バグギングを実際のデータに応用した例としては、Stock and Watson (2004), Inoue and Kilian (2005) 等があり、今後もこのような応用例は益々増加していくであろうと予想される。

しかし、この手法の有効性にもかかわらず、実際の応用例は多くなく、日本の経済データに応用した例は現在のところ存在しない。さらに、バグギングを理論的に分析した研究は Bülmann and Yu (2002) による単純なモデルを除いてほとんど存在しない。

2. 研究の目的

本研究では、まずバグギングを様々なモデルの下で応用することを考え、理論的にその特性を分析する。また、バグギングにより得られる推定量と関わりの深い推定量として、いくつかの縮小推定量の特性を理論的に分析する。次に、得られた結果を実際のデータに応用し、その予測のパフォーマンスを評価する。バグギングを用いた予測は、既存の手法よりも良いパフォーマンスを持つことが期待される。

バグギングの特性の多くが未知であり、また、その有効性にもかかわらず、日本の経済データへの応用がまだ行われていないことから、以上のような研究は非常に大きな意味を持つものと考えられる。

3. 研究の方法

様々なモデルおよび推定量に対してバグギングを応用する方法を考える。その上で、可能なものに関しては、推定量の分布、モーメントの導出等の解析的な分析を行う。解析的な分析において数値的な比較が必要と判断されるものについては、Fortran または C 言語等のコンピュータ言語を用いた数値計算によって数値的な分析を行う。解析的な分析が完全に不可能であると判断された場合には、モンテカルロ法によるコンピューター・シミュレーションを用いて分析を行う。解析的な分析が不可能な複雑なモデルにでも、コンピューター・シミュレーションによる分析は、コンピュータの処理能力に依存するものの、かなりの程度で可能である。

4. 研究成果

まず、サーベイの結果より以下の2点が明らかになった。

(1) ブートストラップ法が大標本において漸近的に有効である(つまり、推定量の分布がブートストラップ法で近似できる)場合には、バグギングを行っても推定量の性質は漸近的には改善されないことが分かった。しかしながら、ブートストラップ法が漸近的に有効である場合でも、小標本においては推定量の平均自乗誤差が改善される可能性がある。

(2) ブートストラップ法が漸近的に有効でない場合でも、サブサンプリング等のリサンプリング法を用いれば推定量の分布を近似できる可能性がある。このようなケースにおけるバグギングの効果に関しては、いくつかの文献で研究が行われているものの、現在のところ未知である部分が多い。このような、サブサンプリングが有効な場合でも、バグギングを用いることで、小標本において推定量の平均自乗誤差が改善される可能性がある。

以上のサーベイの結果をふまえた上で以下の研究結果が得られた。

(3) Andrews (2000) で提案されたブートストラップ法による分布の近似が漸近的に有効でない予備検定推定量を例にとり、バグギングの効果を検証した。結果として、この推定量に対しても、バグギングを行う事で平均自乗誤差を改善できる可能性がある事が示された。

(4) バグギングを実際に応用できる可能性

のある例として、カーネル密度推定法にリサンプリング法を応用することを考えた。

カーネル密度推定法においては、バンド幅と呼ばれるパラメータの選択が非常に重要であることが知られている。バンド幅は、通常、推定量の平均自乗誤差が最小化されるような値を用いるが、このような値を用いた場合、カーネル密度推定量はバイアスを持つことが知られている。しかし、ブートストラップ法を用いてカーネル密度推定量の分布を近似することを考えると、通常のブートストラップ法ではカーネル密度推定量のバイアスを近似することができないため、ブートストラップ法は有効でない。このことから、カーネル密度推定法に関してはバグギングによる効果を得られる可能性があるように思える。しかし、実際は、ブートストラップ法の繰り返しによって得られるカーネル密度推定量の平均値（ブートストラップ推定量）がカーネル密度推定量そのものに一致してしまうという理由から、この問題においてはバグギングを用いても推定量の性質を改善することができないことが分かった。

(5) (4) の結果をふまえた上で、さらにリサンプリング法をカーネル密度推定量に応用することを考えた。カーネル密度推定法においては、バンド幅と呼ばれるパラメータの選択が非常に重要であることが知られている。バンド幅の選択方法は、いくつかの方法があるが、多くの文献においてクロス・バリデーションによる選択が推奨されている。クロス・バリデーションで選択されたバンド幅は、カーネル密度推定量の平均自乗誤差を最小化する値に確率収束することが知られているが、この場合、信頼区間は推定量のバイアスを受けることになる。したがって、推定量のバイアスが信頼区間にどのような影響を及ぼすか調べるため、シミュレーションによる分析を行った。シミュレーションの結果から、①バイアスを理論的に取り除けば、ブートストラップ法による近似はかなり良好であること、②バイアス修正を行ったとしても、バイアス項の推定精度の低さのため、信頼区間の精度はかなり低くなってしまうこと、③バイアスを無視した信頼区間の精度はそれほど低くないことが分かった。以上の結果から、バンド幅の選択にクロス・バリデーションを用いた場合、バイアス項を無視するのは現実的な方法であると考えられた。

しかしながら、バイアス項を無視する方法は、バイアスが非常に大きいときは有効でないことが予想されるので、更に正確な結果が期待できる方法（例えば、スムーズド・ブートストラップを応用する方法）等を検討する必要がある。

さらに、バグギングによる推定量と深い関わ

りのある縮小推定量について以下の分析を行った。

(6) リッジ推定量(Huang, 1999)に予備検定を取り入れることを考え、LINEX 損失関数の下でその小標本特性を分析した。コンピュータによる数値計算を用いた分析結果から、この推定量を用いれば、LINEX 損失関数を評価基準として用いた場合でも、多くの場合に通常の最小2乗推定量よりも精度の高い推定量が得られることが示された。また、数値計算の結果は、予備検定を取り入れることにより、推定量の精度を高められる可能性が高いことを示している。

(7) 必要な回帰係数が誤って省かれてしまった回帰モデルを考え、この場合の Stein 型推定量の小標本特性を分析した。コンピュータによる数値計算の結果から、評価基準として Pitman nearness と呼ばれる基準を用い、さらに攪乱項が多変量 t 分布に従っていた場合でも、定式化の誤りの程度が低ければ、最小2乗推定量よりも Stein 型推定量の方が精度の高い推定量が得られることが示された。

以上の結果のほとんどは、理論分析の結果得られた物であり、実際のデータにおいてどういった結果が得られるかは未知の部分が多い。実際のデータに対する応用と、その結果の分析については、今後更に研究していきたい。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 4 件)

①Akio Namba and Kazuhiro Ohtani, Risk Performance of a Pre-test Ridge Regression Estimator under the LINEX Loss Function when Each Individual Regression Coefficient is Estimated, Journal of Statistical Computation and Simulation, 2009, 近刊, 査読有

②難波 明生, バンド幅をクロス・バリデーションで選んだ場合のカーネル密度推定による信頼区間に関するシミュレーション分析, 2008, 国民経済雑誌, 第 197 巻, 2008, pp73-86, 査読無

③Akio Namba and Kazuhiro Ohtani, Risk comparison of the Stein-rule estimator in a linear regression model with omitted relevant regressors and multivariate- t errors under the Pitman nearness criterion, Statistical Papers, 48, 2007, pp151-162, 査読有

④難波 明生，単純な予備検定推定量に対するバックギングの効果について，国民経済雑誌，第 193 巻，2006，pp85-95，査読無

6. 研究組織

(1) 研究代表者

難波 明生 (NAMBA AKIO)

神戸大学・大学院経済学研究科・准教授

研究者番号：60324901

(2) 研究分担者

(3) 連携研究者