

機関番号：12601

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23530249

研究課題名(和文)有限混合モデルの統計的推測理論の開発

研究課題名(英文)Development of inference methods for finite mixture models

研究代表者

下津 克己(Shimotsu, Katsumi)

東京大学・経済学研究科(研究院)・教授

研究者番号：50547510

交付決定額(研究期間全体)：(直接経費) 3,800,000円、(間接経費) 1,140,000円

研究成果の概要(和文)：本研究は、有限混合モデルにおける統計的推測の理論を発展させた。有限混合モデルの分析においては、モデルを構成する要素の個数をデータから決定することが、特に重要である。しかしながら、要素の個数に関する統計的推測は、有限混合モデルの特殊な構造のため、非常に困難であることが知られている。

本研究は、有限混合モデルの尤度関数を分析する新しいアプローチを構築し、それに基づき、有限混合モデルの要素の個数に関する、幅広いモデルに適用可能である新しい統計的推測の手法を構築した。さらに、コンピュータ・シミュレーションによって、本研究によって提唱された手法が、実用上有用であることが確認された。

研究成果の概要(英文)：This research project developed a new method for inference for finite mixture models. Finite mixture models are very popular in applications, but their inference is known to be difficult. Specifically, when one tests for the number of components in a mixture model, the likelihood ratio test statistic does not have the standard chi-square asymptotic distribution. We developed a new approach to analyzing the likelihood function of finite mixture models, and constructed inference procedures that are applicable to a wide class of models and have an easily computable asymptotic distribution. Further, we analyzed the finite sample performance of the proposed procedures by computer simulations.

研究分野：経済学

科研費の分科・細目：経済統計学

キーワード：有限混合モデル 計量経済学 統計的推測 漸近理論

### 1. 研究開始当初の背景

有限混合モデルは、観測されない多様性を柔軟にモデルにすることが可能なため、経済学、生物学、医学などの多様な分野の応用において非常に幅広く使われている。経済学では、有限混合モデルは、主に、観測されない異質性・個人効果をモデルするために使用される。有限混合モデルにおいては、モデルを構成する要素の個数は、観測されない異質性を持つ集団の数を決定するため、モデルの特定化において重要な要素である。さらに、要素の個数を誤って設定した場合には、モデルの重要なパラメータの推定値が、経済学的に無意味な値をとり得ることが知られている(Heckman and Singer, 1984)。

実際の応用では、経済理論が要素数に関する情報を提供することは稀であるため、要素数はデータとモデルを比較することにより決定されなければならない。したがって、要素数を決定する実用的な統計的手法を開発することは重要な課題である。

多くの統計的仮説検定においては、尤度比検定が有効な手法である。しかしながら、有限混合モデルは数学的に特殊な構造を持つため、有限混合モデルの要素数を決定するために尤度比検定を用いた場合には、尤度比検定等計量の漸近分布は通常のカイ2乗分布に従わない。尤度比検定統計量の漸近分布は、多次元ガウス過程の汎関数となり、その分布をコンピュータ・シミュレーションで求めることは非常に困難であるため、尤度比検定の漸近分布に基づいた検定を実際に行うのはほとんど不可能である。実際の応用では、多くの研究者が、尤度比検定統計量の値を報告するが、その値を何の分布と比較してよいか分からないため、統計的に正しい推測ができない状況にある。したがって、有限混合モデルの要素数を検定する実用的な手法を開発することは、目下の重要な課題となっている。

### 2. 研究の目的

本研究の目的は、有限混合モデルの要素数に関する統計的推測の新しい手法を開発すること、ならびにそのパフォーマンスをシミュレーションで検証することである。

### 3. 研究の方法

有限混合モデルの要素の個数に関する統計的推測、すなわち要素の数が  $m$  個である帰無仮説を  $m+1$  個である対立仮説に対して検定することが困難であるのは、以下の理由による。要素の数が  $m+1$  個のモデルが、「要素の数が  $m$  個である」という帰無仮説を満たすのは、(1) 1つの要素がゼロの実現確率を持つ場合、(2) 2つの要素のパラメータの値が等しい場合、の2つの場合である。両方の場合において、通常尤度比検定の正規条件は満たされない。ケース(1)においては、ゼロの実現確率をもつ要素のパラメータは任意の値をとり得るため、識別されない。すなわち、

帰無仮説の下でパラメータが一意的に識別されない。さらに、ケース(1)を検定する尤度比検定統計量の漸近分布は、多次元ガウス過程の汎関数となる。ケース(2)においては、パラメータの値が等しい二つの要素は任意の実現確率を持つため、これらの要素の実現確率は識別されない。よって、帰無仮説の下でパラメータが一意的に識別されない。このため、尤度関数に対する通常分析は適用不可能となる。

本研究の鍵となる発見は、ケース(2)を分析するためのパラメータの変換である。ケース(2)では、帰無仮説の下でパラメータが一意的に識別されないため、フィッシャー情報行列が特異となる。このため、従来の研究は、複雑な分析を用いて、パラメータがスカラーの場合のみに限定し、尤度比検定統計量の漸近分布を導出していた。従来の手法は、非常に複雑であるため、パラメータがベクトルの場合には応用不可能であった。本研究は、あるパラメータの変換を行うと、特定のパラメータに関してのスコアがゼロとなるため、フィッシャー情報行列が特異となる方向を特定できること、さらに、そのようなパラメータに関しては、対数尤度関数の2次の導関数がスコアの役割を果たすことを発見した。この発見により、有限混合モデルでは、対数尤度関数に4次のテーラー展開を適用することで、対数尤度関数が漸近的にパラメータの多項式の2次関数として近似できる。

以上の発見に加え、各要素の実現確率が0ならびに1の値をとり得ない場合には、ケース(2)のみを分析すればよいことに着目すると、各要素の実現確率が0ならびに1の値をとらない仮定の下で、要素の数が  $m$  個である帰無仮説を  $m+1$  個である対立仮説に対して検定する尤度比検定統計量の漸近分布の導出が可能となる。

実際の検定においては、Li et al. (2009) が提唱した EM test が、実行の簡単さ、漸近分布の簡単さ、仮定の弱さ等の点で、尤度比検定統計量よりも優れていることを示している。本研究では、Li et al. (2009) の EM approach を拡張し、要素の数が  $m$  個である帰無仮説を  $m+1$  個である対立仮説に対して検定する場合の modified EM test を構築する。

### 4. 研究成果

本研究は、以下の3つの主要な研究成果をあげた。(1)有限混合モデルの対数尤度関数をパラメータの多項式の2次関数によって漸近的に近似する理論の構築。(2)有限混合モデルの要素数に関する新しい検定統計量である modified EM test の構築とその漸近分布の導出。(3) modified EM test の有限標本下でのパフォーマンスのシミュレーションによる検証。

(1) 有限混合モデルの対数尤度関数の新しい漸近近似理論の構築

本研究は、有限混合モデルの対数尤度関数をパラメータの多項式の2次関数によって漸近的に近似する、新しい理論を構築した。有限混合モデルの対数尤度関数の漸近近似は、近年まで未解決の問題であった。本研究は、あるパラメータの変換を行うと、特定のパラメータに関してのスコアがゼロとなるため、フィッシャー情報行列が特異となる方向を特定できること、さらに、そのようなパラメータに関しては、対数尤度関数の2次の導関数がスコアの役割を果たすことを発見した。この発見により、有限混合モデルでは、対数尤度関数に4次のテーラー展開を適用することが適切であることが判明し、その結果、対数尤度関数が漸近的にパラメータの多項式の2次関数として近似できることが示された。

(2) 有限混合モデルの要素数に関する新しい検定統計量である modified EM test の構築とその漸近分布の導出

前述した、有限混合モデルの対数尤度関数のパラメータの多項式の2次関数による漸近近似に基づき、さらに、Li et al. (2009) によって提唱された EM test を拡張することにより、modified EM test を構築した。modified EM test は、Li et al. (2009) の EM test と比較すると、パラメータがベクトルである場合も使用可能であり、罰則項を必要としないという利点を持つ。多くの実証研究において、パラメータはベクトルであり、実際の応用での罰則項の選択は難しい問題であるため、modified EM test のこの利点は実際上重要である。

modified EM test の漸近分布は、 $m$  個の確率変数の最大値で表され、さらに各々の確率変数は、正規分布に従う確率変数を多項式からなる錐に投影することにより得られることが示された。この漸近分布は、既存研究で示された、パラメータがスカラーである場合の尤度比検定統計量の漸近分布を一般化する点も、特筆すべき事実である。

(3) modified EM test の有限標本下でのパフォーマンスのシミュレーションによる検証

実証研究で多用される、Weibull duration mixture model を用いて、広範なコンピュータ・シミュレーションを行い、modified EM test のパフォーマンスを検証した。その結果、

帰無仮説が正しい場合は、modified EM test の漸近分布は、その有限標本下での分布の良い近似となっていること、対立仮説が正しい場合は、modified EM test は、帰無仮説を棄却する能力があること、modified EM test を行うのに必要な計算時間は過大ではなく、modified EM test は十分実用に供することのできるものであることが確認された。

参考文献

Heckman, J. and Singer, B. (1984). "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica*, 52, 271-320.

Li, P., Chen, J., and Marriott, P. (2009). "Non-finite Fisher Information and Homogeneity: An EM Approach." *Biometrika*, 96, 411-426.

## 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6件)

Hiroyuki Kasahara, Tatsuyoshi Okimoto, Katsumi Shimotsu. Modified Quasi-Likelihood Ratio Test for Regime Switching. *Japanese Economic Review*, 65. 査読有. 2014, 25-41.

Hiroyuki Kasahara, Michio Suzuki, Katsumi Shimotsu. Does an R&D Tax Credit Affect R&D Expenditure? The Japanese Tax Credit Reform in 2003. *Journal of the Japanese and International Economies*, 31. 査読有. 2014, 72-97.

Hiroyuki Kasahara, Katsumi Shimotsu. Nonparametric Identification and Estimation of the Number of Components in Multivariate Mixtures. *Journal of the Royal Statistical Society: Series B*, 76. 査読有. 2014, 97-111.

Hiroyuki Kasahara, Katsumi Shimotsu. Testing the Number of Components in Finite Mixture Models. CIRJE Discussion Paper, University of Tokyo, F-867. 査読無. 2012, 1-37.

Hiroyuki Kasahara, Katsumi Shimotsu. Sequential Estimation of Structural Models with a Fixed Point Constraint. *Econometrica*, 80. 査読有. 2012, 2303-2319.

Katsumi Shimotsu. Exact Local Whittle Estimation of Fractionally Cointegrated Systems. *Journal of Econometrics*, 169. 査読有. 2012, 266-278.

〔学会発表〕(計7件)

下津克己、Recent Advances in Statistical Theory and Applications for High Dimensional Data Analysis and Related Topics, Testing the Number of Components

in Finite Mixture Models, 2013 年 9 月 6 日、  
小樽商科大学

下津克己、2013 International Conference  
in Financial Econometrics、Testing the  
Number of Components in Finite Mixture  
Models、2013 年 7 月 7 日、山東大学(中国)

下津克己、日本経済学会 2013 年度春季大  
会、Modified Quasi-Likelihood Ratio Test  
for Regime Switching、2013 年 6 月 23 日、  
富山大学

下津克己、Conference in Econometrics:  
“ Partial Identification, Weak  
Identification, and Related Econometric  
Problems ”、Testing the Number of  
Components in Finite Mixture Models、  
2013 年 6 月 6 日、Yale University (米国)

下津克己、29th Annual Meeting of  
Canadian Econometric Study Group、  
Testing the Number of Components in  
Finite Mixture Models、2012 年 10 月 27 日、  
キングストン(カナダ)

下津克己、京都大学経済研究所研究会、  
Testing the Number of Components in  
Finite Mixture Models、2012 年 8 月 29 日、  
京都大学

下津克己、28th Annual Meeting of  
Canadian Econometric Study Group、Local  
Quadratic Approximation in Finite  
Mixture Models and Testing the Number of  
Components、2011 年 10 月 22 日、トロント

## 6 . 研究組織

### (1)研究代表者

下津 克己 (SHIMOTSU KATSUMI)  
東京大学・大学院経済学研究科・教授  
研究者番号：50547510  
研究者番号：