

海外特別研究員最終報告書

独立行政法人 日本学術振興会 理事長 殿

採用年度 平成 29 年度

受付番号 60

氏 名 植松 良公

(氏名は必ず自署すること)

海外特別研究員としての派遣期間を終了しましたので、下記のとおり報告いたします。

なお、下記及び別紙記載の内容については相違ありません。

記

1. 用務地（派遣先国名）用務地： ロサンゼルス （国名： アメリカ合衆国 ）

2. 研究課題名（和文）※研究課題名は申請時のものと違わないように記載すること。

高次元マクロ計量経済学：理論と応用

3. 派遣期間：平成 29 年 4 月 1 日 ～ 平成 30 年 8 月 10 日

4. 受入機関名及び部局名

南カリフォルニア大学マーシャル経営大学院データサイエンス&オペレーション学科

5. 所期の目的の遂行状況及び成果…書式任意 **書式任意（A4 判相当 3 ページ以上、英語で記入可）**

(研究・調査実施状況及びその成果の発表・関係学会への参加状況等)

(注)「6. 研究発表」以降については様式 10-別紙 1~4 に記入の上、併せて提出すること。

米国派遣に際しての所期の目的は、大規模マクロ経済データが利用可能な場合の、有効な統計学・計量経済学的方法論の提案とその理論解析、そして実データへの応用であった。この目的に照らし、派遣期間中（2017 年 4 月から 2018 年 8 月まで）の研究成果として以下の 2 つが挙げられる。

1. ファクター構造に基づく高次元変数選択と偽発見率 (FDR) コントロール手法の提案

ある変数 Y について、それを説明しうる数多くの変数の集合 $X=(X_1, \dots, X_p)$ があるとする。ここでの問題は、いかにして Y にとって真に重要な変数の部分集合を選ぶかである。そのための統計学的な指標となるのが、FDR と検出力 (power) である。FDR とは、 Y を説明するとして選ばれた変数のうち、真には不要である変数の数の比の期待値で定義される。これは「第一種の誤り」を図る指標の一つとして知られ、低い方が望ましい。一方の検出力とは、 Y を説明するとして選ばれた変数のうち、真に必要な変数の数の比の期待値で定義される。これは「第二種の誤り」を図る指標であり、高い方が望ましい。一般に両者にはトレードオフの関係があるため、FDR をゼロに抑えつつ 100% の検出力が得られるような手法は存在しない。可能な限り正確に FDR を低い水準（例えば 10% や 20%）にコントロールしつつも、より高い検出力を保持するような変数選択の方法が望まれる。

これまで用いられてきた FDR コントロールの手法としておそらく最もポピュラーなものは、Benjamini and Hochberg (1995) によって提案されたものである。これは p -値のランキングに基づくもので、 Y を説明しうる変数の集合の次元 p が小さいとき、良好な結果を示す。しかし p が高次元のとき、その挙動は非常に不安定になることが知られている。その理由は、一般に高次元下での誤差分散の推定が不安定なことに起因する。その結果、 p -値の計算精度にも問題が生じ、Benjamini and Hochberg (1995) タイプの FDR コントロールはほぼ不可能となる。こうした問題点を背景に、Barber and Candès (2015) は、 p -値の計算を必要としない全く新しいタイプの FDR コントロール手法を提案した。これは Y とは無関係でかつ変数集合 X を模倣する Knockoffs と呼ばれる人工的な変数組を構成することで、 X との比較を通して Y にとって重要な X の部分集合を選択する。（具体的に

は、例えば、 Y を X とその Knockoffs で作られる拡大行列に Lasso 回帰し、その係数を比較することによって X の部分集合を選択する。) これはその後、Barber and Candès (2017) や Candès et al. (2018) で拡張された。特に後者によって最近提案されたフレームワークは Model-X Knockoffs と呼ばれ、Knockoffs による総計的推測の理論的な基礎を与える。しかし Model-X Knockoffs は X と Y の真の分布に関する完全な情報を必要とするため、実データに対して直接用いることは困難であることが多い。経済データを想定する場合、その傾向は顕著である。一般的な統計的推測の手順は、何らかの統計モデルを想定した上で、実データからそれを推定することから始める。このようにモデル推定が必要とされる場合、Model-X Knockoffs に基づいた変数選択は、何ら FDR コントロールを保証しない点で注意が必要となる。

我々の研究で想定するモデルは、(1) Y は X の部分集合の線形回帰モデルで記述され、(2) X は観測不可能な潜在ファクターによって生成される、と仮定する。ここで(1)の回帰係数と、(2)の確率的ファクター、確定的ファクター・ローディング、そしてファクター・モデルの誤差項の分布の持つパラメータは未知とする。また、ファクター数は有限である一方、 X の次元 p と観測数 n はともに発散すると仮定する。これは計量経済学において非常に自然なモデリングであり、高次元データ環境の下での様々な経済現象を記述しうる。いま、(2)のパラメータ、すなわちファクターの実現値、真のファクター・ローディング、そして誤差項のパラメータがすべて既知であるとする。この場合は、Knockoffs をこのコモン・コンポーネントに誤差項の独立コピーを足す形で定義することによって、Model-X Knockoffs のフレームワークを直接適応することができる。ここではこの理想的な Knockoffs を Oracle knockoffs と呼ぶことにする。もちろん現実にはこれらの4パラメータは未知であるため、Oracle knockoffs は利用不可能である。一方でこれに代えて、各パラメータを実データから推定した上で、つまりコモン・コンポーネントの推定量と推定された誤差分布からのランダム・サンプリングを用いて、実行可能な Knockoffs を構成することを考える。ここではこれを Empirical Knockoffs と呼ぶ。最大の問題点は、この Empirical knockoffs が Model-X Knockoffs framework の規定する Knockoffs の定義を満たさないことである。直観的には、各推定量に一致性があれば、漸近的には Empirical knockoffs は Oracle knockoffs に近似することが期待されるが、フォーマルな証明は簡単ではない。

実際の統計理論的な正当化のためは、Empirical Knockoffs から計算される FDR が、Oracle Knockoffs から計算される FDR と漸近同等となることを示せばよい。その際の数学的に最も難しい点は、FDR が階段関数の期待値として定義される点である。一般的に、有界かつ連続な関数の期待値は非常に扱いやすく、多くの確率論的な収束定理が知られている。その一方で、FDR のような不連続関数の期待値に関しては収束定理が適応できず、大変扱いにくい。この難点を克服するために、我々はアルゴリズムの数理的な安定性を仮定した。データの摂動によるわずかな変化が変数選択に大きな影響を与えないと仮定することにより、漸近的に不連続性が無視でき、結果的に FDR の同等性が証明できる。つまり、Empirical knockoffs による FDR コントロールが漸近的に可能であることを示す。これは、未知な分布を伴う高次元モデルにおいて、初めての FDR コントロール可能な変数選択手法となる。我々はさらに一定の条件のもと、この手法の検出力が通常の Lasso と同じ下限を持つことを証明した。多くの場合 Lasso の検出力は 100% となるので、それに対してロスがないことを意味する。結果的に、マクロ経済分析において非常に汎用性の高いモデルに関して、検出力の高い FDR コントロール付きの変数選択手法を提案できた。

さらにモンテカルロ・シミュレーションと実データを用いた実証分析例により、この手法の有用性を検証した。シミュレーションでは、大多数の場合で FDR が正確にコントロールされていることが分かった一方で、さらに Barber and Candès (2017) の手法に対して検出力が大幅に高いことも確認できた。また、モデルの特定化を誤った場合でも、良好な FDR コントロールと高い検出力が得られるなど、頑健性の確認もできた。実証では大規模マクロ経済データを用いて、各種変数の予測問題を考察した。一般によく用いられるファクター・モデルや Lasso と比較して、同様に高いパフォーマンスを示すことが分かった。

2018 年 6 月にシンガポールで開催された Institute of Mathematical Statistics の国際学会の Invited Paper Session と、8 月に小樽で開催された SWET2018 の計量経済学セッションにて、この研究についての報告をした。現在は論文の最終確認をしている段階で、間もなく専門誌に投稿する予定である。また今後の研究方針として、因果推論への FDR コントロールの応用のほか、共変量に離散変数を含めるための拡張や、系列相関を持つファクターを許すための理論展開を考えている。

2. SOFAR の応用によるウィーク・ファクター・モデルの推定法の確立とその統計的理論解析

近年、大規模マクロ経済データを効率的にモデリングするための方法として、ファクター・モデルの研究が盛んである。これは、データ行列のある種の低ランク近似であり、推定すべきパラメータを大幅に減らすことができる。いま、 N を変数の数、 T を観測数とする。ウィーク・ファクター・モデルとは、ファクター・ローディングの内積の発散レートが $O(N)$ に満たないようなファクター・モデルのことをいう。一般のファクター・モデルは簡単化のため発散レートを $O(N)$ と仮定しているが、いくつかの実証面からの観察により、それでは大きすぎるという指摘があるため、そのギャップを埋めるものとしてウィーク・ファクターが注目を集めている。

モデルがウィーク・ファクター構造を持つための十分条件として、ファクター・ローディング行列がスパースであればよい。スパース性が強まるほど、各ローディングの内積の発散レートは $O(N)$ よりも遅くなる。非ゼロ要素の個数が有限個であれば、これは有界つまり $O(1)$ となる。スパースなパラメータを推定するだけなら Lasso など既知の手法を適応できる。しかし、ファクター・モデルはファクターが直交しないといけないため、ある種の制約付きスパース推定を行わなければならない。実証面からの有用性が指摘される一方、ウィーク・ファクター構造の実際の推定は非常に難しい。実際に現在のところ、非常に限られた状況を除いてウィーク・ファクター・モデルを推定する方法は提案されていない。

この推定を可能にするのが、私たちが 2016 年から 2017 年にかけて行った Sparse Orthogonal Factor Regression (SOFAR) に関する研究である。この研究では、高次元多変量回帰の係数行列の特異値分解について、その低ランク近似を行いつつ左右特異値ベクトルの行列をスパース性かつ直交性を同時に満たすような推定方法を考案した。論文中では、その推定量に関する非漸近的誤差上限を導出するとともに、そのアルゴリズムが局所解に収束することを示した。ここで、推定された左特異値ベクトルからなるスパース行列は共変量の変数選択（次元削減）を行う一方、右から推定された右特異値ベクトルの行列を掛けることで、被説明変数の変数選択（次元削減）も可能となる。いま、多変量回帰モデルの共変量を単位行列と置いて SOFAR 問題を解けば、スパースなファクター・ローディング推定量を持つファクター・モデル、つまりウィーク・ファクター・モデルを推定することができる。しかし、SOFAR の理論解析の中では左右両方の特異値ベクトルの行列がスパースであることを仮定しているが、モデルの特性上、ファクター・ローディング、つまり特異値からなる対角行列と右特異ベクトルからなる行列の積のみがスパースであると仮定している。そのため、証明の方法は全く新しいものとなる。その他にも SOFAR と異なる点として、ウィーク・ファクター・モデルのファクター数、つまりランクは事前に決める方針を取ったことが挙げられる。

ファクター・モデルにおいて理論上重要なのは、いかに正確にファクター数を推定するかという点である。通常ファクター・モデルに関しては、Bai and Ng (2002) や Ahn and Horenstein (2013) の方法がよく知られているが、ウィーク・ファクター・モデルには直接適用できない。一方、Onatski (2010) による方法では、ファクター数を決定するためにはシグナルの特異値が誤差行列の特異値よりも早く発散すれば十分であるため、我々の想定するウィーク・ファクター構造のもとでもよく機能することが分かった。次に、ファクターとスパース・ファクター・ローディングの SOFAR 推定量の非漸近的な誤差上限を導出した。さらに現在は、ペナルティ項にウェイト付けすることで得られる adaptive SOFAR 推定量を考察し、それによるスパース・ローディング推定のファクター選択に関する一貫性を証明している。そのためにはまず、非ゼロ要素部分に制約した SOFAR 問題の最適化を通じてそれら推定量の収束レートを導出する必要がある。これが得られれば、あとはゼロ要素を加えたものを目的の推定量と置き、これが制約なしの SOFAR 最適化問題から得られる KKT 条件を満たすことを確認すればよい。通常 Lasso や凹ペナルティ項付き回帰の推定量に関してはその証明はよく知られているが、スパースなファクター・モデルについては同時に直交条件も満たさなければならない。単純ではない。

こうした統計理論の構築に加え、シミュレーションと実証分析により我々の手法の有用性を確認した。シミュレーションでは、かなりスパース性の強い、つまりかなり弱いウィーク・ファクター構造の場合でも、Onatski のファクター数選択が漸近的に正確に機能することが分かった。サンプルサイズがそれほど大きくない場合でも、比較的良好な結果が得られた。また、ファクターとスパース・ローディングの推定精度に関しても、SOFAR 推定が非常にうまくいくことが分かった。実証分析では、マクロ経済データを用いた予測を行った。その結果、通常ファクター・モデルよりもウィーク・ファクターを仮定した方が、予測精度が上昇することが分かった。また S&P500 に含まれる銘柄についても、ウィーク・ファクターの存在が確認された。

2018 年 3 月に開催された東北大学での研究会で、この研究について報告した。現在は上記の理論構築を完成させるために、さらなる研究を進めている。