

高次元時系列データ分析の最近の展開

早川 和彦*

Recent Development of High-Dimensional Time Series Analysis

Kazuhiko Hayakawa*

本稿では高次元時系列データ分析の最近の発展を紹介する。伝統的な時系列分析では数種類(例えば 4, 5 種類)の変数に基づいた分析を行うが, 高次元時系列データ分析では数十から数百種類の時系列データに基づいた分析を行う。このようなデータを分析する際, 伝統的な方法はそのまま適用できないため, 新しい分析手法が必要となる。本稿ではその手法の簡単な紹介と, 高次元時系列データがどのような経済分析で利用されているのかを紹介する。

This paper provides a survey of recent development of high-dimensional time series analysis. Compared with the traditional time series analysis where only a few variables are considered, several dozens or hundreds kinds of data are investigated in high-dimensional time series analysis. When analysing such a dataset, the traditional statistical methods are not always applicable, and new statistical methods are required. This paper gives a brief survey on such new statistical methods to analyse high-dimensional data, and then introduces several economic applications where the use of high-dimensional time series data is advantageous.

キーワード: 高次元時系列データ, ファクターモデル, 主成分分析, 変数選択, 経済予測

1. はじめに

近年, 情報通信技術の発展により, 多くのデータが収集され利用可能になっている。社会科学の分野に注目すると, 例えばマーケティングでは, 消費者の大量の購買データから購買パターンを見つけ, それを販売戦略に有効活用するという試みが行われている。また, 経済学の分野でも POS データを用いたインフレ率の研究や高頻度データを用いたファイナンスの研究が盛んに行われている。本稿ではその中でも特に多種類のマクロ時系列データ(高次元時系列データ)を用いた近年の研究の紹介を主に行う。伝統的な多変量時系列分析では, データの利用可能性や理論的制約などにより, 数種類の時系列データを分析するだけであったが, 近年は数十から数百種類の時系列データの分析が活発に行われている。

* 広島大学大学院社会科学研究所: 〒 739-8525 広島県東広島市鏡山 1-2-1 (E-mail: kazuhaya@hiroshima-u.ac.jp).

Bernanke and Boivin (2003) はこのように多種類の時系列データが利用できる環境のことを「データが豊かな環境」(data rich environment)と呼んでいる。

このような大量のデータを用いて分析を行おうという発想は、情報量が非常に多いという点においては、統計学的に非常に自然なものである。しかしながら、データの大量化に伴って新たな問題が生じるという側面もある。例えば、データの種類が多い場合、これまで用いられてきた統計手法は、多くの場合そのまま利用することができないため、新たな手法を開発する必要がある。本稿の目的の1つは、高次元データを扱うために開発された統計手法の紹介を行うことである。もう1つの目的は、近年、高次元時系列データを用いたマクロ実証分析が盛んに行われるようになってきており、その簡単な紹介をすることである。

高次元時系列データ分析に関連するサーベイはいくつかある。Reichlin (2003), Breitung and Eickmeier (2006) は高次元時系列データの統計手法とその応用についてサーベイしている。Bai and Ng (2008b), Breitung and Choi (2011) は高次元時系列データの分析によく用いられる高次元ファクターモデルに関する統計手法をサーベイしており、Stock and Watson (2011) はダイナミックファクターモデルに焦点を当てたサーベイをしている。また、次節で説明する高次元時系列データを用いた予測手法に焦点を当てたサーベイとしては Watson (2003), Stock and Watson (2006), Eklund and Kapetanios (2008) などがある。本稿は高次元時系列データの統計手法とその応用についてサーベイするため、内容は Reichlin (2003), Breitung and Eickmeier (2006) に近いが、これらの論文より、より最近の研究を紹介している。

本稿の構成は次のようになっている。第2節において、大規模ファクターモデルの統計手法の最近の理論的發展を紹介する。第3節では高次元時系列データが利用されている経済分析例を紹介する。最後に第4節で結論とする。

2. ファクターモデル

高次元時系列データの分析で重要な役割を果たすのがファクターモデルである。ファクターモデルは統計学の分野では古くから利用されており、経済学においても, Geweke (1977), Sargent and Sims (1977) など用いられている。時点 t における第 i 番目のデータを x_{it} , ($i = 1, \dots, N$; $t = 1, \dots, T$) と表す。 T は時系列方向のサンプルサイズで, N は変数の種類の総数を表している。ここで, x_{it} に次のようなファクター構造を仮定しよう。

$$x_{it} = \lambda_i' \mathbf{F}_t + e_{it}, \quad (i = 1, \dots, N; t = 1, \dots, T). \quad (2.1)$$

λ_i, \mathbf{F}_t は $r \times 1$ であり, それぞれファクター負荷ベクトル, 静学的ファクターベクトルと呼ばれる。 e_{it} は誤差項である。このモデルでは x_{it} だけがデータとして観測可能で, $\lambda_i, \mathbf{F}_t, e_{it}$

は観測不能である。また、このモデルは x_{it} と \mathbf{F}_t の関係が静学的であるため、静学的ファクターモデルと呼ばれ、 r は静学的ファクター数である。ただし、 \mathbf{F}_t 自体は静学的である必要はなく、 $\mathbf{A}(L)\mathbf{F}_t = \mathbf{u}_t$ のように動学的になっても良い。ただし $\mathbf{A}(L)$ はラグ多項式であり、 \mathbf{u}_t は *iid* 変数である。モデル (2.1) は e_{it} に置かれる仮定に応じて呼び方が異なる (Chamberlain and Rothschild (1983))。 e_{it} にクロスセクション間の相関がない場合、すなわち、 $E(e_{it}e_{jt}) = 0, (i \neq j)$ のとき、(2.1) は厳密なファクターモデル (strict factor model) と呼ばれる。一方、 e_{it} に“弱い”クロスセクション間の相関を許す場合、より厳密に言うと $N^{-1} \sum_{i=1}^N \sum_{j=1}^N \sup_t |E(e_{it}e_{jt})| \leq M < \infty$ を満たすとき、(2.1) は近似的ファクターモデル (approximate factor model) と呼ばれる。

静学的ファクターモデル (2.1) に対して、次のようなモデルを動学的ファクターモデルと呼ぶ。

$$x_{it} = \boldsymbol{\lambda}'_i(L)\mathbf{f}_t + e_{it}. \quad (2.2)$$

ただし、 $\boldsymbol{\lambda}_i(L) = \lambda_{i0} - \lambda_{i1}L - \dots - \lambda_{is}L^s$ であり、 $\boldsymbol{\lambda}_{ij}, (j = 0, \dots, s)$ は $q \times 1$ ベクトルである。ラグ次数 s が有限の場合をダイナミックファクターモデル、 s が無限の場合を一般化ダイナミックファクターモデルと呼んだりする。また、どちらの場合も \mathbf{f}_t は $\mathbf{f}_t = \mathbf{C}(L)\boldsymbol{\varepsilon}_t$ という動学構造を持つとする。ただし、 $\boldsymbol{\varepsilon}_t$ は $q \times 1$ の *iid* 変数である。通常、 q のことをダイナミックファクター数と呼ぶ。ダイナミックファクターモデル (2.2) は次のように静学的ファクターモデルに書き直すことができる。

$$x_{it} = \boldsymbol{\Lambda}'_i\mathbf{F}_t + e_{it}$$

ただし、 $\boldsymbol{\Lambda}_i = (\boldsymbol{\lambda}'_{i0}, \boldsymbol{\lambda}'_{i1}, \dots, \boldsymbol{\lambda}'_{is})'$ 、 $\mathbf{F}_t = (\mathbf{f}'_t, \mathbf{f}'_{t-1}, \dots, \mathbf{f}'_{t-s})'$ である。したがって、静学的ファクター数 r とダイナミックファクター数 q の間には $r = q(s+1)$ という関係があることになる。ファクターモデルはファイナンスやパネルデータ分析など計量経済学の様々な分野で用いられているが、ここでは高次元時系列データの分析ツールという側面から考える¹⁾。

ファクターモデルを用いて統計分析をする際、変数の種類 N の大きさによっていくつかの点が異なる。1つ目は推定手法である。例えば N が小さいダイナミックファクターモデルでは、通常、カルマンフィルターを用いて推定するが、 N が大きい場合はパラメータ数が非常に大きくなってしまい、推定が極めて困難になる。そのため N が大きい場合は、後

¹⁾ 高次元のファクターモデルはパネルデータの分析でも頻繁に用いられている。伝統的なパネルデータモデルではクロスセクション主体が互いに独立であるということ想定してきたが、この仮定は現実的ではないため、クロスセクション間に相関を許したモデルの分析が多く行われている。そのときに用いられるのがファクターモデルである。近年のサーベイについては千木良他 (2011) を参照されたい。

で簡単に紹介するが、主成分分析を用いて推定することが多い。2つ目は漸近分析に使われるフレームワークである。 N が小さい場合は N を固定して $T \rightarrow \infty$ とするが、 N が大きい場合は N も T も大きくなる漸近理論を用いる²⁾。3つ目は2つ目と関連しているが、ファクターの漸近的結果である。 N が小さい場合は \mathbf{F}_t は一致性を持たないが、 N が大きい場合は \mathbf{F}_t は一致性を持つ。特に、この最後の特徴は高次元時系列分析においては非常に重要になってくる。Burns and Mitchell (1946)は「景気」を「多くの変数に共通する変動」としているが、ファクター \mathbf{F}_t がまさにこの景気の定義に合致しているからである³⁾。以下では N が大きいときのファクターモデルの推定とファクター数の推定について説明する。

2.1 大規模ファクターモデルの推定

大規模ファクターモデルの推定は基本的に主成分分析に基づいて行われる。Stock and Watson (2002) (以下, SW), Bai (2003) は時間領域での主成分分析を提案し, Forni *et al.* (2000, 2004, 2005) (以下, FHLR) は周波数領域での主成分分析を提案している。時間領域の主成分分析はデータの共分散行列の固有ベクトルを計算するだけであるため、計算は非常に簡単である。一方、周波数領域の主成分分析はスペクトル密度の推定などが必要になるため、計算の負荷が大きい。特定の周期の成分を抽出したりすることが可能であるため、応用性は高い。時間領域と周波数領域の主成分分析のどちらが優れているかははっきりしていない。また、主成分分析の代替的な方法として Kapetanios and Marcellino (2009) は sub space アルゴリズムを利用した推定方法を提案している。この方法はモデルをパラメトリックに状態空間表現する必要があるという点で若干制約的であるが、Kapetanios and Marcellino (2009) はシミュレーションで時間領域・周波数領域の主成分分析と同じようなパフォーマンスを持つことを示している。一方、Doz *et al.* (2012) は N が小さい場合に良く用いられる最尤推定量について考察しており、最尤法によって推定されたファクターは $N \rightarrow \infty$ のときに一致性を持つことを示している。

より最近の研究では主成分分析の改善が試みられている。誤差項 e_{it} に系列相関やクロスセクション相関、不均一分散がある場合、主成分分析はその構造を無視しているため、効率的な推定方法ではない。この問題を解決する方法がいくつか提案されている。Breitung and Tenhofen (2011), Choi (2012), Bai and Liao (2013) は一般化最小二乗法の考え方を適用した一般化主成分分析を提案しており、Bai and Li (2012a,b), Bai and Liao (2012) は最尤法を用いた効率的な推定方法を提案している。

²⁾ N と T が両方とも大きくなる漸近理論はパネルデータ分析でも頻繁に利用される。

³⁾ Bai and Ng (2006b) は抽出されたファクターと実際のデータとの関係性を統計的に考察する方法を提案している。

2.2 ファクター数の選択

ファクターモデルを利用するときには必ずファクター数を決めなければならない。大規模ファクターモデルにおけるファクター数の選択についても多くの研究が行われている。まず、大規模ファクターモデルの枠組みで、静学ファクター数 r の選択を最初に考察したのは Bai and Ng (2002) であり、彼らは情報量基準を用いてファクター数を一致推定する方法を提案している。より最近の研究としては Onatski (2009, 2010), Kapetanios (2010), Ahn and Horenstein (2013) などがある。また、ダイナミックファクター数 q の推定方法は、Amengual and Watson (2007), Hallin and Liska (2007), Bai and Ng (2007), Breitung and Pigorsch (2013) によって提案されている。

上で述べたように、抽出されたファクターは観測できない景気を表していると解釈できるため、実証分析を行ううえで非常に重要なツールになっている。次節ではファクターを用いた経済分析の例を紹介する。

3. 高次元時系列データを用いた経済分析

大規模データから抽出された共通ファクターを使った実証分析は多岐にわたる。本節ではその中から代表的なものを紹介する⁴⁾。特に、経済予測に使われることが多いため、経済予測については比較的詳しく説明する。

3.1 Diffusion Index 予測

経済時系列データの予測は従来、Box-Jenkins 流のアプローチに基づいて行われてきた。しかしながら、SW の先駆的な研究以降、高次元時系列データを用いて経済予測するアプローチが盛んに用いられるようになった。SW は主成分分析を用いて、高次元時系列データから数個のファクターを抽出し、それを用いて経済予測を行うと、単純な時系列モデルに基づいた場合よりもパフォーマンスが良いことを示している。この方法は Diffusion Index (DI) 予測と呼ばれている。DI 予測の手順を簡単に示そう。

予測したい変数を y_t とし、その h 期先の値を y_{t+h} とする。たとえば y_t がインフレ率の時は y_{t+h} は h 期先のインフレ率となる。そして、 y_{t+h} を t 期に利用可能な N 個の変数 x_{1t}, \dots, x_{Nt} を用いて予測したいとする。そのために、次のようなモデルを考える。

$$y_{t+h} = \beta_0 + \beta_1 x_{1t} + \dots + \beta_N x_{Nt} + \gamma_1 y_{t-1} + \dots + \gamma_p y_{t-p} + u_{t+h} \quad (3.1)$$

変数の数 N が小さい時は、情報量基準などで最も当てはまりの良いモデルを選ぶことが可能であるが、 N が大きい時 (例えば $N = 100$)、全ての組み合わせを考えることはほぼ不可

⁴⁾ Stock and Watson (2014) は高次元時系列データを用いて景気転換点の予測問題を考察しているが、共通ファクターに基づいた分析をしていないため、ここでは紹介しない。

能である。そこで、次元縮小として (2.1) のようなファクター構造を考える。(2.1) を用いると、(3.1) 式は次のように表せる。

$$y_{t+h} = \beta_0 + \boldsymbol{\delta}' \mathbf{F}_t + \gamma_1 y_{t-1} + \cdots + \gamma_p y_{t-p} + \varepsilon_{t+h}$$

実際は、 \mathbf{F}_t はデータとして観測されないので x_{it} に主成分分析を適用して、推定値 $\hat{\mathbf{F}}_t$ を計算し、次のようなファクター拡張回帰 (factor augmented regression) モデルを推定する。

$$y_{t+h} = \beta_0 + \boldsymbol{\delta}' \hat{\mathbf{F}}_t + \gamma_1 y_{t-1} + \cdots + \gamma_p y_{t-p} + \varepsilon_{t+h} \quad (3.2)$$

これが SW によって提案された DI 予測である。DI 予測の特徴的な点は、高次元時系列データの情報を数個のファクターに集約し、情報量を保ちながらモデルを小さくすることが可能であり、予測力の向上につながるという点にある。

3.1.1 DI 予測の評価

SW はアメリカのデータを用いて、DI 予測は単純な時系列モデルによる予測よりも精度が高いことを報告している。DI 予測の有効性については各国で様々な検討が行われている。Cheung and Demers (2007) はカナダの GDP 成長率とコアインフレ率、Marcellino *et al.* (2003) はユーロ圏の 11 カ国の主要変数、Artis *et al.* (2005) はイギリスの主要変数、Schumacher and Dreger (2002), Schumacher (2007) はドイツの主要変数、Bruneau *et al.* (2007) はフランスのインフレ率、Banerjee *et al.* (2009) はユーロエリアとスロベニアのインフレ率、失業率、GDP 成長率等、Shintani (2005), 北村・小池 (2002), 飯星 (2009) は日本の主要変数の予測のパフォーマンスを評価している。また、Eickmeier and Ng (2011) は世界各国のデータを含んだ大規模データを使うことでニュージーランドの GDP 成長率の予測精度が改善されるのかを検証しており、他の地域の変数を含めることは予測パフォーマンスの向上に効果的であることを示している。Wang (2009) は DI 予測と dynamic stochastic general equilibrium (DSGE) モデルに基づいた予測を比較しており、DI 予測は生産高とインフレの短期予測に、DSGE モデルはインフレの長期予測に優れていることを示している。

また、ファクター抽出に使われる統計手法の選択という観点から比較を行っている研究もある。Boivin and Ng (2005) はファクターの推定を SW の方法と FHLR の方法を使った場合、予測モデルで逐次法、直接法、無制約法、ノンパラメトリック法を採用したときの DI 予測のパフォーマンスを比較している。結果として、簡単な data generating process (DGP) の場合はそれほど違いはなかったが、複雑な DGP の場合は SW の無制約法のパフォーマンスが良かったと報告している。一方、D'Agostino and Giannone (2012) はファクター抽出の方法として SW と FHLR を使った DI 予測のパフォーマンスを比較しており、Boivin and Ng (2005) と異なり、両方法のパフォーマンスはほとんど同じであることを確認している。また、Forni *et al.* (2003) は FHLR と SW の方法を用いてインフレ率と

industrial production (IP) の予測パフォーマンスを評価し, SW と FHLR は同じようなパフォーマンスを持つこと, 多変量に基づいた予測は 1 変量の場合よりもパフォーマンスが良いケースが特にインフレの場合に多いこと, 金融変数はインフレの予測に役立つが, IP の予測には役立たないことを示している.

3.1.2 DI 予測の改善

DI 予測の問題点や改善方法に関連して様々な議論がある. 以下ではそれらを簡単に説明する. まず, 予測に使う変数の種類 N についてであるが, 理論的には N は大きい方が好ましい. しかしながら, 実際の予測では必ずしも大きい N が望ましくないことを Boivin and Ng (2006) は示している. 実際, 彼らは $N = 40$ と $N = 147$ の場合を比べて $N = 40$ の方が予測精度が高いことを示している. そして, その理由として, ファクター抽出に予測には役に立たない不要なデータが含まれているからであると述べている. 次に, モデルの定式化についてであるが, 通常ファクターモデルではファクター負荷 λ_i は時間を通じて一定であると仮定している. この仮定を緩め, ファクター負荷の構造変化の可能性を考慮した DI 予測の議論が Stock and Watson (2009), Banerjee *et al.* (2009) によって行われている⁵⁾. また, DI 予測では, 通常 x_{it} に主成分分析を適用するが, Bai and Ng (2008a) は x_{it} と x_{it}^2 の系列に主成分分析を使う方法, あるいはファクターの 2 乗を説明変数に使う方法を考察している. Ludvigson and Ng (2007) は実際にこの方法を使って超過収益の実証分析をしており, 2 乗のデータから得られたファクターを使うよりもファクターの 2 乗を説明変数に使う方が説明力が高いことを示している. Groen and Kapetanios (2013) はファクター拡張回帰モデルにおける情報量基準を提案している. 推定されたファクターを説明変数に加えると, ファクターの推定誤差が存在するため通常の情報量基準は適用できないことに注目し, その誤差を考慮した新しい情報量基準を提案している. Armah and Swanson (2010) は大規模パネルから推定されたファクターを使うと予測誤差の分散が大きくなるという Bai and Ng (2006a) の問題点に注目し, Bai and Ng (2006b) で提案されたファクターと観測データの関連性をチェックする検定で選ばれた観測変数を, ファクターの代わりに説明変数に使うことで予測のパフォーマンスが上がることを示している.

DI 予測は多種類の時系列データが持つ情報を数個のファクターに集約するという主成分分析の考え方に基づいているが, 高次元時系列データを用いて経済予測を行う方法はこれだけではない. 以下では代替的な方法として次元縮小アプローチ, 変数選択アプローチ, その他のアプローチを紹介する.

⁵⁾ Breitung and Eickmeier (2011), Bates *et al.* (2013) は一般的なファクターモデルの枠組みでファクター負荷の構造変化について議論している.

3.1.3 代替的な方法

次元縮小アプローチ 上で説明した主成分分析を用いる方法は次元縮小アプローチの1つであるが、主成分分析以外にも次元縮小する方法が提案されている。主成分分析の問題点の1つは、ファクターの抽出に関して、予測したい変数 y_t の情報を一切使っていない点にある。つまり、 y_t がインフレ率であっても GDP 成長率であっても同じファクターを用いて予測を行うということである。しかしながら、予測したい変数に応じてファクターも変わると考える方が自然である。したがって、予測したい変数 y_t の情報を取り入れながらファクターを推定することが望ましい。この問題に対して、Lin and Tsay (2005) や Groen and Kapetanios (2008) は partial least squares を用いることを提案している。また、Heij *et al.* (2008) はファクターの推定と予測を同時に行う principal covariance regression を考察している。

変数選択アプローチ 予測回帰モデル (2.2) 式をみると、これは N 種類のデータのうちのどの変数が予測に役立つのかを調べるという変数選択の問題として捉えることもできる。 N が小さい場合は情報量基準などで最も良いモデルを選ぶことができるが、 N が大きい場合、すべてのモデルの組み合わせを考えることはほとんど不可能である。しかしながら、 N が大きいときの変数選択の問題は、近年の統計学の文献で様々な方法が提案されている。代表的な方法が Tibshirani (1996) による least absolute shrinkage and selection operator (LASSO) である。Bai and Ng (2008a) や De Mol *et al.* (2008) は LASSO, elastic net で選択された x_{it} を予測に使う方法を提案している。Bai and Ng (2008a) はさらに、選択された系列に主成分分析を適用してファクターを抽出することを提案している。この2段階法は、1段階目で予測したい変数に関連のある変数だけを取り出すため、予測には役に立たない不要なデータが取り除かれている。そのため、この方法を用いると Boivin and Ng (2006) が指摘した問題点を回避できる。

主成分分析以外にも変数選択の手法は多く存在する。Inoue and Kilian (2008) は bootstrap aggregation (bagging) を使って説明変数の選択を行う方法を提案しており、同一のウェイトを用いた予測結合などより、かなりパフォーマンスが良く、ベイジアン縮小推定や LASSO と同じくらいのパフォーマンスを持つことを示している。Bai and Ng (2009) は機械学習の手法であるブースティングを使った DI 予測の改善を提案しており、Kapetanios (2007) は simulated annealing と genetic algorithm と呼ばれる変数選択の手法を用いて多くの変数の中から情報量基準を最小にする変数の組み合わせを求める方法を提案している。

その他のアプローチ ここまでは次元縮小アプローチと変数選択アプローチを説明してきたが、これら以外にも高次元時系列データを用いて予測を行う方法がいくつか提案されて

いる。以下ではそれらのうち代表的な方法を簡単に紹介する。

まず、変数選択アプローチの代替的な方法として、Kapetanios *et al.* (2008) や Cheng and Hansen (2014) はモデル平均化アプローチを提案している。Kapetanios *et al.* (2008) はベイジアンモデル平均化法と頻度論的モデル平均化法を予測の評価に使い、両推定量とも同じような予測パフォーマンスを持つことを示している。また、Cheng and Hansen (2014) は Mallows の基準と cross validation 基準を最小にするようなウェイトを用いたモデル平均化法を DI 予測の枠組みで提案している。Stock and Watson (2012) はベイズ平均法や経験ベイズ法、baggingなどを特殊ケースとして含む一般化縮小表現を提案し、この表現に基づいた予測手法は、一般的にダイナミックファクターモデルによる予測よりも劣ることを示している。

また、これまでの手法はすべて1つの変数だけの予測問題を扱っていたが、複数の変数を同時に予測するという場合も考えられる。このような問題に対して、Carriero *et al.* (2011) は通常の reduced rank 回帰、ベイジアンベクトル自己回帰 (VAR) モデル、ベイジアン reduced rank 回帰、reduced rank posterior を用いたベイジアン VAR モデル、多変量ブースティングに基づいた予測のパフォーマンスを比較している。そして、rank reduction と縮小法を同時に使う方が別々に使う場合よりもパフォーマンスが良いこと、特に reduced rank posterior を用いたベイジアン VAR モデルとベイジアン reduced rank 回帰がかなり正確な予測結果を与えることを示している。また、Bańbura *et al.* (2010) はベイジアン縮小法を使った VAR モデルは N が大きい場合でも妥当な手法であり、 N が大きくなるにつれて予測精度を改善することを示している。

3.2 ファクター拡張 VAR モデル

Bernanke *et al.* (2005) はファクター拡張 (factor augmented, FA) VAR モデルを提案し、金融政策の効果の分析を行っている。通常の VAR モデルは大きく分けて次の3つの問題がある。1. 数個の変数しか利用しないので情報量が少なく、それが推定結果のバイアスにつながる、2. 「経済行動」や「物価水準」を表す変数として鉱工業生産指数や消費者物価指数を使うことが多いが、少し恣意的であり、観測誤差やデータ改訂などがある、3. 動学的な反応は VAR に含まれている変数のみ計算できる。FAVAR モデルではこれらの問題に対処している。すなわち、1. FAVAR モデルは通常の VAR モデルを包含しているので観測できないファクターを考慮することで追加的な情報が得られるかを確認できる、2. 経済変動を描写するファクターを使っているため、鉱工業生産指数や消費者物価指数が「経済活動」や「物価水準」といった概念と一致しているという仮定を置く必要はない、3. 主に分析する変数だけではなく、ファクター抽出に用いた変数の応答関数も計算できる。推定には主成分分析と通常の VAR モデルの推定を組み合わせる2ステップ法とギブズサン

リングを使ったベイズの両方を用いているが、推定結果はほとんど同じであることを示している。

FAVAR モデルを簡単に説明しよう。 \mathbf{Y}_t を $m \times 1$ の観測できる変数、 \mathbf{F}_t を $r \times 1$ の観測できないファクターとする。このとき、 $(\mathbf{F}'_t, \mathbf{Y}'_t)'$ の動学的構造が以下のようになっていると仮定する。

$$\Phi(L) \begin{bmatrix} \mathbf{F}_t \\ \mathbf{Y}_t \end{bmatrix} = \mathbf{v}_t \quad (3.3)$$

ただし、 $\Phi(L) = \mathbf{I} - \Phi_1 L - \dots - \Phi_p L^p$ はラグ多項式であり、 \mathbf{v}_t は *iid* 変数である。 Bernanke *et al.* (2005) はこのモデルのことを FAVAR モデルと呼んでいる。このモデルにおいて、もし \mathbf{Y}_t の式の中の \mathbf{F}_t のラグ変数に関連する項の係数がすべて 0 の場合、 \mathbf{Y}_t に関しては通常の VAR モデルと同じになる。したがって、FAVAR モデルは通常の VAR モデルを特殊ケースとして含むことになる。この特徴のため、通常の VAR モデルとの比較や \mathbf{F}_t を含むことによって得られる追加的な効果を簡単に検証することが可能となる。実際の分析では、 \mathbf{F}_t は観測できないため、高次元時系列データに主成分分析を適用して得られたファクターの推定値 $\hat{\mathbf{F}}_t$ を用いる。 Bernanke *et al.* (2005) はこの手法を用いて金融政策の評価を行っている。また、 Shibamoto (2007) この手法を用いて日本の金融政策の効果を評価をしている。

3.3 金融政策の評価と DSGE モデルの推定

Bernanke and Boivin (2003) は、中央銀行ではいくつもの変数を見て政策立案しているにもかかわらず、伝統的な金融政策を評価するモデルでは数個の代表的な変数だけを取り込んでいるという点を指摘し、SW で提案された DI モデルが金融政策の評価において有効であるかを議論している。そして、政策反応関数を推定し、中央銀行の行動がインフレ率と産出高の予測値だけによって説明されるどうかの検定方法を提案している。また、データを収集しベンチマークとなる政策を自動的に計算するエキスパートシステムの可能性にも言及している。

Favero *et al.* (2005) は forward looking 型テイラールールモデルのファクターを使った推定について考察している。特にファクター抽出の方法として SW と FHLR の方法を比較し、両者とも同じようなパフォーマンスであることを確認している。また、DSGE モデルでは期待変数が説明変数に用いられることが多いが、一般的に期待変数は観測できないため、政策評価を行う場合は、それを推定する必要がある。しかしながら、通常の方法では識別の弱さ、欠落変数によるモデルの定式化ミス、パラメータの不安定性などの問題が生じてしまう。 Beyer *et al.* (2008) は、この問題に対して、 Bai and Ng (2010)、 Kapetanios and Marcellino (2010) で提案された factor-GMM を用いて対処することを提案している。

日本のデータを用いた研究としては、飯星 (2009) が forward looking 型テイラールールモデルの推定を行っている。

3.4 Nowcasting

GDP は経済活動の最も重要な総合的指標であるが、通常、四半期ベースで、しかもかなりの遅れを伴って公表される。この問題を高次元時系列データを用いた nowcasting で克服しようとする研究が行われている。nowcasting とは、一般的には文字通り「現在を予測する」ということであるが、ここでは、月次データを用いた当該四半期の GDP 成長率の予測、と定義する。例えば、2月までの月次データを用いて第1四半期の GDP 成長率を予測するということである。nowcasting を行うときには、大きく分けて2つの問題が生じる。1つはデータの公表時期がそれぞれ異なるという問題であり、このようなデータは ragged-edge データと呼ばれている。もう1つの問題点は四半期データや月次データなど、公表頻度の異なるデータが混在しているという問題である。これらの問題に対処した研究として、Giannone *et al.* (2008) がある。Giannone *et al.* (2008) は伝統的な bridge 方程式を使う方法では、数種類の変数しか扱えないが、実際には多くの情報変数があり、それらを使わないのは適切ではないと批判し、月次ベースの高次元時系列データからファクターを抽出して、それを伝統的な bridge 方程式に取り入れる方法を提案している。Angelini *et al.* (2011) は伝統的な bridge 方程式と Giannone *et al.* (2008) によって提案された主成分分析で抽出されたファクターを使った bridge 方程式のパフォーマンスを比較しており、ファクターを用いた方が精度が高いことを示している。Schumacher and Breitung (2008) は主成分分析と EM アルゴリズムを使って月次データと四半期データを同時に用いて月次 GDP の成長率の推計を行っている。また、上記の2つの問題を解決する代替的な方法として、Marcellino and Schumacher (2010) はファクター MIDAS(mixed-frequency data sampling) という方法を提案している⁶⁾。

3.5 景気動向指数の作成

欧米では高次元ダイナミックファクターモデルから得られた共通ファクターを用いて景気動向指数を作成・公表する試みが行われている。例えば、ヨーロッパでは Altissimo *et al.* (2001) によって開発された EuroCOIN と呼ばれる一致指数が CEPR (Centre for Economic Policy Research) によって公表されていた。2007年9月からは Altissimo *et al.* (2010) によって提案された New EuroCOIN が CEPR から公表されている。New EuroCOIN の推計にはユーロ圏諸国の生産、物価、賃金、雇用、小売販売額など計145種類のデータが用いられている。

⁶⁾ MIDAS については Ghysels *et al.* (2007) や Andreou *et al.* (2010) によって考察されている。

またアメリカでもシカゴ連邦準備銀行が Chicago Fed National Activity Index (CFNAI) という一致指数を大規模ダイナミックファクターモデルを用いて作成している。CFNAI は生産高, 所得, 雇用, 労働時間, 新規住宅着工数など計 85 種類のマクロ時系列データに主成分分析を適用することで計算されている⁷⁾。これらの指標の日本への適用は早川・小林 (2011) で行われている。

3.6 コアインフレ指数の作成

高次元時系列データを用いてコアインフレ指数を作成しようという研究も行われている。コアインフレ指数の明確な定義は定まっていないが, 標準的な定義は, 物価指数のうち, 一時的な要因による物価の変動を除去した物価の変動を表す指数のことである。日本では, 生鮮食品を除く消費者物価指数がコアインフレ指数として利用されている。Kapetanios (2004) はコアインフレ指数を多くの価格指数の共通部分と定義し, 73 種類の価格指数に Kapetanios and Marcellino (2009) で提案された sub space アルゴリズムを用いてファクターを推定し, 第 1 ファクターをコアインフレ指数として見なしている。Cristadoro *et al.* (2005) は FHLR で提案された周波数領域の主成分分析を, 約 450 種類のデータに適用してインフレ指数の長期成分を抽出し, それをユーロエリアのコアインフレ指数として提案している。Giannone and Matheson (2007) は Cristadoro *et al.* (2005) と同様の分析をニュージーランドのデータを用いて行っている。

3.7 イールドカーブの予測

Moench (2008) は, これまでのイールドカーブの研究は数種類のマクロ変数だけを考慮して金利の期間構造のモデル化を行っているが, 実際には中央銀行では数多くの変数を見ているため正しいモデル化ではないと批判している。そして, 多くのマクロ変数から抽出されたファクターを使って金利の期間構造を推計する方法を提案している。具体的には短期金利の動学構造を FAVAR モデルで描写し, 期間構造を無裁定から導かれるパラメータ制約を用いて導出している。Moench (2008) は提案した方法を no-arbitrage FAVAR モデルと呼んでおり, この方法によるイールドの中長期予想は既存の方法よりも精度が高いことを示している。

Exterkate (2008) は Nelson-Siegel モデルに基づいてイールドカーブの予測をする際に, ファクターの抽出方法が重要であり, 1. 経済理論によってグループ分けされた変数群を使う方法, 2. Heij *et al.* (2007) によって提案された principal covariate regression モデルを使う方法, 3. multi-response sparse regression (MRSR) モデルを使う方法, 4. Bai and Ng (2008a) によって提案された変数選択の方法の 4 つの場合を比較している。実証分析の結

⁷⁾ CFNAI の厳密な計算手順については Evans *et al.* (2002) を参照されたい。

果からは、1.の方法は中長期のイールドを推定するときに有効であること、2.の方法は短期イールドの予測に有効であること、3.の方法は他の方法ほど予測精度が良くないこと、4.の方法はそれほど長くない満期のイールドの予測に有効であることを示している。

3.8 株価収益率の予測可能性

リスク・リターンの関係を分析する伝統的な方法は、株式市場の超過収益の条件付期待値や条件付分散を推定するときに、比較的少ない情報を用いて分析している。しかしながら、少ない情報を用いると、リスク・リターンの関係にバイアスをもたらしてしまう。Ludvigson and Ng (2007)はこの問題を克服するために、高次元時系列データからファクターを抽出し、それを用いて条件付期待値や条件付分散を推定する方法を提案している。結果として、彼女らは同時点の条件付平均と条件付分散は、強い正の相関関係を持っていると述べている。

また、Ludvigson and Ng (2010)は債券市場プレミアムに循環的な変動があるかどうかを高次元時系列データを用いて検証している。結果として、先物相場やイールドスプレッドに含まれている予測力に加え、実物ファクターとインフレファクターがアメリカの国債市場の将来超過リターンの予測に重要な役割を果たすことを示している。そして、マクロファクターの情報を使わなければリスクプレミアムは非循環的であるが、マクロファクターを用いた場合、リスクプレミアムは明確な半景気循環的要素を持っており、経済理論の結果と合致すると結論付けている。

3.9 通貨危機のモニタリングと予測

Inoue and Rossi (2008)は通貨危機のモニタリングと予測の問題について議論しており、多くの経済変数を使った新しい早期警戒システムを提案している。そして、多くの変数を使うことで通貨危機の予測パフォーマンスを改善できるのか、通貨危機を予測するのにどの変数が最も重要であるかを考察している。結果として、DI予測は為替レートの水準そのものをうまく推定していないが、通貨危機の確率の予測には有効な方法であることを示している。また、通貨危機の予測に有効な変数は経済成長率、利子率、マネーサプライ伸び率であることを示している。

3.10 国際間景気変動の分析

高次元時系列データは国際間景気変動の連動性や伝播の分析などにも用いられている。Helbling and Bayoumi (2003)は高次元時系列データを用いてG7参加国間の国際間景気変動を分析し、国際的なショックが2000-2001年の景気の後退の主な原因であった、という仮説を支持する結果を得ている。また、Eickmeier (2007)は大規模ダイナミックファクターモデルを用いてアメリカ経済のショックがドイツ経済にどのような影響を及ぼすのかを考察しており、貿易と金融政策が最も関連性が強い伝播経路であり、金融市場や信用チャ

ネルについては明確な関連性が得られなかったことを示している。

4. まとめ

本稿では近年の高次元時系列データの計量経済分析に関連する研究のサーベイを行った。まず、高次元時系列データの分析手法として、高次元ファクターモデルの推定手法等の紹介を行った。そして、従来の数種類の時系列データに基づいた多変量時系列分析と異なり、高次元時系列データを使うことで、共通ファクターをより正確に推定することが可能であること、そのファクターは Burns and Mitchell (1946) の景気の定義と合致することを説明した。また、その推定したファクターは DI 予測や FAVAR モデル、景気予測などの経済分析で使われていることを紹介した。

最後に、本稿はサーベイ論文であるため、実際の分析は行っていない。しかしながら、本稿で紹介したような高次元時系列データを用いた実証研究はこれからも盛んに行われると考えられる。特に日本のデータを用いた研究はまだ限られており、これからの研究の蓄積が必要であると考えられる。

謝辞

本稿の作成にあたり、査読者から非常に丁寧なコメントをいただきました。また、谷崎久志編集長に大変お世話になりました。ここに深く感謝申し上げます。本研究は科学研究費補助金 (25780153) から研究助成を受けております。

参 考 文 献

- Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors, *Econometrica*, **81**, 1203–1227.
- Altissimo, F., Bassanetti, A., Cristadoro, R., Forni, M., Lippi, M., Reichlin, L. and Veronese, G. (2001). Euro-Coin: A real time coincident indicator of the Euro area business cycle, CEPR working paper, No. 3108.
- Altissimo, F., Cristadoro, R., Forni, M., Lippi, M. and Veronese, G. (2010). New Eurocoin: Tracking economic growth in real time, *Rev. Econ. Stat.*, **92**, 1024–1034.
- Amengual, D. and Watson, M. (2007). Consistent estimation of the number of dynamic factors in large N and T panel, *J. Bus. Econ. Stat.*, **25**, 91–96.
- Andreou, E., Ghysels, E. and Kourtellis, A. (2010). Regression models with mixed sampling frequencies, *J. Econom.*, **158**, 246–261.
- Angelini, E., Camba-Mendez, G., Giannone, D., Reichlin, L. and Rünstler, G. (2011). Short-term forecasts of Euro area GDP growth, *Econom. J.*, **14**, C25–C44.
- Armah, N. A. and Swanson, N. R. (2010). Seeing inside the black box: Using diffusion index methodology to construct factor proxies in largescale macroeconomic time series environments, *Econom. Rev.*, **29**, 476–510.
- Artis, M., Banerjee, A. and Marcellino, M. (2005). Factor forecasts for the UK, *J. Forecast.*, **24**, 279–298.
- Bai, J. (2003). Inferential theory for factor models of large dimensions, *Econometrica*, **71**, 135–171.
- Bai, J. and Li, K. (2012a). Maximum likelihood estimation and inference for approximate factor models of high dimension, working paper.
- Bai, J. and Li, K. (2012b). Statistical analysis of factor models of high dimension, *Ann. Stat.*, **40**, 436–465.

- Bai, J. and Liao, Y. (2012). Efficient estimation of approximate factor models via regularized maximum likelihood, working paper.
- Bai, J. and Liao, Y. (2013). Generalized principal components for panel data and factor models, working paper.
- Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models, *Econometrica*, **70**, 191–221.
- Bai, J. and Ng, S. (2006a). Confidence intervals for diffusion index forecasts and inference with factor-augmented regressions, *Econometrica*, **74**, 1133–1150.
- Bai, J. and Ng, S. (2006b). Evaluating latent and observed factors in macroeconomics and finance, *J. Econom.*, **113**, 507–537.
- Bai, J. and Ng, S. (2007). Determining the number of primitive shocks, *J. Bus. Econ. Stat.*, **25**, 52–60.
- Bai, J. and Ng, S. (2008a). Forecasting economic time series using targeted predictors, *J. Econom.*, **146**, 304–317.
- Bai, J. and Ng, S. (2008b). Large dimensional factor analysis, *Found. Trends Econom.*, **3**, 89–163.
- Bai, J. and Ng, S. (2009). Boosting diffusion indices, *J. Appl. Econom.*, **24**, 607–629.
- Bai, J. and Ng, S. (2010). Instrumental variable estimation in a data rich environment, *Econom. Theory*, **26**, 1577–1606.
- Bañbura, M., Giannone, D. and Reichlin, L. (2010). Large Bayesian vector auto regressions, *J. Appl. Econom.*, **25**, 71–92.
- Banerjee, A., Masten, I. and Massimiliano, M. (2009). Forecasting macroeconomic variables using diffusion indexes in short samples with structural change, in *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, Castle, J. L. and N. Shephard (eds.), Oxford University Press, Chap. 9, 227–254.
- Bates, B. J., Plagborg-Møller, M., Stock, J. H. and Watson, M. W. (2013). Consistent factor estimation in dynamic factor models with structural instability, *J. Econom.*, **177**, 289–304.
- Bernanke, B. S. and Boivin, J. (2003). Monetary policy in a data-rich environment, *J. Monet. Econ.*, **50**, 525–546.
- Bernanke, B. S., Boivin, J. and Eliasch, P. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach, *Q. J. Econ.*, **120**, 387–422.
- Beyer, A., Farmer, R. E. A., Henry, J. and Marcellino, M. (2008). Factor analysis in a model with rational expectations, *Econom. J.*, **11**, 271–286.
- Boivin, J. and Ng, S. (2005). Understanding and comparing factor based forecasts, *Int. J. Cent. Bank.*, **1**, 117–152.
- Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis?, *J. Econom.*, **132**, 169–194.
- Breitung, J. and Choi, I. (2011). Factor models, in *Handbook of Research Methods and Applications in Empirical Macroeconomics*, Edward Elgar, Chap. 11, 249–265.
- Breitung, J. and Eickmeier, S. (2006). Dynamic factor models, *Allgemeines Statistisches Archiv*, **90**, 27–42.
- Breitung, J. and Eickmeier, S. (2011). Testing for structural breaks in dynamic factor models, *J. Econom.*, **163**, 71–84.
- Breitung, J. and Pigorsch, U. (2013). A canonical correlation approach for selecting the number of dynamic factors, *Oxford Bull. Econ. Stat.*, **75**, 23–36.
- Breitung, J. and Tenhofen, J. (2011). GLS estimation of dynamic factor models, *J. Am. Stat. Assoc.*, **106**, 1150–1166.
- Bruneau, C., De Bandt, O., Flageollet, A. and Michaux, E. (2007). Forecasting inflation using economic indicators: The case of France, *J. Forecast.*, **26**, 1–22.
- Burns, A. F. and Mitchell, W. C. (1946). *Measuring Business Cycles*, National Bureau of Economic Research.
- Carriero, A., Kapetanios, G. and Marcellino, M. (2011). Forecasting large datasets with reduced rank multivariate models, *J. Appl. Econom.*, **26**, 735–761.
- Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure and mean-variance analysis in large asset markets, *Econometrica*, **51**, 1281–1304.

- Cheng, X. and Hansen, B. E. (2014). Forecasting with factor-augmented regression: A frequentist model averaging approach, forthcoming in *J. Econom.*
- Cheung, C. and Demers, F. (2007). Evaluating forecasts from factor models for canadian GDP growth and core inflation, working paper.
- 千木良弘朗, 早川和彦, 山本拓 (2011). 『動学的パネルデータ分析』 知泉書館.
- Choi, I. (2012). Efficient estimation of factor models, *Econom. Theory*, **28**, 274–308.
- Cristadoro, R., Forni, M., Reichlin, L. and Veronese, G. (2005). A core inflation indicator for the Euro area, *J. Money Credit Bank.*, **37**, 539–560.
- D'Agostino, A. and Giannone, D. (2012). Comparing alternative predictors based on large-panel dynamic factor models, *Oxford Bull. Econ. Stat.*, **74**, 306–326.
- De Mol, C., Giannone, D. and Reichlin, L. (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?, *J. Econom.*, **146**, 318–328.
- Doz, C., Giannone, D. and Reichlin, L. (2012). A quasi-maximum likelihood approach for large, approximate dynamic factor models, *Rev. Econ. Stat.*, **94**, 1014–1024.
- Eickmeier, S. (2007). Business cycle transmission from the US to Germany-A structural factor approach, *Eur. Econ. Rev.*, **51**, 521–551.
- Eickmeier, S. and Ng, T. (2011). Forecasting national activity using lots of international predictors: An application to New Zealand, *Int. J. Forecast.*, **27**, 496–511.
- Eklund, J. and Kapetanios, G. (2008). A review of forecasting techniques for large datasets, *Natl. Inst. Econ. Rev.*, **203**, 109–115.
- Evans, C. L., Liu, C. T. and Pham-Kanter, G. (2002). The 2001 recession and the Chicago fed national activity index: Identifying business cycle turning points, in *Economic Perspectives, Federal Reserve Bank of Chicago, Third Quarter*, 26–43.
- Exterkate, P. (2008). Forecasting the yield curve in a data-rich environment, working paper.
- Favero, C., Marcellino, M. and Neglia, F. (2005). Principal components at work: The empirical analysis of monetary policy with large datasets, *J. Appl. Econom.*, **20**, 603–620.
- Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000). The generalized factor model: Identification and estimation, *Rev. Econ. Stat.*, **82**, 540–554.
- Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2003). Do financial variables help in forecasting inflation and real activity in the Euro area?, *J. Monet. Econ.*, **50**, 1243–1255.
- Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2004). The generalized factor model: Consistency and rates, *J. Econom.*, **119**, 231–255.
- Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting, *J. Am. Stat. Assoc.*, **100**, 830–840.
- Geweke, J. (1977). The dynamic factor analysis of economic time series, in *Latent Variables in Socio Economic Models*, Aigner, D. J. and A. Goldberger (eds.), North Holland.
- Ghysels, E., Sinko, A. and Valkanov, R. (2007). MIDAS regressions: Further results and new directions, *Econom. Rev.*, **26**, 53–90.
- Giannone, D. and Matheson, T. D. (2007). A new core inflation indicator for New Zealand, *Int. J. Cent. Bank.*, **3**, 145–180.
- Giannone, D., Reichlin, L. and Small, D. (2008). Nowcasting: The realtime informational content of macroeconomic data, *J. Monet. Econ.*, **55**, 665–676.
- Groen, J. J. J. and Kapetanios, G. (2008). Revisiting useful approaches to data-rich macroeconomic forecasting, working paper.
- Groen, J. J. J. and Kapetanios, G. (2013). Model selection criteria for factor-augmented regressions, *Oxford Bull. Econ. Stat.*, **75**, 37–63.
- Hallin, M. and Liska, R. (2007). Determining the number of factors in the general dynamic factor model, *J. Am. Stat. Assoc.*, **102**, 603–617.

- 早川和彦, 小林庸平 (2011). 「大規模マクロデータを用いた景気動向分析」 浅子和美, 渡部敏明 (編) 『ファイナンス・景気循環の計量分析』 ミネルヴァ書房.
- Heij, C., Groenen, P. J. F. and van Dijk, D. (2007). Forecast comparison of principal component regression and principal covariate regression, *Comput. Stat. Data Anal.*, **51**, 3612–3625.
- Heij, C., van Dijk, D. and Groenen, P. J. F. (2008). Macroeconomic forecasting with matched principal components, *Int. J. Forecast.*, **24**, 87–100.
- Helbling, T. and Bayoumi, T. (2003). Are they all in the same boat? The 2000–2001 growth slowdown and the G-7 business cycle linkages, IMF working paper 03-46, Washington, International Monetary Fund.
- 飯屋博邦 (2009). 「主成分分析によるマクロ経済パネルデータの共通ファクターの抽出とその利用」 ESRI Discussion Paper Series No. 219.
- Inoue, A. and Kilian, L. (2008). How useful is bagging in forecasting economic time series? A case study of U.S. CPI inflation, *J. Am. Stat. Assoc.*, **103**, 511–522.
- Inoue, A. and Rossi, B. (2008). Monitoring and forecasting currency crises, *J. Money Credit Bank.*, **40**, 523–534.
- Kapetanios, G. (2004). A note on modelling core inflation for the UK using a new dynamic factor estimation method and a large disaggregated price index dataset, *Econ. Lett.*, **85**, 63–69.
- Kapetanios, G. (2007). Variable selection in regression models using nonstandard optimisation of information criteria, *Comput. Stat. Data Anal.*, **52**, 4–15.
- Kapetanios, G. (2010). A testing procedure for determining the number of factors in approximate factor models with large datasets, *J. Bus. Econ. Stat.*, **28**, 397–409.
- Kapetanios, G. and Marcellino, M. (2009). A parametric estimation method for dynamic factor models of large dimensions, *J. Time Ser. Anal.*, **30**, 208–238.
- Kapetanios, G. and Marcellino, M. (2010). Factor-GMM estimation with large sets of possibly weak instruments, *Comput. Stat. Data Anal.*, **54**, 2655–2675.
- Kapetanios, G., Labhard, V. and Price, S. (2008). Forecasting using Bayesian and information theoretic model averaging: An application to UK inflation, *J. Bus. Econ. Stat.*, **26**, 33–41.
- 北村富行, 小池良司 (2002). 「多くの情報変数を用いた予測方法の有用性について」 『金融研究』 第 21 巻, 101–142 頁.
- Lin, J.-L. and Tsay, R. S. (2005). Comparison of forecasting methods with many predictors, working paper.
- Ludvigson, S. and Ng, S. (2007). The empirical risk return relation: A factor analysis approach, *J. Financ. Econ.*, **83**, 171–222.
- Ludvigson, S. and Ng, S. (2010). Macro factors in bond risk premia, *Rev. Financ. Stud.*, **22**, 5027–5067.
- Marcellino, M. and Schumacher, C. (2010). Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP, *Oxford Bull. Econ. Stat.*, **72**, 518–550.
- Marcellino, M., Stock, J. H. and Watson, M. W. (2003). Macroeconomic forecasting in the Euro area: Country specific versus Euro wide information, *Eur. Econ. Rev.*, **47**, 1–18.
- Moench, E. (2008). Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach, *J. Econom.*, **146**, 26–43.
- Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models, *Econometrica*, **77**, 1447–1479.
- Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues, *Rev. Econ. Stat.*, **92**, 1004–1016.
- Reichlin, L. (2003). Factor models in large cross sections of time series, in *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, Dewatripoint, S. T. M. and L. P. Hansen (eds.), Vol. 3, Cambridge University Press, Chap. 2, 47–86.
- Sargent, T. and Sims, C. (1977). Business cycle modeling without pretending to have too much a priori economic theory, in *New Methods in Business Cycles Research*, Sims, C. (ed.), Minneapolis, Federal Reserve Bank of Minneapolis.
- Schumacher, C. (2007). Forecasting German GDP using alternative factor models based on large datasets, *J. Forecast.*, **26**, 271–302.

- Schumacher, C. and Breitung, J. (2008). Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data, *Int. J. Forecast.*, **24**, 386–398.
- Schumacher, C. and Dreger, C. (2002). Estimating large-scale factor models for economic activity in Germany: Do they outperform simpler models?, working paper.
- Shibamoto, M. (2007). An analysis of monetary policy shocks in Japan: A factor augmented vector autoregressive approach, *Jpn. Econ. Rev.*, **58**, 484–503.
- Shintani, M. (2005). Nonlinear forecasting analysis using diffusion indexes: An application to Japan, *J. Money Credit Bank.*, **37**, 517–538.
- Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes, *J. Bus. Econ. Stat.*, **20**, 147–162.
- Stock, J. H. and Watson, M. W. (2006). Forecasting with many predictors, in *Handbook of Economic Forecasting*, Elliott, G. and A. Timmermann (eds.), Vol. 1, North Holland, Chap. 10, 515–554.
- Stock, J. H. and Watson, M. W. (2009). Forecasting in dynamic factor models subject to structural instability, in *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, Castle, J. L. and N. Shephard (eds.), Oxford University Press, Chap. 7, 173–205.
- Stock, J. H. and Watson, M. W. (2011). Dynamic factor models, in *Oxford Handbook on Economic Forecasting*, Clements, M. J, Hendry, D. F., (ed.), Oxford University Press, Chap. 2, 35–59.
- Stock, J. H. and Watson, M. W. (2012). Generalized shrinkage methods for forecasting using many predictors, *J. Bus. Econ. Stat.*, **30**, 481–493.
- Stock, J. H. and Watson, M. W. (2014). Estimating turning points using large data sets, *J. Econom.*, **178**, 368–381.
- Tibshirani, R. J. (1996). Regression shrinkage and selection via lasso, *J. R. Stat. Soc. Ser. B*, **58**, 267–288.
- Wang, M.-C. (2009). Comparing the DSGE model with the factor model: An out-of-sample forecasting experiment, *J. Forecast.*, **28**, 167–182.
- Watson, M. W. (2003). Macroeconomic forecasting using many predictors, in *Advances in Economics and Econometrics, Theory and Applications, Eighth World Congress*, M. Dewatripont, L. Hansen and S. Turnovsky (eds.), Vol. 3, Chap. 3, 87–115.