金融市場間のグループ・依存構造を考慮した大規模金融データの分析手法について

Melbourne Business School Associate Professor　安　道　知　寛

―　目　　次　―

要　　旨

金融情報処理技術の発展により、高頻度取引データなどに代表される大規模金融データが観測、蓄積されている。日々蓄積されている様々な情報は、金融市場の特徴を把握するための重要な材料で、対象となる金融市場の特徴を的確に把握できる可能性を秘めている。本報告では金融市場間のグループ・依存構造を考慮した大規模金融データの分析手法について検討し、その応用例を紹介する。

1．はじめに

金融情報処理技術の発展により、高頻度取引データなどに代表される大規模金融データが観測、蓄積されている。そのような情報量の爆発的な増加は、ますます激しく変化する世の中の特徴であり、今後も、金融関連企業のみならず社会全体が急速に変化すると予想される。日々蓄積されている様々な情報は、金融市場の特徴を把握するための重要な材料で、その材料を、「どのように実務（例えば、資産運用、トレーディング、企業価値評価、リスク管理など）に生かすのか？」という問いは、現場実務のみならず、学術的にも検討するべき課題である。その問いを検討するに際しては、まず、対象となる金融市場の特徴を的確に把握する必要がある。その目的のために「計量的手法」を活用する場合、既存の計量的モデリング手法をさらに複雑化させる方向（例えば、金融資産収益率の分布に対する仮定を正規分布と比べ厚い裾を持つ特徴へと拡張するなど）で金融市場の特徴を部分的に取り込むことが考えられる。しかし、計量的モデリング手法を活用する際の根本的な問題は、金融市場に対する仮定が現実の金融市場と比較してどの程度整合的であるかという点にあると考える。

１点目として（統計的な意味での）独立性についてまず検討したい。計量的手法を利用する際、特に統計数理分野や数理ファイナンス分野の理論を利用する際には、数理的な取り扱いの容易さから「独立性」を仮定する場合が多いように思う。しかし、金融イベント

において独立性が完全に保障される場面がどこにあるのだろうか？主観的ではあるが、金融市場は、幾多の経済イベント、多様な投資家の行動などを含め、様々な要因が複雑に絡み合っているように感じられる。そのような現象に対して「独立性」を仮定することについて、それは現実の金融市場と比較してどの程度整合的であるのか？この問いに正面から対峙した場合、的確な説明をできるケースでは、独立性についてのハードルは越えていると考えられる。

２点目として（統計的な意味での）計量モデルに利用する変数について検討したい。ここでは簡単のため、古典的な回帰分析を例に考えたい。回帰分析では、分析対象となるある変数をいくつかの変数で説明（説明変数）しようとする。現実の金融市場の解明に重要な説明変数が手元にない、もしくは十分に活用していない場合において、アクセス可能な説明変数のみで回帰分析を実行するとどのような結果にあるであろうか？理論的には既知のとおり、ほとんどの場合において推定結果にバイアスが生じる。現実的には、データベースなどへのアクセス制限から、重要な説明変数の取得が難しい場合が生じる。そのような場合、どのように分析するかは実際の分析作業に従事する担当者の判断によるであろう。

３点目として、将来時点における金融市場の（市場参加者の行動、規制など）構造が、過去・現在の延長線上にない可能性がある点についてである。金融市場の予測可能性については、従来からファイナンス分野において議論されている（例えば、Ohno and Ando（2015）及びそこで引用されている参考文献をなど参照されたい）。予測可能性についての統一的見解はでておらず、トレーディングなどに代表されるような、金融市場の分析結果を将来に向けての意思決定に利用する際、その判断が非常に難しい。言い換えれば、１点目、２点目の課題について十分なケアができたとして、「どのような場合に現在の分析結果が将来時点においても的確であるのか？」という問いに対して、統一的な回答が難しいように思われる。

その他にも様々な問題が考えられるが、本報告では、上の１点目、２点目の課題を（限定的ではあるが）解決することを目的とする。特に、大規模金融データに内在する構造を柔軟に把握するための手法について検討し、Ando and Bai（2015）によって開発された計量的手法について解説していく。Ando and Bai（2015）の計量的手法は、以下のような特徴がある。

特徴１：数万～数十万銘柄の金融資産収益率（それぞれの銘柄について）に、影響を与えると考えられる様々な経済要因・金融市場要因等の中から、実質的に影響を与える少数の要因を自動的に特定する。

特徴２：市場に影響を与える観測されない要因（例えば、などは観測可能であるが、金融市場参加者のセンチメントなどは直接観測することができない）を特定する。

特徴３：資産収益率の分布を特定せず、さらに、金融資産間の依存性を明示的に仮定しないことで、金融市場の特徴を柔軟に把握することができる。

特徴４：資産収益率の特徴が似通った金融銘柄を自動的に特定する。

特徴５：特定した市場に影響を与える観測されない要因に対して経済的解釈をおこなうことができる。

特徴１は、大規模な変数選択をそれぞれの銘柄について実行することに対応する。それぞれの銘柄に影響を与える経済要因・金融市場要因などは異なることから、銘柄ごとに変数選択を実行できる手法は実務において有用であろう。一般に、変数選択問題は膨大な計算量を必要とするが、Ando and Bai（2015）は効率的な計算手法を構築している。特徴２は、潜在共通ファクターの同定、及びその個数の選択が可能となったことを意味する。特

徴３は、時系列方向、及び銘柄間の依存性を許容している。Ando and Bai（2015）は特徴１で触れた経済・金融市場要因、及び特徴２で触れた市場全体に影響を与える観測されない要因などに依存性を持たせ、さらに、説明できない銘柄個別にも時間方向・銘柄間の依存性を許容するという、非常に弱い仮定に基づき手法を開発している。これは、既存の計量的手法において頻繁に仮定される、独立性という強い制約から解放されることを意味する。また、特徴４は、個別銘柄のクラスタリングを自動的に実行できるということを意味する。特に、クラスター数、個別銘柄のグルーピングの一致性なども理論的問題となるが、Ando and Bai（2015）ではそれらについても検討・解決している。さらに特徴5であるが、Ando and Bai（2015）は潜在共通ファクターの同定後、その同定された潜在共通ファクターを、それらと関連するであろうと予想される経済変数に回帰させて解釈をおこなうことを提案している。これは金融実務においても非常に重要であろう。

本報告の構成は以下のとおりである。２節においては、Ando and Bai（2015）のモデルについて解説し、モデルの推定法について３節で触れる。４節では、Ando and Bai（2015）により導かれた推定されたモデルの漸近理論について解説する。５節では実証分析への応用例に触れる。最後に、６節で総括がおこなわれる。

2.　計量モデル

いま、$t = 1, \cdots, T$ を時系列方向に関するインデックス、$i = 1, \cdots, N$ をパネルの観察数に関するインデックスとする。また、金融資産のグループ数を S と仮定し、資産 i がどのグループに属するかを記述するパラメータとして $g_i \in \{1, \cdots, S\}$ を考える。一般に S、及び $G = \{g_1, \cdots, g_N\}$ は未知のパラメータであるので、ここでも未知とする。Ando and Bai（2015）は金融資産 i の時刻 t における収益率を記述するため以下の高次元パネルデータモデルを考案している。

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta}_i + \boldsymbol{f}_{g_i,t}'\boldsymbol{\lambda}_{g_i,i} + \varepsilon_{i,t},$$
$$i = 1, \ldots, N,\ t = 1, \ldots, T. \tag{1}$$

ここで、$\boldsymbol{x}_i$は観測される説明変数であり、企業固有の情報、金融市場レベルの情報、マクロ経済レベルの情報などを含んだものである。また、観察できないr_{g_i}次元のファクターリターン$\boldsymbol{f}_{g_i,t}$も収益率を説明する変数として考えている。ここで、$\boldsymbol{\lambda}_{g_i,t}$がファクターリターン$\boldsymbol{f}_{g_i,t}$に対する収益率の反応度である。(1)式のモデルは

$$\boldsymbol{y}_i = X_i\boldsymbol{\beta}_i + F_{g_i}\boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_{g_i,i}, \quad i = 1, \ldots, N$$

と表現できる。ここで、

$$\boldsymbol{y}_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix},\ X_i = \begin{pmatrix} \boldsymbol{x}_{i1}' \\ \boldsymbol{x}_{i2}' \\ \vdots \\ \boldsymbol{x}_{iT}' \end{pmatrix},$$

$$F_j = \begin{pmatrix} \boldsymbol{f}_{j1}' \\ \boldsymbol{f}_{j2}' \\ \vdots \\ \boldsymbol{f}_{jT}' \end{pmatrix},\ \boldsymbol{\varepsilon}_i = \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iT} \end{pmatrix}$$

である。

一般に誤差項 ε_i には独立性を仮定してきた。しかし、金融市場の分析を想定した場合、銘柄間の相互依存関係は避けられない問題である。現在、金融市場で観察される価格情報を個々の銘柄ごとに分析する様々な手法は存在し、また、ガウシアンコピュラに代表されるようなコプラアプローチにより銘柄間の相互依存構造を扱う手法もある。しかしながら、個々の銘柄ごとに分析する場合、金融市場における依存関係などの重要な性質は無視され、コピュラアプローチではそのパラメ

ータ推計の難しさから比較的小規模なデータの分析（例えば、数百銘柄の同時分析）にすら制約がある。そのため、金融資産間の相互依存、極端なイベント発生の可能性などをはじめとする金融市場の性質を考慮しつつも、数千・数万のオーダーの個別銘柄からなる大規模市場データを実際に分析する問題は非常にチャレンジングな課題である。Ando and Bai（2015）では、誤差項 ε_i に Bai（2009）のアイデアを利用し、誤差項には時系列方向、及び観察単位の両方向に依存性を持たせている。

ファクターの構造$F_{g_i}\boldsymbol{\lambda}_i$には、Dynamic exact factor model (Geweke, 1977; Sargent and Sims, 1977), Static approximate factor model (Chamberlain and Rothschild, 1983), Generalized dynamic factor model (Forni et al., 2000; Forni and Lippi, 2001; Amengual and Watson, 2007; Hallin and Liska, 2007), Bayesian factor models (Aguilar and West, 2000; Lopes and West, 2004; Lopes et al., 2008; Ando, 2009; Bhattacharya and Dunson, 2011; Tsay and Ando, 2012）など様々なファクターモデルを利用できる。Ando and Bai（2015）は Static approximate factor model の構造を仮定している。

Ando and Bai（2015）に関連するパネルデータの研究としては、Ando and Bai (2014a, 2014b, 2014c), Bai (2009), Bonhomme and Manresa (2012), Diebold et al. (2008), Kose et al. (2008), Lin and Ng (2012), Moench and Ng (2011), Moench et al. (2012), Pesaran (2006). Sun (2005). Wang (2010), などがある。

3.　モデル推定

グループ数 S、ファクター構造 $\{F_1,\cdots, F_S\}$ のそれぞれの次元 $\{r_1,\cdots, r_S\}$ を固定した元で(1)式のモデルに含まれるパラメータは、β_i、G, $\{F_1,\cdots, F_S\}$、$\Lambda=(\lambda_1,\cdots,\lambda_N)'$ である。それは以下の目的関数を、ファクター構造（一意性のため）に対する制約$F_j'F_j/T = I$，$\Lambda_j'\Lambda_j = \Gamma_j$（ここで Γ_j はある対角行列）のもとで最小化することで推定される。

$$L(\boldsymbol{\beta}_1, .., \boldsymbol{\beta}_N, G, F_1, ..., F_S, \Lambda_1, ..., \Lambda_S)$$

$$= \sum_{i=1}^{N} \|\boldsymbol{y}_i - X_i\boldsymbol{\beta}_i - F_{g_i}\boldsymbol{\lambda}_{g_i,i}\|^2 + T\sum_{i=1}^{N} p(\boldsymbol{\beta}_i),$$

第２項目 $p(\beta_i)$ はパラメータ β_i に対する罰則であり、Ando and Bai（2015）は SCAD 罰則（Fan and Li（2001）を利用している。それは以下で与えられる。

$$p_{\kappa_i,\gamma}(|\beta_{ij}|) =$$

$$\begin{cases} \kappa_i|\beta_{ij}| & (|\beta_{ij}| \le \kappa_i) \\ \dfrac{\gamma\kappa_i|\beta_{ij}| - 0.5(\beta_{ij}^2 + \kappa_i^2)}{\gamma - 1} & (\kappa_i < |\beta_{ij}| \le \gamma\kappa_i) \\ \dfrac{\kappa_i^2(\gamma^2 - 1)}{2(\gamma - 1)} & (\gamma\kappa_i < |\beta_{ij}|) \end{cases},$$

ただし $\kappa_i > 0$, $\gamma > 2$である。

SCAD 罰則の代替として the lasso method (Tibshirani, 1996)、its variants (Zou, 2006; Yuan and Lin, 2006, Park and Casella, 2008), least-angle regression (Efron et al., 2004), elastic net (Zou and Hastie, 2005), the minimax concave penalty method (MCP; Zhang 2012), the Dantzig selector (Candes and Tao, 2007）なども利用可能である。日本語の書籍としては、安道（2014）などを参照されたい。

最適化においては $\{\beta_1, \cdots, \beta_N\}$, G 及び $\{F_1, \cdots, F_S, \Lambda_1, \cdots, \Lambda_S\}$ を繰り返し最適化していけばよい（詳細は Ando and Bai（2015）に記述されている）。例えば、$\{\beta_1, \cdots\cdots, \beta_N\}$, Gが与えられたもとで、グループ j に属する資産にのみ着目し、$T\times N_j$- 次元行列 $W_j=(w_{j1}, \cdots, w_{jN})$ を以下のように定義する。

$$\boldsymbol{w}_{ji} = \boldsymbol{y}_i - X_i\boldsymbol{\beta}_j.$$

　このとき、⑴式のパネルデータモデルはファクターモデルに帰着する。

$$\boldsymbol{w}_{ji} = F_j \boldsymbol{\lambda}_{j,i} + \boldsymbol{\varepsilon}_i.$$

　パラメータ $\{F_j, \Lambda_j\}$ の推定は、目的関数

$$\mathrm{tr}\left\{ \left(W_j - F_j \Lambda_j'\right)\left(W_j - F_j \Lambda_j'\right)' \right\}$$

をファクター構造に対する制約（$F_j' F_j / T = I$, $\Lambda_j' \Lambda_j = \Gamma_j$）のもとで最小化することとなる。また、$\{F_j, \Lambda_j\}$, g_i が与えられたもとでは β_i の推定は

$$(\boldsymbol{y}_i - X_i \boldsymbol{\beta}_i - F_{g_i} \boldsymbol{\lambda}_{g_i,i})'$$
$$(\boldsymbol{y}_i - X_i \boldsymbol{\beta}_i - F_{g_i} \boldsymbol{\lambda}_{g_i,i}) + Tp(\boldsymbol{\beta}_i)$$

の最小化による。結局、ある初期値から出発し、いま述べた方法で $\{\beta_1, \cdots\cdots, \beta_N\}$, G 及び $\{F_1, \cdots, F_S, \Lambda_1, \cdots, \Lambda_S\}$ を繰り返し最適化していけばよい。グループ数 S、ファクターの次元 $k_1, \cdots, k_S$、及び正則化パラメータ $\kappa_1, \cdots, \kappa_N$ の選択が問題となるが、以下の Ando and Bai（2015）のモデル選択基準により最適な値を選択できる。

$$\mathrm{PIC}(S, k_1, ..., k_S, \kappa_1, ..., \kappa_N)$$

$$= \frac{1}{NT} \sum_{j=1}^{S} \sum_{i; g_i = j} \left\| \boldsymbol{y}_i - X_i \hat{\boldsymbol{\beta}}_i - \hat{F}_{\hat{g}_i} \hat{\boldsymbol{\lambda}}_{\hat{g}_i,i} \right\|^2 + \frac{1}{N} \sum_{i=1}^{N} \hat{\sigma}^2 \log(T) \hat{p}_i + \sum_{j=1}^{G} k_j \times \hat{\sigma}^2 \left(\frac{T + N_j}{T N_j} \right) \log(T N_j) \tag{2}$$

　ここで $\hat{p}_i$ は推定された $\hat{\boldsymbol{\beta}}_i$ のうち0でない成分数、$\hat{\sigma}^2$ は $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} E(\varepsilon_{it}^2)$ の一致推定量である。⑵式で与えられた PIC を最小とするように $\{S, k_1, \cdots, k_S, \kappa_1, \cdots, \kappa_N\}$ は選択される。モデル選択の実行手順として、Ando and Bai（2015）は以下のアルゴリズムを考案している。

実行手順

Step 1.　候補となる正則化パラメータの値 $\{\kappa_1, \cdots, \kappa_N\}$、グループ数 S それぞれのグループに関するファクターの次元 $\{r_1, \cdots, r_S\}$. を用意する。

Step 2.　グループ数 S、各グループに関するファクターの次元 $\{r_1, \cdots, r_S\}$ を固定する。

Step 3.　グループ数 S それぞれのグループに関するファクターの次元 $\{r_1, \cdots, r_S\}$ の下で、正則化パラメータ κ_i $(i = 1, \cdots, N)$ を⑵式で与えられた PIC を最小とするように選択する。

Step 4.　現在与えられているグループ数 S 及び正則化パラメータ κ_i $(i = 1, \cdots, N)$ の下で、それぞれのグループに関するファクターの次元 $\{r_1, \cdots, r_S\}$ を⑵式で与えられた PIC を最小とするように選択する。

Step 5.　Step 3及び Step 4を収束するまで繰り返し、最終的な PIC の値を保存する。

Step 6.　グループ数 S を変えて Step 2～Steps 5を実行する。

Step 7.　保存した PIC の値を最小とするような正則化パラメータの値 $\{\kappa_1, \cdots, \kappa_N\}$、グループ数 S それぞれのグループに関するファクターの次元 $\{r_1, \cdots, r_S\}$. を選択する。

4．漸近理論

4.1　仮　　定

Assumption A: Group-specific pervasive factors

　$E\|\boldsymbol{f}_{j,t}\|^4 < \infty$ $j = 1, ..., S$,とし $T^{-1}\sum_{t=1}^{T} \boldsymbol{f}_{j,t} \boldsymbol{f}_{j,t}' \to \Sigma_{F_j}$ $(T \to \infty$,とする。ここで Σ_{F_j} は $r_j \times r_j$ 次元の正定値行列である。$\boldsymbol{f}_{j,t}$ と $\boldsymbol{f}_{k,t}$ $(j \neq k)$ は相関しているが、完全に相関しないとする。

Assumption B: Factor loadings

(B1)：$\Lambda_j = [\boldsymbol{\lambda}_{j,1},\ldots,\boldsymbol{\lambda}_{j,N_j}]'$は以下を満たすとする$E\|\boldsymbol{\lambda}_{j,i}^4\| < \infty$, $\|\boldsymbol{\lambda}_{j,i}\| > 0\|N_j^{-1}\Lambda_j'\Lambda_j - \Sigma_{\Lambda_j}\| \to \mathbf{0}$ as $N_j \to \infty$、ここでΣ_{Λ_j}は$r_j \times r_j$次元の正定値行列である。

(B2)：すべての i、j について $\boldsymbol{f}_{j,t}'\boldsymbol{\lambda}_{j,i}$ は strongly mixing processes とし mixing coefficients は $r(t) \leq \exp(-a_1 t^{b_1})$ を満たし、裾確率は$P(|\boldsymbol{f}_{j,t}'\boldsymbol{\lambda}_{j,i}| > z) \leq \exp\{1-(z/b_2)^{a_2}\}$を満たすとする。ここで a_1、a_2、b_1、b_2は正の定数である。

Assumption C: Error terms

誤差項 ε_t の期待値は0とするが、時系列方向、及び観察単位の両方向に依存性を持つとする。また正の定数 $C < \infty$があり全ての N T について以下が成り立つとする。

(C1)：$E[\varepsilon_{it}] = 0$ for all i and t;

(C2)：$E[\varepsilon_{it}\varepsilon_{js}] = \tau_{ij,ts}$ with $|\tau_{ij,ts}| \leq |\tau_{ij}|$ for some τ_{ij} for all (t, s), and $N^{-1}\sum_{i,j=1}^N |\tau_{ij}| < C$; and $|\tau_{ij,ts}| \leq |\eta_{ts}|$ for some η_{ts} for all (i, j), and $T^{-1}\sum_{t,s=1}^N |\eta_{ts}| < C$. In addition, $(TN)^{-1}\sum_{i\ j\ t\ s=1} |\tau_{ij,ts}| < C$.

(C3)：For every (s, t), $E[|N^{-1/2}\sum_{i=1}^N(\varepsilon_{is}\varepsilon_{it} - E[\varepsilon_{is}\varepsilon_{it}])|^4] < C$.

(C4)：$T^{-2}N^{-1}\sum_{t,s,u,v}\sum_{i,j}|\mathrm{cov}(\varepsilon_{is}\varepsilon_{it}, \varepsilon_{js}\varepsilon_{jt})| < C$.

(C5)：$T^{-1}N^{-2}\sum_{t,s}\sum_{i,j,k,l}|\mathrm{cov}(\varepsilon_{it}\varepsilon_{jt}, \varepsilon_{ks}\varepsilon_{lt})| < C$.

(C6)：すべての i について、ε_{it} は strongly mixing processes とし mixing coefficients は $r(t) \leq \exp(-a_1 t^{b_1})$ を満たし、裾確率は $P(|\boldsymbol{f}_{j,t}'\boldsymbol{\lambda}_{j,i}| > z) \leq \exp\{1-(z/b_2)^{a_2}\}$を満たすとする。ここで a_1、a_2、b_1、b_2は正の定数である。

(C7)：ε_{it} は$\boldsymbol{x}_{js}$, $\boldsymbol{\lambda}_{j,i}$ $\boldsymbol{f}_{j,s}$とは独立とする。

Assumption D: Observable observable risk factors

(D1)：x_{it} は以下を満たすとする。$\max_{1\leq i\leq N} T^{-1}\|X_i\|^2 = O_p(N^{\alpha})$ with $\alpha < 1/8$. また $N/T^2 \to 0$とする。

(D2)：以下を定義する $D_j = \frac{1}{NT}\sum_{i;g_i=j} X_i' M_{F_j} X_i$, $E_j = \mathrm{diag}\{E_{j1},\ldots,E_{jS}\}$, $L_j = (L_{j1}',\ldots,L_{jS}')'$, ここで $E_{jk} = \frac{1}{N}\sum_{i;g_i=j,g_i^0=k}(\boldsymbol{\lambda}_{k,i}^0 \boldsymbol{\lambda}_{k,i}^{0\,\prime}) \otimes I_T$, $L_{jk} = \sum_{i;g_i=j,g_i^0=k}\frac{1}{NT}\boldsymbol{\lambda}_{k,i}^0 \otimes M_{F_j}X_i$とする。ここで$g_i^0$は真のグループ$\boldsymbol{\lambda}_{k,i}^0$は真のファクターローディングとする。いま $A = \{F_j : F_j'F_j/T = I, j = 1,\ldots,S\}$として以下を仮定する。

$$\sum_{j=1}^{S}(D_j - L_j'E_j^- L_j)$$

は全ての（F_1, ⋯, F_S）$\in AG$ について正定値行列である。ここでE_j^-は一般化逆行列である。

Assumption E: Number of units in each group

グループ j に属する銘柄数 N_j は$0 < \underline{a} < N_j/N < \bar{a} < 1$を満たす。

Assumption F: Central limit theory

いま$X_{i,\beta_i^0\neq 0}$を真のパラメータ$\boldsymbol{\beta}_i^0$が0でない部分に対応する X_i の部分行列とする。以下を仮定する。

$$\frac{1}{\sqrt{T}}X_{i,\beta_i^0\neq 0}' M_{F_g^0}\boldsymbol{\varepsilon}_i \to_d N(\mathbf{0}, J_i(F_g^0)),$$

ここで$J_i(F_g^0)$は

$$\frac{1}{T}X_{i,\beta_i^0\neq 0}' M_{F_g^0} E[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'] M_{F_g^0} X_{i,\beta_i^0\neq 0}.$$

の収束先（T →∞）である。

Assumption G:

いま$\Omega_{k\ell}=E[\boldsymbol{\varepsilon}_k\boldsymbol{\varepsilon}_\ell']$とし、

$$B_{NT,i}=\frac{1}{N^2T}\sum_{k\neq i}^{N}\sum_{\ell\neq i}^{N}X_k'M_{F_g^0}\Omega_{k\ell}M_{F_g^0}X_\ell=o_p(1).$$

を仮定する。

4.2　漸近理論

本節では Ando and Bai（2015）により導かれた漸近理論についての結果を紹介する。いま$\{F_j^0,\ j=1,...,S\}$を真のファクターとする。

Theorem 1 : Ando and Bai (2015)
Consistency.

Assumptions A–E を仮定する $\kappa_i\to 0$、$T\times\kappa_i\to\infty$のもとで推定されたファクター$\{\hat{F}_j,\ j=1,\cdots,S\}$は一致性をもつ。

$$T^{-1}\|\hat{F}_j-F_j^0H_j\|^2=o_p(1),\quad j=1,...,S, \tag{3}$$

ここで$H_j^{-1}=V_{j,N_jT}(F_j^0\hat{F}_j/T)^{-1}(\Lambda_j^{0\prime}\Lambda_j^0/N_j)^{-1}$,
としV_{j,N_jT}

$$\left[\frac{1}{N_jT}\sum_{i;\hat{g}_i=j}^{N_j}(\boldsymbol{y}_i-X_i\hat{\boldsymbol{\beta}}_i)(\boldsymbol{y}_i-X_i\hat{\boldsymbol{\beta}}_i)'\right]\hat{F}_j$$
$$=\hat{F}_jV_{j,N_jT}.$$

を満たす。

Theorem 2 : Ando and Bai (2015)
Consistency of the estimator of group membership.

Theorem 1 の仮定が成り立つとする。このとき全ての $\tau>0$について

$$P\left(\sup_{i\in\{1,...,N\}}\left|\hat{g}_i-g_i^0\right|>0\right)=o(1)+o(N/T^\tau).$$

が成り立つ。

つまり、資産収益率の特徴が似通った金融銘柄を自動的に特定することが理論的に保証されている。

次に変数選択の一致性、及び漸近正規性について述べる。いま$\boldsymbol{\beta}_i^0=(\boldsymbol{\beta}_{i10}{}',\boldsymbol{\beta}_{i20}{}')'$を真のパラメータの値とし$\hat{\boldsymbol{\beta}}_i=(\hat{\boldsymbol{\beta}}_{i1}',\hat{\boldsymbol{\beta}}_{i2}')'$を推定されたパラメータとし$\boldsymbol{\beta}_{i20}$のみ 0 とする。

Theorem 3 : Ando and Bai (2015)
Asymptotic normality and variable selection consistency.

仮定 $A\sim G$ が成り立つとし、$\kappa_i\to 0$、$T\times\kappa_i\to\infty$とする。このとき$\sqrt{T}(\hat{\boldsymbol{\beta}}_{i1}-\boldsymbol{\beta}_{i10})$は漸近的に平均0分散共分散行列$R_i(F_{g_i^0}^0)$の正規分布に従う。

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_{i1}-\boldsymbol{\beta}_{i10})\to_d N\left(\mathbf{0},R_i(F_{g_i^0}^0)\right),$$

ここで

$$R_i(F_{g_i^0}^0)=D_i(F_{g_i^0}^0)^{-1}J_i(F_{g_i^0}^0)D_i(F_{g_i^0}^0)^{-1},$$

とし$D_i(F_{g_i^0}^0)$は

$$\frac{1}{T}\left(X_{i,\beta_i^0\neq 0}'M_{F_{g_i^0}^0}X_{i,\beta_i^0\neq 0}+\Sigma_i(\kappa_i)\right),$$

の収束先（$T\to\infty$）である。ここで、$X_{i,\beta_i^0\neq 0}$はX_iの部分行列であり、真のパラメータβ_i^0の0でない成分に対応する。ただし

$$\begin{aligned}\Sigma_i(\kappa_i)\quad=\quad&\mathrm{diag}\left\{p'_{\kappa_i,\gamma}(|\beta_{i1}^0|)/|\beta_{i1}^0|,\right.\\&\left.\ldots,p'_{\kappa_i,\gamma}(|\beta_{iq_i}^0|)/|\beta_{iq_i}^0|\right\}.\end{aligned}$$

とする。さらに変数選択の一致性がなりたつ。

$$P(\hat{\boldsymbol{\beta}}_{i,2}=\mathbf{0})\to 1,\quad N,T\to\infty.$$

それぞれの銘柄に影響を与える経済要因－金融市場要因などは異なるが、銘柄ごとに変数選択を実行できることが示された。

Theorem 4 : Ando and Bai (2015)
Consistent model identification.

Theorem 3 の仮定が成り立つとする。このとき *PIC* は真のグループ数、真のファクター数、及び真の説明変数を同定する。

つまり、大規模金融データの適切なクラスタリングが理論的に保証されることを意味する。

5. 実証分析への応用例

本節では、金融市場開のグループ－依存構造を考慮した大規模金融データの実証分析例について報告する。Ando and Bai（2014b）は(1)式の β_i をグループ内で共通として中国A株市場－B株市場に属する1,000銘柄以上の株式（無リスク資産に対する超過）収益率に関する分析をおこなった。2002年～2010年までの月次収益率を利用して、A株市場・B株市場のグループ数について検討し、その結果、6つのグループに分類されることが明らかとなった。また、それぞれのグループが影響を受けるファクターは違うという実証分析を得ている。表1は、特定された$\hat{g}_i$ $i=1, \cdots, N$と以下の分類による2×2テーブルである。(1) Location of stock exchanges（市場）、(2) Types of share（A株・B株）、及び(3) Industry（産業分類）特定された$\hat{g}_i$は市場の場所、産業分類とはあまり関連が見られないが、A株・B株という株式の種類には関連が見られる。つまり、A株であるか、B株であるかという要因が重要であることを示唆している。詳細については Ando and Bai (2014b) を参照されたい。

また、Ando and Bai（2015）は6,000銘柄以上の国際金融関連株式について分析をおこ

表1　Ando and Bai（2014b）特定された $\hat{g}_i$ $i=1,...,N$と以下の分類による 2×2テーブル
(1) Location of stock exchanges, (2) Types of share, and (3) Industry.

	Classification	$G1$	$G2$	$G3$	$G4$	$G5$	$G6$
(1)	Location of stock exchanges						
	Shanghai stock exchange	179	67	132	77	105	81
	Shenzhen stock exchange	125	29	94	64	95	93
(2)	Types of share						
	A-shares	211	95	224	141	196	172
	B-shares	93	1	2	0	4	2
(3)	Category based on Industry						
	Chemicals, Construction, Manufacturing	76	15	70	36	53	49
	Food, Beverages, Personal Goods	40	14	24	21	25	13
	Gas, Metals, Mining, Oil	42	16	16	17	17	26
	Banks, Financial Services, Real Estate	30	6	25	15	23	17
	Retails	29	18	26	19	19	21
	Utilities	17	8	16	6	19	9
	Pharmaceuticals, Health	24	6	21	10	16	12
	Information Technology	27	8	21	9	19	11
	Others	11	4	4	5	7	13

なっている。特に、2007年以降の米国サブプライム問題に端を発する金融危機について検討し、以下の５つの期開について分析をおこなっている。

Period 1 :
May 1 2006 to December 31, 2006
Period 2 :
May 1 2007 to December 31, 2007
Period 3 :
February 1 2008 to August 31, 2008
Period 4 :
September 1 2008 to March 31, 2009
Period 5 :
May 1 2009 to December 31, 2009

その結果、Period 1にはグループ数は11であったが、リーマンブラザーズ破綻など深刻なイベントを含む Period 4にはグループ数は17へと増加し、市場の不確実性が高まっているという報告をおこなっている。また、グローバルファクターよりも米国ファクター、EU ファクターの影響が Period 4においては非常に大きいという実証結果を得ている。詳細については Ando and Bai（2015）を参照されたい。このように、金融市場間のグループ・依存構造を考慮した大規模金融データの分析手法を用いることによって新たな知見を得ることができる。

6．結　　論

本報告では、Ando and Bai（2015）によって提案された計量的手法について解説し、その応用例についても報告した。現実の金融市場の構造を把握可能とするために、Ando and Bai（2015）の手法は柔軟な・精徴化された仮定のもとで計量的モデリングを実行する。そのため、大規模金融データを効率的・的確に分析するうえで非常に有用な道具となる。さらに、的確かつ効率的な分析をおこなえるため、グローバル金融市場のモニタリング、新しい資産運用戦略－ペアトレード戦略の提案、金融市場の変化に対してロバストなリスク管理ツールの構築、グローバルな観点からの類似企業探索など、金融市場参加者の様々な分析ニーズへ対応することが可能となる。さらに、金融当局者においても、システミツクリスクの事前計測、各金融機関の市場リスクエクスポージャー把握など様々な場面において、Ando and Bai（2015）の手法は新しい情報を提供するであろう。今後も金融市場の様々な特徴を考慮した計量的手法の開発が望まれる。

謝　辞

本報告で触れた計量的手法などについては、以下の国際会議、研究集会、大学セミナーなどにおいて参加者の方々に有益なコメントをいただきました。The 20 th international panel data conference（Tokyo）, the 11th International Symposium on Econometric Theory and Applications（Hitotsubashi University）、第16回ノンパラメトリツク統計解析とベイズ統計（於：慶磨義塾大学）、統計科学の新地平：ノンパラメトリツク統計解析とベイズ統計（於：東京大学）、Department Seminar at University of Sydney, Department Seminar at UCLA. ここに記して御礼申し上げます。

【参考文献】

(1) Amengual, D. and M. W. Watson. 2007. Consistent estimation of the number of dynamic factors in a large N and T panel. Journal of Business and Economic Statistics 25: 91-96.
(2) Ando, T.（2013）. Inference and model selection for large panel data models with a factor structure. Working Paper.
(3) 安道知寛（2014）。高次元データ分析の方法。朝倉書店。
(4) Ando, T. and Bai, J. 2014a. Asset pricing with a general multifactor structure.

Journal of Financial Econometrics, forthcoming.

(5) Ando, T. and Bai, J. 2014b. Panel data models with grouped factor structure under unknown group membership. *Journal of Applied Econometrics*, forthcoming.

(6) Ando, T. and Bai, J. 2014c. Selecting the regularization parameters in high-dimensional panel data models: consistency and efficiency. *Econometric Reviews*, forthcoming.

(7) Ando, T. and Bai, J. 2014d. A simple new test for slope homogeniety in panel data models with interactive effects. Working Paper.

(8) Ando, T. and Bai, J. 2015. The subprime financial crisis and spillover effects on global financial markets: A high-dimensional panel data clustering approach. Working Paper.

(9) Ando, T. and Tsay, R. 2015. On model selection for large panel data with interactive effects. Working Paper.

(10) Bai, J. 2009. Panel data models with interactive fixed effects. *Econometrica* 77, 1229–1279.

(11) Bai, J. and Ng, S. 2002. Determining the number of factors in approximate factor models. *Econometrica*, 70, 191–221.

(12) Baca, S.P., Garbe, B.L., Weiss, R.A., 2000. The rise of sector effects in major equity markets. *Financial Analysts Journal* 56, 34–40.

(13) Beckers, S., Connor, G., Curds, R., 1996. National versus global influences on equity returns. *Financial Analysts Journal* 52, 31–39.

(14) Bester, A. and Hansen, C. 2012. Grouped effects estimators in fixed effects models. *Journal of Econometrics*, Forthcoming.

(15) Bonhomme, S. and Manresa, E. 2014. Grouped patterns of heterogeneity in panel data. *Econometrica*, forthcoming.

(16) Candes, E. and Tao, T. (2007), The Dantzig selector: statistical estimation when p is much larger than n, Annals of Statistics, 35, 2313–2351.

(17) Cavaglia, S., Brightman, C., Aked, M., 2000. The increasing importance of industry factors. *Financial Analysts Journal* 56, 41–54.

(18) Chamberlain, G. and M. Rothschild. 1983. Arbitrage, factor structure and mean-variance analysis in large asset markets. Econometrica 51: 1305–1324.

(19) Connor, G. and Korajzcyk, R. 1986. Performance measurement with the arbitrage pricing theory: a new framework for analysis. *Journal of Financial Economics*, 15, 373–394.

(20) Dimitriou, D. and Kenourgios, D. (2013). Financial crises and dynamic linkages among international currencies. *Journal of International Financial Markets, Institutions and Money*, 26, 319–332.

(21) Dornbusch, R., Park,Y. and Claessens,S. 2000. Contagion:understanding how it spreads. The World Bank Research Observer 15, 177–197.

(22) Fama, E. F. and French, K. R. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33, 3–56.

(23) Fan, J. and Li, R. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96, 1348–1361.

(24) Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized

likelihood and its oracle properties. *Journal of the American Statistical Association* 96, 1348–1361.

(25) Fan, J. and Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *Annals of Statistics*, 32, 928–961.

(26) Forbes, K. J. and Rigobon, R. (2002). No contagion, only interdependence: Measuring stock market comovements. *Journal of Finance*, 57, 2223–2261.

(27) Forni, M., M. Hallin, M. Lippi, and L. Reichlin. 2000. The generalized dynamic factor model: identication and estimation. Review of Economics and Statistics, 82: 540–554.

(28) Forni, M. and M. Lippi. 2001. The generalized factor model: representation theory. Econometric Theory 17: 1113–1141.

(29) Geweke, J. 1977. The dynamic factor analysis of economic time series. In: Aigner, D. J., Goldberger, A. S. (eds), Latent Variables in Socio-Economic Models. Amsterdam: North-Holland, pp. 365–383

(30) Griffin, J. M. 2002. Are the Fama and French Factors Global or Country Specific? *Review of Financial Studies* 15:783–803.

(31) Griffin, J.M., Karolyi, G.A., 1998. Another look at the role of the industrial structure of markets for international diversification strategies. *Journal of Financial Economics* 50, 351–373.

(32) Hallin, M. and R. Liska. 2007. The generalized dynamic factor model: determining the number of factors. Journal of the American Statistical Association 102: 603–617.

(33) Hamao, Y., Masulis, R., and Ng, V. (1990). Correlation in price changes and volatility across international stock markets. *Review of Financial Studies*, 3, 281–307.

(34) Heston, S.L., Rouwenhorst, K.G., 1994. Does industrial structure explain the benefits of industrial diversification? *Journal of Financial Economics* 36, 3–27.

(35) Heston, S.L., Rouwenhorst, K.G., 1995. Industry and country effects in international stock returns. *Journal of Portfolio Management,* 21, 53–58.

(36) Hou, K, Karolyi, G.A. and Kho, B.-C. (2011) What Factors Drive Global Stock Returns? *Review of Financial Studies* 24:2527–2574.

(37) Kaminsky, G.,Reinhardt,C. and Vegh,C. 2003. The unholy trinity of financial contagion. *Journal of Economic Perspectives* 17, 51–74.

(38) Kuo, W., Satchell, S.E., 2001. Global equity styles and industry effects: the pre-eminence of value relative to size. *Journal of International Financial Markets, Institutions and Money*, 11, 1–28.

(39) Lee, B. -S., Rui, O. M., and Wang, S. S. (2004). Information transmission between the NASDAQ and ASIAN second board markets. *Journal of Banking and Finance*, 28, 1637–1670.

(40) Lin, C. and Ng. S. 2012. Estimation of Panel Data Models with Parameter Heterogeneity When Group Membership is Unknown. *Journal of Econometric Methods*, 1, 42–55.

(41) Longstaff, F.A. 2010 The subprime credit crisis and contagion in financial markets. *Journal of Financial Economics* 97, 436–450.

(42) Mallows, C. L. 1973. Some comments on *Cp. Technometrics* 15, 661–675.

(43) Ohno, S. and Ando, T. 2015 Stock return predictability: A factor augmented predictive regression system approach. *Econometric Reviews*, forthcoming.
(44) Park, T. and Casella, G. (2008), The Bayesian Lasso, Journal of the American Statistical Association, 103, 681-686.
(45) Pesaran, M. H. 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74, 967-1012.
(46) Pesaran, M.H., Pick, A., 2007. Econometric issues in the analysis of contagion. *Journal of Economic Dynamics & Control* 31, 1245-1277.
(47) Roll, R., 1992. Industrial structure and the comparative behavior of international stock market indices. *Journal of Finance* 47 (1), 3.42.
(48) Sargent, T. J. and C. A. Sims. 1977. Business cycle modeling without pretending to have too much a priori economic theory. In: Sims C. et al. (eds), New Methods in Business Cycle Research. Federal Reserve Bank of Minneapolis, Minneapolis.
(49) Stock, J. H. and Watson, M. W. 2002. Forecasting using principal components from a large number of observable risk factors. *Journal of the American Statistical Association*, 97, 1167-1179.
(50) Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, B58, 267-288.
(51) Wang, P. 2010. Large dimensional factor models with a multi-level factor structure. Working paper, Department of Economics, HKUST.
(52) Yuan, M. and Lin, Y. (2006), Model selection and estimation in regression with grouped variables, Journal of the Royal Statistical Society B, 68, 49-67
(53) Zhang, C. H. 2010. Nearly unbiased variable selection under minimax concave penalty. *Annals of Statistics*, 38, 894-942.
(54) Zou, H. (2006), The adaptive Lasso and its oracle properties, Journal of the American Statistical Association, 101, 1418-1429.
(55) Zou, H. and Hastie, T. (2005) Regularization and variable selection via the Elastic Net, Journal of the Royal Statistical Society B, 67, 301-320.

（あんどう　ともひろ）