パネルデータ分析の最新の手法について
ファクターモデルと非線形モデル

奥井　亮

平成24年6月24日改訂版

このスライドについて

- このスライドは、平成２４年６月に日本経済学会春季大会(北海道大学)で行った、チュートリアルセッションで使用したスライドを改訂したものです。学会参加者の皆様、また北村行伸先生にはお世話になりました。記して感謝します。

- スライドの使用は自由ですし、適宜、修正して自分用のスライドに作りかえていただいて、かまいませんが、私は、一切の責任をおいません。

- このスライドは、今後とも私のホームページ
`http://www.kier.kyoto-u.ac.jp/~okui/`
に掲載を続ける予定です。また、参考文献の一覧も別ファイルとしてホームページに掲載しています。

はじめに

- 近年のパネルデータ分析の研究成果について紹介するチュートリアルを担当させていただきます。
- パネルデータ分析の基礎的な知識は前提とさせていただきます。
- 証明などは追わず、研究成果の概観と、基本的な考え方を中心に講義を行います。従いまして、より詳しい理論や計算方法などは、参考文献にあげました論文などで調べていただければと思います。

内容

以下の２つのトピックを扱います。

１. 大きなパネルデータにおけるファクターモデル

２. 固定効果の入った非線形モデルの推定とバイアス修正法

近年は、他にも、ノンパラメトリックな識別問題と推定法、横断面での相関の取り扱い方や、構造モデルの推定など、盛んに研究がおこなれ、また実証上も重要な問題も多くありますが、上の二つに絞って話をさせていただきます。

トピックその1

大きなパネルデータにおけるファクターモデル

モデル

パネルデータとして、(y_{it}, x_{it})が観測できるとします。y_{it}とx_{it}は

$$y_{it} = x_{it}'\beta + \lambda_i' F_t + w_{it} \tag{1}$$

という関係があるとします。ここで、観測できない部分は$\lambda_i' F_t + w_{it}$であり、

- F_tをファクター
- λ_iをファクターローディング

と呼びます。

パネルデータモデルの有用性

なぜファクターモデルが有用なのかをみるために、まず、教科書的なパネルデータの有用性の議論から見ていきましょう。

パネルデータは、観測できないが説明変数と相関のある変数が、ある構造をもっている場合、その変数を観測することなしに、欠落変数のバイアスを避けることができるのが、重要な利点です。

個人効果の入ったモデル

例えば、そのような観測できない変数が時間を通して一定であれば、固定効果モデル

$$y_{it} = x'_{it}\beta + \mu_i + w_{it} \tag{2}$$

としてモデル化します。そして、$\dot{y}_{it} = y_{it} - \bar{y}_i$ などと変数変換をすると、

$$\dot{y}_{it} = \dot{x}'_{it}\beta + \dot{w}_{it} \tag{3}$$

とし μ_i を消すことができるので、μ_i を観測できないことから来る欠落変数のバイアスを避けることができます。

個人効果と時間効果

さらに、欠落変数のうち、各個人への影響は同じだが、時間とともに変化するものがあれば、時間効果としてη_tを

$$y_{it} = x'_{it}\beta + \mu_i + \eta_t + w_{it} \tag{4}$$

のように、モデルに加えます。推定においては、$\ddot{y}_{it} = y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}$というように変換をすると、

$$\ddot{y}_{it} = \ddot{x}'_{it}\beta + \ddot{w}_{it} \tag{5}$$

となり、μ_iもη_tも消え、欠落変数のバイアスを避けることができます。

ファクターモデルの有用性

- ファクターモデルは、

$$y_{it} = x'_{it}\beta + \lambda'_i F_t + w_{it} \tag{6}$$

  ですので、個人効果や時間効果を入れたモデルのさらなる一般化であり、かなり広い範囲のバイアスをもたらすような欠落変数の影響を取り除くのに有用であると考えられます。

- また、横断面での相関や自己相関を表現するモデルとしても有用です。

- また、問題によっては、ファクターの値自体に興味がある場合があります。

ファクターそのものに興味がある場合

- 例えば、ある一つの経済指標を表すのに多くの変数が利用可能な場合、i はそれぞれの指標を表すものとして、

$$y_{it} = \lambda_i' F_t + w_{it} \tag{7}$$

としてモデル化し、小さい次元のベクトルである F_t を推定し、それを目的とする経済指標として使用するということが考えられます。こちらの方がファクターモデルの主な使われ方かも知れません。

- マクロ実証分析の例では、Stock and Watson (1999) やBernanke, Boivin and Eliasz (2005) 等があります。

ファクターの推定

まず、

$$y_{it} = \lambda_i' F_t + w_{it} \tag{8}$$

というモデルを考え、ファクターとファクターローディングを推定する方法を紹介します。

おそらく最もよく使われている推定法は、Stock and Watson (2002) による、主成分分析による方法です。Y を y_{it} を並べた $T \times N$ の行列とします。λ_i と F_t の次元を r とし、それらを並べた行列として Λ と F を定義します。Λ は $N \times r$ で F は $T \times r$ の行列です。

主成分分析法

- $\tilde{F}$ を YY' の固有値が大きいものから r までの固有ベクトルとします。

- $\tilde{\Lambda} = \tilde{F}'Y/T$ とします。

- これは、

$$\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \lambda_i' F_t)^2 \tag{9}$$

  を $F'F/T = I$ という標準化の下で、Λ と F について最小した解と同じになります。

漸近理論

- 漸近理論は、N と T の両者が無限に行く、2重漸近分析を使用します。

- Stock and Watson (2002) は、一致性を証明しました。

- 漸近分布は Bai (2003) によって導出されています。

- 漸近理論は、w_{it} に弱い相関を、時系列方向にも横断面方向にも許容しています。

識別問題

- 追加的な条件なしでは、Λ と F は識別不能です。これは、$Y = F\Lambda' + w$ というモデルのため、任意の正則行列 A をとって $F^* = FA$ かつ $\Lambda^* = \Lambda A^{-1}$ としても観測上同値なモデルが得られるためです。

- 良く使われる識別条件は、$F'F/T = I$ かつ $\Lambda'\Lambda$ が対角行列というものです。この条件は、主成分分析による推定と整合的です。

- 他にもさまざまな識別条件を考えることができます。それらは、Bai and Ng (2011) で議論されています。

ファクターの数の選び方

推定にあたっては、ファクターの数rを決める必要があります。

- Bai and Ng (2002) はファクターの数を選ぶための情報量基準を提唱しました。今のところ、この方法が最もよく使われているのではないでしょうか。

- Onatski (2009) は、ファクターの数を検定する方法を開発しました。

有効な推定量

- w_{it}がi.i.d.でない場合には、主成分分析推定量は、有効になりません。

- Breitung and Tenhofen (2011)、Choi (2011)やBai and Li (2012)などにより、主成分分析法よりも、有効な推定量が提唱されています。

- Iwakura and Okui (2012)では、ファクターの有効性限界を導出し、それらの推定量が有効になる条件を導いています。

回帰モデルへの応用

ファクター構造をもつ回帰モデルの推定へ主成分分析を応用することができます。初めに紹介した

$$y_{it} = x_{it}'\beta + \lambda_i' F_t + w_{it} \tag{10}$$

というモデルを考えます。

Bai (2009) が推定法を提唱しています。方法は、

$$\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - x_{it}'\beta - \lambda_i' F_t)^2 \tag{11}$$

を β と Λ、それに F について最小化することで行います。

やはり、$F'F/T = I$ で $\Lambda'\Lambda$ が対角行列であるという制約をおいて最大化します。

推定法に関する注意

- 実際の推定は数値計算上のトリックが必要となります。Bai (2009) を参照して下さい。
- また推定量は、漸近的にバイアスがでることがありますので、バイアス修正推定量を考えたほうがよいでしょう。
- Bai (2009) は x_{it} が強外生であると仮定しています。この仮定を外し先決変数の場合も取り扱ったものとして、Moon and Weidner (2010) があります。

他の推定法

- Pesaran (2006) は、モデルに $(\bar{y}_t, \bar{x}_t)$ を追加的な回帰変数として入れ、

$$y_{it} = x'_{it}\beta + \bar{y}_t\rho + \bar{x}'_t\eta + v_{it} \tag{12}$$

というモデルを推定することで、ファクターの影響を取り除ける場合があることを示しました。必要となる条件は強いですが、簡便な方法であるため、実用上非常に便利です。

- Sarafidis and Yamagata (2010) はファクターを推定し、$M = I - F(F'F)F^{-1}$ という行列をかけてファクターとの相関をなくした操作変数を用いた操作変数推定量を提唱しました。

ファクターモデルのまとめ

- ファクターモデルは、時間を通じて変化しないが個人ごとに異なる、あるいは個人間で同じだが時間を通じて変化する変数が引き起こす、欠落変数のバイアスを考慮するのに有用なモデルです。

- また推定法は、主成分分析など、簡便な方法も多く提唱されており、十分に実用に耐えるものです。

- ファクター自体も、多くの変数の情報を、少ない変数でまとめることができるため、色々な実証研究で有用でしょう。

トピックその２

固定効果の入った非線形モデルの推定とバイアス修正法

モデル

$\{y_{it}, x_{it}\}$ というパネルデータがあり、次の尤度をもたらすモデルを考えます。

$$f_{it}(y_{it}|\theta, \alpha_i) = f(y_{it}|x_{i1}, \ldots, x_{iT}; \theta, \alpha_i) \tag{13}$$

- y_{it}: 内生変数。x_{it}: 外生変数。
- θ: 推定したいパラメーター。
- α_i: 固定効果。
- N が大きくて、T がそれほど大きくない状況を考えます。ただ T が小さくても問題です。$T \geq 7$ ぐらいの場合を考えてください。

例

例として、２項選択モデルを考えましょう。つまり $y_{it} \in \{0,1\}$ の場合です。パラメトリックなモデルの場合、F をなにかの分布関数として、

$$E(y_{it}|x_{it}) = F(\theta' x_{it} + \alpha_i) \tag{14}$$

とかけます。

また、尤度は、

$$f_{it}(y_{it}|\theta, \alpha_i) = F(\theta' x_{it} + \alpha_i)^{y_{it}}(1 - F(\theta' x_{it} + \alpha_i))^{1-y_{it}} \tag{15}$$

となります。

固定効果推定量

ここでいう、固定効果推定量は、α_iを母数として扱った、最尤推定量のことです。

つまり、次の$\widehat{\theta}_T$として、定義されます。

$$\widehat{\theta}_T = \arg\max_{\theta} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \log f(y_{it}|\theta, \widehat{\alpha}_i(\theta)) \tag{16}$$

ここで、

$$\widehat{\alpha}_i(\theta) = \arg\max_{\alpha} \sum_{t=1}^{T} \log f(y_{it}|\theta, \alpha) \tag{17}$$

です。つまり、それぞれのα_iを母数として取り扱うということです。

問題

問題は、$N \to \infty$ で T を固定とした漸近理論を考えたときに、$\widehat{\theta}_T$ は、一致性を持たないことです。

なぜこのような問題が起こるのか考えて見ましょう。まず、目的関数は、

$$L_T(\theta) = E\left(\sum_{t=1}^{T} \log f(y_{it}|\theta, \widehat{\alpha}_i(\theta))\right) \tag{18}$$

に収束します。ここでのポイントは、$\widehat{\alpha}_i(\theta)$ はそのままで、何かに収束するということはないということです。さて、通常の議論から、

$$\widehat{\theta}_T \to_p \theta_T = \arg\max_{\theta} L_T(\theta) \tag{19}$$

を示すことが出来ます。

Incidental Paramteres Problem

一般的に、$\theta_T \neq \theta_0$ です。したがって、$\hat{\theta}_T$ は、一致性を持ちません。

これまでの議論でわかるとおり、問題は、$\hat{\alpha}_i$ の推定誤差が漸近的にも残ってしまうということです。ちなみに、

$$\theta_0 = \arg\max_{\theta} E\left(\sum_{t=1}^{T} \log f(y_{it}|\theta, \alpha_i)\right) \tag{20}$$

です。

この問題はIncidental Parameters Problem (Neyman and Scott (1948))と呼ばれます。

ちなみに、線形モデルの固定効果推定量は、$\hat{\theta}_T$ が一致性を持つ特殊な例です。

解決法

この固定効果推定量の不一致性の問題には２つの解決法があります。

- T が固定されていても、一致性を持つ推定量を考える。
- T が固定されているときの一致性はあきらめて、代わりに、T がそれほど大きくなくてもうまくいくように、バイアス修正をかける。

T が固定されている場合の一致性

T が固定されている状況で、一致性をもつような推定量は、これまでいろいろ研究されてきました。Arellano and Honore (2001) にまとめがあります。ただこの場合の問題は、

- 推定できないモデルが存在します。
- また、推定可能な場合でも、推定量の導出のやり方は、モデルに依存します。
- ある種のモデルでは、推定量の収束速度が $\sqrt{N}$ より遅くなります。たとえば、Honore and Kyriazidou (2000) などです。

バイアス修正法の意義

一方、ここで紹介するバイアス修正法は、T が固定されている状況での一致性はないものの、以下のような利点があります。

- このバイアス修正法は、適用範囲が広く、多くのモデルで使用可能です。
- 最尤法のバイアス修正なので、推定量の分散は小さくなります。
- また、ジャックナイフなどの自動的にバイアス修正をかける方法もありますので、実際に使用するのも簡単です。

バイアス修正の文献

- ここで紹介する内容は、Hahn and Newey (2004) で紹介されている方法です。Li, Lindsay and Waterman (2003) にその原点があります。

- また、Arellano and Hahn (2007) にまとめがあります。このチュートリアルも、Arellano and Hahn (2007) に大きくよっています。

- 特定のモデルに焦点を絞った分析としては、２項選択モデルの場合を扱った Fernandez-Val (2009) などがあります。

バイアスその一

ここでは、N と T が同じ速度で、無限に行くような、漸近理論を考えます。さて、このとき、通常のモデルでは、

$$\sqrt{NT}(\widehat{\theta}_T - \theta_T) \to_d N(0, \Omega) \tag{21}$$

となります。Ω は漸近分散ですが、ここの議論には直接関係しません。

さて、θ_T は θ_0 では、ないのですが、T が無限にいくとき、θ_0 に収束します。通常のモデルなら、

$$\theta_T = \theta_0 + \frac{B}{T} + O\left(\frac{1}{T^2}\right) \tag{22}$$

と書くことが出来ます。

バイアスその二

今仮に、$0 < \lim N/T < \infty$としましょう。すると、

$$\begin{aligned} \sqrt{NT}(\hat{\theta}_T - \theta_0) &= \sqrt{NT}(\hat{\theta}_T - \theta_T) + \sqrt{NT}\frac{B}{T} + O\left(\sqrt{\frac{N}{T^3}}\right) \qquad (23) \\ &\to_d N\left(B \lim \sqrt{\frac{N}{T}}, \Omega\right) \qquad (24) \end{aligned}$$

となります。つまり、漸近分布を考えると、Tが無限にいく状況でも、漸近分布の平均は、0にならず、バイアスがかかってることが、わかります。

この、バイアス項Bを取り除くことが、ここでの目的になります。

バイアス修正

もし、B を推定することが出来たとしましょう。そのような推定量を $\hat{B}$ とします。$\hat{B} \to_p B$ と仮定します。

そして、バイアス修正推定量を

$$\tilde{\theta}_T = \hat{\theta}_T - \frac{\hat{B}}{T} \tag{25}$$

と定義します。すると、

$$\begin{aligned} \sqrt{NT}(\tilde{\theta}_T - \theta_0) &= \sqrt{NT}(\hat{\theta}_T - \theta_T) + \sqrt{NT}\frac{B - \hat{B}}{T} + O\left(\sqrt{\frac{N}{T^3}}\right) && (26) \\ &\to_d N(0, \Omega) && (27) \end{aligned}$$

となり、バイアスがなくなることがわかります。

バイアス修正法に関するポイント

- このバイアス修正法では、漸近分散はΩのままで、変わりません。つまり、漸近的には、分散に影響を与えず、バイアスを修正できることがわかります。そのような例は、実はかなり稀です。通常は、バイアスを修正すると、分散が大きくなります。

- また最尤推定の場合には、Iwakura (2012) が$\tilde{\theta}_T$が有効推定量であることを示しました。

- 実際にどうやって、バイアスを推定するかですが、Hahn and Newey (2004) や、Arellano and Hahn (2007) に式が載っています。

他のアプローチ

- 目的関数に修正をかける方法。Arellano and Hahn (2007) を参照。

- スコア（一次の条件）に修正をかける方法。Arellano and Hahn (2007) を参照。

- Arellano and Bonhomme (2009) に固定効果に関して積分を取る方法で、バイアスが出ないようになるような方法が、紹介されています。

- Bonhomme (2011) は関数解析の手法を使い、固定効果を含むモーメント条件から固定効果の影響を取り除く方法と、その方法が可能であるための条件を提示しています。

ジャックナイフ法

バイアス修正法として、ここでは、ジャックナイフ法を考えます。この方法は、いちいちバイアスの式を計算する必要がないため、簡単にいろいろなモデルに応用可能ですし、有用です。ただ、推定量をいくつも計算する必要があるため、計算時間はかかるかもしれません。

まず、$\widehat{\theta}^{(t)}$ を t 期のデータを抜いて計算した推定量とします。つまり、

$$\widehat{\theta}^{(t)} = \arg\max_{\theta} \frac{1}{N} \sum_{i=1}^{N} \sum_{s \neq t} \log f(y_{is}|\theta, \widehat{\alpha}_i(\theta)), \tag{28}$$

$$\widehat{\alpha}_i^{(t)}(\theta) = \arg\max_{\alpha} \sum_{s \neq t} \log f(y_{is}|\theta, \alpha) \tag{29}$$

です。

ジャックナイフによるバイアス修正

そして、バイアス修正推定量を、

$$\breve{\theta}_T = T\widehat{\theta}_T - (T-1)\sum_{t=1}^{T}\widehat{\theta}^{(t)}/T \tag{30}$$

と定義します。

この推定量は、バイアスを次の $\breve{B}$ で修正した推定量です。つまり、

$$\breve{\theta}_T = \widehat{\theta}_T - \frac{\breve{B}}{T}, \tag{31}$$

$$\frac{\breve{B}}{T} = (T-1)\left(\frac{1}{T}\sum_{t=1}^{T}\widehat{\theta}^{(t)} - \widehat{\theta}_T\right) \tag{32}$$

として、バイアスを推定しているわけです。

ジャックナイフ法の説明

なぜこのジャックナイフ法はうまくいくのかを考えて見ましょう。

まず、$\widehat{\theta}$ は B/T のバイアスがかかっています。同じように $\widehat{\theta}^{(t)}$ にはそれぞれ、$B/(T-1)$ のバイアスがかかっています。したがって、

$$\left(\frac{1}{T}\sum_{t=1}^{T}\widehat{\theta}^{(t)}-\widehat{\theta}\right) \tag{33}$$

は

$$\left(\theta_0+\frac{B}{T-1}\right)-\left(\theta_0+\frac{B}{T}\right)=\frac{B}{T(T-1)} \tag{34}$$

の推定量として使えるということがわかります。

追記

- ここでは、1 期だけを省いて、パラメーターを推定する、ジャックナイフを考えましたが、２つの期を省いて行うジャックナイフも考えることができます。しかし、そうすると分散が増えてしまいます。

- また、ジャックナイフ法は、バイアスの式などが必要ないので、既存の計量パッケージでも、計算することが可能です。

- ここでは、母数の推定を考えましたが、実証研究では、限界効果 (marginal effect) の推定も重要になります。限界効果のバイアス修正をかけた推定法も同じように考えることができます。

動学モデルの場合

- 動学モデルの場合には、バイアスの式はもっと複雑になり、またバイアスの推定法も変わります。Hahn and Kuersteiner (2007) と、Hahn and Kuersteiner (2011) を参照してください。

- ジャックナイフ法も、そのままでは使えなくなります。動学モデルでも使えるジャックナイフ法の話は、Dhaene and Jochmans (2010) にでています。

- 具体的なモデルを扱ったものとしては、2項選択モデルで、動学的要素の入ったモデルを扱った、Fernandez-Val (2009) やCarro (2007) や、分散の時系列構造を考慮したモデルを扱ったHospido (2012) があります。

まとめ

- このチュートリアルでは、非線形モデルの固定効果推定を考えました。
- 固定効果推定量は、T が大きくないときには、バイアスを無視することは、できないものの、バイアス修正が可能であることがわかっています。
- 特にジャックナイフ法は、一般的に使えるので、有用です。
- ただ、動学モデルの場合は、ここで紹介した方法はそのまま使えませんので、注意してください。